The Design and Empirical Evaluation of the Core-Satellite Framework for Urban Passenger Data Collection

by

Patrick Maung Soe Win Htun Loa

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science

Department of Civil and Mineral Engineering
University of Toronto

The Design and Empirical Evaluation of the Core-Satellite Framework for Urban Passenger Data Collection

Patrick Loa

Master of Applied Science

Department of Civil and Mineral Engineering
University of Toronto

2019

Abstract

Household travel surveys play a vital role in the operation and development of transportation infrastructure. The urban passenger travel data obtained through household travel surveys play a crucial role in the planning of transportation networks and form the basis of the methods used to forecast the utilization of infrastructure. Traditional household travel survey methods are growing obsolete, owing to technological trends and changing data needs. This thesis proposes a modification to the core-satellite data collection paradigm proposed by Goulias, Pendyala, & Bhat (2011), that aims to take a more holistic approach to the collection of urban passenger travel data. The proposed framework presents a method for combining several purpose-specific surveys to create a basis for the analysis of travel behaviour that is greater than the sum of its parts. This thesis also presents two empirical studies that utilize the different types of data outlined in the expanded framework.

Acknowledgments

First and foremost, I would like to thank my supervisor, Dr. Khandker Nurul Habib, for all of his guidance, encouragement, and support throughout my time as a MASc student. His dedication, expertise, and continued willingness to expand his knowledge serve as a standard that I strive to emulate. I would also like to thank Dr. Eric J. Miller for his guidance and for providing me with my first opportunity to work in academia, a door that I believed was closed after my first two years as an undergraduate.

I would like to thank Dr. Adam Weiss and Dr. Sami Hasnine for their mentorship and guidance. I hope to reach the standards of expertise and kindness that I have witnessed from these two outstanding academicians and friends. I would like to acknowledge Zhujun Li and Albert Lo, two people with whom I have spent more time in the lab and office than I am willing to admit and people with whom I am always delighted to spend time. I want to extend my thanks to the friends that I have made throughout my Master's – Siva, Brittany, Sanjana, Graham, Alaa, Paul, Catalina, Tony, Wenting, Luna, Jason, Salehin, Saeed, Alireza, Ran, Chris, and anyone else that I may have missed. Spending time with you in the office, in the lab, on our food adventures, and on conference trips have made these past two years infinitely more exciting and enjoyable. I also want to thank my friends Mushfik Alam, Micheal Zhen, and Ahnaf Tahmid, I can always count on you to help me relax when work is getting me down.

Finally, I would like to thank my parents, Khin, and Melvin for their unwavering and continued support throughout my journey. They made innumerable sacrifices to ensure that my brother and I had every opportunity to succeed and ensure that we will not have to struggle the way that they did. Their work ethic and selflessness are traits that I hope to be able to live up to. I also want to acknowledge my brother Albert, with whom I have had many foods and travel adventures and with whom I hope to have many more. This thesis is dedicated to my grandfather, who unfortunately passed away during my Master's; I'm on my way to becoming a doctor, just not quite the kind that you had in mind.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Acronyms

ABS	Address-Based sampling
API	Application Programming Interface
APTA	American Public Transit Association
CAPI	Computer-Assisted Personal Interview
CATI	Computer-Assisted Telephone Interview
CAWI	Computer-Assisted Web Interview
CDR	Call Detail Record
CPO	Cellphone-Only
CSDCP	Core-Satellite Data Collection Paradigm
DA	Dissemination Area
EPOI	Enhanced Points of Interest
GGHA	Greater Golden Horseshoe Area
GIS	Geographic Information System
GPS	Global Positioning System
GTFS	General Transit Feed Specification
LOS	Level of Service
LUDI	Land Use Density Indicator
MNL	Multinomial Logit
MPO	Metropolitan Planning Organization
MSA	Metropolitan Statistical Area
NCR	National Capital Region
OD	Origin-Destination
PCA	Principal Component Analysis
PCE	Passenger Car Equivalent
POI	Point of Interest
RDD	Random Digit Dialing
SMTO	StudentMoveTO
SP	Stated Preference
TAZ	Traffic Analysis Zone
TB	Time Budget
TILOS	Tool for Incorporating Level of Service

TTC	Toronto Transit Commission
TTS	Transportation Tomorrow Survey
UTTRI	University of Toronto Transportation Research Institute
VoIP	Voice over Internet Protocol

Technical Report and Conference Paper Acknowledgements

Sections of this thesis are based on research modified from technical reports that have been created for various projects. The references are as follows:

Srikukenthiran, S., **Loa, P.**, Hossain, S., Chung, B., Habib, K.N., Miller, E.J. "Transportation Tomorrow Survey 2.0: Final Report". Report to the Transportation Information Steering Committee, Toronto. University of Toronto Transportation Research Institute, Dec. 2018.

Loa, P., Hossain, S., Habib, K.M.N. "Satellite Survey Design for the Collection of Cycling Demand in the National Capital Region – Survey Design and Recommendations". Report to the TRANS Committee. University of Toronto Transportation Research Institute, Dec. 2018.

Loa, P., Ladan, T., Hossain, S., Habib, K.M.N. "Core-Satellite Survey Design Accommodating Active Mode Demand in the National Capital Region – Review and Recommendations". Report to the TRANS Committee. University of Toronto Transportation Research Institute, Nov. 2017.

Ladan, T., **Loa, P.**, Habib, K.M.N. "Data for Planning Active Travel Modes: A Literature Review". Report to the TRANS Committee. University of Toronto Transportation Research Institute, Nov. 2017.

Loa, P., Srikukenthiran, S., Miller, E.J., Habib, K.M.N. "Core-Satellite Survey Design Framework: A Conceptual Framework". Report to the Transportation Information Steering Committee, Toronto. University of Toronto Transportation Research Institute, Nov. 2017.

Loa, P., Srikukenthiran, S., Miller, E.J., Habib, K.M.N. "Current state of web-based survey methods". Report to the Transportation Information Steering Committee, Toronto. University of Toronto Transportation Research Institute, Sept. 2015.

In this thesis, the following chapters have been reproduced with modification from papers that have been submitted for presentation and publication:

Chapter 5 APPLICATION OF SATELITE SURVEY DATA

Loa, P., Habib, K.M.N. "How Does Toronto Transit Treat Post-Secondary Students Making Discretionary Trips in the City? A Choice Model-Based Analysis of Accessibility by Transit in

Toronto. Submitted for publication and presentation at the 99th annual meeting of the Transportation Research Board.

