UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

UTTRI

THE CORE-SATELLITE SURVEY DESIGN PARADIGM: A CONCEPTUAL FRAMEWORK

2017-11-24

Transportation Tomorrow Survey 2.0

Patrick Loa, Siva Srikukenthiran
Khandker Nurul Habib & Eric J. Miller

EXECUTIVE SUMMARY

Data can be a powerful tool for engineers, planners, and policy-makers alike when trying to understand the choices that people make and the factors that influence these decisions. As one of the more popular methods of collecting data, particularly when larger sample sizes are desired, the design of a survey should be informed by the capability that the resulting data are attempting to enable. This consideration should be reflected in each component of the survey, from the choice of sampling frame to the design of the survey instrument, and the manner in which participants are recruited.

This report summarizes the findings of the first stage of an investigation into the feasibility of conducting a survey using the core-satellite survey paradigm. More traditional travel survey methods have become increasingly inadequate, both in terms of their ability to represent the target population and in terms of the changing set of capabilities that the users of these data desire to enable. The former has stemmed from the mounting challenges associated with forming a comprehensive sampling frame, as well as the growing obsolescence of landline telephones. The latter stems from the desire to model travel behaviour at more disaggregate levels, which increases both the amount and level of detail of data required. In an attempt to address these issues, the core-satellite paradigm proposes a more targeted approach to data collection.

The core-satellite survey paradigm is comprised of three components: the core survey, the satellite survey(s) and complementary datasets. The core survey is analogous to the traditional travel survey – a relatively large-sample survey of the population. The satellite surveys and complementary datasets serve to supplement and enrich the core. Within the core-satellite survey paradigm, the data collected through the core survey should be a reflection of the capabilities that the surveyors deem to be fundamental, such as the ability to understand the travel behaviour of the target population. Conversely, satellite surveys are typically used specifically to target the members of a sub-population (characterized by sociodemographic characteristics and/ or behaviour), and aim to collect relatively smaller samples. The key benefit to the core-satellite approach is the ability to distribute response burden among respondents, allowing more detailed questions to be posed to the members of the sub-population of interest.

When conducting a survey in this manner, a key consideration is the need to be able to link the data collected between the core and satellite surveys together. This necessitates that each survey collects information to facilitate this linkage, such as the sociodemographic information. Furthermore, the design of the core and satellite surveys should ensure that an adequate level of compatibility between the resulting datasets in the three contexts in which data exist – spatial, temporal, and semantic.

The conduct of the Transportation Tomorrow Survey (TTS) in the Greater Golden Horseshoe Area (GGHA) provides agencies within the survey area with a unique opportunity. Representing a 5% sample of households within the GGHA, the TTS can act as the core survey within the core-satellite survey paradigm. The conduct of the TTS can allow an individual agency to focus its data collection efforts on the capability that they want to enable, while being able to rely on the data from the TTS to enable a basic level of capability. The ability to exploit the availability of data from the TTS is particularly useful for agencies who do not have the resources to conduct their own surveys on the same scale.

In this report, the best-practices for five types of satellite surveys are discussed: transit on-board surveys, active mode user surveys, post-secondary student surveys, employee surveys, and attitudinal and stated preference surveys. While each of these surveys have their own unique characteristics, there exists a set of fundamental considerations when designing a satellite survey. Because satellite surveys primarily serve to supplement or enrich the core dataset, these surveys should be designed to be compatible with the core. Additionally, because satellite surveys take a more targeted approach to data collection, the design of these surveys should account for the characteristics of the respondents, as well as the context in which the survey will be conducted.

While this paradigm has the potential to produce a more representative and comprehensive dataset, additional work is needed to fully develop the paradigm. A key area of future work is the evaluation of the effects of incompatibility between two datasets on the ability of the satellite survey dataset to supplement the core dataset. Additional work is also required to develop a methodology to implement the core-satellite survey paradigm in practice.

The Core-Satellite Survey Design Paradigm: A Conceptual Framework

TRANSPORTATION TOMORROW SURVEY 2.0

Table of Contents

Glossary of Terms

Term	Description
AADBT	Average Annual Daily Bicycle Traffic
AAPE	Average Absolute Percent Error
APTA	American Public Transportation Association
APC	Automated Passenger Counter
AVL	Automatic Vehicle Location
CAPI	Computer-Assisted Personal Interview
CATI	Computer-Assisted telephone interview
CBD	Central Business District
CDR	Call Detail Records
DOT	Department of Transportation
FMS	Future Mobility Survey
FSA	Forward Sortation Area
FTA	Federal Transit Administration
GGHA	Greater Golden Horseshoe Area
HTS	Household Travel Survey
IPF	Iterative Proportional Fitting
ITS	Individual Travel Survey
MPO	Municipal Planning Organization
NHTS	National Household Travel Survey
RP	Revealed Preference
SP	Stated Preference
TM	Targeted Marketing
TAZ	Traffic Analysis Zone
TTS	Transportation Tomorrow Survey

1 INTRODUCTION

On its own, a set of data has no inherent value, it is a simply a collection of numerical values organized in a specified manner. The value of data lies in its ability to enable a capability, whether it be the capability to gain insights into the behaviour of the target population, or the capability to make better-informed assumptions about the future, data should be regarded as a means to an end, rather than the end itself. As a means of collecting data, surveys should be regarded in the same manner. The sole purpose of conducting a survey is to collect specific information about the characteristics and/ or behaviour of a particular population, in order to gain insights or make inferences about the population. It is this paradigm that leads to a key requirement for the design of any type of survey – that the administration of the survey should be specifically tailored to the population that is being targeted and the capability that the data is attempting to enable. Considerations include the choice of the sampling frame from which the sample is drawn, the design of the survey instrument, and the manner in which members of the sample are recruited.

Among the numerous types of errors and biases that are inherent in each set of data, such as measurement error, coverage error, sampling bias, and non-response bias (Habib 2014), the impacts of these errors and biases can be mitigated and in some cases avoided in the design phase of a survey. The impacts of coverage error (the gap between the target population and the sampling frame) and sampling bias (the gap between the sampling frame and sample) (Habib 2014), can be mitigated through a few methods to ensure that the sample is adequately representative of the target population. These include the selection of a sampling frame that is as representative as possible and the selection of a sampling technique that allows for the categorization of sampling units. For example, given the distribution of the ages and genders of the residents of a census tract, applying a stratified random sampling method, in which the population is divided into groups based on their age and gender, can help to ensure that the sample is adequately representative. Non-response error, which stems from differences in respondents versus non-respondents (Habib 2014), can be addressed in the recruitment phase to improve response rates. Within the context of travel survey data, measurement errors can be reduced through proper design of the survey instrument.

The design of the survey instrument has become increasingly important in recent years, as traditional travel survey methods have yielded increasingly inadequate results. Traditionally, household travel surveys have been conducted over the landline phone, often with the use of computer-assisted telephone interview (CATI) software, with households predominantly used as the sampling unit. Over time, two key changes have led the data collected through traditional household travel surveys to become increasingly inadequate. The first key change is declining landline telephone ownership, which has had an adverse effect on the sampling frames from which these surveys draw samples. As landline telephone ownership has steadily declined, the traditional use of lists of residential (landline) telephone numbers as a survey sample frame has gradually increased the potential for coverage error. In the 2013 edition of the Residential Telephone Service Survey, 30% of Canadian households did not own a landline telephone, while 60% of households whose members are all younger than 35 years old did not own a landline (Statistics Canada 2014). The second change is the desire in transportation demand modelling to perform analysis at more disaggregate levels, increasing the detail and amount of data needed. The increased difficulty in collecting all the data required for these types of analyses in a single survey is difficult. In addition to this, increased difficulties experienced when trying to develop or obtain a comprehensive sampling frame has led to a desire to base the design of future surveys on the core-satellite survey paradigm.

The core-satellite paradigm aims to address the issues associated with traditional travel survey methods by taking a more targeted approach to data collection. This framework for data collection is comprised of two key components: the core survey, and one or more satellite surveys. The core survey is akin to the large-scale travel surveys of the general population that are currently conducted (such as the Transportation Tomorrow Survey in the Greater Golden Horseshoe Area). They are characterized by their relatively large sample sizes, as well as the utilization of the data collected through said surveys as the basis upon which basic transportation planning capabilities are built. In contrast, satellite surveys target specific sub-populations (characterized by sociodemographic characteristics and/ or behaviour), and typically attempt to obtain a relatively small number of samples. One of the strengths of the core-satellite approach is the ability to distribute response burden among survey respondents. The conduct of satellite surveys allows more detailed questions to be posed to respondents that belong to a sub-population of interest, without burdening other respondents for whom the questions are irrelevant. Satellite surveys predominantly serve to supplement or enrich in the core dataset by addressing gaps and collecting more detailed information from the target population. A more detailed discussion of the core-satellite survey paradigm is presented in Chapter 2. As part of the TTS 2.0 project carried out by the University of Toronto, the ability to incorporate the principles of the core-satellite survey paradigm into the Transportation Tomorrow Survey (TTS) is currently being investigated.

In the Greater Golden Horseshoe Area (GGHA), the Transportation Tomorrow Survey (TTS) has been conducted since 1986 (U of T Data Management Group, 2014); the study area of the 2016 TTS is shown in Figure 1. The 2011 version of the TTS was one of the largest travel surveys ever undertaken, with data collected from almost 160,000 households and over 400,000 persons in the GGHA. The periodic conduct of the TTS in the GGHA presents a unique opportunity for agencies in the area to supplement their own data collection efforts with data from a large-sample regional travel survey. This is provided that their data was collected in a manner that allows for direct comparison and combination (i.e. as satellite surveys). With this in mind, a literature review pertaining to a number of different satellite surveys is presented, and the requirements to ensure that the data collected through these surveys are compatible with the data collected through the TTS are discussed.

FIGURE 1: THE BOUNDARIES OF THE 2016 TRANSPORTATION TOMORROW SURVEY

This report provides an overview of conducting surveys within the core-satellite survey paradigm. Chapter The Core-Satellite Survey Paradigmdescribes the core-satellite survey paradigm, including the roles that each of its components plays, general design principles, and common types of satellite surveys, before presenting the conduct of the Utah Travel Study as a case study. Chapter Opportunities to Utilize TTS Datapresents the opportunities and motivations for ensuring that data collected through satellite surveys be compatible with data collected through the TTS. Chapter Best-Practices for Satellite Surveys discusses the conduct of several types of surveys that have the potential to be used a satellite surveys, namely transit on-board surveys, surveys of active mode users, post-secondary student surveys, and workplace surveys. Chapter 5 provides an overview of the different types of passive data, including the advantages and disadvantages of using each type of data, and a discussion of the applications of passive data that do not fall into one of the four categories of satellite surveys. Finally, Chapter 6 provides recommendations for future work.

2 THE CORE-SATELLITE SURVEY PARADIGM

This chapter presents the principles of the core-satellite survey paradigm. First, the motivation of conducting surveys within the core-satellite survey paradigm is presented. Next, the roles of the core and satellite surveys are identified and discussed before general principles of the design of satellite surveys are presented. Common types of satellite surveys (transit onboard, active mode user, post-secondary student, and employee surveys) are discussed before the Utah Travel Study is presented as a case study of an integrated approach to the design of a survey using the core-satellite approach.

2.1 Motivation

The core-satellite survey paradigm arose in response to two key shortcomings of more traditional travel survey methods: the growing inadequacy of the data collected through said methods, and the need for more detailed data to support the analysis and modelling of transportation networks at a more disaggregate level. These shortcomings stem from a number of issues, the most fundamental of which is the inability of a commonly-used sampling frame to adequately represent the general population. Travel surveys tend to utilize address-based sampling (ABS) or random-digit dialing (RDD) techniques to select members of the sample, with the latter becoming increasingly problematic. While a list of residential (landline) phone numbers used to be able to provide sufficient coverage of households in a survey area, a number of developments have resulted in lists of residential phone numbers becoming inadequate for use as a standalone sampling frame. Increases in smartphone ownership have coincided with decreasing levels of residential landline ownership (creating so-called "mobile phone-only households") which, combined with do-not-call lists and call screening services, have significantly impaired the ability to obtain a representative sample of the population using RDD-based sampling methods. In addition to issues with the sampling and recruitment stages of the survey, there are issues pertaining to the collection of data that must also be addressed.

The core-satellite survey paradigm aims to address both traditional and emerging concerns regarding the collection of data through surveys. One of the key sources of error in surveys is non-response bias, which arises when the information provided by respondents systematically differs from that of non-respondents (Habib 2014). Non-response can take two forms: item (question) non-response and unit (survey) non-response (Tourangeau, Groves and Redline 2010). While recruitment techniques may contribute to unit non-response, response burden can affect both item and unit non-response. The emerging need for more detailed data, in support of the development models at a more disaggregate level (such as activity-based models), has the potential to further exacerbate non-response. Attempts to obtain more detailed data have the potential to manifest themselves as the inclusion of more items within a questionnaire or an increase in the complexity of questions posed to survey participants, both of which have the potential to increase the burden experienced by respondents. The core-satellite survey paradigm aims to address the potential increase in response burden by shifting the burden away from the respondents to whom the additional questions do not pertain. While this approach to data collection is one means of addressing issues stemming from response burden, each stage of the survey should be designed with response burden in mind.

Response burden generally has an adverse effect on both response rates and data quality. Response burden arises due to the effort required to complete a questionnaire, which stems from both the length of the questionnaire and the overall design of the survey instrument (Rolstad, Adler and Ryden 2011).

While Rolstad, Adler, and Ryden (2011) found a relatively weak association between questionnaire length and response burden, the effects of questionnaire either had no effect or an adverse effect on response burden. The length of questions posed to respondents and the number of response options are presented have also been shown to increase burden to respondents (Yan and Tourangeau 2008). The utilization of satellite surveys can address the burden experienced by respondents to the core survey, both by allowing the core survey to focus solely on the collection of key data (as defined by the surveyors), thus reducing the length of the questionnaire, and potentially by allowing for the use of a smaller set of more general response options. Thus, the use of satellite surveys has the potential to positively impact the quality of both the data collected through the core survey, as well as the quality of the overall dataset. The division of data collection efforts into core and satellite surveys allows surveyors to reduce the burden associated with completing the core survey. This has the potential to improve response rates and reduce the costs associated with data collection.

2.2 The Roles of Core and Satellite Surveys

The application of the core—satellite paradigm allows surveyors to take a more targeted approach to data collection, particularly by using satellite surveys to focus on obtaining information pertaining to a particular subpopulation or behaviour. As with all surveys, the design of each component of the data collection effort should reflect the population being targeted and the capability that the surveyors wish to enable. One of the key motivators for the conduct of satellite surveys is their ability to address gaps in the core dataset, through the collection of information that would otherwise "not be feasible and/or cost-effective to collect as part of the core" (Miller, et al. 2011). Thus, satellite surveys can be used to collect information from a specific subpopulation, such as a demographic group (e.g. post-secondary students) or persons who engage in a certain activity (e.g. transit users, cyclists), without adding additional burden to respondents who are not part of the population of interest. The nature of satellite surveys underscores the need to both tailor the design of the survey to suit the members of the target population, as well as design the data collection effort in a holistic manner; this allows surveyors to exploit the advantages of the core-satellite survey paradigm, reduce redundancy, and to take a proactive approach to addressing possible issues. By taking a core-satellite approach to survey design, both the core and satellite surveys can be designed to reflect the designed to reflect their respective roles in the data collection process and to complement one another.

Within this framework, the core survey is a relatively large-sample survey that collects "primary information" pertaining to the key behaviour of the target population, as defined by the surveyors (Miller, et al. 2011). The core survey is analogous to a standalone survey, in that it aims to collect all of the data that are required to enable the desired capability; however, the key difference is that the use of one or more satellite surveys reduces the onus placed on the core survey to collect all of the necessary data. Thus, the core-satellite paradigm allows the core survey to focus solely on collecting the data needed to facilitate the capabilities that the surveyors deem to be fundamental, which tends to manifest itself in shorter questionnaires compared to standalone surveys. Core surveys are used to collect data that are fundamental to policy and/ or planning needs; generally, this will include attributes that can allow this core dataset to be linked to data collected through the various satellite surveys. They also are conducted with sample sizes large enough to allow for statistically-valid inferences about the population as a whole (Miller, et al. 2011).

Satellite surveys primarily serve to enrich and/ or supplement the core dataset by addressing gaps in the core dataset or by collecting data pertaining to specific subpopulations or behaviours. Compared to the core survey, satellite surveys are smaller-sample, focused surveys that aim to gather additional information about specific behaviours of interest (Miller, et al. 2011). The collection of data in this manner allow questions that are only relevant to a subset of respondents to be removed from the core survey, thus reducing its length, while also providing surveyors with a means of asking the members of the targeted subpopulation more detailed questions than if they were to be asked in the core survey. Furthermore, the ability to administer satellite surveys independent of the core survey allows surveyors to include more questions than if the data were being collected through the core survey. Satellite surveys can also be used to obtain data for specific models, through the use of additional questions, attitudinal questions, and the sampling of subpopulations (Miller, et al. 2011). The benefits of conducting satellite surveys include the ability to distribute survey burden among different groups of participants than the core survey (Goulias, Pendyala and Bhat 2011), which can have a positive impact on response rates, and by extension, survey costs. Examples of satellite surveys include surveys of cyclists, multi-day trip diaries, residential mobility, and dwelling type choice surveys, and car ownership surveys (Goulias, Pendyala and Bhat 2011). Regardless of the nature of the satellite survey, it must collect data that can facilitate linkage to the core data set (Miller, et al. 2011). The linkage between the core and satellite datasets is discussed in Chapter Opportunities to Utilize TTS Data.

The third component of the core-satellite paradigm is the complementary dataset, which can take the form of data collected through surveys or through passive data collection methods (such as cordon counts and smart card records). Complementary datasets mainly serve to augment both the core and satellite surveys; however, they may not contain data that can facilitate linkage to these surveys (Miller, et al. 2011). Complementary datasets include land use infrastructure data, travel time and cost data, indicators of industry-specific presence, the availability of firms by time-of-day, smartcard transaction records, and cordon counts (Goulias, Pendyala and Bhat 2011). An interesting development in the field of complementary surveys is the use of targeted marketing (TM) data as a means of obtaining up-to-date demographic and socioeconomic information. Interest in the use of "non-traditional" datasets has been motivated by the increase in large third-party datasets, including mobile phone signal traces, GPS data, transit smart card data, and credit card spending patterns (Kressner and Garrow 2014). Kressner and Garrow (2014) evaluated the viability of using TM data to expand travel survey data in lieu of census data and found that the distributions of age, gender, household income, and the presence of children were similar between the TM dataset and census data. Where possible, the conduct of complementary surveys should be guided by the expected applications of the collected data, and the gaps that the surveyors are attempting to address.

The definition of a gap in a dataset stems from the desired application(s) of said dataset. A gap in a dataset can only exist in relation to a specific application of said data, as data itself has no inherent value. Thus, the inability to use a specific dataset to facilitate the desired capability is indicative of a gap, which can exist for several reasons. First and foremost, gaps can arise from the use of an existing dataset for a purpose that was not accounted for by the original surveyors. This can be the result of cost constraints or differing goals of the original survey. Gaps can also arise due to the underrepresentation of a particular sub-population in the dataset, which may be the product of the sample frame used, the sampling technique used, or the number of samples obtained by the original surveyors. When engaging in data collection program, gaps should not exist, as the requirements of the dataset (as defined by the desired capabilities) should be used to inform the design of the core and/ or satellite surveys. Although the design of a satellite survey will predominantly be influenced by the capability that it is meant to enable, there is a fundamental set of design principles that should be kept in mind where designing a satellite survey.

2.3 General Principles of Satellite Survey Design

In general, the core-satellite survey paradigm emphasizes the content of the survey, rather the use of any particular mode (Miller, et al. 2011), and thus a satellite survey should be designed to enable a capability that the core survey cannot, or to supplement the core survey by addressing a gap in the core dataset. This gap may be the result of a desire to enable a new capability, or the desire to further develop an existing capability. Therefore, the definition of core and satellite data depends entirely on the needs of the surveying agency (Miller, et al. 2011), which should, in turn, be informed by the capability that they are seeking to enable. Because satellite surveys inherently serve to supplement or enrich the core dataset, these surveys should be designed to ensure that the collected data is compatible with the core dataset. As satellite surveys take a more targeted approach to data collection, their design should account for the characteristics of the respondents as well as the context in which the survey will be completed.

This sentiment is echoed by Rolstad, Adler, and Ryden (2011), who argue that the choice of survey instrument should be informed by the manner in which respondents perceive the survey. The authors cite a pair of studies in which survey respondents were asked to indicate their preferred survey instruments; in both cases, respondents preferred the instrument that was more pertinent to their specific context over the instrument that was shorter (Rolstad, Adler and Ryden 2011). Factors such as education, technical aptitude, question characteristics, and the number of response options have been shown to affect response times (Yan and Tourangeau 2008), which means that these characteristics also have the potential to affect response burden and the perceived quality of the survey.

One of the key design considerations is the process through which responses are chosen by survey participants, as this process can affect both non-response bias and measurement error. The process of survey response can be distilled into four components: comprehension of the question, retrieval of relevant information, use of said information to answer the questions, and the reporting of the answer (Yan and Tourangeau 2008). The wording of questions should be reflective of the target audience, in order to reduce the effort required to comprehend the question and to reduce the likelihood of measurement errors due to the misinterpretation of questions. Similarly, the survey instrument should be designed to improve the ease with which responses can be reported, and should account for factors such as technical aptitude, literacy, and constraints on the time of the respondent. In the case of attitudinal questions, the order in which questions are presented should also be taken into consideration. Attitudinal questions require a certain level of consideration, which requires information to be retrieved from within oneself and integrated when a question is asked; as these considerations tend to be organized by topic, the retrieval process can be expedited when questions are grouped by topic (Yan and Tourangeau 2008). Given the different populations that a satellite survey can target, a number of conventions exist for specific satellite surveys.

2.4 Common Types of Satellite Surveys

2.4.1 Transit On-Board Surveys

Transit on-board surveys are data collection efforts conducted by transit agencies to obtain data from their customers. The specific data collected by an agency is often dictated by the needs of the MPO and the agency itself, and typically falls into three categories: passenger demographics, travel behaviours, and customer satisfaction data; occasionally these are supplemented by stated preference (SP) and attitudinal data (Agrawal, et al. 2017). These data are typically used by transit agencies and MPOs for the purposes of travel modelling, long-range or area-planning, operations planning, scheduling, marketing, and customer communications (Simas-Olivera and Casas 2010). While transit on-board surveys have traditionally been conducted by distributing paper questionnaires on-board transit vehicles, new technologies such as the personal digital assistant (PDA) and tablet computers have led to the development of new survey instruments. The most notable of these are GPS-enabled tablet-based systems that are used to conduct face-to-face interviews with a subset of transit riders and to collect stop-level boarding and alighting data.

2.4.2 Active Mode User Surveys

Active mode trips (also referred to as non-motorized trips) refer to trips made by bicycle or on foot. Two general types of active mode user surveys exist surveys exclusively targeting current users, and those targeting both current and future users. Surveys solely targeting current active mode users, such as the 2006 *Downtown Commuter Cyclist Survey* in the city of Calgary (The City of Calgary 2007) and the Veloland Schweiz survey of the Swiss national cycling network (Richardson 2006), can be conducted as intercept surveys and may include the distribution of a paper questionnaire that can be completed and mailed back to the surveyors. Conversely, surveys that target both current and potential active mode users, such as the *Cycling in Cities* survey conducted by the University of British Columbia, may utilize a more probabilistic approach to sampling, in order to ensure that people who do not currently use active modes regularly are included. Aside from trip information and respondent characteristics, active mode user surveys will typically include attitudinal or SP questions to identify barriers to the use of active modes, gauge latent demand, gain insights into perceptions of cycling and pedestrian facilities. Passive data, such as GPS data and crowdsourced maps, are being used to study the factors that affect the route choice decisions of cyclists, including the preference of facilities.

2.4.3 Post-Secondary Student Surveys

Post-secondary students are typically underrepresented in traditional household travel surveys, in part due to the sampling frames traditionally used (Verreault and Morency 2016). This underrepresentation has led to the relatively inadequate representation of post-secondary students in regional travel demand models (Son, Khattak and Wang, et al. 2013), which is problematic, as post-secondary institutions can have a significant impact on travel demand in surrounding areas. Post-secondary student travel surveys tend to be administered through web-based surveys, in order to exploit the availability of email lists maintained by post-secondary institutions, and the relative ease with which students can be recruited through their institutional emails. The uses of post-secondary student travel data can be divided into three categories: to understand travel behaviour (including through the estimation of models), to develop and assess the impacts of policy changes (through the use of models), and to supplement traditional household travel data that underrepresents post-secondary students. Data collected through this type of survey are typically used to develop special generator models for post-secondary institutions and post-secondary student sub-models for incorporation into regional travel demand models.

2.4.4 Employee Surveys

Employee surveys (also referred to as workplace surveys), usually take the form of an intercept survey or a centralized survey of employees. Workplace surveys have two sets of sampling units: business establishments and individuals, with the sampling of both units needed to be addressed in the survey sampling plan. Because non-response has a significant impact on the accuracy of the survey results, partnering with or gaining support from the local Chamber of Commerce or business associations is recommended to establish credibility, and to improves acceptance rates on the part of businesses and survey response rates on the part of employees (Southwell, Zhang and Sharp 2014). Data from employee surveys are typically used to develop trip attraction rates and trip rates for particular types of trips, for inclusion in regional travel demand models.

2.5 Case Study: The Utah Travel Study

The *Utah Travel Study*, conducted by Research Systems Group Inc. in 2012, is an excellent example of an integrated design of a survey under the core-satellite survey paradigm. The core survey (the *main household diary*) utilized the U.S. Postal Service's *Computerized Delivery Sequence* (CDS) as the sampling frame, with households being chosen through a stratified random sampling process, with strata being defined based on geographic location and land use classification (Research Systems Group, Inc., 2013). The final survey results represented a 0.9% sample of households in the study area (124,888 households), with participation rates varying from 0.7 to 2.0%.

The core survey utilized advanced mailers to recruit households in the sample, assigned each household a unique eight-digit password, and gave recipients the option to complete the survey online or over the phone. The questionnaire was used to collect data about household characteristics (e.g. number of adults, housing tenure, household income and dwelling type), person-level characteristics (age, gender, race, ethnicity, education, and employment status), and vehicular information (household vehicle ownership, year, make, model, fuel type, and number of miles driven). Adult respondents were asked to report the trips that they and their children made during their assigned travel day. The information that was collected included the locations that they visited, the start and end times of each trip, the purpose of each trip, and the travel mode. Active mode users were asked to indicate whether they used a dedicated sidewalk or bike path, while respondents who were students or employed were asked about their typical commute, including the number of days that the same trip is made, the typical times that they arrive at and leave work or school, and the mode that they typically use. The *Utah Travel Study* utilized two different types of satellite surveys, integrated and independent. Integrated satellites utilized the unique eight-digit password to link the data collected to those collected through the core, while independent satellites separately targeted specific subpopulations in the survey area.

The integrated satellite surveys recruited participants in two ways. First, a proportion of the original sample was selected to receive one of three debrief surveys, which were seamlessly appended to the end of the core survey. Second, a proportion of respondents who indicated a willingness to participate in future surveys were invited to complete an additional survey. Regardless of the manner in which they were recruited, respondents to the integrated satellite surveys were asked to enter their household-specific eight-digit password, in order to link their responses to the data collected through the core survey. Recruited households were randomly selected to participate in one of three debrief surveys, the *attitudinal survey*, the *long-distance survey*, or the *bicycle and pedestrian survey*, which are later discussed in further detail. In the attitudinal survey, respondents were asked to rank the degree to which they agreed with statements regarding transportation planning priorities. The *long-distance survey* asked respondents about the major cities that they had visited in the past year to determine demand for high-speed rail, in addition to asking respondents to report the origin and destination, purpose, mode, departure data, and number of travellers for each long-distance trip that was made by the respondent in the past year.

The independent satellite surveys each used a sampling technique that was specific to the target population. Two independent satellite surveys were conducted: the *college travel diary*, and the *Dixie Sun Transit On-board Survey*. Residential addresses and demographic information from respondents were collected through the two satellites in order to facilitate linkages with other datasets, as well as to facilitate comparisons between household characteristics. The surveys also allowed for the analysis of non-response bias, and for the adjustment of the sampling plan to target specific types of households with lower response rates. The *college travel diary* was conducted in partnership with eight post-secondary institutions in Utah, with invitations being sent via email. Respondents were asked questions regarding their college affiliation, demographic characteristics, travel habits, and attitudes. The travel diary only considered off-campus trips, with students who did not make any such trips being asked why. The *Dixie Sun Transit On-Board Survey* was conducted onboard Dixie Sun Transit vehicles, with the design aiming to maximize the extent to which the survey was compatible with the previous on-board survey that was conducted by the Utah Transit Authority (UTA). The survey was designed to be short enough to be completed on a single sheet of paper; the questionnaire included questions pertaining to the expansion of the SunTran system, trip details, SunTran services, and respondent demographics.

Two satellite surveys that were conducted as part of the Utah Travel Study were conducted as both an independent and integrated satellite survey. The *bicycle and pedestrian survey* had two components: the *debrief survey* (which was seamlessly appended to the end of the core survey), and the *barriers survey*. The *debrief survey* asked about walking and cycling habits and included attitudinal questions about the respondents' perceptions of existing walking and cycling infrastructure. The *barriers survey* was administered to core survey respondents who had indicated a willingness to participate in future surveys, in addition to a set of local businesses and community organizations that were likely to have an interest in walking or cycling issues. The *barriers survey* included five components: household details and screening, pedestrian habits and barriers, bicycle habits and barriers, attitudes, and demographics. The survey mainly focuses on identifying "problem areas" and physical barriers to the use of active modes. The *residential choice survey,* a combined revealed preference/stated preference survey, was administered to both core survey respondents who indicated a willingness to participate in future surveys and respondents to the UTA on-board survey who indicated the same.

3 OPPORTUNITIES TO UTILIZE TTS DATA

This section discusses the benefits of ensuring that travel survey data collected by agencies in the GGHA are compatible with data collected through the TTS. First, an overview of the conduct, data collected, and data expansion techniques is provided. Next presented are the potential data that could be collected through five types of satellite surveys: transit on-board, active mode user, post-secondary student, workplace surveys, and attitudinal and stated preference surveys. A summary of the types of surveys collected by agencies in the GGHA is presented, followed by presenting an argument for why the data collected by these agencies should be compatible with TTS data. The principles for ensuring that datasets are compatible are then discussed before, finally, the conduct of the NHTS and its Add-On program are presented as a case study.

3.1 Overview of the TTS

3.1.1 Survey Conduct

The 2011 *Transportation Tomorrow Survey* (TTS) was conducted by the Data Management Group at the University of Toronto on behalf of 23 local, regional, provincial, and transit agencies in the Greater Golden Horseshoe Area. The 2011 edition of the survey represented one of the largest travel surveys ever undertaken, with data being collected on almost 160,000 households, over 410,000 respondents, and over 850,000 trips. The sampling frame for the 2011 survey was a list of residential phone numbers maintained by the Cornerstone Group of Companies, a company that specializes in the maintenance of residential landline and address listings. The sampling procedure utilized a stratified random sampling approach, with strata being defined based on the type of dwelling and the forward sortation area (FSA) in which the household was located. The survey aimed to obtain a 5% sample of households in the study area; however, multi-family units (e.g. apartments) were sampled at a higher rate (from 10 to 15%) due to the relatively low response rates anticipated for the residents of these dwellings. The final dataset represented a 5.1% sample of households in the GGHA, with sampling rates varying from 3.8 to 10.9% in different municipalities.

A mailing house was contracted to send advance letters to households that were included in the sample, notifying recipients of the time and date for the household to expect a call from a survey interviewer, typically a week later. The letter also included a unique web access code that recipients could use to complete the survey online. For households that completed the survey over the phone, a single member of the household was asked to act as a proxy for all other household members and provide responses on their behalf. While demographic data was collected from all household members, trip information was only collected for respondents over the age of 11. (U of T Data Management Group 2014)

3.1.2 Data Collected and Data Expansion

In 2011, the TTS collected data pertaining to household, person, trip, and transit trip characteristics, as outlined in Table 1 below (U of T Data Management Group 2014). The two primary sources of bias in the dataset were the sample frame and the timing of the sample selection, as the list of residential phone numbers was not representative of the survey population. The data collected through the survey were expanded to represent the target population in two stages. The data were first expanded based on the distribution of dwelling units within each FSA, with the requirement that a minimum of 500 interviews had to have taken place with households in a given FSA for it to qualify; those FSA that did not qualify were grouped with adjacent FSAs. The second stage of data expansion involved the expansion of data based on the distribution of age cohorts in each municipality, with age groups being combined where necessary to ensure that each group had at least 1,000 members. Age-specific expansion factors were applied to the person and trip data associated with each survey respondent. (U of T Data Management Group 2013)

TABLE 1: SUMMARY OF THE DATA COLLECTED IN THE TTS

Data Type	Data Collected
Household Data	• Home location • Dwelling unit • Number of persons • Number of personal vehicles
Person Data	• Gender • Age • Driver's license ownership • Usual work location • Availability of free parking at work • Student status • Employment status • Occupation • Usual school location • Origin of the first trip
Trip Data	• Destination location • Trip purpose • Start time • Mode • Number of persons in vehicle • Use of highway 401
Transit Trip Data	• Egress mode • Access mode • Sequence of transit routes used

3.2 Potential Satellite Data

The choice of data collected through a satellite survey should be informed by the conduct of the core survey, to ensure that the satellite data is compatible (as outlined in Section Ensuring Compatibility). The definition and collection of satellite survey data must also be informed by the purpose of the satellite survey and the capabilities that the surveyors hope to enable. This section outlines potential data that could be collected through four subpopulations: transit users, active mode users, post-secondary students, and employees.

3.2.1 Applications for Transit Users

For transit users, potential satellite data include *attitudinal and SP data* pertaining to the customers' perceptions of transit service, in addition to information on the o*rigin, destination, boarding stop, alighting stop, access mode, and egress mode* for transit trips if they are not already collected. In terms of complementary datasets, the use of *APC, AVL, and smart card transaction data* have increased in prominence, as attempts have been made to more-precisely measure demand and to estimate route-level OD matrices.

3.2.2 Applications for Active Mode Users

As active mode users tend to be underrepresented in traditional household travel surveys, an example of satellite data could be as simple as socioeconomic and demographic information, combined with the collection of a travel diary. Satellite data can also take the form of attitudinal and stated preference (SP) data, in order to better understand the route choice behaviour of cyclists, to identify barriers to the use of active modes, and to develop models to forecast the potential impacts of policies and infrastructure improvements. Complementary datasets, such as GPS data and passive count data, are now being used to understand the factors that affect the route choices of cyclists and to quantify the impact of weather and seasonality through the estimation of models.

3.2.3 Applications for Post-Secondary Students

Similar to satellite surveys of active mode users, the underrepresentation of post-secondary students in traditional household travel surveys means that basic survey data corresponding to post-secondary students (i.e. socioeconomic and demographic information, combined with a travel diary) can be considered as satellite data. The collection of attitudinal and SP data through satellite surveys can also facilitate the development of post-secondary student behavioural models that can be used to forecast and understand the impacts of different policies.

3.2.4 Applications for Employees

When conducted as a satellite survey, data from employee surveys are predominantly used to develop trip attraction rates and purpose-specific trip rates. Because there is the need to ensure that these data are compatible with data collected through the core survey, employee surveys must also collect some form of "basic" survey data, (i.e. socioeconomic and/ or demographic information). The stratification of these trip rates based on area type, establishment type, and community size (Sperry, et al. 2015) necessitates the collection of complementary data pertaining to sampled establishments, including the geographic location of each business, the industry sector(s) in which the business operates, and the population of the surrounding areas. When conducting intercept surveys with the intention of expanding data from the sample, it is imperative to collect counts of all persons entering and leaving the establishment in order to develop data expansion factors (Southwell, Zhang and Sharp 2014).

3.3 Surveys Conducted by TTS Agencies

A selection of the data collected by agencies that participate in the TTS through various types of surveys is presented in Appendix A. The list of data collected represents a subset of the surveys conducted by these agencies, based on published survey reports. While there is a certain degree of commonality between different surveys, by and large, the data that is collected through each survey is determined by the purpose of the survey and the goals of the surveyors.

3.4 Motivation for Using TTS Data

Although it requires additional effort, designing one's survey to ensure compatibility with the data collected through the TTS can be beneficial for agencies in the GGHA. As the data fusion approach tends to primarily focus on the data needs of conventional models (Stopher and Greaves 2007), designing a survey to act as a satellite survey of the TTS allows an agency to focus its data collection efforts on the capability that it is trying to enable, while only collecting data that is required to facilitate linkage between the survey dataset and that of the TTS. One of the key benefits that stems from ensuring that the survey data collected by an agency is compatible with TTS data is the ability to exploit the availability of data collected for a relatively large proportion of the residents of the GGHA. The ability to exploit this availability is particularly useful for agencies who do not have the resources to conduct their own surveys on the same scale, which can be used to help meet their planning and policy needs. Thus, the data from the TTS could conceivably be used to enable a basic set of capabilities, such as the development of a traditional four-stage model, while satellite surveys could be used to develop or improve sub-models for particular subpopulations, such as post-secondary student trip generation models or cyclist route choice models. Furthermore, satellite surveys can also be used to collect additional information on modes of transportation for which little or no information is collected in the core survey. Before satellite survey data can be fused with TTS data, the design of the satellite survey(s) must ensure that the datasets are compatible.