Chapter 6: COMBINING PASSIVE AND SURVEY DATA

Loa, P., Hossain, S., Habib, K.M.N. “Understanding the Influence of Built Environment and Zonal Attributes on the Relationship between Public Transit and Ride-Hailing Services”. Paper Presented at Conference on “Research and Applications on the Use of Massive Passive Data for Public Transport”, July 8 – 10, 2019. Paris, France.

Chapter 1
Introduction

Urban passenger travel data play a critical role in the design, planning, and operation of transportation systems. These data provide a wealth of information that is used to understand the utilization of transportation facilities and services and are also used to forecast the usage of facilities and services in the future. Travel demand data form the basis for evidence-based decision-making, which helps to ensure that scarce resources are directed to projects and initiatives that have the greatest potential benefits. Urban passenger travel data is most frequently collected through household travel surveys, wherein sampled households are asked to provide, among other things, a record of the trips that each household member made on a specific date. In the Canadian context, household travel surveys play a central role in transportation planning activities, as they tend to be the only source of data that can be used to support travel demand modelling and policy analysis (Miller & Habib, 2014). Due to trends in both technology and travel demand analysis, traditional approaches to the conduct of household travel surveys are producing increasingly inadequate passenger travel data. Addressing this inadequacy is paramount, as passenger travel data form the basis of both the operation and planning of transportation infrastructure.

One of the key drivers of the growing inadequacy of the data collected using traditional household travel survey methods is the declining ownership and use of landline telephones. As of 2017, 63% of Canadian households reported that they continue to receive landline telephone service, compared to 70% in 2013 (Statistics Canada, 2018; Statistics Canada, 2014). By comparison, the same Statistics Canada study found that 89.5% of Canadian households owned at least one cellphone as of 2017. The adverse impacts of this decline on the efficacy of household travel surveys stem from the traditional reliance of these surveys on lists of landline telephone numbers as the primary sample frame, i.e. the lists from which survey participants are sampled. In addition, landline telephone ownership has declined at a greater rate among younger households than older households. Consequently, lists of landline telephone numbers are continuing to become less representative of the population. This trend has resulted in older members of the population being over-represented in household travel surveys and younger members of the population and students being under-represented (Miller & Habib, 2014). The declining ability of lists of landline telephone numbers to sufficiently represent the population has also created another issue – the need to identify alternative sample frames.

The other driver of the growing inadequacy of traditional household travel surveys is the changing data needs of both practitioners and academicians. These changes stem from both the desire to investigate travel behaviour at more disaggregate levels and the desire to investigate aspects of travel behaviour that have traditionally been overlooked in operational travel demand models. Examples of the contemporary and emerging aspects of travel demand analysis that are driving changing data needs include activity-based models and the study of seasonal variations in travel, long-distance travel, and the influence of attitudinal factors on travel behaviour. The desire to understand these aspects of travel behaviour creates the desire for more data and for these data to be collected at a greater level of detail than can be accommodated by a traditional household travel survey. Taken together, the factors that have led to the growing inadequacy of household travel surveys creates the need to rethink the existing travel survey paradigm.

1.1 Motivation and Project Overview

The issues that have plagued the conduct of household travel surveys is not unique to any one location or survey. In the Greater Golden Horseshoe Area (GGHA), the regional household travel survey, the Transportation Tomorrow Survey (TTS), has experienced issues related to both representation and changing data needs. The TTS has been conducted once every five years since 1986. As part of the TTS, one adult member of each sampled household is asked to provide information on the characteristics of the household, socio-economic information about its residents, and a one-day travel diary for each resident over the age of 11. The most recent iteration of the TTS was conducted in 2016, where information was collected from over 160,000 households and 395,000 persons, representing more than 706,000 trips (R.A. Malatest & Associated Ltd., 2018). The survey area of the 2016 TTS is shown in Figure 1. The 2016 TTS took a primarily address-based approach to sampling, wherein addresses were selected from a Canada Post database of all addresses in the study area (Data Management Group, 2018). The sampled households were given the chance to respond to the survey online or by calling into a toll-free number, while households for which the landline phone number was found could also wait to receive a call from an interviewer (Data Management Group, 2018).

Figure 1: The Survey Area for the 2016 Transportation Tomorrow Survey

The impact of relying on lists of landline phone numbers as the primary sample frame is apparent in the two previous iterations of the TTS, particularly as it relates to the ages of the respondents. As shown in Figure 2 and Figure 3, the proportion of TTS respondents under the age of 45 is less than that of the Canadian Census in both 2011 and 2016. This disparity is particularly significant for persons aged 15 to 34, which speaks to one of the key issues of using lists of landline phone numbers as a sample frame. Also, persons over the age of 50 are over-represented in the both the 2011 and 2016 TTS. It should be noted that the 2011 Census does not provide detailed age information for persons over the age of 85. This disparity between the age distribution of TTS respondents and the residents of the survey is problematic, as it can bring about sampling and non-response bias in the survey data and results. Given the role that household travel surveys play in the planning and design of transportation infrastructure, these biases can threaten the robustness and validity of the resulting travel demand models.