3.5 Ensuring Compatibility

The benefits of conducting a survey according to the core-satellite survey paradigm are maximized when the data collected through the satellite surveys can be linked (or "fused") together with the core, which necessitates a certain level of compatibility. The compatibility of two or more datasets is based on the context in which the data were collected; data exist in three contexts (Miller, et al. 2011).:

- **The Spatial Context:** the spatial indicators used
- **The Temporal Context:** the time period during which data was collected

- **The Semantic Context:** the manner in which variables and categories are defined

Thus, in order to ensure compatibility, the different sets of data must either refer to a similar spatial context, a similar temporal context, or a similar semantic context. Furthermore, the increasing need for data at finer levels of resolution places an additional strain of data collection efforts and typically results in at least one compatibility issue that must be resolved prior to the data fusion process (Bayart & Morency, 2008). Citing the work of Judson (2006), Bayart and Morency (2008) argue that there are three principles for the integration of databases:

- **The Latent Variable Principle:** the recognition that the estimand exists, but may not always be directly observed;
- **The Uncertainty Principle:** the understanding that the data contributing to the estimate are not themselves without flaws; and
- **The Modelling Principle:** the formalization of the relationship between the estimand and the data source(s)

In their most basic form, data fusion methods aim to enrich survey data in order to better meet data needs (Bayart and Morency 2008). This is done by utilizing multiple datasets to supplement one another and compensate for gaps in each individual dataset in order to produce a new, more comprehensive dataset (Miller, et al. 2011). A basic requirement of data fusion is the existence of at least two datasets that collectively contain all of the required information, with at least one common variable among the datasets to allow a suitable statistical matching method to impute the value(s) of the variable(s) of interest (Miller, et al. 2011). Miller et al. (2011) argue that the key impediment to ensuring that an acceptable degree of compatibility exists between multiple data sets is the need to address the different levels of aggregation that exist between different sets of data before they can be fused together; Bayart and Morency (2008) argue that the basic issue with the data fusion process is the fact that statistical inferences are being made about the joint distribution of two variables without being able to observe the distribution directly.

Where possible, the survey design should make every effort to ensure that the different datasets that are to be collected are compatible prior to the start of the data collection process (D'Orazio, Di Zio and Scanu 2006). Ensuring that different data sets are sufficiently compatible in the spatial context is relatively simple, as the survey areas are defined by geographic boundaries. Compatibility in the semantic context can be assured by maintaining a consistent definition of each key term and value. This process becomes more difficult when attempting to ensure that several sets of data are sufficiently compatible. The greatest amount of uncertainty exists when attempting to ensure compatibility in the temporal context. The only sure-fire means of ensuring compatibility in this context is to conduct the core and satellite surveys. Aside from that, the impact of the time lag between the conduct of the core and satellite survey(s) has not been empirically evaluated. As a rule of thumb, shorter time lags are preferable, although the occurrence of a paradigm-shifting event is a greater impediment to the compatibility of different datasets than the time lag.

While the data fusion process requires that the datasets must be compatible in at least one dimension (context), compatibility may not initially exist when one attempts to fuse a set of disparate datasets together. In such situations, there are several methods to ensure that the data that are to be collected are compatible with an existing dataset. D'Orazio, Di Zio and Scanu (2006) cite the work of Van der

Laan (2000) when identifying a number of methods to “harmonize” datasets (i.e. to improve the degree to which they are compatible):

- Harmonizing the definition of spatial and temporal units
- Harmonizing the reference periods
- Ensuring that the target populations are compatible
- Harmonizing the definitions of the variables
- Harmonizing the definitions of classifications
- Adjusting for measurement errors
- Adjusting for missing data, and
- Deriving variables in a consistent manner

In cases where the definition of one or more variables or categories are inconsistent (i.e. the semantic contexts of the datasets are not compatible), a few approaches are possible; variables can be re-coded or re-categorized, variables can be substituted with a new set of variables, and variables that cannot be harmonized can be identified in order to avoid their use as a common variable in the data fusion process (D'Orazio, Di Zio and Scanu 2006). Once the datasets have reached an adequate level of compatibility, there are a number of methods to fuse the datasets.

3.6 Overview of Data Fusion Methods

In addition to the three contexts in which data exist (spatial, temporal, and semantic), there are three additional contexts into which the process of data fusion can be categorized (Miller, et al. 2011):

- The Mixed Context: where the different biases that arise due to different survey modes are accounted for;
- The Survey Mode Context; and
- The Data Type Context

The mixed context is becoming increasingly prominent, as mixed-mode surveys are becoming more prevalent, due to their potential to reduce response bias, however, it also has the potential to affect the compatibility of the data collected through the different survey modes (Bayart and Morency 2008). The consideration of the survey mode(s) being offered as part of the decision to participate in the survey has the potential to reduce the randomness of the sample, which may introduce bias into the results (Bayart and Morency 2008), in addition to potentially affecting the quality of the data if respondents were reluctant to participate in the first place (Tourangeau, Groves and Redline 2010).

Within the context of transportation survey data, the most commonly-used data fusion method is the weighting and expansion of survey data to represent the entirety of the target population, typically through the use of census data to obtain population control totals. Bayart and Morency (2008) provide an excellent overview of data fusion methods; in general, two broad categories of data fusion methods exist:

- **Micro-Level Data Fusion**: a synthetic file containing all required data is constructed; and
- **Macro-Level Data Fusion**: the initial files are used to estimate the joint distribution

Within these classifications, there are three general approaches to data fusion:

- **Exact Matching:** records can be matched without uncertainty, typically through the use of a unique identifier (such as how the *Utah Travel Study* used the unique eight-digit identifier to link data collected through the *main household diary* to data collected through the integrated satellite surveys)
- **Explicit Models:** a model is used to connect variables of interest between two sets of data on the basis of a set of common variables
- **Implicit Models:** an existing record in the donor file that is as similar to a record in the receptor file as possible is found; can be constrained to limit the number of times that a particular record in the donor file is used, or unconstrained using a weighted combinatorial procedure

The common (matching) variables can be identified through a number of means, based on the type of variables being examined. The identification of common variables can be carried out through methods such as the calculation of correlation coefficients or the use of regression analysis for continuous variables, the Pearson chi-squared statistic for categorical variables or Somers' D for ordinal variables, (D'Orazio, Di Zio and Scanu 2006).

In the literature, a prominent example of surveys being designed to be compatible with an existing dataset is seen in studies that compare their results to those based on the data collected through the National Household Travel Survey (NHTS), described in Section 3.7. Son, et al. (2013) used data from the 2009 NHTS Add-On survey (conducted using random-digit dialing), and data from the 2008 National Capital Region Household Travel Survey (conducted using address-based sampling) to identify the impact of sampling on non-coverage, non-response, and measurement errors. In this case, the spatial and temporal contexts of the two surveys were compatible because they corresponded to a similar target population, and because they were conducted during similar time periods (March 2008 – May 2009 vs. Feb 2007 – Apr 2008) (Son, Khattak and Wang, et al. 2013). The conduct of the NHTS was also taken into account by Khattak et al. (2011) when attempting to compare the travel behaviour of university students in Virginia to that of the residents of the surrounding population. A web-based survey tool was developed to conduct a travel survey of students at four universities in Virginia and was designed to mirror the survey instrument used by the NHTS (Khattak, et al. 2011). In this case, semantic compatibility between the two datasets was desired in order to facilitate comparisons between the datasets.

3.7 Case Study: The NHTS and Add-On Survey

3.7.1 The NHTS

The National Household Travel Survey (NHTS) is a survey of the civilian, non-institutionalized population of the U.S. that is conducted via telephone and employs a stratified cluster sampling method. Using households as a sampling unit, clusters are formed using "100-banks," i.e. phone numbers with the same first eight digits, which were screened to ensure that they were residential phone numbers. The survey collects data pertaining to household demographics, vehicle ownership, personal demographic information, and travel day information, as well as data specifically pertaining to employed respondents. In the most recent iteration of the NHTS, the questionnaire included, among other things, questions pertaining to driver information, land use data, attitudes towards transit, internet usage, and the typical number of trips made that spanned multiple days. Trip information collected through the NHTS include trip purpose, mode, duration, time-of-day, travel day, interstate usage, and the specific household vehicle that was used to complete the trip. For a detailed summary of all data collected through the NHTS, see Appendix B. (U.S. Department of Transportation - Federal Highway Administration 2011)

Each household in the sample was contacted by phone, and a household member over the age of 16 was asked to act as a proxy for the rest of the members of the household. Travel days were assigned for all seven days of the week, with the aim of ensuring that travel across the entire year was captured. In total, data was collected from over 150,000 households and over 350,000 people. Person-level control totals that were obtained from the 2008 American Community Survey, which were stratified by geographic region, were used to expand person-level data. Household weights were derived based on the selection probability of each household and the calculation of a non-response adjustment factor. The data collected through the NHTS were used to quantify travel behaviour, analyze changes in travel characteristics over time, relate travel behaviour to traveller demographics, and to study the relationship between demographics and travel over time. (U.S. Department of Transportation - Federal Highway Administration 2011)

3.7.2 The NHTS Add-On Survey

The NHTS *Add-On program* is an excellent example of an agency taking advantage of the conduct of a large-scale travel survey to focus its resources on collecting more specific and detailed data to enable a desired set of capabilities. Conducted in conjunction with the NHTS, the *Add-On program* allows state departments of transportation (DOT) and municipal planning organizations (MPOs) to purchase additional samples within their jurisdictions (U.S. Department of Transportation - Federal Highway Administration 2011). Originally meant to complement the data that was collected through the NHTS (Ivey, Badoe and Edwards 2012), the 2009 edition of the *Add-On program* allowed its 20 “add-on partners” to add up to five additional questionnaire items to be asked specifically within their respective jurisdictions. Upon purchasing an add-on sample, partners were required to define the geographic extent of the study area as well as their desired sample size (Ivey, Badoe and Edwards 2012). Unlike the rest of the NHTS survey area, trip-origin and destination addresses were collected in add-on jurisdictions and geocoded, with assistance from CATI software (U.S. Department of Transportation - Federal Highway Administration 2011).

In general, the sample sizes and coverage of the NHTS is not large enough to allow for analysis at a finer spatial resolution than the census subdivision level, making it relatively inadequate for local use (Ivey, Badoe and Edwards 2012). The purchase of an add-on sample allows partners to stratify the survey area and specify required samples sizes at a relatively finer level of spatial resolution (the smallest likely being the census block level, as the weighting factors are partially based on census data), thus facilitating analysis and other applications at the local level. The data collected within the jurisdictions of add-on partners were weighted based on state sample selection probabilities, demographics, and other adjustment factors, with samples in add-on areas being selected randomly on the basis of telephone exchanges (the middle three digits in the telephone number) (Ivey, Badoe and Edwards 2012).

In a survey of agencies that had participated in the *add-on program* in the past, conducted by Ivey, Badoe and Edwards (2012), several motivations were highlighted for participating in the program.

These included the opportunity to use a survey instrument that had already been validated, the chance to obtain a set of data that was consistent with the data from the national sample, and the ability of the program to enable economies of scale. Although the *add-on program* is not a satellite survey but rather a more focused version of the NHTS, the applications of add-on data highlight the benefits of conducting satellite surveys. The alignment of a data collection effort with that of a validated survey instrument allows agencies to focus on its resources on the collection of data to enable a specific set of capabilities. This is particularly useful for smaller agencies, who can rely on the data from the core survey to enable a basic set of capabilities. Furthermore, the alignment of data collection method also gives agencies the opportunity to obtain a set of data that is consistent with data from a larger sample (a regional sample in the case of the TTS). Taking the time and effort to ensure that survey data is compatible with TTS data provides the surveyors with the ability to utilize from a regional travel survey with a relatively large sample size. The benefits of ensuring compatibility include obtaining data which was collected using a survey tool that has already been validated, at a lower cost than if the agency were to conduct a similar survey.

4 BEST-PRACTICES FOR SATELLITE SURVEYS

This section presents a review of the literature identifying the industry standards and best-practices for a selection of surveys that show promise as potential satellite surveys. The conduct of five types of potential satellite surveys are discussed: transit on-board surveys, active mode user surveys, post-secondary student surveys, employee surveys, and attitudinal and stated preference surveys. For each type of survey, the sampling methodologies are presented, the considerations that inform the choice of survey mode are discussed, and the applications of data are identified.

4.1 Transit On-Board Surveys

Transit on-board surveys are typically conducted by transit agencies as a means of collecting information on the characteristics of their customers, such as their demographics, travel patterns, and perceptions of the service that they are being provided. Transit agencies generally prefer to conduct customer surveys onboard their vehicles and in their facilities, as this type of survey tends to yield more accurate, reliable, and detailed information on their customers, compared to other survey modes. This section presents an overview of the conduct of transit on-board surveys, including the sampling techniques used by agencies, the considerations that inform the choice of survey mode, and the typical applications of data collected through transit on-board surveys.

4.1.1 Sampling Techniques

A key difference between transit on-board surveys and traditional travel surveys is that the former typically use transit trips as the sampling unit, rather than households or individuals. This discrepancy is due, in part, to the sampling techniques that are employed by transit agencies when conducting on-board surveys. For transit on-board surveys, sampling can take place on two levels: the route trip-level and the individual trip-level.

The 2009 *IndyGo on-board survey* is an example of a transit user survey that sampled on the route trip-level. When developing a sampling plan for trips, three categories of routes were defined: high-, mid-, and low-ridership, with target values of standard error and sample size defined for each (Bernardin, Lochmueller and Associates 2010). Route trips were stratified based on route direction, time of day, and route number. The sampling plan aimed for a 10% sample of the estimated daily ridership on fixed and circulator routes, and a 20% sample of the estimated daily ridership of express routes (Bernardin, Lochmueller and Associates 2010). On each sampled bus route, all passengers over the age of 15, as well as children who were travelling with guardians, were given a questionnaire, with participants being offered a one-day pass in exchange for completing the survey.

This approach to sampling is relatively common, as transit on-board surveys have been conducted as intercept surveys, with surveyors onboard transit vehicles distributing self-administered paper questionnaires to all passengers onboard sampled vehicles. The conduct of surveys in this manner necessitates the use of transit trips as sampling units, rather than the transit users themselves, because ensuring that customers were not given multiple surveys would prove to be a challenging task for the surveyors, particularly if the on-board survey is being conducted over multiple days. Another consideration is ensuring equal representation of various groups. The United States Federal Transit Administration (FTA) has directed agencies to ensure that minorities and low-income riders participate in these surveys, as they have traditionally been underrepresented (Agrawal, et al. 2017). While the distribution of self-administered questionnaires has been the norm, the use of computer-assisted personal interview (CAPI) software installed on tablet computers has emerged as a viable option to conduct an on-board survey (McHugh, et al. 2017). The utilization of tablet computers to conduct an on-board survey also necessitates the random selection of passengers to participate in the survey, rather than all passengers that board the vehicle.

4.1.2 Choice of Survey Instrument

The choice of survey instrument should be guided by two factors: the characteristics of the respondents, and the context in which the survey is received by respondents. With respect to the characteristics of respondents, the choice of survey instrument needs to account for the languages spoken by potential respondents, the distribution of ages of potential respondents, and amount of time that potential respondents will be spending on vehicles. The traditional conduct of transit on-board surveys, i.e. intercept surveys distributed on board transit vehicles, has informed the instrument that has been typically used to conduct these surveys – paper questionnaires.

A key consideration regarding the distribution of paper questionnaires is the manner in which they are returned to the surveyors. The 2009 *IndyGo on-board survey* addressed this issue by including a business reply mail permit with the questionnaire to allow respondents to complete the questionnaire at their leisure (Bernardin, Lochmueller and Associates 2010), while the on-board surveys conducted by TriMet in Portland, OR, prior to 2015 gave respondents the option to deposit completed surveys onboard transit vehicles, or mail them back to the surveyors (McHugh, et al. 2017). The ability to return questionnaires by mail removes the constraint of needing to complete the questionnaire before disembarking from the transit vehicle, which helps to address issues associated with collecting data from people completing short transit trips (Simas-Olivera and Casas 2010). If respondents are expected to complete the questionnaire while onboard the transit vehicle, the questionnaire should be designed to be as simple as possible in order to ensure that it is completed. The expected time to complete the questionnaire is also a key consideration if the survey is to be conducted using a tablet-based computer-assisted personal interview (CAPI) system.

Over the years, the emergence of new technologies such as the personal digital assistant (PDA) and the tablet computer has created new means of conducting transit on-board surveys. It has reached the point that the FTA has recognized the use of electronic handheld tablets as a legitimate and preferred method for collecting survey data (McHugh, et al. 2017). Compared to the traditional conduct of transit on-board surveys, the use of tablet-assisted surveys allows data to be coded in real time, offers the ability to monitor data collection, reduces the need to oversample to meet quotas (McHugh, et al. 2017). Furthermore, this approach reduces costs associated with printing survey materials and postage and has the potential to reduce costs associated with the hiring of a consultant to oversee the survey if an out-of-the-box system is used (McHugh, et al. 2017).

Following the conduct of the 2012 *Fare Survey*, TriMet decided to conduct the next iteration of the survey (in 2016) entirely by using tablet computers. McHugh et al. (2017) found that compared to 2012, for the same number of sample hours, the 2016 survey yielded a response rate that was 48 percentage points higher, yielded more completed surveys, obtained more information, offered better QA, resulted in a better resource distribution, and had fewer data errors. The use of technology-based survey methods tends to affect the sampling technique employed by the surveyors. Instead of being administered to all riders on the vehicle, the survey is instead taken of a sample of transit riders, typically using a random number generator to select riders to approach (McHugh, et al. 2017). The increasing ubiquity of internet access, combined with the relatively stagnant budgets of public transit agencies, has led some to investigate the feasibility of using web-based survey tools as a means of reducing the costs associated with collected data from transit users (Cummins, et al. 2013).

Agrawal et al. (2017) studied the impact of the manner in which a survey to bus passengers is administered on the quality of data. The overall response and completion rates, completion rates for particular questions, respondent demographics, and labour costs per complete survey were used as performance metrics. Three administration modes were considered: self-complete paper surveys, self-complete computer surveys, and tablet-assisted on-board interviews. The authors found that while respondents were more likely to accept web survey invitations, the paper self-complete surveys had higher response rates, and offered the best cost-to-completion ratio. The online and tablet surveys had fewer instances of skipped questions, with paper survey respondents skipping questions at a much higher rate (71% of questions skipped vs. 47%) (Agrawal, et al. 2017).

In order to quantify the effects of three different on-board survey design techniques on the overall response rate and unit costs, Memarian, Jeong and Uhm (2012) carried out a pilot test on the Tulsa Transit System. The authors independently tested the effects of three aspects of the survey design process:

- The length of the questionnaire (6, 14, and 29 questions)
- The incentive structure (free two-day tickets versus the chance to win a 31-day pass)
- The role of the surveyors (male, female, and designated drop-off boxes

The highest response rates were seen for the shortest survey, the survey that offered the chance to win a monthly pass, and for surveys distributed in person, with the length of the questionnaire being found to have a significant impact on both response rates and unit costs (Memarian, Jeong and Uhm 2012).

4.1.3 Data Collected and Applications

There is a relatively standard set of data that is collected through transit on-board surveys. In their review of 150 transit on-board surveys, the American Public Transportation Association (APTA) identified a number of variables collected by a majority of agencies: age, gender, ethnicity, household income, household size, occupation, vehicle availability, vehicle ownership. With respect to transportation data, agencies tend to collect information on access and egress mode, alternative modes, duration and frequency of transit usage, frequency of transit travel, transfer frequency, and trip purpose (American Public Transportation Association 2007). Drawing from the needs associated with the capabilities that transit agencies and MPOs want to enable, agencies usually use transit on-board surveys to collect three types of data: passenger demographics, travel behaviours, and customer satisfaction (Agrawal, et al. 2017). In the case of the 2009 *IndyGO transit on-board survey*, the survey team solicited three types of data from respondents: origin-destination travel patterns, access and egress modes, and rider demographics (Bernardin, Lochmueller and Associates 2010).

A key consideration when choosing the types of data to collect is whether or not the data will be expanded in order to represent the target population, which stems from the desired applications of the data. In order to facilitate the expansion of data, in the 2009 *IndyGo transit on-board survey,* the boarding stop of each respondent was obtained through the use of serial numbers on each questionnaire that was distributed. This information, along with the proportion of trips sampled and the average daily ridership, was used to weight the sample (Bernardin, Lochmueller and Associates 2010). Data collected through on-board surveys are typically used for travel demand modeling, long-range and area-wide planning, route planning and scheduling, service design, marketing, and customer communications (Memarian, Jeong and Uhm 2012).

In addition to recruiting passengers onboard transit vehicles, transit on-board surveys may also place an onus on surveyors to record boarding and alighting data, which can induce burden. Bernardin, Lochmueller and Associates (2010) attempted to address this issue by providing surveyors with "GPS-enhanced palm devices". The devices recorded temporal information, and utilized GPS sensors to assist surveyors with the task of collecting boarding and alighting counts at each stop. These counts are used by some surveyors to expand survey data.

4.1.4 Ensuring Compatibility with the TTS

While ensuring that the spatial context of a transit on-board survey is a relatively straightforward process, a key consideration is to ensure compatibility in the semantic context. In the TTS, transit trips are defined by the coordinates of the trip origin and destination, the coordinates of transit access and egress point, access mode, egress mode, and the specific transit routes used (U of T Data Management Group 2013). In order to ensure compatibility of trip data collected through a transit on-board survey with TTS data, there should be a common set of variables that define transit trips in the satellite survey. Furthermore, the collection of demographic data such as age, dwelling time, and residential location should also be collected, to facilitate the calculation of data weights if the satellite data are to be expanded in the same manner as TTS data (U of T Data Management Group 2013). With respect to the temporal context, the core and satellite surveys should be conducted in similar operational contexts, meaning that the service attributes (headways, span of service, etc.) should be relatively similar.

4.2 Active Mode User Surveys

The term "active mode users" refers to members of the population that complete trips by walking or cycling. In traditional household travel surveys, trips made by active mode users tend to be underreported, which has been attributed to several factors. One of the foremost causes of underreporting is the lack of a clarification regarding the reporting of short trips (Edwards, et al. 2012), and the tendency for respondents to neglect short and discretionary trips when reporting their travel (Son, Khattak and Chen, et al. 2012). This issue, combined with the tendency for active mode trips to be made over relatively short distances (typically less than three miles) (Edwards, et al. 2012), results in the underreporting of trips made using active modes and the underrepresentation of active mode users in household survey data.

The conduct of a satellite aimed at active mode users can help to supplement these gaps. It also provides an opportunity to develop a deeper understanding of the behaviour and travel patterns of active mode users, in a manner that may not be feasible in done through a traditional household travel survey. The first portion of this section discusses the key components of active mode user surveys conducted via questionnaire, before presenting the manners in which passive data have been used to study the behaviour of active mode users.

4.2.1 Sampling and Recruitment Techniques

The sampling and recruitment techniques employed by the administrators of active mode user surveys, predominantly cyclist surveys, vary based on the goals of the survey and the population that is being targeted by the surveyors. When surveyors are interested specifically in the behavior and travel patterns of cyclists, participants are typically recruited through intercept surveys or through snowball sampling.

One example of a survey that specifically targeted active mode users is the Veloland Schweiz survey of the Swiss national cycling network. The survey aimed to determine the usage of the national cycling network, and was conducted from 1999 to 2002, as described by (Richardson 2006). At 16 locations across the network, surveyors manually counted cyclists and pedestrians, conducted track-side interviews, and distributed questionnaires; these tasks were performed at each location on two weekdays (typically Wednesdays) and two weekends (typically Sundays) (Richardson 2006). Surveyors were instructed to record the time that the n^{th} cyclist arrived at the check point, and an estimate of their age, gender, and the size of the group with whom they were travelling. The surveyors were also instructed to request that respondents stop in order to complete the track-side interview; this interview collected information on their current home location (town/ postcode/ county) and whether or not the trip was overnight (Richardson 2006). Survey sites were chosen based on the stratification of the network by space, time, and population subgroups.

In cases where the surveyors are interested in the factors that influence the decision to walk or bike, the sampling frame that is used must contain both cyclists and non-cyclists. These types of surveys can employ both probability sampling, non-probability sampling, or a combination of both. For example, Twaddle, Hall and Bracic (2010) set out to investigate the demography of persons who made cycling trips to the University of Calgary campus. This study was conducted to determine if the majority of said trips were made by males (as was the case for Downtown Calgary), and to identify the barriers that prevent women from cycling. Conducted from Apr 13 to May 1, 2009, participants were recruited through the university's registrar's office, and using news articles on the web pages of the University of Calgary, its Office of Sustainability, and of a local bicycle shop. The survey was also aggressively marketed to students on campus, where seventy-five posters were placed around campus and on bike racks, and business cards containing survey invitations were taped to the handlebars of bicycles (Twaddle, Hall and Bracic 2010).

The *Cycling in Cities* survey carried out by the University of British Columbia took a probability-based sampling approach when attempting to identify the factors that affect the decision to cycle. During each of the three survey periods (Feb-Apr, May-July, and Sept-Dec), a population-based random sample of 4,000 residential phone numbers was selected from a phone directory, with sampled households sent an introductory letter that outlined the purpose of the survey (Winters, et al. 2011).

The study targeted cyclists and potential cyclists, the latter defined as people who had not cycled in the past year but would consider it in the future. The survey was comprised of 73 items, that represented a broad array of motivators and deterrents. The *bicycle and pedestrian* survey, conducted as part of the *Utah Travel Study*, utilized a combination of probability and non-probability sampling when attempting to identify barriers to the use of active modes. The participants of the *barriers survey* were recruited from the subset of *main household diary* respondents who indicated a willingness to participate in future surveys, representing 40% of said subset. The *barriers survey* focused on identifying "problem areas" and physical barriers to the use of active modes, in order to identify the needs of cyclists and pedestrians (Research Systems Group, Inc., 2013). The survey team also identified local businesses and organizations that were likely to have an interest in walking and biking issues, who were invited to forward the survey to their members, customers, and constituents (Research Systems Group, Inc., 2013). The approach yielded over 5,000 completed and valid responses from adults; for reference, the main household diary contains records corresponding to 18,700 adult respondents (Research Systems Group, Inc., 2013).

The different approaches taken to sampling and recruiting active mode users highlights key factors that must be considered when selecting sampling and recruitment methods. First and foremost, the desired application of the data will determine whether a probability or non-probability sampling method can be used. The former is necessary when attempting to use sample statistics to make inferences about the population as a whole, while the latter can be adequate when attempting to gain insights into the behaviours of the members of a subpopulation. When specifically targeting active mode users, the use of intercept surveys and other targeted sampling methods provide a reliable means of recruiting members of this subpopulation. Conversely, attempts to identify the barriers to and motivators of the use of active modes must employ a sampling technique that is capable of recruiting both users and non-users. The inclusion of non-users in the sample is particularly beneficial when attempting to identify and quantify latent demand for active modes. In such cases, the choice of sampling technique should resemble that of a general population survey, to ensure that both users and non-users have the opportunity to be selected.

4.2.2 Choice of Survey Instrument

The choice of survey instrument should be guided by the goals of the survey, the data to be collected, and the characteristics of the target population. In the case of active mode users, trips tend to take place over a relatively short distance, meaning that they are more likely to be neglected and consequently underreported (Son, Khattak and Chen, et al. 2012), introducing several considerations into the decision-making process. The underreporting of trips made by active modes can be partially attributed to the lack of well-defined guidelines for the definition of walking and cycling trips, the tendency to omit instructions to include short trips, the inability to process multi-modal trips (Edwards, et al. 2012), incomplete recall, survey length, memory decay, and a failure to understand or follow survey instructions (Son, Khattak and Chen, et al. 2012). These issues can be divided into two categories: those that can and those who cannot be addressed through the choice of survey instrument.

Issues pertaining to the lack of instructions and the failure to follow instructions can be addressed through the manner in which questions and instructions are worded, which should cater to the characteristics and expectations of the respondents. The effects of memory decay can be addressed either through the design of the survey, or the choice of the survey instrument. While household travel surveys typically assign a specific date for which travel must be reported by respondents, Son, Khattak and Chen, et al. (2012) found that allowing respondents to report their travel for the previous day had a positive impact on the number of complete responses that were received. Issues stemming from the length of the survey can be addressed by utilizing a survey instrument that allows for the integration of branching (or “skip logic”), found in CATI or web-surveys. Finally, incomplete recall can be addressed through the choice of a survey mode that facilitates the use of prompts to jog the memories of respondents or to verify the validity of responses, such as with CATI or web-based surveys. The *Future Mobility Survey* (FMS), a smartphone-based travel diary conducted using GPS data, presented maps and visuals to respondents in order to assist them with the process of validating the entries of their travel diary (Cottrill, et al. 2013). The design of the survey instrument should also account for the range of distances over which trips take place. For example, 70% of active mode trips reported in the 2001 NHTS took place over a distance of 3 mi (4.8 km) or less (Edwards, et al. 2012). The relatively short distances over which these trips take place may require the ability to prompt respondents for additional information, to ensure that these trips are not overlooked. A pilot test should be conducted prior to the conduct of the full survey, in order to identify issues with the design of the survey instrument and the questionnaire and to modify the survey instrument to address these issues.

The survey instrument must also account for the contexts in which the members of the sample are invited to and expected to complete the survey. For example, because the surveyors recruited participants through an intercept survey, the collection of basic information such as the volume and direction of cyclists was left to the surveyors. On the other hand, the more detailed questionnaire was distributed together with an envelope, which allowed its recipients to complete the survey at their leisure before returning it. In their survey of students at the University of Calgary, Twaddle, Hall and Bracic (2010) provided respondents with a chance to win a $150 gift card to a local bike shop, in part due to the manner in which respondents were recruited. While survey invitations were sent via email and posted on a number of web pages, the advertisement of the survey using posters and business cards introduces an additional burden to potential respondents who were recruited in this manner. Thus, the manner in which survey participants are recruited, and the additional effort that may be required to access the survey, must be taken into account when choosing a survey instrument (and determining the survey’s incentive structure). In an attempt to reduce the burden to respondents, Twaddle, Hall and Bracic (2010) conducted the survey to take advantage of the distribution of survey invitations through email, and exploit the ability to include an online mapping tool as an alternative to describing their route in words.

4.2.3 Data Collected and Applications

A wide variety of data pertaining to active mode users can be collected; however, it is important to tailor the type of data collected to the goals of the survey and the capability that the surveyors are attempting to enable. The collection and applications of five types of data are presented in this section: pedestrian- and cyclist-specific data, multi-modal trip data, attitudinal and stated preference (SP) data, data pertaining to seasonality and weather effects, and passive data. For each of these types of data, the issues related to data collection are identified, the manner in which the data are collected is presented, and the applications of said data are discussed.

4.2.3.1 PEDESTRIAN DATA VS. CYCLIST DATA

The ability to collect pedestrian and cyclist data using the same data collection method depends on the data that the surveyors are aiming to collect, and by extension, the capability that the surveyors are aiming to enable. If a better understanding of demand is desired, a different approach for pedestrians and cyclists may be required. If intersection counts are desired, pedestrian and cyclist counts can be obtained using a single data collection procedure. If facility-level volumes are desired, two methods exist to simultaneously count both pedestrians and cyclists – video recording and manual counts. Passive methods of counting cyclists (such as screenline counts) tend to be inadequate, particularly when multiple persons cross the screenline simultaneously. The manual collection of cyclist and pedestrian volumes has increased in prominence in recent years. The U.S. has experienced a surge in the number of communities that are using volunteers to perform manual pedestrian and cyclist counts at intersections (Lowry, et al. 2016).

Despite this trend, there is a dearth of standardized guidelines that outline the manner in which manual counts should be conducted. In an attempt to address this issue, the National Bicycle and Pedestrian Documentation (NBPD) project, a collaboration of Alta Planning + Design and the ITE recommends that counts be conducted once per season, for two hours in the morning (7:00 to 9:00 AM) and afternoon (4:00 to 6:00 PM) on a Tuesday, Wednesday, or Thursday (Alta Planning + Design 2016). The manual collection of inherently introduces the potential for human error which can, in turn, produce erroneous results. In an attempt to quantify the effect of the human error, Lowry et al. (2016) conducted a pilot test in Moscow, Idaho which compared counts collected by volunteers to the "ground truth" (obtained by video recording). The authors found that the use of the "four-movement leaving" and "twelve-movement" techniques did not have a significant impact on the accuracy of the counts, nor did the requirement for volunteers to record up to seven additional pieces of information about each traveller.

When both cyclists and pedestrians are included in the same sample, the design of the survey should account for differences in the demographic characteristics of the two populations. A key detail to be mindful of is the fact that the likelihood of a person being a cyclist decreases after the age of 40 (Sears, et al. 2012). This may result in differences in technical aptitude and other key characteristics between pedestrians and cyclists. The nature of this discrepancy should be explored, as it may necessitate the use of different survey instruments to collect data from pedestrians and cyclists. When attempting to identify physical barriers, motivators for and deterrents to the use of active modes, surveys are typically administered to members of the general population, with either a sole focus on cycling or a focus on both cycling and walking. Ultimately, the decision of whether to collect pedestrian and cyclist data using the same methods should be based on the data requirements and the characteristics of the target population.

4.2.3.2 ACCOUNTING FOR MULTI-MODAL TRIPS

At the most basic level, the ability to account for multi-modal trips or any nature is facilitated by the definition of a "trip" that is presented to respondents. Some studies have defined the concept of a "trip" as a means of going from one location to the next; however, this definition is still somewhat ambiguous and may lead respondents to only report their primary mode. The inclusion of a minimum-distance threshold is also unlikely to address these issues, in part because estimates of trip length are often inaccurate (Edwards, et al. 2012). The majority of the literature focuses solely on trips made by active modes, with Sears, et al. (2012) finding that only 2% of reported bicycle commuting trips made by their panel members over a two-year period were multimodal. In addition to the need to recall multi-modal trips, the manner in which information on multi-modal trips is collected is another issue that must be addressed. This is of particular concern when recording information on transit trips, where the need to report the transit access and egress modes should be emphasized.

The desired applications of the dataset should inform the extent to which data regarding multi-modal travel with active modes are collected. Transit may be the simplest multi-modal trip about which data can be collected, as household travel surveys such as the TTS typically ask respondents about their access and egress modes when transit trips are reported. If basic information regarding the use of multiple modes to complete a particular trip is desired, prompts can be used to determine whether a given trip was multi-modal within an existing location- or activity-based framework. In such cases, the ability to collect information on multi-modal trips can be enabled through the design of the questionnaire.

The desire to gain more detailed information on multi-modal trips poses a more complex challenge, which must be addressed on multiple fronts. The definition of a trip may have to be modified to include the consideration of the mode of travel, framing trips as a means of going from one *location* to another, but not necessarily from one *activity* to another. Given the often-inaccurate estimates of trip lengths, the desire to collect detailed data on multi-modal trips lends itself to the use of CATI software or a web-based survey, as these survey instruments are able to geocode locations and validate said data to ensure that it meets relevant standards or criteria (e.g. the location is within the survey area). The desire to complete a multi-modal trip may vary by gender, as a survey of students at the University of Calgary revealed that women reported a desire for bicycles to be allowed on LRT vehicles significantly more frequently than men (Winters, et al. 2011).

4.2.3.3 STATED-PREFERENCE AND ATTITUDINAL DATA

Stated-preference data and attitudinal data are typically collected when the surveyors are interested in gaining insights into the perceptions of respondents. SP data are typically used to determine the factors that affect the preferences of the respondents, based on the selection of an alternative from a set of hypothetical options (Habib 2014). A common use of SP data is to identify the factors that affect the choice of facilities that cyclists use. Majumdar and Mitra (2017) took a willingness-to-pay approach when attempting to quantify cyclists' perceptions of key attributes related to their choice of route. Conducted in both Kharagpur and Asansol, the survey collected trip and socioeconomic characteristics and included stated preference questions that focused on attributes associated with cycling trips. Using the SP data that were collected, the authors found that the perceived level of risk had the primary consideration when choosing a route (Majumdar and Mitra 2017).

Tilahun, Levinson and Krizek (2007) utilized a web-based adaptive stated preference (ASP) survey when attempting to evaluate individual preferences for different cycling environments. Administered to the staff at the University of Minnesota, each respondent was presented with a choice between two different facilities to utilize for a commute to work, with the travel time associated with the use of each facility being presented. The use of the web-based ASP allowed the authors to incorporate two unique features in their study:

- The use of an iterative process to identify the "tipping point" after which the choice of facilities changes, and
- The ability to present video clips that are reflective of the seasonal conditions

The results of the ASP survey were used to gain insights into the respondents' preferences for facilities and the factors that affected facility choice. The authors found that the improvement of bicycle lanes had the greatest impact on cycling and that designated bicycle lanes were the most desired piece of cycling infrastructure. The results also indicated that women tend to choose facilities that they perceive to be safer and that respondents from households with more than two members are less likely to choose a safer facility, which may be indicative of tighter time constraints (Tilahun, Levinson and Krizek 2007).

Attitudinal data has predominantly been used to gain insights into the factors affecting the decision to cycle, including identifying barriers and motivators and developing a better understanding of the impact of weather. Surveys of this nature tend to take two forms: those that ask respondents to identify motivators and barriers, and those that ask respondents to indicate the impact that a set of motivators and barriers has on their behavior. In both cases, it is important to ensure that the set of response options is both mutually-exclusive and mutually exhaustive while ensuring that the number of response options does not increase the burden experienced by respondents. Although the collection of information through textboxes is also an option, the need to parse through individual responses can prove to be a cumbersome task, particularly if the responses must also be categorized.

The survey conducted at the University of Maryland by Akar and Clifton (2009) is an example of a study that asked respondents to identify barriers and motivators to cycling. In an attempt to increase the bicycle mode share for trips made to the University of Maryland, the authors administered a web-based survey to staff, students, and faculty. Respondents were asked to indicate motivators and barriers to cycling and collected information on the frequency with which commuting trips were made using a given mode and the associated travel time, socioeconomic and residential characteristic information, and how far respondents lived from campus. The data were used to identify potential methods to encourage cycling (Akar and Clifton 2009). Similarly, the *barriers survey*, a component of the *Utah Travel Study*, asked both cyclists and pedestrians to identify barriers to cycling, in addition to their perceptions of existing walking and cycling infrastructure, the degree to which they agreed with various statements pertaining to walking and cycling, their ranking of a set of transportation planning priorities, and their preferences for certain types of cycling infrastructure (Research Systems Group, Inc., 2013).