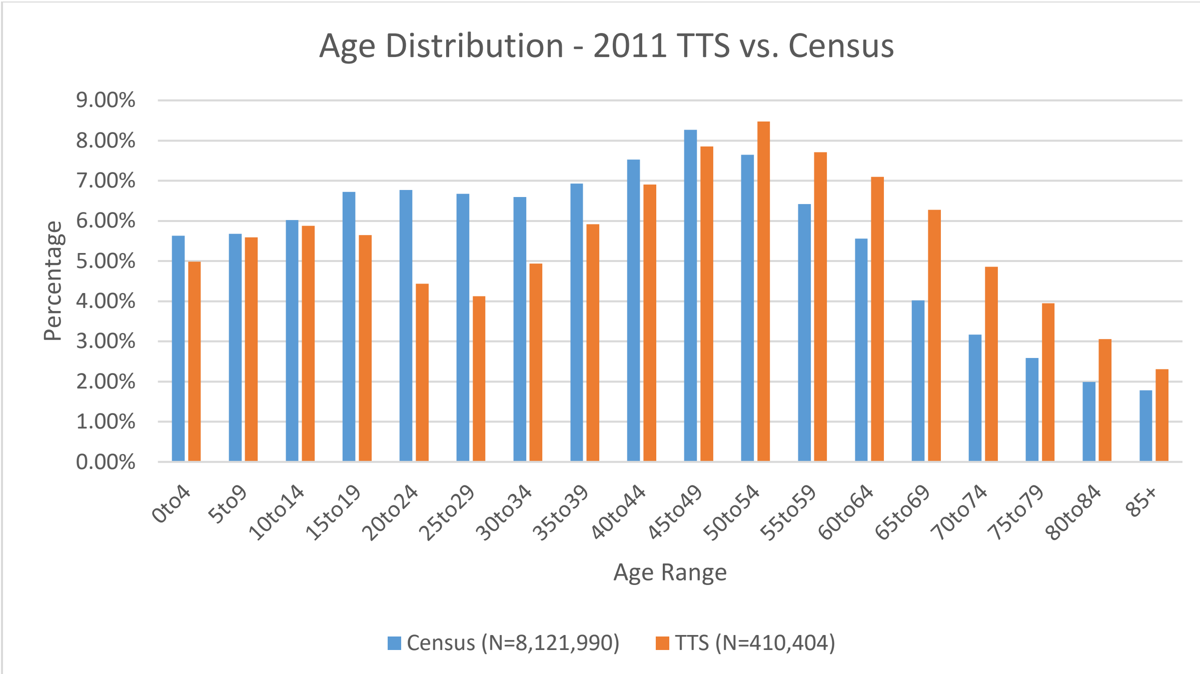


Figure 2: 2011 TTS Respondent Age vs. 2011 Census

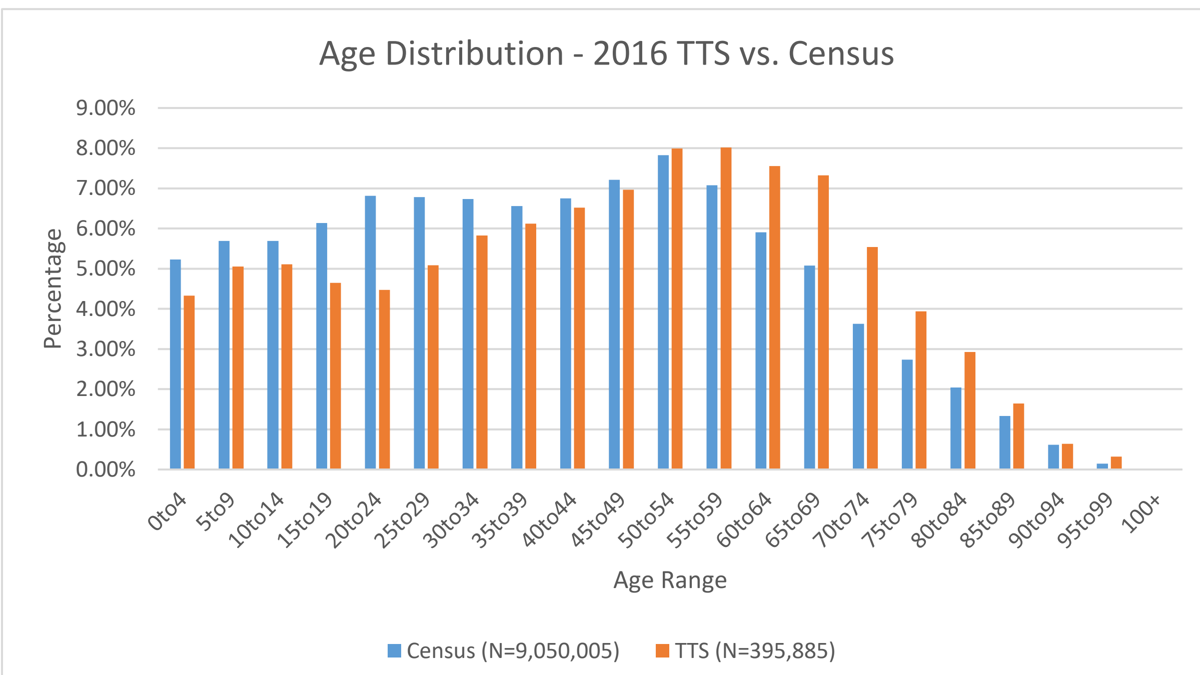


Figure 3: 2016 TTS Respondent Age vs. 2016 Census

In response to the challenges facing the TTS, the University of Toronto Transportation Research Institute (UTTRI) initiated the ‘TTS 2.0’ project, outlined in (Miller & Habib, 2014). The project included three major components: methodological research to identify the state-of-the-art in travel data collection methods; the field testing of prospective data collection methodologies to determine their efficiency, efficacy, and practicality; and to make recommendations regarding the design of future iterations of the TTS (Srikukenthiran, et al., 2018). The TTS 2.0 project focused on addressing four key aspects of the design and conduct of the TTS: the choice of sample frame(s), the choice of sampling unit(s), the choice of survey mode(s) and the associated burdens, and the choice of survey instrument (Miller & Habib, 2014). In addition, the project took two approaches to its evaluation of the existing survey framework – evaluating different approaches based on survey modes, and the application of the core-satellite framework, as described by (Goulias, Pendyala, & Bhat, 2011). Part of the work presented in this thesis was carried out as part of the TTS 2.0 project. The focus of this thesis is the empirical design and application of the core-satellite survey framework within the context of urban passenger data.

1.2 Chapters Summary

The remainder of this thesis is organized as follows. Chapter 2 presents a literature on the issues that stem from the use of traditional household travel survey methods. Chapter 3 presents the core-satellite survey framework for urban passenger data collection; each component of the framework is discussed, and common types of satellite surveys are presented. Chapter 4 presents a case study outlining work undertaken to design a satellite survey pertaining to cycling for the National Capital Region (NCR) in Canada. Chapter 5 outlines an empirical study of the location choice behaviour of university students in Toronto, using data obtained from a satellite survey, StudentMoveTO (SMTO). Chapter 6 discusses the findings of an empirical study of the relationship between ride hailing services (such as those offered by Uber and Lyft) and public transit, combining passive data provided by the City of Toronto and data from the TTS. Finally, Chapter 7 summarizes findings and identifies areas for future work.

Chapter 2
Literature Review

The declining effectiveness of household travel surveys has led to a wholesale re-examination of the traditional travel survey paradigm. In the past, household travel surveys were conducted through either telephone interviews or computer-assisted telephone interviews (CATI), with sampled households being selected from a register of landline telephone numbers. Given the ubiquity of landline telephones at the time, these registers represented a fairly comprehensive list of the households within a study area. Consequently, household travel surveys that were carried out in this manner were able to provide a sufficient representation of the target population. The trend of declining landline ownership has driven the desire to modernize the travel survey paradigm to both make use of contemporary and emerging technologies and to address outdated aspects of the paradigm. This chapter presents a literature review on key aspects of the household travel survey paradigm. First, the issues with traditional household travel survey methods are discussed. Then, the key factors that must be address when designing a travel survey framework are presented.