Conversely, the *Cycling in Cities* survey conducted by the University of British Columbia represents a survey that focused solely on soliciting the attitudes of respondents to a number of barriers and motivators. The questionnaire that was administered to current and potential cyclists consisted of 73 items, which included a broad array of potential motivators and deterrents, which were identified from other surveys and qualitative and quantitative research, and chosen in conjunction with local policy-makers and advocates (Winters, et al. 2011). In an attempt to identify whether the results of the 2006 *Downtown commuter Cyclist Survey* in Calgary, i.e. that 75% of cyclists who commuted downtown were male, also applied to trips made to the University of Calgary, the questionnaire included a 15-item list of possible on-route and trip-end improvements. Respondents who were categorized as "possible cyclists" were asked to rank their top three on-route and trip-end improvements from the 15-item list that was presented to them (Twaddle, Hall and Bracic 2010).

4.2.3.4 ADDRESSING SEASONALITY AND WEATHER EFFECTS

There is a general consensus in the literature that temperature, precipitation, and the time of year have an impact on the decision to cycle. Temperature, precipitation, humidity, and wind are typically associated with the decision to cycle, while the impact of the season is typically more prominent in areas with a less-temperate climate, and the effect of temperature tends to be non-linear in nature (Nosal, Miranda-Moreno and Krstulic 2014). There appears to be a dichotomy in terms of the degree to which weather affects cyclists, as regular cyclists tend to be less sensitive to variations in weather and precipitation than occasional cyclists (Godefroy and Morency 2012). Existing impediments to the ability to quantify the impacts of weather on cycling volumes include the lack of disaggregate hourly data for bicycle flows over long periods of time, the complex relationship between weather and the decision to bike, the impact of perception and recent experience (typically over the past six days), and the inability to quantify the impacts of weather on cycling flows before, during, and after inclement weather (Miranda-Moreno and Nosal 2011).

The manner in which effects of seasonality and weather are addressed varies based on the goal of the survey. There have been a number of different approaches to better understand the effect of weather conditions and seasonality on cycling. Typically, information regarding weather conditions and the time of the year are combined with counts, as well as travel and attitudinal data. Additionally, the effects of weather and seasonality on cycling data have been addressed through sampling techniques and the calculation of weighting and adjustment factors. A common approach to studying the relationship between cyclist volumes, weather conditions, and seasonality is the use of count data, coupled with meteorological data, to estimate models. A prerequisite to the ability to account for daily, monthly, and seasonal variations is the continuous collection of bicycle and weather data (Miranda-Moreno and Nosal 2011).

A more conventional approach to using counts to identify the effects of weather conditions is described by Miranda-Moreno and Nosal (2011). They attempted to examine and empirically quantify the effects of weather on bicycle ridership, and to investigate whether cycling volumes increased in Montreal when controlling for temporal and weather effects. Hourly count data collected from five permanent counters were combined with the corresponding weather data and used to estimate an absolute ridership model, which related volumes to temporal and weather variables, and a relative ridership model, that related deviations from average cycling volumes to hourly deviations in weather conditions. The results of the modelling exercise were used to identify the factors that influence absolute and relative ridership. The authors found that temperature and humidity had a significant impact on absolute ridership, as did the presence of precipitation in the past three hours, while temperature, humidity, and precipitation had a significant impact on relative ridership. The results also implied that the effects of temperature increases vary on a season-to-season basis. Temperature increases above monthly average temperature had a positive impact on volumes in colder months, but an adverse impact in the warmer months (Miranda-Moreno and Nosal 2011).

The effects of seasonality and weather on the collection of data pertaining to active mode users can also be addressed through several approaches. In their panel study of the influence of weather on the choice to commute by bicycle, Sears et al. (2012) addressed seasonality effects by surveying panel members once per season. Panel members were required to commute at least two miles each way, and to have commuted by bicycle at least more than twice annually. Each panel member was given a unique identifier to facilitate the collection of longitudinal data over a two-year period, in which travel for a full week was reported once every season. The longitudinal travel data was combined with average temperatures, wind speeds, and the total amount of precipitation, collected from weather stations serving the communities from which the panel members were recruited. The panel data were used to, among other things, identify differences in the propensity of male and female respondents to cycle in the summer and winter (Sears, et al. 2012). The *Cycling in Cities* survey took a similar approach to data collection, by conducting the survey in three seasonal waves (Feb-Apr, May-July, and Sept-Dec), in an attempt to prevent bias due to seasonality effects (Winters, et al. 2011).

Attempts to address seasonality and weather effects have also been made through the development of weights and adjustment factors. In the Veloland Schweiz survey of the Swiss cycling network, seasonality and weather were accounted for through the computation of the data expansion factors. In addition to factors that accounted for response and acceptance rate, factors that related the maximum temperature and daily rainfall to the corresponding annual average factor were utilized to derive data expansion factors (Richardson 2006). There is also a large volume of research pertaining to the use of bicycle and pedestrian count data to develop temporal and seasonal adjustment factors that can be used to adjust short-duration counts into annual volumes (Lowry, et al. 2016).

4.2.3.5 APPLICATIONS OF PASSIVE DATA

The two most frequently used types of passive data are passive counts and GPS data. Passive counts have predominantly been used to study the effects of weather and seasonality on cycling volumes, while GPS data have typically been used to study the factors that influence the route choices of cyclists. In practice, a common approach to account for the effects of weather when estimating average annual daily bicycle traffic (AADBT) at a given site has been to extrapolate the values of short-term counts, taken at said site, using the values of long-term counts taken at other locations (Nosal, Miranda-Moreno and Krstulic 2014). This approach has been shown to produce inaccurate results, as it fails to account for the sensitivity of cyclists to weather, special events, or other exogenous factors. In response to this inadequacy, Nosal, Miranda-Moreno and Krstulic (2014) used long-term counts collected at stations in Ottawa and Montreal to evaluate the accuracy of four estimation methods:

- **The Traditional Method:** expansion factors for each month and day of the week are computer for an entire year
- **The Day-by-Month Method:** expansion factors for each day of the week are computed separately for each month
- **The Weather Model Method:** a model relating deviations from average cyclist counts are related to deviations from average weather conditions are used to adjust short-term counts
- **The Disaggregate Factor Method:** expansion factors are computed for each data using raw daily counts and the annual daily average

Each long-term counter was randomly assigned to represent either a short-term or long-term counter, with each of the four methods being used to estimate the AADBT at sites designated as "short-term". Using the data collected between April to November, the accuracy of the estimates of AADBT were evaluated on the basis of average absolute percent error (APE). With the exception of a single site, the disaggregate method produced the lowest values of AAPE, followed by the weather control method. The authors also found that increasing the duration of short-term data collected produces more accurate results, with the benefit peaking at 20 days of data collection (Nosal, Miranda-Moreno and Krstulic 2014). Within the core-satellite survey paradigm, passive count data is categorized as a complementary dataset which can supplement the data collected through the core or satellite surveys. One example of this supplementation is the use of AADBT estimates to derive data expansion factors.

The use of GPS data to identify and understand the factors influencing the route choice process has generally been limited to the study of the behaviour of utilitarian cyclists. This has been in part due to their behaviour being more easily explained using traditional econometric approaches. Using GPS data corresponding to a total of 724 bicycle trips, Cassello and Usyukov (2014) attempted to gain insights into the route choice decisions of cyclists. For each OD pair, a shortest-path algorithm was used to identify alternative routes using GIS software, each with different characteristics. This information was used to develop a route choice model, which considered the length of each link, the posted speed limit, auto volumes, gradient, and the presence or absence of cycling facilities. Using this model, "trips for which investments in cycling facilities along the shortest path may produce the greatest return on investment" were identified, based on a comparison of the chosen path and the most likely path (Cassello and Usyukov 2014).

The increasing availability and prevalence of third-party data from companies such as Strava, MapMyRide, and CycleMaps provide planners and engineers with the opportunity to study the travel behaviour and route choice of cyclists (Khatri, et al. 2016). An emerging research area in the use of GPS data stems from the availability and use of data from bike share users, particularly from bikeshare systems that do not require customers to return bikes to designated stations. In the US, there has been a 46% increase in the number of commuting trips made by bicycle from 2005 and 2013, due in part to growing public concern about lack of physical activity, increased auto dependency, and congestion. Khatri, et al. (2016) utilized GPS data from the users of the Grid Bikeshare system in Phoenix, AZ to study the route choices of bikeshare customers making utilitarian trips. Using data corresponding to 9,101 utilitarian cycling trips, route choice models were derived for casual and registered bike share users. Taking a number of traffic and facility-specific attributes into account, the authors identified the factors that affected types of facilities that each type of cyclist chose to use (Khatri, et al. 2016).

The work of Broach and Dill (2016), who used GPS data to study the behaviour of active mode users is relatively unique, as GPS-based studies tend to focus on the behaviour of cyclists. Regional travel models to tend inadequately represent active mode choice, due widespread use of the four-stage model and the definition of level-of-service attributes at the zonal level. The definition of travel with respect to zonal centroids is particularly problematic for trips made by active modes, as it can lead to bicycle and pedestrian facilities being overlooked. Broach and Dill (2016) used GPS data collected from 330 households in Portland from 2010 to 2013 to infer the effect of the built environment on the propensity to walk and bike.

To a certain extent, GPS data can act as the revealed-preference (RP) equivalent to attitudinal and stated-preference surveys when attempting to identify potential motivators and deterrents. Compared to attitudinal and SP data, GPS data provides a reflection of the choices that a respondent has made, rather than the choice that a respondent makes when presented with a hypothetical scenario. In this sense, GPS data can complement the data obtained through attitudinal and SP surveys and can be used to gain insights into the behaviour of active mode users. One means of exploiting this relationship is the combination of facility choice information from GPS and SP data, akin to the classic case of the fusion of RP and SP data. Similar to the work of Broach and Dill (2016), GPS data can also be used to evaluate the impact of the built environment on the decision to use active modes (or not to use active modes).

4.2.4 Ensuring Compatibility with the TTS

Designing a satellite survey to follow the spatial context of the TTS is relatively straightforward, as the survey area of the TTS represents a geographic boundary of the survey area of the satellite survey. The key challenge is ensuring that the definition of trips is compatible to that of the TTS, which poses a challenge for multi-modal trips made using non-motorized modes. In the TTS, trips are defined based on their start and end times, trip purpose, origin and destination purpose, trip origin and destination coordinates, and the *primary* mode of travel (U of T Data Management Group 2013). If the non-motorized portion of the trip is not considered to be the primary mode, the possible values of the trip purpose and destination purpose variables are not equipped to handle the change of mode as a purpose. One means of addressing this compatibility issue is to consider trips that were made either predominantly or solely using active modes when combining satellite data with TTS data. This may, however, be inadequate if one of the goals of the survey is to gain insights into the multimodal travel behaviour of active mode users. It may also require that the sample size of the satellite survey be large enough in order to draw statistically-valid inferences from the resulting dataset. With regards to the temporal context, it is important to ensure that the choice set of pedestrian and cycling facilities is relatively similar for respondents to the core and satellite surveys. Any significant changes to cycling or pedestrian facilities between the conduct of the two surveys may result in the two sets of data referring to different built environments.

4.3 Post-Secondary Student Surveys

The key motivation for conducting post-secondary student travel surveys is to better understand the effect that the travel behaviour of students has in the areas surrounding post-secondary institutions. Despite the fact that universities (and colleges) can have a significant impact on travel demand in a region, post-secondary staff and students tend to be underrepresented in traditional household travel surveys (Garikapati, et al. 2016), making it difficult to quantify and understand their impact on the travel demand of surrounding areas. Traditional household surveys have done an increasingly poor job of obtaining a representative sample of its target population, in part because of the significant extent to which younger members of the population are underrepresented. This underrepresentation is the result of a number of factors, such as the use of lists of residential phone numbers as a sampling frame (Verreault and Morency 2016). To a certain extent, the underrepresentation of younger members of the population stems from their characteristics and attitudes. Some have argued that the underrepresentation of students stems from a lack of engagement of students in civic processes. The propensity of students to ignore surveys creates difficulties keeping sampling frames up-to-date when the students are relatively transient (Volosin, et al. 2013), and students reside in dormitories (Khattak, et al. 2011).

The underrepresentation of post-secondary students (along with staff and faculty) in travel surveys manifests itself in a discrepancy in the manner in which students are represented in regional travel demand models and their actual behaviour (Wang, Khattak and Son 2012). This trend of underrepresentation also creates problems for MPOs when attempting to incorporate the use of a university campus sub-model into their existing modelling frameworks, due to a lack of both data and frameworks that are able to adequately represent student behaviour (Garikapati, et al. 2016). A common assumption in transportation research is to treat university students as a member of the general population, or as a low-income one-person household, despite the fact that this is often not the case in reality. Students have a set of mandatory trips and access to subsidized services (Huegy, et al. 2014). University campuses also often represent a unique set of land uses, as they tend to be more livable, better-able to facilitate the use of alternative modes, have higher densities than other environments, and offer mixed travel modes (Khattak, et al. 2011). In particular, trip distances, activity durations, travel mode choice, and the temporal distribution of trip making behaviour can vary widely between low-income households and university students, due to the impact of the land use characteristics of the campus on student travel behaviour (Huegy, et al. 2014). The conduct of a survey that specifically targets post-secondary students can help to mitigate the impact of a lack of data on the development of university campus sub-models in regional travel demand models, and can be used to gain insights into the travel behaviour of students.

4.3.1 Sampling Techniques

It is important to be mindful of the fact that post-secondary student surveys have two possible sampling units: institutions and students. Post-secondary student surveys perhaps have access to one of the most comprehensive sampling frames of any type of survey, as enrollment in a post-secondary institution requires students to provide personal information in order to complete the registration process. The increasing prevalence of internet access, particularly on the campuses of post-secondary institutions, make email an effective and efficient means of reaching the entirety of a student population, particularly when students are required to maintain an email address within the school's domain. Furthermore, the availability of student information, such as student status (e.g. graduate vs undergraduate, part-time vs full-time), allows for the organization of students into strata. The use of email list servers to invite students to participate in travel surveys is a fairly common sampling technique (Akar and Clifton 2009, Verreault and Morency 2016), and is somewhat able to mitigate the impact of the transient nature of student populations on their (under)representation in travel surveys. When a stratified random sampling methodology is applied, strata tend to be defined based on student status (undergraduate vs. graduate) and housing status (on-campus vs. off-campus) (Son, Khattak and Chen, et al. 2012); the sampling of undergraduate students is a key consideration, as the literature indicates a tendency for this type of student to display lower response rates (Khattak, et al. 2011).

When studying the travel patterns of students in North Carolina, Huegy et al. (2014) invited post-secondary institutions to participate in the study based on three criteria:

- **The Influence of the Campus on the Community**, defined as the campus population relative to the population of within a 50-minute drive time of the campus;
- **The Working Population**, based on the proportion of part-time undergraduate students; and
- **The Location of the School**, to ensure that each of the four regions into which the state was divided were covered

A somewhat unconventional sampling methodology was employed by Miralles-Guasch and Domene (2010) at the Autonomous University of Barcelona (UAB). In an attempt to study the effects of the recently-implemented UAB mobility plan, which strove to introduce sustainability as a primary consideration in mobility management (among other things), a stratified sampling method was used to choose locations on campus at which intercept surveys would be conducted. The locations were chosen based on the stratification of locations based on their ability to capture respondents belonging to particular community groups, the type of work-study centre, and the distribution of the genders of students, to ensure that members from all faculties were chosen (Miralles-Guasch and Domene 2010).

4.3.2 Choice of Survey Instrument

The availability of email list servers as a sampling frame, combined with the propensity for post-secondary students to be relatively comfortable using web browsers, tends to lead the majority of post-secondary student surveys to be conducted using web-based tools. The use of web-based surveys tends to reduce the time and costs associated with the administration of the survey (Akar and Clifton 2009). The preference to conduct such surveys using web-based tools can, in part, be attributed to the opportunity to exploit existing technology and infrastructure in order to reduce burden, improve usability, and to ensure that the survey is accessibility-compliant (Volosin, et al. 2013). The consideration of response burden has a significant impact on the tendency to conduct post-secondary student surveys using web-based tools.

Beginning with the recruitment of participants, the manner in which students are recruited should be taken into account when choosing the survey instrument, as the need to use a specific web address to access the survey has the potential to increase response burden. When students are recruited through invitation emails or advertisements on web pages, the inclusion of hyperlinks improves the ease with which potential respondents can access the survey (Akar and Clifton 2009). Conversely, recruitment methods that require potential respondents to type in the web address of the survey manually, or require the capture of a QR code will likely have an adverse impact on response burden, and by extension, response rates.

Huegy et al. (2014) argue that voluntary surveys should strive to provide a questionnaire that is simple and easy to complete, as reduced response burden, along with the provision in incentives and good advertising, tend to result in good response rates. The use of web-based surveys is conducive to the reduction of the burden associated with completing the questionnaire, as it allows for the integration of map-based interfaces, which tend to be less burdensome than typing in an address. Response burden should be considered throughout the design of a survey, as it can affect both item (question) and unit (survey) no-response. Excessive response burden can have an adverse impact on both response and completion rates, which has the potential to induce both measurement and non-response errors in a dataset.

In their study of the transfer of the conduct of the NHTS from a telephone-based survey to a web-based survey, carried out using students attending two universities in Virginia, Son, Khattak and Chen, et al. (2012) describe the measures that were taken to improve response and completion rates between the first and second waves of their survey. Called the University Student Travel Survey (USTS), the study was carried out in two waves: first in 2009, then in 2010. The conduct of the second wave stemmed from the proportion of incomplete responses that were received during the first wave.

The redesign of the survey aimed to reduce response burden, reduce the number of incomplete responses, and to recall and report their trips. To achieve these goals, several sections of the survey were removed, and an emphasis was placed on gathering information that is important for travel demand modelling purposes. The survey interface was changed to be clearer, the travel diary was made more user-friendly. Internal logic was used to verify the number of trips that respondents reported was accurate, and students who made a single trip were asked to explain their lack of a return trip. A number of these changes were facilitated by the use of a web-based survey interface and resulted in a greater completion rate and higher reported trip rates at both schools. The authors argue that the survey should be short enough to ensure that the travel and socioeconomic variables that are required for travel demand modelling are obtained from respondents, as respondents tend to be willing to provide said information through web-based surveys, as long as the survey appears to have a focus on “key travel issues” (Son, Khattak and Chen, et al. 2012).

4.3.3 Data Collected and Applications

As post-secondary students have traditionally been underrepresented in household travel surveys, the most common use of post-secondary student travel survey data is to better understand the travel patterns and behaviour of these students. This application of data can be segmented into two categories: the use of survey data to understand the travel characteristics of students (e.g. modal splits, trip rates, trip purpose, travel times), and the use of these data to estimate travel demand and mode choice models.

An example of a post-secondary student survey is the study conducted by Miralles-Guasch and Domene (2010). Intercept surveys were conducted on the campus of the Autonomous University of Barcelona (UAB) to identify the impacts of the UAB’s mobility plan on student transportation. The survey team utilized a 19-item questionnaire to collect information on mode shares, travel times, mode choice motivations, perceptions of public transit, the respondent’s preferred mode, and barriers to the use of the preferred mode. The authors found that 54% of university community members would prefer to commute through different means, including 65% of student respondents (Miralles-Guasch and Domene 2010).

This type of survey has also been used to compare and contrast the travel behaviours of students with those of the general population in the area surrounding university campuses. One such study was conducted by Khattak et al. (2011) in four universities in Virginia. The web-based survey used by the authors included questions regarding personal characteristics, the commute to and from the university campus, and attitudes towards walking and cycling. A travel diary was also used, in order to compare the travel characteristics of students (e.g. mode choice, trip rates) to those of the general population. The authors found significant differences in travel behaviour of the two groups (Khattak, et al. 2011).

Travel survey data corresponding to post-secondary students are also used to estimate travel demand models, in order to more accurately represent their travel behaviour in regional travel demand models. Garikapati et al. (2016) utilized data collected through the Arizona State University *student travel survey*, in conjunction with their framework for modelling university student travel. Using these data, they developed a university sub-model to be implemented in the regional travel demand model for the Albuquerque metropolitan area. This sub-module distinguished between intra- and non-intracampus trips, and explicitly represented students, staff, and faculty members (Garikapati, et al. 2016).

In their attempt to model the association between the travel behaviour of university students and their personal characteristics, residential location (on- vs. off-campus), and academic status (full- vs. part-time), Wang, Khattak and Son (2012) conducted a web-based survey of students at Old Dominion University in Virginia. The student trip rate models estimated from these data formed the basis of the trip generation models used by the regional travel demand model for university-dominated traffic analysis zones. The data collected over the course of two years from university students in North Carolina by Huegy et al. (2014) were used to develop trip generation models for the home-based university, home-based other, and university-based non-home trips, as well as trip attraction models and mode choice models. These all served to support the development of travel demand models for communities that included university campuses (Huegy, et al. 2014).

The use of post-secondary student travel data to develop and to inform the development of components of travel demand models has a number of short- and long-term benefits. Post-secondary student travel data are of interest to MPOs, who use the data to understand and model the behaviour of this particular segment of the population (Garikapati, et al. 2016). The modelling of the behaviour of post-secondary students provides a means of gaining insights into the factors that affect various aspects of their travel characteristics (e.g. trip rates, mode choice, trip purposes), and creates the opportunity to forecast and analyze the potential impacts of policies geared specifically towards post-secondary students. The survey conducted by Akar and Clifton (2009) at the University of Maryland is an example of an attempt to understand student travel behaviour in order to identify potential methods to increase the proportion of trips made to campus by bike. The results of the survey were used to gain insights into the factors that affect the decision to cycle, which in turn can be used to develop policies that aim to increase the number of cycling trips made by students.

Post-secondary student data can also be used to directly address the underrepresentation of students in the dataset of a household travel survey. Verreault and Morency (2016) attempted to use the data collected through a travel survey of university students in the city of Sherbrooke to address this discrepancy. Post-secondary students in the Sherbrooke area were invited to participate in the survey through email lists maintained by participating schools; in addition to collecting data on the travel behaviour of respondents, the household structure (age and gender of the other members of each respondent's household) were also collected, to facilitate the fusion of data with the Sherbrooke household travel survey. The project strove to recruit two types of respondents: those who were included in the reference population (the 2011 census), but excluded from the sampling frame ["Sample 1B"], and respondents who were not included in the reference population and live in a dormitory ["Sample 2A"], with data from these respondents being integrated with data from the household travel survey. Data from respondents in Sample 1B were weighted based on household structure, home location, respondent age group, and gender, using data from the 2011 Canadian census. Data from respondents in Sample 2A were weighted based on the total number of students in each school, under the assumption that students living in dormitories participated in the survey in the same proportions as other students. The integration of the data from Samples 1B and 2A with the data from the Sherbrooke household travel survey resulted in the sampling rate for the 20-24 age group exceeding the 10% target, however, the 25-29 age group was still underrepresented. Overall, the integration of the three datasets had a positive effect on sampling coverage and data quality, with reduced weighting factors and increased overall and transit-specific trip rates being observed. (Verreault and Morency 2016)

4.3.4 Ensuring Compatibility with the TTS

As was the case with other types of potential satellite surveys, the issues associated with ensuring sufficient compatibility between the satellite survey data and TTS data varies on a context-to-context basis. Compatibility in the spatial context can be assured by selecting post-secondary institutions that are within the TTS survey area. With respect to the temporal context, conducting the satellite survey when classes are in the session can help ensure that the sample can provide an adequate representation of the student population. Compatibility in the semantic context can primarily be addressed by ensuring that the definitions of the trip and personal attributes are similar to those used in the TTS. The key challenge in ensuring the compatibility of the two datasets in this context is the difference in sampling units (individual versus household). Verreault and Morency (2016) attempted to address this issue by collecting information on the structure of the household of each respondent (defined as the age and gender of all other household members) in order to facilitate the fusion of post-secondary student survey data with household travel survey data.

4.4 Employee Surveys

Employee surveys (also referred to as workplace surveys) are, in essence, a specific type of special generator survey, which predominantly aim to collect data pertaining to trips made to workplaces and similar establishments. While there has not a large amount of research into the conduct of workplace surveys, Chapter 18 of the Travel Survey Manual[1] (Southwell, Zhang and Sharp 2014) provides a detailed set of guidelines regarding the conduct of workplace surveys, many of which are presented here.

4.4.1 Typical Methodology

The conduct of workplace surveys began in the mid-1980s, as planning agencies began to develop regional models. Prior to the collection of data, the surveyors typically obtain data on the employers in the study area and available information on commuting patterns. Southwell, Zhang and Sharp (2014) provide an excellent summary of the procedures applied when collecting data pertaining to employees and visitors separately:

- Call each establishment in the sample to determine if it is still in business, to verify the address, and to establish a contact
- Send a recruiting letter to each employer
- Interview and recruit the employer, and establish a contact person
- Schedule the survey day
- Schedule a personal site visit
- Remind the contact person at each business to deliver the employee questionnaire

[1] Published by the Transportation Research Board

4.4.2 Sampling Techniques

Due to the nature of workplace surveys, sampling can occur with up to two types of sampling units: establishments and individuals. The sampling of establishments typically takes a stratified random sampling approach, with strata being defined based on the area type or its location within the study area, the industry section of the establishment, the number of employees, and whether it is free-standing. Once establishments are selected, the sampling plan of workplace survey can have up to two target populations: employees and/ or visitors. Southwell, Zhang and Sharp (2014) emphasize that the conduct of intercept surveys necessitates the simultaneous collection of counts at each of the building's entrances, in order to facilitate the expansion of survey data in the post-processing stage. When workplace surveys are conducted as an intercept survey, rather than as a census, a common approach to the recruitment of participants is to interview every n^{th} visitor.

4.4.3 Recruitment Strategies

Similar to the selection of sampling units, the process of recruitment typically takes place in two stages – the recruitment of employers and the recruitment of employees. As non-response is a key consideration when attempting to improve the accuracy of the survey results, Southwell, Zhang and Sharp (2014) argue that this issue is best addressed in the recruitment phase of the survey. The authors identify a number of accuracy-enhancing measures, including the solicitation of support from the local Chamber of Commerce and/or business associations, contacting the most senior managers possible, and focusing on larger firms. When attempting to recruit an establishment, the survey team should draft a letter requesting permission from employers conduct the survey at their establishment; however, it should be sent by a third-party (such as the Chamber of Commerce or a business association) in order to increase the credibility of the invitation. Once permission has been obtained for a particular establishment, the survey team should also interview a knowledgeable staff member in order to devise a site plan, that outlines proposed cordon locations and potential exits. The survey team should also produce a cover letter to distribute to employees, that invites employees to participate in the survey, reiterates their employer's support for the survey, and provides contact information for a liaison at the company who can provide assistance if required. The most effective means of persuading employees to participate in the survey is to stress the support of the employer (Southwell, Zhang and Sharp 2014).

4.4.4 Choice of Survey Instrument

As with any survey, the choice of survey instrument should be guided by the goals of the survey and the characteristics of the target population. Establishment surveys tend to be carried out in one of three ways: visitor and employee intercept surveys with random selection, a centralized employee survey, or a combination of visitor intercept and centralized employee survey. When an intercept survey is administered as part of a workplace survey, the most commonly-used instruments are the personal distribution of self-administered survey forms, personal interviews using pen-and-paper methods, and the use of computer-assisted personal interview software (Southwell, Zhang and Sharp 2014). When conducting a centralized employee survey, Petrunoff et al. (2013) found that the use of a web-based survey to obtain travel information from the employees of an establishment produced statistically-similar results to those collected through a travel diary, with respect to the modes and travel times reported. Sperry et al. (2015) provides an excellent summary of the workplace survey that is conducted by the Texas DOT (TxDOT), as part of their *Travel Survey Program*.

4.4.5 Case Study: The TxDOT Travel Survey Program

In Texas, urban areas with a population of more than 50,000 people are required to establish MPOs, whose responsibilities include the development of travel demand models to support the development of travel plans. The TxDOT *Travel Survey Program* conducted once every ten years, consists of a household travel survey, a workplace survey, and a commercial vehicle survey, an external station survey, and a state-wide border crossing survey. The workplace survey has three main components:

- **Establishment Surveys:** collects basic information about each establishment, including the type of establishment, operating hours, and the number of employees
- **Traffic Counts:** conducted for either vehicles or persons entering and exiting the establishment, depending on whether the business is free-standing
- **Intercept Interview Surveys:** personal interviews are conducted with travellers at each establishment to collect information about trip characteristics (origin and destination, purpose of travel, modes, number of persons in travel party, and residential location)

The DOT used a list of businesses obtained from the Texas Employment Commission as the sampling frame for the survey, with establishments stratified based on the industry section (as defined by its North American Industry Classification System code) and urban area; a quota was established for each type of business in each urban area. Traffic counts and intercept surveys were conducted during business hours, plus one hour before opening time and one hour after closing time). The workplace interview data were expanded to match the observed traffic counts during each day of the survey; trip data were expanded to match the person and vehicle counts. The total number of trips were classified separately for employees and visitors at freestanding and non-freestanding establishments, and distinguished between the residents and non-residents of each area. (Sperry, et al. 2015)

4.4.6 Applications of Data

In general, the data collected through establishment surveys fall into one of three categories: establishment information, employee information, and visitor information, with establishment surveys themselves being composed of up to five components (Southwell, Zhang and Sharp 2014):

- The collection of employer information,
- Surveys of employees,
- Surveys of visitors,
- Person and vehicle counts, or
- Counts of delivery persons and/ or vehicles

The applications of workplace survey data include the study of traffic impacts, congestion management, and trip reduction programs. The most common application of workplace survey data is to determine trip attraction rates, which are used as inputs in travel demand models. As the focus of household travel surveys has shifted to providing travel model outputs at finer levels of spatial resolution, the trip attraction rates collected through traditional methods, such as household travel surveys, external vehicle surveys, and commercial vehicle surveys, have become increasingly inadequate (Southwell, Zhang and Sharp 2014). Sperry et al. (2015) used data collected through the Texas DOT's workplace travel survey over the past 13 years, containing 63,000 trips from 5,100 establishments in 15 urban areas, to devise trip attraction rates for home-based work, non-home-based work, and non-home-based trips, which were cross-classified based on area-type (CBD, CBD fringe, Urban, Suburban, Rural), establishment type (basic, service, retail, and education), and community size (small, medium, large, and metropolis). Count data are typically used to expand workplace survey data.

In the GTHA, the Smart Commute program is a collaboration between Metrolinx and municipalities in the GTHA that aims to assist in the identification of alternative to commuting by automobile. The services provided by Metrolinx through the Smart Commute program include:

- The administration of site assessments and surveys to understand the commuting behaviour of employees,
- The development of customized action plans that encourage employees to explore the use of alternative modes, and
- Providing tools to businesses to facilitate exclusive carpool ride-matching programs, emergency ride home programs, discounted transit pass programs, tele-network programs and flexible work arrangements, walking and cycling programs, and events and promotions

The program is meant to support travel demand management initiatives pertaining to land use policy, walking, and cycling through partnerships with stakeholders. The Smart Commute program provides resources to businesses looking to explore carpooling, cycling, transit, and walking options, in addition to telecommuting and flexible work arrangements. (Metrolinx 2014)

4.4.7 Ensuring Compatibility with the TTS

A key factor to consider when attempting to ensure sufficient compatibility is the difference in sampling units between employee surveys and the TTS. To a certain extent, this incompatibility between an individual travel survey (ITS) and household travel survey (HTS) can be addressed one of two ways. Employee surveys could require that the respondent also provide travel diaries for other members of their household, which could potentially act as an "update" to the TTS between survey years. Another means of addressing the ITS-HTS issue was discussed in Chapter 4.2.4. Collecting information pertaining to the household structure of each employee would help to facilitate linkages to TTS dataset through household attributes (e.g. household size, vehicle ownership, number of children, etc.).

Aside from the ITS-HTS issue, ensuring that the sampled workplaces are within the TTS study area is a relatively simple means of ensuring spatial compatibility with TTS data. In order to ensure semantic compatibility with the TTS, occupation definitions should be similar to those used by the TTS (either by being exactly the same or by having categories that are a subset of exactly one occupational category). Similar measures should be taken to ensure that the definition of trip purposes has a kind of compatibility with that of the TTS. Ensuring sufficient compatibility in the temporal context is less straightforward, mainly due to the influence of the home locations of employees. For other types of surveys, the addition or elimination of infrastructure between the conduct of the core and satellite surveys can have an adverse impact on the temporal compatibility of the two datasets. In the case of employee surveys, the inability to obtain the home locations of the respondents prior to the collection of data precludes the ability to understand the effects of infrastructure changes on the compatibility of the data collected through the satellite survey.

4.5 Attitudinal and Stated-Preference Surveys

The design of attitudinal and stated preference surveys should be informed by the overall goals of the survey. The power of these types of the survey is that they provide the ability to gain insights into the perceptions of respondents, as well as the ability to gain insights into the factors that affect the decision-making process of respondents. The choice of sampling frame and recruitment technique should follow from the goals of the survey, as the target population will be based on the goals that the surveyors are trying to achieve. If the survey is attempting to gain insights into the behaviour of the members of a specific sub-population, then a non-probability sampling method may be adequate. Conversely, if the survey is attempting to understand and begin to quantify latent demand, the sampling frame must include a broader range of potential respondents.

Following from the identification of the target population, the survey instrument and questionnaire should be designed with two key considerations in mind: the characteristics of the target population and response burden. While some survey modes are more conducive to the use of burden-reducing measures than others, the ability of the members of the target population to understand and effectively use the mode must also be considered. The need to navigate the learning curve associated with a survey mode may have a negative impact on response burden, depending on the aptitude of the respondents. Throughout the survey design process, the survey administrator must be aware that their decisions may represent an inherent set of trade-offs between usability, response burden, costs, and response rates. The design of a survey is more of an art than a science, and maintaining an awareness of these trade-offs can help to mitigate the adverse impact that they can have on respondents.

5 THE ROLE OF PASSIVE DATA

The increasing prevalence and availability of passive data, i.e. data that are collected without explicit input from subjects (Matsuda, et al. 1998), has created new opportunities to gain insights into the factors that influence travel behaviour, and to collect data that is not or cannot be collected through traditional travel survey methods. In this chapter, the advantages and issues associated with the use of five types of passive data are discussed, in order to prevalence: GPS data, cellular data, smart card data, Bluetooth data, and targeted marketing (TM) data. For each type of passive data, its place within the core-satellite survey paradigm is briefly discussed.

As the applications of some of these types of data have already been discussed, this chapter concludes with the presentation of other interesting types of applications.

5.1 GPS Data

GPS data are among the most commonly-used data in transportation research. With an abundance of different coordinate systems that can be used to project GPS data onto shapefiles of the road network and other transportation infrastructure using GIS software. The collection of GPS data typically involves the periodic recording of the latitude and longitude reading transmitted by a GPS sensor, along with a timestamp corresponding to the time at which the data were transmitted. Due to the costs associated with the distribution of GPS units, their use in data collection applications may be limited to satellite surveys. The increasing prevalence of smartphones helps to mitigate this issue, however the use of smartphones to collect data has to yield biased samples. This bias is the product of the learning curve associated with these methods, which tends to favour younger respondents, who tend to be more tech-savvy. At this point, the use of GPS data should be limited to satellite surveys or complementary datasets.

5.1.1 Advantages

The main advantage of using GPS data is the ability to collect information that is more detailed and accurate than the information that is reported by respondents. The ability of a GPS sensor to accurately collect information on routes, trip start and end times and trip lengths provides the ability to mitigate the tendency to neglect to report short and discretionary trips (Dumont, Shalaby and Roorda 2012). This ability to obtain detailed information pertaining to the routes used by respondents is particularly useful when studying the behaviour of active mode users, as their route choice decisions tend to less utilitarian than those of drivers. The use of GPS devices as a means of producing a travel diary also allows data to be collected over the course of multiple days, as the marginal burden associated with collecting data for additional days is relatively low (Dumont, Shalaby and Roorda 2012). The Future Mobility Survey (FMS) that was pilot tested in Singapore is an example of a smartphone-based travel survey that required respondents to validate the trip characteristics imputed from the GPS data; when comparing trip characteristics obtained via GPS data to those obtained through a traditional household travel survey, Zhao et al. (2015) found that:

- Travel times reported by respondents tend to be rounded to the nearest five or ten minutes

- People tend to report a simplified (typical) travel day
- Short activities are underreported
- Travel times are estimated for short trip
- There are significant variations in day-to-day activities, which cannot be captured in a one-day survey

The use of GPS data to construct a travel diary allows surveyors to exploit the fact that people tendency for people to be able to recognize their past activities and locations visited, while mitigating the impact of the inability to recall trip start and end times for more than the past few days (Cottrill, et al. 2013).

The use of GPS devices to obtain travel information is particularly useful when developing an activity-based model, as issues with traditional travel survey data, such as missing or miscoded information, the underreporting of trips, the systematic underreporting of data after the first day in the data collection period, and inaccurate travel time information, can be mitigated through the passive collection of location data (Simas Olivera, et al. 2011). The increasing availability of GPS data collected through third-party smartphone applications, such as Strava (Grover, et al. 2016), provides new opportunities to study the behaviour of active mode users, particularly given the ability to use GIS software to associate facility-level characteristics (such as vehicle volumes, gradients, speed limits, etc.) with routes.

5.1.2 Issues

There a number of issues with the collection of GPS data that affect its quality of the data collected and the reliability of the results obtained solely through the use of GPS data. GPS devices are plagued by the issues of cold starts, short-duration trips, and the canyoning effect in dense urban environments (Shen and Stopher 2014). The use of GPS technology has yet to address the difficulties associated with the collection of information on short-duration activities, travel companions, trip purpose when travelling to dense urban areas, fares paid, and parking information (Simas Olivera, et al. 2011). Costs and privacy are key considerations when using GPS devices to obtain travel information (Ge and Fukuda 2016).