2.1 Issues with Traditional Travel Survey Methods

Traditionally, passenger travel data has been obtained through personal interviews, either face-to-face or over the telephone. In the past, telephone interviews were a reliable means of securing travel information from a representative sample of a population. The widespread ownership and use of landline telephones meant that registers of landline telephone numbers were able to capture the majority of the residents of a survey area, which reduced the extent of coverage errors in the survey data. Shifts in technology, including the replacement of traditional phone services with cellphones and Voice over Internet Protocol (VoIP) services, has reduced the proportion of the population that is captured by registers of landline telephone numbers. The growing prevalence of cellphones has been particularly detrimental to the efficacy of traditional household travel survey methods. In 2012, 15.5% of Canadian households owned at least one cellphone and no landline telephones; by 2017 this percentage had increased to 36.0% (Statistics Canada, 2018). The growing proportion of so-called cellphone-only (CPO) households is the main driver of the growing inadequacy of the conventional household travel survey paradigm. This issue has not gone unnoticed by survey administrators, who have turned to alternative survey modes and sample frames to augment the data obtained through computer-assisted telephone interviews and the samples selected using registers of landline telephone numbers.

Many large-scale surveys have turned to address-based sampling (ABS) to address the now well-known issues with the sampling of households via their landline telephone number. Address-based sampling involves the selection of households based on their mailing address, which is usually obtained from the organization that provides postal services to the survey area. In Canada, the Canada Complete™ Consumer Masterfile is often used (Canada Post, 2015). The 2016 iteration of the TTS utilized the ABS approach to sampling households, with attempts being made to match addresses to landline telephone numbers (Data Management Group, 2018). This sample was augmented by a sample of telephone numbers (corresponding to both landlines and cellphones) obtained from a survey sampling firm (Data Management Group, 2018).

These additional efforts appear to be warranted, given the differences that exist between households that own landlines and CPO households. This difference was observed by Son, Khattak, & Kim (2013) in their study of the Washington Metropolitan Area. Using household travel survey data obtained from the residents of the area, it was found that, compared to households that owned a landline telephone, CPO households were more likely to be single-person households, to live in multi-family or rental housing, to be between the ages of 19 and 34, and to belong to a minority group. In addition, respondents belonging to CPO households displayed a greater propensity for transit use and were more likely to be employed (Son, Khattak, & Kim, 2013). These findings highlight the need to account for the increasing proportion of CPO households when designing and administering household travel surveys.

In addition to the issues stemming from the declining ownership of landline telephones, the inadequacy of traditional household travel surveys can be attributed to changing approaches to the analysis of travel behaviour. Perhaps the most significant of these changes is the shift towards the activity-based approach to travel demand analysis, where linkages between activity participation and travel are explicitly modelled (Castiglione, Bradley, & Gliebe, 2015). In addition to this shift, there has been an increased desire to account for the role that attitudes play in travel behaviour, to understand the utilization of emerging modes of travel, and to model travel for non-commuting purposes. Household travel surveys are often geared towards understanding travel patterns on a typical weekday and tend to rely on travel diaries to obtain information pertaining to travel behaviours. As a result, accommodating the desire to expand the scope of travel demand analysis beyond weekday commuting behaviour within the traditional household travel survey paradigm would require the inclusion of additional questions.

Increasing the length of the questionnaire may have an adverse impact on both response rates and the quality of the survey data, due its potential to increase the burden experienced by the respondents. Although there is not a strong association between questionnaire length and response burden, increasing the length of a questionnaire has, at best, a neutral impact on burden (Rolstad, Adler, & Ryden, 2011). Consequently, studies that aim to investigate the more overlooked aspects of travel behaviour tend to obtain their data using one-off, purpose-built surveys. The ability to integrate these surveys with household travel survey would create the opportunity to exploit the availability of a set of travel data that has been obtained from a relatively large sample of the population.

The need to re-examine the traditional household travel survey paradigm that has been brought on by declining landline ownership and changing data needs also provides the opportunity to address issues that have plagued household travel surveys. Common issues with self-reported travel surveys include the tendency to report a typical day of travel, round travel times, and omit short activities. In addition, long-distance, short-distance, and discretionary trips tend to be under-reported in household travel surveys, while their cross-sectional nature means that they cannot capture day-to-day variations in travel and activity patterns (Stopher & Greaves, 2007). In household travel surveys, approximately 10 to 35% of trips may be unreported, although the exact value depends on sociodemographic and trip characteristics (Son, Khattak, Chen, & Wang, 2012). This under-reporting can partially be attributed to proxy bias, which arises when one household member has to report the travel of another (Chung, Srikukenthiran, Habib, & Miller, 2016).

Contemporary technologies, such as Global Positioning Systems (GPS) and smartphones, have the potential to address the factors that have traditionally led to trip under-reporting. In particular, the passive collection of spatiotemporal information has the potential to mitigate the impacts of incomplete recall, memory decay, and carelessness (Son, Khattak, Chen, & Wang, 2012). The re-examination of the travel survey paradigm creates the opportunity to modernize the conduct of household travel surveys, which includes the integration of contemporary technologies. The incorporation of newer technologies into travel surveys facilitates the collection of data in greater volumes and at greater levels of precision than traditional survey modes. The availability of these new types of data can provide a basis for novel approaches to the analysis of travel behaviour.

2.2 Survey Design Considerations

Several key factors must be considered when designing a large-scale household travel survey. One of the fundamental aspects of the design of a survey relates to sampling. Specifically, the choice of sample frame(s) from which survey participants are selected has a significant impact on the quality of the survey data. The selection of the sample frame is critical, as coverage errors can be induced if members of the target population are systematically omitted. The choice of survey mode(s) also has a significant impact on the outcomes of the survey, as the method(s) of data collection influences the burden experienced by the respondents and the types of information that can be obtained. In addition to these factors, the design of a survey is often influenced by considerations related to response rates and the potential for biases in the survey data. The following section will discuss these aspects of survey design in greater detail.

2.2.1 Sampling Strategies and Sample Frames

Sampling plays a critical role in the outcome of any survey. A crucial decision that must be made before the administration of a survey is the selection of the sample frame, which represents the members of the target population who have a chance to be included in the survey sample (Habib K. , 2014). The source(s) used to construct the sample frame of a survey play an important role in the control of coverage errors, which stem from the omission of members of the target population from the sample frame (Habib K. , 2014). In the past, the near-universal ownership of landline telephones meant that registers of telephone numbers were a sufficiently comprehensive sample frame in and of themselves. The rise of CPO households has created the need to seek out alternative sample frames to address the omission of these households from registers of landline telephone numbers. The adoption of different sample frames has also led to the use of different approaches to sampling, and the consideration of different sampling units.

Declining rates of landline ownership have led many large-scale household travel surveys to modify their approach to sampling. Traditionally, these surveys have sampled households from a register of landline telephone numbers or have used random digit dialing (RDD), wherein a random subset of all possible four-digit numbers within existing telephone exchanges are selected (Waksberg, 1978). The increasing prevalence of CPO households has led to the adoption of address-based sampling (ABS) strategies to either augment or replace landline-based samples. Compared to traditional approaches to sampling, using ABS allows specific geographic areas to be targeted and can yield additional information about the household, including the name of one of its residents (Bradley, Bergman, Lee, Greene, & Childress, 2015). The ability to obtain this information, particularly if it can be used to personalize the survey invitation, has the potential to improve response rates (Parsons, 2007; Weiner, Puniello, & Noland, 2016). As a result of the increasing prevalence of ABS methods, registers of residential addresses, such as the Canada Complete™ Consumer Masterfile, are now being used more frequently in household travel surveys.

The continued development and restructuring of the household travel survey paradigm has led some to question whether these surveys should still be conducted at the household level. The emergence of GPS- and smartphone-based methods of data collection, as well as the increase in the number of sources of passive data, has only increased the salience of this question. The choice of sampling unit is intimately related to the choice of sample frame, as it is meant to be a list of sampling units in the study area. The choice of individuals as the sampling unit of a survey, rather than households, necessitates the use of different approaches to sampling. For example, transit on-board surveys are often carried out as intercept surveys, wherein passengers are given paper questionnaires or are asked to participate in computer-assisted personal interviews (CAPI) (McHugh, Dong, Recker, & Shank, 2017). Post-secondary student surveys, wherein the survey is distributed using emailing lists, are another example of an individual-based survey (Verreault & Morency, 2016; Akar & Clifton, 2009). In addition to the sample frame, the choice of sampling unit can also impact the selection of the survey mode(s).

2.2.2 Survey Mode Selection

The selection of the survey mode(s) is a critical component of the design of any survey. In the context of a survey, the mode refers to the method through which the survey is implemented (Srikukenthiran, et al., 2018). Each survey mode comes with its own inherent strengths and weaknesses, both of which can affect the quality and quantity of the collected data. Broadly speaking, the choice of survey mode should be influenced by the characteristics of the target population. In addition, the choice of survey mode should consider the context in which respondents will be recruited for the survey. Contemporary travel surveys tend to be carried out using one of several modes: computer-assisted telephone interview (CATI), computer-assisted personal interview (CAPI), computer-assisted web interview (CAWI), and smartphone-based surveys. The pros and cons of using each survey mode are presented in Table 1 below. For a more detailed discussion of the pros and cons of each of these survey modes, as well as the recommended usage of each mode, is presented in Chapter 5 of (Srikukenthiran, et al., 2018).

Table 1: Pros and Cons of Prominent Survey Modes

Survey Mode	Pros	Cons
Computer-Assisted Telephone Interview (CATI)	• Presence of interviewer allows more complex questions to be incorporated • Reduces response burden • Phone ownership (cellphone and landline) is still ubiquitous	• Potential for social desirability bias (i.e. providing answers that are more socially desirable) • Use of proxies to obtain information can induce proxy bias • Tends to over-represent older members of the population
Computer-Assisted Personal Interview (CAPI)	• Provides comparatively high response rates and better data quality • Can address issues of trip under-reporting • Creates the potential for long and detailed interviews	• High marginal and per-completion costs • Potential for social desirability bias • Efficiency is significantly influenced by built-environment characteristics
Computer-Assisted Web Interview (CAWI)	• Allows the data collection process to be monitored in real-time • Provides respondents with anonymity • Has the potential to reduce response burden	• The design of the survey interface can affect response burden and measurement error • Must account for different levels of aptitude • Tends to over-represent younger and wealthier members of the population
Smartphone-Based Surveys	• Able to collect precise spatiotemporal data • Allows context-specific questions to be posed to respondents in real time • Can address issues related to memory loss	• Inferred information must be validated by respondents • Inferring trip information requires information that is comprehensive and accurate • Tends to over-represent younger members of the target population

Broadly speaking, none of these survey modes are without their faults. While the more traditional CATI-based survey has its well-documented issues with representation, web- and smartphone-based methods have their own representation issues. Specifically, these modes tend to over-represent younger and wealthier members of the population. In addition, persons with lower levels of education and those with lower incomes tend to be less likely to participate in smartphone-based surveys (Nitsche, Widhalm, Breuss, Brändle, & Maurer, 2014). Aside from differences in the types of respondents that each survey mode tends to capture, the characteristics of these modes tend to produce differences in responses.

The influence of the survey mode on the responses provided to a survey is referred to as survey mode bias. Broadly speaking, survey mode biases are the result of how respondents experience the survey. Survey mode bias is partially the result of the behaviours of survey respondents and their expectations. Specifically, survey respondents tend to behave as if they are engaged in a conversation; this approach tends to lead respondents to seek out visual and non-verbal cues, particularly when they are unsure about how to proceed (de Bruijne & Wijnant, 2013). The visual heuristics that are applied once a respondent has found a visual or non-verbal cue, are often the reason why the design and layout of a questionnaire or interface can affect responses (Couper, Tourangeau, Conrad, & Crawford, 2004; Toepoel & Dillman, 2011). The design of certain survey modes can affect the burden experienced by the respondents, which in turn can lead them to seek out cues to determine how to proceed. Respondent-friendly designs can help to mitigate the impacts of survey mode bias, and can also be addressed by ensuring that instructions and questions are understandable, clear, concise, and free of jargon (Dillman & Smyth, 2007). Besides, respondent-friendly designs can help address measurement errors, which arise due to differences in the true value of an attribute and what value reported by the respondent (Habib K. , 2014). For a more detailed discussion of survey mode bias, see Chapter 6 of (Srikukenthiran, et al., 2018).

Herein lie two key benefits of utilizing a survey mode that includes interviewers – the ability to provide assistance and clarification to respondents and eliminating the need for respondents to navigate the survey interface. In addition, the ability of an interviewer to prompt respondents for additional information in real time has the potential to address the under-reporting of trips and rounding of travel times that tend to be present in self-reported surveys (Zhao, et al., 2015). However, survey modes that include an interviewer are not without their faults. The need to hire and train interviewers creates additional costs that may not be incurred with smartphone- and web-based surveys, making them costly to conduct on a large scale. In addition, the presence of an interviewer can lead to social desirability bias, where respondents may provide answers that are more socially acceptable (Stern, Bilgen, & Dillman, 2014).