A common issue with the use of GPS devices to produce a travel diary is the need to impute information, such as travel mode and trip purpose, based on data collected through other sensors, such as the accelerometer. The amount of uncertainty that this approach introduces necessitates the validation of the imputed information by the respondents, which has the potential to introduce additional burden. The two main types of errors in the data collected through Future Mobility Survey (FMS), a smartphone-based travel survey tool that was pilot tested in Singapore, were validation errors and data gaps (Zhao, et al. 2015), highlighting the additional potential for error that the inclusion of a validation stage introduces. In order to address this issue, as well as reduce response burden, the FMS utilized AI and machine learning techniques in an attempt to improve the accuracy of the imputed information (Cottrill, et al. 2013). Due to the passive nature of GPS data, additional information pertaining to respondent characteristics and travel behaviour must be solicited from users.

5.2 Cellular Data

The use of cellular data in transportation research is a means of exploiting the fact that cellular service providers already collect location data for billing purposes. These so-called call detail records (CDRs) are recorded every time cellular service is used and includes location information as well as a time stamp. While CDRs can provide spatio-temporal information, their applications within the core-satellite survey framework should be limited to complementary datasets. The inability to link records to respondents precludes the ability to identify the extent to which the data represents the population, and impairs the ability to link CDRs with respondents of the core and satellite survey.

5.2.1 Advantages

Call detail records are similar to GPS data, in that they provide a record of changes in location over a period of time, however, unlike GPS data, CDRs are not recorded periodically. Despite this issue, CDRs contain a wealth of spatiotemporal information, which can be used to gain insights into mobility patterns and complement the data collected through household travel surveys (Alexander, et al. 2015). Depending on the cellular service provider from which CDRs are obtained, this set of records has the potential to provide travel information from a larger proportion of the population than would be obtained via household travel survey, with locations that are generally more accurate than those reported by respondents to travel surveys.

5.2.2 Issues

While call detail records are a promising set of data from the perspective of transportation research, there several issues with using CDRs as the sole source of data for analysis. The key issue is that these records are collected by individual cellphone service providers, meaning that each set of CDRs can only provide information on a subset of cellphone users, who in turn are a subset of the population as a whole. Inherent to the manner in which CDRs are produced is the requirement that cellular service is used in order to for the location of the phone (and the associated timestamps), meaning that people who frequently use these services will be overrepresented in the dataset. Another issue associated with the use of that CDR data is the lack of associated demographic and trip information (Alexander, et al. 2015), and unlike GPS data, there are no means of obtaining this information. With respect to trip information, this means that a set of algorithms must be developed to infer trip origins and destination, travel modes and trip purposes (Ge and Fukuda 2016), likely in the absence of any means of validating the imputed information.

5.3 Smart Card Data

The number of transit agencies that have adopted the use of the smart fare card (“smart card”) is growing, as are smart card penetration rates (defined as the percentage of trips for which fares were paid using a smart card, rather than traditional fare media). With respect to smart card usage, the fare policies employed by agencies tend to fall into two categories: tap-on only (a customer taps their card on a reader when first boarding a vehicle), or tap-on and tap-off (a customer must tap their card on a reader when both boarding and alighting from a vehicle). The fare policy employed by an agency can affect the advantages and issues associated with using smart card data. Smart card data predominantly fall into the category of complementary data within the core-satellite framework, however, they also have the potential to be used as satellite survey data. The extent to which smart card data can be used as satellite survey data depends on the information that is provided by users when registering for a card, as well as the ability to obtain said information. User information is required to facilitate linkage to the core dataset, and in the case of the TTS, would require both personal and household attributes. Currently, the most practical usage of smart card data is to provide an independent record of transit trips, which can be used as reference values.

5.3.1 Advantages

One of the main advantages of using smart card data for transportation analysis is its ability to facilitate the use of the individual as the sampling unit, rather than the trip, which is common when conducting transit on-board surveys. One of the key benefits of using smart card data is the ability to obtain detailed spatiotemporal information on demand at for each transit route (Ji, Mishalani and McCord 2015). This information has a variety of applications, including the adjustment of service delivery and providing a means of precisely measuring demand, both at the route (Morency, Trepanier and Agard 2007) and passenger levels (Agard, Morency and Trepanier 2006).

5.3.2 Issues

The main issue with the use of smart card data is that low smart card penetration rates (i.e. the percentage of transit trip paid for using a smart card) generally produce results that are not representative of transit users as a whole. Also, as with GPS and cellular data, smart cards have the potential to produce a dataset that is devoid of necessary demographic and trip information. While the lack of demographic information can be addressed by requiring card owners to provide said information and to register their cards, this cannot account for instances where card owners are lent to others. The issue of missing trip information is somewhat more difficult to address, particularly when the agency does not require that customers tap-off when alighting from transit vehicles, as this requires that the alighting stop is imputed (Munizaga and Palma 2012). Because smart card data solely pertain to trips made by transit, a number of important pieces of information about each trip are potentially lost. Solely relying on smart card data means that information pertaining to the origin and destination of the trip are not available, nor is information pertaining to the mode(s) used to access the boarding stop from the trip origin and egress to the trip destination from the alighting stop. furthermore, the utility of smart card data can only fully be realized when an automatic vehicle location (AVL) system is in place. The presence of an AVL system is needed to associate the location of each tap-on transaction with the boarding stop of travellers (and the alighting stop if tap-offs are required). The lack of a single standard for the design of infrastructure that enables smart card applications is a relatively minor issue, however, it can introduce additional burden when analyzing data from different transit agencies (Pelletier, Trepanier and Morency 2011).

5.3.3 Applications of Smart Card Data

While the two most commonly-used sources of passive data are smart fare card (“smart card” data and data collected through automated passenger counter (APC) systems), the most fundamental type of passive data is collected using automated vehicle location (AVL) systems. The combination of APC and smart card data with AVL data allows count and transaction data to be contextualized, particularly through the association of APC and smart card data with a boarding and alighting stop.

A common use of smart card and APC data is the development of route-level OD matrices, although these methods typically use one of these types of data, rather than both. When using APC data, methodologies to develop route-level OD matrices typically apply an iterative proportional fitting (IPF)-based approach, in which a base or null OD matrix (the probability of a trip being made between any feasible OD pair is equally likely) is updated based on reference boarding and alighting counts that act as marginal totals (Mishalani, Ji and McCord 2011). Using route-level OD data collected through an on-board survey conducted on the campus of Ohio State University in 2009, Mishalani, Ji and McCord (2011) evaluated the effectiveness of their proposed methodology; APC data collected in 2010 was used to update the 2009 OD matrix, and compared the results to the matrix obtained through an on-board survey conducted in 2010 using the Hellinger distance. The authors found that the use of a base OD matrix produces better results than when a null matrix is used, and that the IPF method tends to produce estimates that are as good as or better than the estimates derived from surveys. Ideally, this process is carried out using both APC data and smartcard transaction data, however.

In the absence of APC data, stop-level smartcard transaction data can be used to develop route-level OD matrices. Using smart card transaction and GPS data from Santiago, Chile, Munizaga and Palma (2012) proposed a methodology to estimate a public transit OD matrix. Because transactions were only of the tap-on variety, the alighting locations of transit trips had to be imputed; the imputation algorithm selected a likely alighting stop based on a generalized cost function with a maximum walking distance threshold set to 400 m. In total, 80% of alighting locations were estimated; the inability to identify the remaining alighting locations were attributed to alternative modes being used to address the first-mile-last-mile problem, and errors in the data. The authors found that transactions at bus stops and metro stations were successfully identified at higher rates than transactions at bus stations (Munizaga and Palma 2012).

Given their number, there is potential for assumptions associated with the imputation of alighting stop locations to affect the imputation results. Alsger et al. (2015) used data from TransLink in Brisbane to quantify the effects of three commonly-made assumptions: the transfer time threshold (15 to 90 minutes), the size of buffer zones (acceptable walking distances; tend to vary from 400 to 1,000 m), and the assumption that a person's first boarding location is also their last alighting destination. This analysis was performed using smart card data obtained from TransLink for a single weekday, comprised of 628,479 transactions form 260,803 card holders. The authors found that the transfer time threshold did not have a significant effect on the estimated volumes between OD pairs (based on the Geoffrey E. Havers statistic), the size of the buffer zone had a minimal effect on OD flows (50% of users walked less than 100 m to transfer, and 85% walked less than 650 m), and that of the 109,652 OD pairs that included at least two legs, 82% of passengers returned to their original boarding stops (Alsger, et al. 2015).

5.4 Bluetooth

5.4.1 Advantages

The key benefit to the use of Bluetooth sensors to collect data is the potential to obtain the travel characteristics of a human-populated network, based on the existence of a globally-unique media access control (MAC) number (Malinovskiy, Saunier and Wang 2012). This allows the start and end points of travel along a corridor or between cordons or screenlines to be identified, and the travel time between said points to be imputed. The ubiquity of Bluetooth-equipped devices in consumer electronics and the in-car system has created an opportunity to use data collected through Bluetooth sensor systems to measure vehicular volumes, densities, and flows, in addition to facilitating longitudinal transportation analysis (Friesen and McLeod 2015). Due to the inability to link Bluetooth data to socio-demographic attributes, they can only be used as a complementary dataset within the core-satellite survey paradigm.

5.4.2 Issues

The use of Bluetooth data is rarely seen in transportation research, mainly due to the small sample sizes that are obtained through data collection efforts. Malinovskiy, Saunier and Wang (2012) found that the Bluetooth counters that had been set up at the two ends of a pedestrian corridor only captured 5% of pedestrians using the corridor at one site and 2.25% of pedestrians at another, when studying the potential for Bluetooth sensor data to be used in pedestrian studies. Another key issue regarding Bluetooth data is that devices must be set to be discoverable in order to be tracked by Bluetooth sensors, making them more conducive to the collection of data related to automobile trips (Friesen and McLeod 2015).

5.5 Targeted Marketing Data

Targeted marketing (TM) data is a relatively new set of data that has recently been used in transportation research. Kressner and Garrow (2014) provide an excellent overview of the use of TM data for transportation applications in their paper "*Using Third-Party Data for Travel Demand Modeling – Comparison of Targeted Marketing, Census, and Household Travel Survey Data*", in which they investigate the use of TM data as an up-to-date source of socioeconomic information. TM firms collect a variety of different data, including household demographics, individual demographics, housing and property data, aggregated automotive data, and lifestyle clusters.

5.5.1 Advantages

One of the key advantages of TM data is that they allow MPOs to collect detailed and up-to-date data more frequently than through traditional survey methods, at a fraction of the cost. other advantages of using TM data include the fact that they are available at a more disaggregate level than most census data, updated regularly, are considerably cheaper to collect than tradition data, exist within a national database, contain more information than census data, and can be used in longitudinal studies.

5.5.2 Issues

The issues associated with the use of TM data include the reliance on proprietary algorithms for certain kinds of data, particularly when data must be imputed or clusters must be formed the potential for firms to go out of business, and the inability of TM data to provide a perfect representation of the target population.

5.6 Interesting Applications

5.6.1 The Future Mobility Survey

The Future Mobility Survey (FMS) is a smartphone-based travel survey that was pilot-tested in Singapore, in conjunction with the conduct of the *Household Interview Travel Survey* (HITS) conducted by the Singapore Land Transportation Authority. The recruitment of individuals to participate in the pilot test of the FMS piggybacked off of the HITS and required that participants both owned a smartphone and had some familiarity with web browsers (Zhao, et al. 2015). The FMS collected data in four stages (Cottrill, et al. 2013):

- **Registration**, where basic household demographic information and contact information were obtained,
- **Pre-survey**, where socioeconomic and demographic information were collected,
- Activity diaries, where respondents had to log onto the FMS website in order to validate their trips, and
- **Exit survey**, where additional household information was collected, and respondents were asked to provide feedback

The FMS system had three components: the smartphone app (through which data were collected), the server (which housed the database and where the data processing and machine learning algorithms are applied), and the web interface (where users validated data) (Zhao, et al. 2015). A travel diary was constructed and presented to respondents, where they were asked to correct erroneous information and fill in missing information, with machine-learning algorithms and AI being used to improve the accuracy of the imputed information, and ultimately reduce response burden. The study prioritized the collection of activity, modal, and location data over route information, as it could be imputed from probabilistic map matching or route-planning algorithms after the fact. As a means of reducing the impact of data collection on the battery life of the smartphone, a phased sampling approach was taken, and this constraint helped guide the decision to make the app non-intrusive. The authors found that once the app had been installed, users had the tendency to forget about it, requiring explicit reminders to conduct further interaction. (Cottrill, et al. 2013)

For participants who had validated the travel diary for a sufficient number of days (7.3 on average), the data collected through the FMS and collected through the HITS were compared. Zhao et al. (2015) found that travel times reported during the HITS tend to be rounded to the nearest five or ten minutes, people tend to report their typical travel day, short activities are underreported, travel times are estimated for short trips, and people tend to have large variations in their daily activities. Compared to the data collected though the FMS, the travel times reported in the HITS tends to display a narrower distribution, and significant proportions in-vehicle travel times tend to be reported as 20, 30, and 60 minutes. (Zhao, et al. 2015).

5.6.2 Updating OD Matrices

Ge and Fukuda (2016) proposed a method to utilize CDRs to estimate an OD matrix pertaining to work-related travel. Citing the relative infrequency with which travel surveys are conducted, the non-sampling errors to which survey data are subjected, and the lag time between when the data is collected versus when it is coded, the authors propose a method that utilized CDRs to update an OD matrix obtained through a traditional household travel survey. The methodology estimates work-related OD demand using two sub-modules: one that estimates the time-of-day trip productions and attractions for each zone using reference values, and another that distributes trips to OD pairs using spatial interaction models. In order to protect user privacy, individuals' trace data was aggregated at the hour level and at the mesh level (one grid cell is 500 m x 500 m). As this data was a sample of all cellphone records, it was scaled up based on the penetration rate of smartphones, the market share of the particular mobile phone operator, and the sampling rate. The output of the model was constrained to ensure that the flux of trips is equal to that of people, that the railway capacity was not exceeded, to adhere to the trip production constraints of each zone, and to ensure that trip demand characteristics are not negative. The authors carried out a case study using data from the Tokyo metropolitan area. Given population data at the mesh level from 2012 and a prior OD matrix from the 2008 household travel survey, which formed the basis upon which the mode share and capacity of each mode was calibrated, and from which the reference trip productions and attractions were taken. The authors used the update matrix to make inferences about changes in work-related travel behaviour from 2008 to 2012. (Ge and Fukuda 2016)

5.6.3 Understanding Public Transit Demand

Identifying the inability for public transit agencies to precisely measure demand as the main impediment to their ability to adjust service to meet demand, Morency, Trepanier and Agard (2007) propose a methodology to apply data mining and database management tools to measure the spatial and temporal variability of transit use. Using a set of smart card transaction data, obtained from the Societe de Transport de l'Outatouis, a transit agency that serves Gatineau, Quebec, comprised of 2.2 million boardings made between January and October 2005, the authors utilized a transportation object-oriented modelling methodology. This methodology classified data into four meta-classes of objects: network objects, operations objects, administrative objects, and trips. Spatial variability was defined based on the number of bus stops used and the frequency with which a given stop is used, while temporal variability is examined using the k-means algorithm, to organize the data into clusters. The data were analyzed according to the three different fare types: senior, student, and adult. The results identified a stark contrast in the behaviour of adults compared to students and seniors: the users of the student and senior card types showed a more diverse use of bus stops, and used their most frequently-visited stop less frequently that adults (~50% vs ~60-80%). Similarly, the trips of adult fare cards were clustered around the AM and PM peak periods and on weekdays, while students and seniors displayed more temporal variation. This behaviour was also reflected in the distribution of zero-boarding days: adults did not use transit in large proportions on weekends, while seniors tend to use transit less on weekdays.

6 CONCLUSIONS AND FUTURE WORK

Designing a data collection program according to the core-satellite survey paradigm offers the opportunity to collect more detailed data than would be collected through a traditional household travel survey, due to the ability to conduct multiple surveys that are designed to target a specific subpopulation, with the goal of enabling a specific set of capabilities in mind. Consisting of three components, the core survey, satellite surveys, and complementary datasets, the design of each component should be tailored to suit the characteristics of the target population, facilitate the enabling of the desired capability, and to serve the goals of the survey. Within the GGHA, ensuring that survey data is compatible with the TTS, based either on the spatial, temporal, and/ or semantic contexts, provides the opportunity to utilize data collected through a large-sample travel survey that utilizes a survey method that had already been validated.

This review presented the basis upon which the compatibility of multiple datasets is defined, followed by a discussion of five types of potential satellite surveys: transit on-board surveys, active mode user surveys, post-secondary student surveys, employee/ workplace surveys, and attitudinal and stated preference surveys. For each of these surveys, the sampling techniques used, the considerations that affect the choice of the survey instrument, the applications of the data that are collected, and the steps that can be taken to ensure that the data collected through these surveys is compatible with data collected through the TTS are discussed. Finally, the advantages and issues associated with the use of different types of passive data are examined, and interesting applications are discussed. Overall, the additional effort required to conduct a survey using the core-satellite survey framework can be outweighed by the breadth and depth of the data that can be obtained by applying the said framework.

The next stage of this project is to develop a method to implement the core-satellite survey paradigm in practice. One of the key areas of research is to evaluate the impact of the time lag between the conduct of the core and satellite surveys on the compatibility of the two sets of data in the temporal context. The ideal case would be to conduct the core and satellite surveys in a concurrent manner, however, this may not always be feasible. Additional work is also needed to develop methods of ensuring compatibility in the temporal context. This is particularly important when satellite surveys are conducted between TTS survey years, to identify how soon the satellite survey must be conducted in relation to the conduct of the TTS.

REFERENCES

Agard, B., Morency, C., & Trepanier, M. (2006). Mining Public Transport User Behaviour from Smart Card Data. *IFAC Proceedings Volumes, 39*(3), 399 - 404.

Agrawal, A., Granger-Bevan, S., Newmark, G., & Nixon, H. (2017). Comparing Data Quality and Cost from Three Modes of On-Board Transit Surveys. *Transport Policy, 54*, 70 - 79.

Akar, G., & Clifton, K. (2009). Influence of Individual Perceptions and Bicycle Infrastrcture on Decision to Bike. *Transportation Research Record: Journal of the Transportation Research Board, 2140*(18), 165-172.

Alexander, L., Jiang, S., Murga, M., & Gonzalez, M. (2015). Origin-Destination Trips by Purpose and Time of Day Inferred form Mobile Phone Data. *Transportation Research Part C, 58*, 240 - 250.

Alsger, A., Mesbah, M., Ferreira, L., & Safi, H. (2015). Use of Smart Card Fare Data to Estimate Public Transport Origin-Destination Matrix. *Transportation Research Record: Journal of the Transportation Research Board, 2535*(10), 88 - 96.

Alta Planning + Design. (2016). *National Bicycle and Pedestrian Documentation Project*. Retrieved October 30, 2017, from http://bikepeddocumentation.org/index.php/downloads

American Public Transportation Association. (2007). *A Profile of Public Transportation PAssenger Demographics and Travel Characteristics Reported in On-Board Surveys*. Washington: American Public Transportation Association.

Bayart, C., & Morency, C. (2008). Survey Mode Integration and Data Fusion - Methods and Challenges. *8th International Conference on Survey Methods in Transport*. Annecy.

Bernardin, Lochmueller and Associates. (2010). *2009 IndyGo On-board Transit Survey - Final Report.* Indianapolis: Bernardin, Lochmueller and Associates.