In response to the declining rates of landline ownership, some large-scale travel surveys have begun taking a multi-modal approach to data collection. The decision to incorporate more than one mode into the conduct of a survey is often motivated by a desire to improve response rates. Increasing response rates can reduce the potential for non-response bias, which arises due to the existence of systematic differences between respondents and non-respondents (Habib K. , 2014). This approach to survey administration is seen in the 2016 iteration of the TTS, where responses were obtained using both CATI- and CAWI-based methods. Multi-modal surveys aim to exploit the strengths of the individual survey modes, while mitigating the impacts of the corresponding weaknesses. The main benefits of multi-modal surveys are the potential for each mode to attract different types of respondents (Beebe, et al., 2012) and the ability of some modes to reduce response burden (Bayart & Morency, 2008).

The administration of a multi-modal survey often takes one of two approaches. The first approach is to divide the survey into individual segments, where the survey mode may vary from one segment to the next. This approach is used to limit response burden and improve the quality of the data (Bavdaz, Giesen, Cerne, Lofgren, & Raymond-Blaess, 2015). The other approach is to allow respondents to choose the mode through which they will complete the survey. The idea behind this approach is that some respondents may only be willing to participate in a survey if a specific mode is offered, or conversely, that some respondents may not be willing to use a particular mode. The influence of survey mode on the decision to participate in a survey has the potential to introduce bias into the survey data, due to the potential for self-selection to reduce the randomness of the sample (Bayart & Morency, 2008).

The use of multiple survey modes to collect data also raises the question of how these different sets of data should be combined. Each survey mode comes with its own inherent set of biases, each of which can have varying impacts on the survey data. The unique attributes of each survey mode mean that the mode-specific biases are likely to vary across the different modes. In addition, the impact of the survey mode on the choice of sample frame can lead to differences in coverage and non-response errors that are inherent in the data collected through each mode. Aside from these issues, the use of a multi-modal approach to data collection has produced mixed results. Specifically, including an additional mode into the design of a survey is not guaranteed to address errors related to non-response. Given the potential for multi-modal surveys to address the coverage and participation issues associated with CATI surveys, the issues associated with the harmonization of the data obtained through different modes may be a necessary evil. Furthermore, the inability of any single survey mode to provide an adequate representation of the target population makes multi-modal surveys a viable approach to conducting a travel survey. This comes with the caveat that more work must be done to develop methods to harmonize data obtained from different survey modes.

2.2.3 Response Rates

Many large-scale surveys utilize response rates as a performance metric. In a survey, the response rate is usually defined as the number of interviews completed with each sampling unit, divided by the number of units in the sample (The American Association for Public Opinion Research, 2011). The emphasis placed on response rates partially stems from its influence on non-response bias. Although there is not a well-defined relationship between the two, lower response rates tend to increase the potential for non-response bias (Massey & Tourangeau, 2013). Consequently, improving response rates is one of the more reliable means of reducing non-response bias. Additionally, improving response rates can reduce the number of survey invitations required to obtain the target sample size.

There are a variety of factors that can impact the response rate of a survey, both for those who have and have not begun the survey. The topic of the survey, as well as the pertinence of the topic to prospective respondents, tend to have a major impact on response rates (Sills & Song, 2002; Alsnih, 2006). Increasing the frequency of interactions with the survey invitees, and personalizing survey invitations, can also have a positive impact on response rates (Parsons, 2007; Messer & Dillman, 2011). The socio-economic attributes of the survey invitees can also influence their decision to participate in a survey. For example, larger households or households whose members travel more frequently are generally more likely to be non-respondents, due to the burden associated with having to recall and report the travel of each member (Stopher & Greaves, 2007). The choice of survey mode can also have a significant impact on response rates, in part due to the level of technical literacy inherently required by some modes (de Bruijne & Wijnant, 2013).

Among invitees who have started the survey, terminations pose the greatest threat to response rates. In surveys, the two most common reasons for termination are participant frustration and technical issues. Yan & Tourangeau (2008) argue that the process of responding to a survey can be broken down into four components: comprehending the question, retrieving the relevant information, using the information to answer the question, and reporting the answer. During the comprehension and reporting stages, the design of the survey interface and the wording of questions can place a burden on the respondent, which may motivate them to terminate their participation.

Given the emphasis placed on response rates in the administration of surveys, there have been numerous investigations into methods to improve response rates. For invitees who have already begun the survey, there has been a variety of work done to devise survey design guidelines that aim to improve the usability of the survey interface and the clarity of the questions. For invitees that have not started the survey, prior studies have understandably focused on the role that the invitation plays in the decision to participate. Common approaches to persuade invitees to participate in a survey are to reassure them that the survey will not take too long (Andrews, Nonnecke, & Preece, 2003), to ensure that the invitation is distinct from spam (Pan, 2010), and to offer a paid-in-advance incentive (Millar & Dillman, 2011). Incentives are often offered to encourage participation or to increase the respondent's tolerance for burden (Singer & Ye, 2013).

In particular, the impacts of incentives have been studied extensively, with the results being somewhat mixed. Millar and Dillman (2011) found that offering a $5 cash incentive improved the response rates to their web- and mail-based surveys by 17.9% and 20.6%, respectively. Conversely, the administrators of the 2012 Utah Travel Study offered respondents who had started, but had not completed, the survey a $10 Amazon gift card to complete the survey; this only resulted in 1% of such respondents returning to complete the survey (Resource Systems Group, Inc., 2013). Overall, it appears that incentives tend to have a positive impact on response rates, with money being preferred to gifts and pre-paid incentives being preferred over promised compensation or lotteries (Massey & Tourangeau, 2013; Huegy, et al., 2014). Despite the positive impacts of incentives on response rates, some studies caution that offering incentives may adversely impact data quality. Specifically, respondents who participate in a survey primarily to receive the incentive may rush to complete the survey, which has the potential to produce relatively poor data.

2.3 Summary and Conclusions

The traditional household travel survey framework is becoming increasingly inadequate, due to a combination of technological change and the desire for more travel data. The need to re-examine the conduct of travel surveys also creates the opportunity to modernize the conduct to household travel surveys and exploit the emergence of new technologies. As discussed in this chapter, there are a number of issues that must be addressed when designing a travel survey, including the sampling strategy, the sample frame, and the survey mode. This thesis proposes a new, modular framework for the collection of passenger travel data. The proposed framework aims to integrate the collection of more detailed data, often done through purpose-built surveys, with the conduct of household travel surveys. The integration of these different sources of travel data aims to provide a means of creating a more comprehensive set of passenger travel data that can be used to satiate contemporary data needs and provide a more holistic understanding of travel behaviour.

Chapter 3
The Core-Satellite Framework for Data Collection

The practice of travel behaviour analysis is expanding beyond the traditional four-stage model and the study of weekday commuting behaviour. Contemporary approaches to the analysis of travel behaviour tend to place more of a focus on travel as a consequence of the desire to participate in out-of-home activities. Besides, the analysis of travel behaviour is also beginning to consider factors such as attitudes and stated choice information. Also, the rise of the sharing economy has led to the desire to understand the adoption of emerging modes, such as bike-sharing and ride-hailing, and the factors that influence mobility tool ownership. In response to the constantly evolving data needs for contemporary travel demand analysis, Goulias, Pendyala, & Bhat (2011) proposed the core-satellite data collection paradigm (CSDCP), as shown in Figure 4. This framework was devised to support a modular simulation model system, which was developed to support policy needs in California. The structure of the framework is a reflection of the need to capture both individual and group behaviours, including the spatial, temporal, and social contexts in which they take place (Goulias, Pendyala, & Bhat, 2011).

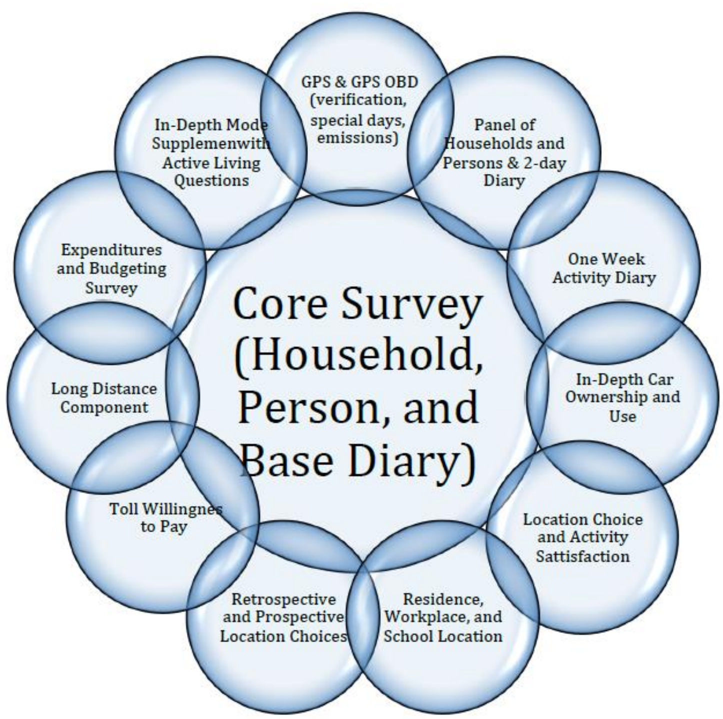


Figure 4: The Core-Satellite Data Collection Paradigm (Goulias, Pendyala, & Bhat, 2011)

This chapter presents a proposed framework for the collection of passenger travel data that is built upon the core-satellite framework proposed by Goulias, Pendyala, & Bhat (2011), i.e. the CSDCP. First, the components of the original core-satellite paradigm are summarized. Next, the components of the proposed framework are presented. Then, the factors that must be considered when designing and conducting satellite surveys are summarized. The role of passive data within the framework is discussed before considerations pertaining to the harmonization of the different sets of data are presented.

3.1 The Core-Satellite Data Collection Paradigm

This section presents an overview of the core-satellite data collection paradigm described in Goulias, Pendyala, & Bhat (2011). First, the factors that led the authors to propose a new passenger travel data collection paradigm are discussed. Then, the main components of the paradigm – the core survey, satellite surveys, and complementary datasets are summarized. The role of each of these components within the survey paradigm is then described.

3.1.1 Motivation

The development of the core-satellite data collection paradigm (CSDCP) arose, in part, as a response to the shortcomings of the traditional household travel survey framework. The CSDCP aims to address both traditional and contemporary issues with survey-based data collection methods. One of these issues is non-response, which can take the form of item (question) non-response and unit (survey) non-response (Tourangeau, Groves, & Redline, 2010). While recruitment techniques can contribute to item non-response, response burden can affect both item and unit non-response. The need for more detailed data, in order to support the development of models at a more disaggregate level, has the potential to further increase rates of non-response. Approaches to obtain more detailed travel information can be done through the inclusion of additional questions or including questions that are relatively more complex; both of these approaches have the potential to increase response burden.

Response burden tends to have an adverse impact on both response rates and data quality. This burden is the product of the effort required to complete a questionnaire, which is influenced by the length of the questionnaire and the design of a survey instrument (Rolstad, Adler, & Ryden, 2011). Although Rolstad, Adler, & Ryden (2011) found a relatively weak association between questionnaire length and response burden, the former has either a neutral or adverse impact on burden. In addition, the length of the questions and the number of response options presented to respondents has been shown to increase the burden experienced by the respondents (Yan & Tourangeau, 2008).The main goal of the CSDCP is to minimize the time and costs required to collect sufficiently detailed passenger travel data (Goulias, Pendyala, & Bhat, 2011).

The core-satellite paradigm aims to achieve this goal by shifting the collection of data from a single, large-scale survey to a relatively large survey (the “core survey”) that is augmented by one or more smaller, purpose-specific (“satellite”) surveys. The existence of satellite surveys within the framework can reduce the burden experienced by the respondents of the core survey by allowing its design to focus on the collection of key data, as defined by the survey administrators. In addition, the incorporation of satellite surveys into the travel survey framework can contribute to shorter questionnaires and can allow for the use of smaller, more general response options. Consequently, the use of satellite surveys has the potential to have a positive impact on the quality of the data obtained through the two types of surveys. The division of the data collection process into core and satellite surveys allows survey administrators to reduce the burden experienced by the respondents, which has the potential to improve response rates and reduce the cost of data collection.

3.1.2 The Roles if the Core and Satellite Surveys

One of the main benefits of the core-satellite data collection paradigm is the ability to take a more targeted approach to data collection, thereby redistributing the burden associated with completing the survey. Applying the core-satellite paradigm to the design of a travel survey allows the design of the core and satellite surveys to be tailored to their respective purposes, allowing the design of the two types of surveys to complement one another.