Cassello, J., & Usyukov, V. (2014). Modeling Cyclists' Route Choice Behaviour Based on GPS Data . *Transportation Research Record: Journal of the Transportation Research Board, 2430*(16), 155 - 161.

Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M., & Zegras, P. (2013). Future Mobility Survey – Experience in Developing a Smartphone-Based Travel Survey in Singapore. *Transportation Research Record: Journal of the Transportation Research Board, 2354*(7), 59 - 67.

Cummins, B., Spitz, G., O'Malley, T., & Campbell, M. (2013). How Close is Close Enough? Statistical Equivalence of Onboard Versus Online Surveys of Transit Customers. *Transportation Research Record: Journal of the Transportation Research Board, 2351*(3), 23 - 29.

D'Orazio, M., Di Zio, M., & Scanu, M. (2006). *Statistical Matching: Theory and Practice*. New York City: John Wiley & Sons.

Dumont, J., Shalaby, A., & Roorda, M. (2012). A GPS-Aided Survey for Assessing Trip Reporting Accuracy and Travel of Students without Telephone Land Lines . *Transportation Planning and Technology, 35*(2), 161 - 173.

Edwards, S., Ivey, S., Lipinski, M., & Golias, M. (2012). Bicycle ad Pedestrian Studies Based on Data from National Household Travel Survey. *Transportation Research Record: Journal of the Transportation research Board, 2299*(16), 150 - 156.

Friesen, M., & McLeod, R. (2015). Bluetooth in Intelligent Transportation Systems - A Survey. *Internation Journal of Intelligent Transportation Systems, 13*, 143 - 153.

Garikapati, V., You, D., Pendyala, R., Patel, T., Kottommannil, J., & Sussman, A. (2016). Design, Development, and Implementation of a University Travel Demand Modeling Framework. *Transportation Research Record: Journal of the Transportation Research Board, 2563*(15), 105 - 113.

Ge, Q., & Fukuda, D. (2016). Updating Origin-Destination Matrices with Aggregated Data of GPS Traces. *Transportation Research C, 69*, 291 - 312.

Godefroy, F., & Morency, C. (2012). Estimating Latent Cycling Trips in Montreal, Canada. *Transportation Research Record:Journal of the Transportation Research Board, 2314*(16), 120 - 128.

Goulias, K. G., Pendyala, R. M., & Bhat, C. R. (2011). *Total Design Data Needs for New Generation Large Scale Activity Microsimulation Models.*

Grover, D., Wygonik, E., Bucossi, S., Bell, A., Piper, S., & Brewer-Colie, K. (2016). Estimating Current and Potential Bicycle Use for Statewide Planning. *Transportation Research Record: Journal of the Transportation Research Board, 2587*(13), 109 - 116.

Habib, K. (2014, September 10). CIV1520: Travel Survey Methods. Toronto, Onatrio, Canada.

Huegy, J., Wang, C., Mei, B., Findley, D., Searcy, S., & Bhadury, J. (2014). *Trip Making Patterns of NC's University Students*. Raleigh: North Carolina DOT.

Ivey, S., Badoe, D., & Edwards, S. (2012). Add-On Program for National Household Travel Survey - Experience of Stakeholders and Best Practices to Maximize Program Benefits. *Transportation Research Record: Journal of the Transportation Research Board, 2291*, 102 - 110.

Ji, Y., Mishalani, R., & McCord, M. (2015). Estimating Trnasit Route OD Flow Matrices from APC Data on Multiple Bus Trips Using the IPF Method with an Iteratively Improving Base - Method and Empirical Evaluation. *Journal of Transportation Enginerring, 140*(5), 1 - 8.

Khatri, R., Cherry, C., Nambisan, S., & Han, L. (2016). Modeling Route Choice of Utilitarian Bikeshare Users with GPS Data. *Transportation Research Record: Journal of the Transportation Research Board, 2587*(17), 141 - 149.

Khattak, A., Wang, X., Son, S., & Agnello, P. (2011). Travel by University Students in Virginia – Is This Travel Different from Travel by the General Population? *Transportation Research Record: Journal of the Transportation Research Board, 2255*(15), 137 - 145.

Kressner, J., & Garrow, L. (2014). Using Third-Party Data for Travel Demand Modeling – Comparison of Targeted Marketing, Census, and Household Travel Survey Data. *Transportation Research Record: Journal of the Transportation Research Board, 2442*(2), 8 - 19.

Lowry, M., McGrath, R., Scruggs, P., & Paul, D. (2016). Practitioner Survey and Measurement Error in Manual Bicycle and Pedestrian Count Programs. *International Journal of Sustainable Transportation, 10*(8), 720 - 729.

Majumdar, B., & Mitra, S. (2017). Valuing Factors Influencing Bicycle Route choice Using a Stated-Preference Survey. *Journal of Urban Planning and Dvelopment, 143*(3), 1 - 11.

Malinovskiy, Y., Saunier, N., & Wang, Y. (2012). Analysis of Pedestrian Travel with Static Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board, 2299*(15), 137 - 149.

Matsuda, Y., Rosenstein, P., Scovitch, C., & Takamura, K. (1998). *Massachusetts Institute of Technology*. Retrieved 07 23, 2017, from Data Collection: Defining the Customer: http://web.mit.edu/ecom/www/Project98/G2/data.htm

McHugh, B., Dong, B., Recker, J., & Shank, V. (2017). Conducting Onboard Transit Rider Surveys with Electronic Handheld Tablets – An Agencywide Consolidated Approach . *Transportation Research Record: Journal of the Transportation Research Board, 2643*(3), 19 - 27.

Memarian, B., Jeong, S., & Uhm, D. (2012). Effects of Survey Techniques on On-Board Survey Performance. *Transport Policy, 21*, 52 - 62.

Metrolinx. (2014, August). *Smart Commute: A Program of METROLINX*. (Metrolinx) Retrieved August 3, 2017, from http://smartcommute.ca/

Miller, E., Habib, K., Lee-Gosselin, M., Morency, C., Roorda, M., & Shalaby, A. (2011). *Changing Practices in Data Collection on the Movement of People*. Quebec City: Lee-Gosselin Associates Ltd.

Miralles-Guasch, C., & Domene, E. (2010). Sustainable transport challenges in a suburban university: The case of the Autonomous University of Barcelona. *Transport Policy, 17*, 454 - 463.

Miranda-Moreno, L., & Nosal, T. (2011). Weather or Not to Cycle - Temporal Trends and Impact of Weather on Cycling in an Urban Environment. *Transportation Research Record: Journal of the Transportation Research Board, 2247*(6), 42 - 52.

Mishalani, R., Ji, Y., & McCord, M. (2011). Effect of Onboard Survey Sample Size on Estimation of Transit Bus Route Passenger Origin-Destination Flow Matrix Using Automatic Passenger Counter Data . *Transportation Research Record: Journal of the Transportation Research Board, 2246*(9), 64 - 73.

Morency, C., Trepanier, M., & Agard, B. (2007). Measuring Transit Use Variability with Smart-Card Data . *Transport Policy, 14*, 193 - 203.

Munizaga, M., & Palma, C. (2012). Estimation of a Disaggregate Multimodal Public Transport Origin-Destination Matrix from Passive Smartcard Data from Santiago, Chile . *Transportation Research Part C, 24*, 9 - 18.

Nosal, T., Miranda-Moreno, L., & Krstulic, Z. (2014). Incorporating Weather - Comparative Analysis of Annual Average Daily Bicyclist Traffic Estimation Methods. *Transportation Research Record: Journal of the Transportation Research Board, 2468*(12), 100 - 110.

Pelletier, M., Trepanier, M., & Morency, C. (2011). Smart Card Data Use in Public Transit - A Literature Review. *Transportation Research C, 19*, 557 - 568.

Research Systems Group, Inc. (2013). *Utah Travel Study*. Salt Lake City: Research Systems Group, Inc.

Richardson, A. (2006). Estimating Bicycle Usage on a National Cycling Netowrk. *Transportation Research Record: Journal of the Transportation Research Board, 1982*, 166 - 173.

Rolstad, S., Adler, J., & Ryden, A. (2011). Response Burden and Questionnaire Length – Is Shorter Better? A Review and Meta-Analysis . *Value in Health, 14*, 1101 - 1108.

Sears, J., Flynn, B., Aultman-Hall, L., & Dana, G. (2012). To Bike or Not to Bike. *Transportation Research Record: Journal of the Transportation Research Board, 2314*(14), 105 - 111.

Shen, L., & Stopher, P. (2014). Using SenseCam to Pursue "Ground Truth" for Global Positioning System Travel Surveys. *Transportation Research C, 42*, 76 - 81.

Simas Olivera, M., Vovsha, P., Wolf, J., Birotker, Y., Givon, D., & Paasche, J. (2011). Global Positioning System-Assisted Prompted Recall Household Travel Survey to Support Development of Advanced Travel Model in Jerusalem, Israel. *Transportation Research Record: Journal of the Transportation Research Board, 2246*(3), 16 - 23.

Simas-Olivera, M., & Casas, J. (2010). Improving Data Quality, Accuracy, and Response in On-Board Surveys. *Transportation Research Record: Journal of the Transportation Research Board, 2183*(5), 41 - 48.

Son, S., Khattak, A., Chen, J., & Wang, X. (2012). Transferring Telephone-Based National Household Travel Survey to the Internet – Application to University Students. *Transportation Research Record: Journal of the Transportation Research Board, 2285*(11), 91 - 99.

Son, S., Khattak, A., Wang, X., Agnello, P., & Chen, J. (2013). Quantifying Key Errors in Household Travel Surveys – Comparison of Random-Digit-Dial Survey and Address-Based Survey . *Transportation Research Record: Journal of the Transportation Research Board, 2354*(2), 9 - 18.

Southwell, F., Zhang, Y., & Sharp, J. (2014). *Chapter 18: Workplace and Establishment Surveys.* (Transportation Research Board) Retrieved July 22, 2017, from http://www.travelsurveymanual.org/Chapter-18.html

Sperry, B., Chigoy, B., Green, L., & Hard, E. (2015). Development of Improved Trip Attraction Rates for Small- and Medium-Sized Travel Demand Models. *95th Transportation Research Board Annual Meeting*. Washington.

Statistics Canada. (2014). *Residential Telephone Service Survey, 2013*. Ottawa: Statistics Canada.

Stopher, P., & Greaves, S. (2007). Household travel surveys: Where are we going? *Transportation Research Part A, 41*, 367 - 381.

The City of Calgary. (2007). *Calgary Downtown Commuter Cyclist Survey Report*. Calgary: The City of Calgary.

Tilahun, N., Levinson, M., & Krizek, K. (2007). Trails, Lanes, or Traffic: Value of Different Bicycle Facilities Using Adaptive Stated-Preference. *Transportation Research*, 287 - 301.

Tourangeau, R., Groves, R., & Redline, C. (2010). Sensitive Topics and Reluctant Respondents – Demonstrating a Link Between Non-Response Bias and Measurement Error . *Public Opinion Quarterly, 74*(3), 413 - 432.

Twaddle, H., Hall, F., & Bracic, B. (2010). Latent Bicycle Commuting Demand and Effects of Gender of Commuter Cycling and Accident Rates. *Transportation Research Record: Journal of the Transportation Research Board, 2190*(4), 2010.

U of T Data Management Group. (2013). *Data Expansion & Validation*. Toronto: U of T Data Management Group.

U of T Data Management Group. (2013). *Transportation Tomorrow Survey 2011 Data Guide*. Toronto: U of T Data Management Group.

U of T Data Management Group. (2014). *Design and Conduct of the Survey*. Toronto: U of T Data Management Group.

U.S. Department of Transportation - Federal Highway Administration . (2011). *2009 National Household Travel Survey - User's Guide*. Washington: U.S. Department of Transportation.

USDOT. (2016). *National Household Travel Survey*. Westat.

Verreault, H., & Morency, C. (2016). Integration of a Phone-Based Housheold Travel Survey and a Web-Based Student Travel Survey. *Transportation, 1007*(10).

Volosin, S., Pendyala, R., Kerrigan, J., Greene, E., Livshits, V., & Samuelson, J. (2013). Measuring the Travel Characteristics of a University Population – Experiences from the Design and Administration of a Web-based Travel Survey . *Travel Survey Methods*.

Wang, X., Khattak, A., & Son, S. (2012). What can be Learned from Analyzing University Student Travel Demand? *Transportation Research Record: Journal of the Transportation Research Board, 2322*, 129 - 137.

Winters, M., Davidson, G., Kao, D., & Teschke, K. (2011). Motivators and Deterrent of Bicycling – Comparing Influences on Decisions to Ride. *Transportation, 38*, 153 - 168.

Yan, T., & Tourangeau, R. (2008). Fast Times and Easy Questions – The Effects of Age, Experience, and Question Complexity on Web Survey Response Times . *Applied Cognitive Psychology, 22*, 51 - 68.

Zhao, F., Pereira, F. C., Ball, R., Kim, Y., Han, Y., Zegras, C., & Ben-Akiva, M. (2015). Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore. *Journal of the Transportation Research Board*, 45–56.

APPENDIX A: SELECTION OF THE DATA COLLECTED BY AGENCIES IN THE GGHA

Variables	2011 TTS (U of T DMG)	Student Move TO (Toronto)	School Travel Planning Tool (Barrie)	2016 Brantford Transit Survey	City of Guelph Cycling Survey (2009)	2015 Transit Priority Project (Guelph Transit)	2015 Transit Operational Study (Guelph Transit)	2015 Transit Optimization Study (Orangeville Transit)	2010 Household Transportation Survey (Peterborough)	2012 GPS Cycling Study (Peel Region)
Household Data										
Dwelling Type	x	x								x
Number of Persons	x	x							x	x
Number of Vehicles	x	x							x	x
Number of driver's licence holders	x								x	x
Number of full-time employees	x									x
Number of part-time employees	x									x
Number of employed persons in the home	x									x
Number of students	x									x
Number of household trips	x									x
Number of children		x								
Home ownership		x								
HH category		x								
Home phone ownership		x								
Household income level		x								
Person Data										
Age	x	x	x	x						x
Sex	x	x	x							x
Driver's license ownership	x	x								x
Car ownership		x								
Availability of a car to borrow		x								
Car-sharing membership		x								
Transit pass ownership	x	x								x
Employment status	x									x
Occupation	x									x
Number of hours worked per week		x								
No work trips made	x									x
Reason for no trips		x								
Student status	x	x								x
Year enrolled		x								

	2011 TTS (U of T DMG)	Student Move TO (Toronto)	School Travel Planning Tool (Barrie)	2016 Brantford Transit Survey	City of Guelph Cycling Survey (2009)	2015 Transit Priority Project (Guelph Transit)	2015 Transit Operational Study (Guelph Transit)	2015 Transit Optimization Study (Orangeville Transit)	2010 Household Transportation Survey (Peterborough)	2012 GPS Cycling Study (Peel Region)
Person Data (cont'd)										
Availability of free parking	x									x
Number of personal trips	x								x	x
Number of transit trips	x									x
Personal living situation		x								
Household category		x								
University attending		x								
Faculty affiliation		x								
Use of ridesharing in past year		x								
Motorized two-wheeler ownership		x								
Presto card ownership		x								
Bicycle ownership		x								
Bike-sharing membership		x								
Number of years at current home		x								
Primary factor for choosing home		x								
Campus registration		x								
Distance from home to main campus		x								
Travel day started away from home		x								
Reasons for walking		x								
Reasons for cycling		x							x	

Variables	2011 TTS (U of T DMG)	Student Move TO (Toronto)	School Travel Planning Tool (Barrie)	2016 Brantford Transit Survey	City of Guelph Cycling Survey (2009)	2015 Transit Priority Project (Guelph Transit)	2015 Transit Operational Study (Guelph Transit)	2015 Transit Optimization Study (Orangeville Transit)	2010 Household Transportation Survey (Peterborough)	2012 GPS Cycling Study (Peel Region)
Trip Data										
Start time	x	x								x
Primary mode	x	x	x				x	x		x
Trip purpose	x	x		x	x		x			x
Origin purpose	x	x								x
Origin UTMX	x	x		x				x		x
Origin UTMY	x	x		x				x		x
Destination purpose	x	x								x
Destination UTMX	x	x		x		x		x		x
Destination UTMY	x	x		x		x		x		x
Destination geo type	x									x
Trip distance	x	x								x
Carpool	x									x
Highway 407	x									x
Alternative modes used					x					
Factors affecting utilitarian cycling					x				x	
End-of-trip amenities					x					
Barriers to utilitarian cycling					x				x	
Trip distance									x	
Typical travel		x								
Visited places		x								
Other modes used		x								
Activity duration		x								
Vehicle occupancy		x								
Departure date		x								
Arrival date		x								
Departure time		x								
Arrival time		x								

Variables	2011 TTS (U of T DMG)	Student Move TO (Toronto)	School Travel Planning Tool (Barrie)	2016 Brantford Transit Survey	City of Guelph Cycling Survey (2009)	2015 Transit Priority Project (Guelph Transit)	2015 Transit Operational Study (Guelph Transit)	2015 Transit Optimization Study (Orangeville Transit)	2010 Household Transportation Survey (Peterborough)	2012 GPS Cycling Study (Peel Region)
Transit Trip Data										
Access mode	x	x								x
Access UTMX	x									x
Access UTMY	x									x
Access distance	x									x
Egress mode	x									x
Egress region	x									x
Egress PD	x									x
Egress UTMX	x									x
Egress UTMY	x									x
Egress distance	x									x
Number of routes	x	x				x	x			x
Route number	x	x				x	x			x
Subway board	x	x								x
Subway alight	x	x								x
GO board	x	x								x
GO alight	x	x								x
Number of GO rail links	x	x								x
Number of GO bus link	x	x								x
Number of subway links	x	x								x
Number of TTC bus links	x	x								x
Number of local transit routes	x	x								x
Number of other transit	x	x								x
Operator code	x									x
Use of TTC	x									x

APPENDIX B: SUMMARY OF DATA COLLECTED THROUGH THE NHTS

Category of Data	Content[2]
Household	• Number of people, drivers, workers, and vehicles • Income • Housing type • Owned or rented • Number of cell phone • Number of other phones • Race of reference person • Hispanic status of reference person • Tract and block group characteristics • Internet use and delivery to households
Person	• Age/sex/relation to reference person • Driver status • Worker status/ primary activity • Internet use • Home deliveries from internet shopping • Travel disability • Effect of disability on mobility • Education level • Immigrant status • Views on transportation • Annual miles driven • Incidence of public transit use in the past month • Incidence of motorcycle use in last month • Incidence of walk and bike trips in past week • School travel
Worker	• Full- or part-time status • More than one job • Occupation • Workplace location • Usual mode to work • Drive alone or carpool • Usual distance to work • Usual time to work • Work from home • Usual arrival time at work

[2] Italicized data are considered to be "core" data

	• Flexibility in work arrival t
Vehicle	• Make/model/age • Annual miles driven • Commercially licensed • How long owned • Odometer reading • Alternative fuel • Primary driver
Daily Travel	• Origin and destination address (add-on locations only) • Time trip started and ended • Distance • Means of transportation • Interstate use • Tolls paid • Trip purpose • Detailed purpose • Travel party size • Last time of travel