Within the CSDCP, the core survey is a relatively large-sample survey that collects information pertaining to key aspects of travel behaviour, as defined by the survey administrators, from the target population (Miller E. , et al., 2011). Here, the core survey is analogous to a standalone travel survey, in that it aims to collect all of the data required to understand the fundamental aspects of the travel behaviour of the target population. The key difference between a typical household travel survey and a core survey is the use of one or more satellite surveys to reduce the onus placed on the core survey to obtain all of the necessary data. Thus, the CSDCP allows the core survey to focus solely on collecting the data required to facilitate the understanding and analysis of fundamental aspects of travel demand and behaviour; this tends to be reflected in questionnaires that are shorter than standalone surveys. Core surveys are used to collect data that are fundamental to policy and/ or planning needs, as well as the data needed for activity-based and other approaches to travel demand forecasting (Goulias, Pendyala, & Bhat, 2011). Core surveys also collect information that will be required for the data obtained from satellite surveys to be linked to the core survey dataset. The sample size of a core survey must be large enough to allow statistically valid inferences to be made about the target population as a whole (Miller E. , et al., 2011).

Satellite surveys, on the other hand, primarily serve to enrich or supplement the core dataset by addressing gaps in the core dataset or by collecting data pertaining to a specific subpopulation, defined either based on socio-demographic attributes or common behaviours. Compared to the core survey, satellite surveys are smaller-sample, focused surveys that aim to obtain additional information on specific behaviours of interest (Miller E. , et al., 2011). Collecting data in this manner allows questions that are only relevant to a subset of respondents to be removed from the core survey. In addition, this approach allows the surveyors to ask questions in a greater level of detail than could be accommodated by the core survey. Furthermore, the use of satellite surveys allows the surveyors to include more questions than if the data were collected through the core survey. Each satellite survey is designed to fulfill a set of objectives pertaining to a specific subpopulation or behaviour and is meant to redistribute response burden amongst the respondents.

One of the key motivators for the conduct of satellite surveys is their ability to address gaps in the core dataset, by collecting information that would otherwise “not be feasible and/or cost-effective to collect as part of the core” (Miller E. , et al., 2011). Thus, satellite surveys can be used to obtain information from a specific subpopulation, such as post-secondary students or cyclists, without adding additional burden to respondents who do not belong to the subpopulation of interest. Incorporating satellite surveys into the data collection framework allows survey burden to be distributed among different groups of respondents, which can have a positive impact on response rates and, by extension, survey costs (Goulias, Pendyala, & Bhat, 2011). Examples of satellite surveys include surveys of cyclists, multi-day trip diaries, residential mobility, and dwelling type choice surveys (Goulias, Pendyala, & Bhat, 2011). Regardless of the specific objective of the satellite survey, it must collect data that can facilitate linkage to the core dataset (Miller E. , et al., 2011). The nature of satellite surveys highlights the need to tailor the design of any survey to suit the members of the target population and to design the data collection effort in a holistic manner. Taking this approach to the design of a data collection program allows survey administrators to exploit the strengths of the core-satellite paradigm, reduce redundancy, and to take a proactive approach to addressing possible issues.

The third component of the core-satellite data collection paradigm is the complementary dataset, which can be in the form of passive data or data obtained through surveys. Within this framework, complementary datasets primarily serve to augment the data obtained through the core and satellite surveys. However, complementary datasets may not contain data that can facilitate linkages to the core or satellite surveys (Miller E. , et al., 2011). Examples of complementary datasets include land use data, infrastructure data, travel time and cost data, indicators of industry-specific presence, smartcard transaction records, and cordon counts (Goulias, Pendyala, & Bhat, 2011). Interest in the use of "non-traditional" datasets has partially been motivated by the increase in the availability of third-party datasets, including mobile phone signal traces, GPS data, transit smartcard data, and credit card spending patterns (Kressner & Garrow, 2014).

The definition of a gap in a dataset is based on the desired application(s) of said dataset. First and foremost, gaps can arise from the use of an existing dataset for a purpose for which it was not meant. This can be the result of cost constraints or a reflection of the goals of the original survey. Gaps can also arise due to the under-representation of a particular subpopulation in the data set. This type of gap can be the result of the sample frame that was used, the sampling technique that was used, or the number of samples obtained by the original surveyors. In order to ensure that gaps do not exist in a dataset, the design of the core and satellite surveys should be informed by the desired applications of the resulting datasets.

3.2 The Proposed Framework

In response to key shortcomings in the original core-satellite data collection paradigm, an expanded core-satellite framework was developed. The factors that motivated the expansion of the original paradigm are described in this section. In addition, the individual components of the expanded framework are presented. For each of these components, their role within the expanded framework is discussed and the factors that influence their implementation are summarized.

3.2.1 Motivation

As discussed in Chapter 2, the traditional household travel survey framework is producing increasingly inadequate data. This inadequacy can be attributed to the representation issues that arise from a continued reliance on landline telephones as both a survey mode and sample frame, and the need for a greater variety of data at greater levels of detail. At the same time, CATI-based surveys are still a means of administering a relatively complex questionnaire without subjecting respondents to excessive amounts of burden. Consequently, any new framework for passenger data collection should look to exploit the strengths of traditional methods, while seeking out new approaches to address both current and long-standing issues with travel surveys. The proposed framework aims to provide a modular approach to the collection of passenger travel data that is capable of addressing both changing data needs and representation issues.

The data collection framework presented in Section 3.2.2 aims to build on the CSDCP proposed by Goulias, Pendyala, & Bhat (2011) and address the shortcomings of the paradigm. The original core-satellite paradigm has two inherent issues. First, it appears that the core survey is considered to be sufficiently representative of the target population, which tends not to be true for household travel surveys. Within the existing framework, the satellite surveys serve solely to supplement or augment the data obtained through the core survey. Consideration is not given to situations where gaps in the core survey need to be addressed. Second, the framework outlined in Goulias, Pendyala, & Bhat (2011) does not distinguish between different types of satellite surveys. This distinction is important, as the purpose of these surveys will influence their design and can affect the extent to which the data can be linked to the core survey (Srikukenthiran, et al., 2018).

3.2.2 Key Modifications to the Core-Satellite Data Collection Paradigm

A key component of the TTS 2.0 project was the adoption and expansion of the core-satellite data collection paradigm for use in the Greater Golden Horseshoe Area. Adjustments were made to the existing CSDCP proposed by Goulias, Pendyala, & Bhat (2011) to distinguish between different types of core and satellite surveys and to accommodate the need for surveys that are meant to address issues with the core survey. For a more detailed description of the expanded core-satellite framework, see Chapter 3 of (Srikukenthiran, et al., 2018). The expanded core-satellite framework is shown in Figure 5. The key feature of the expanded framework is the distinction between different types of core and satellite surveys.

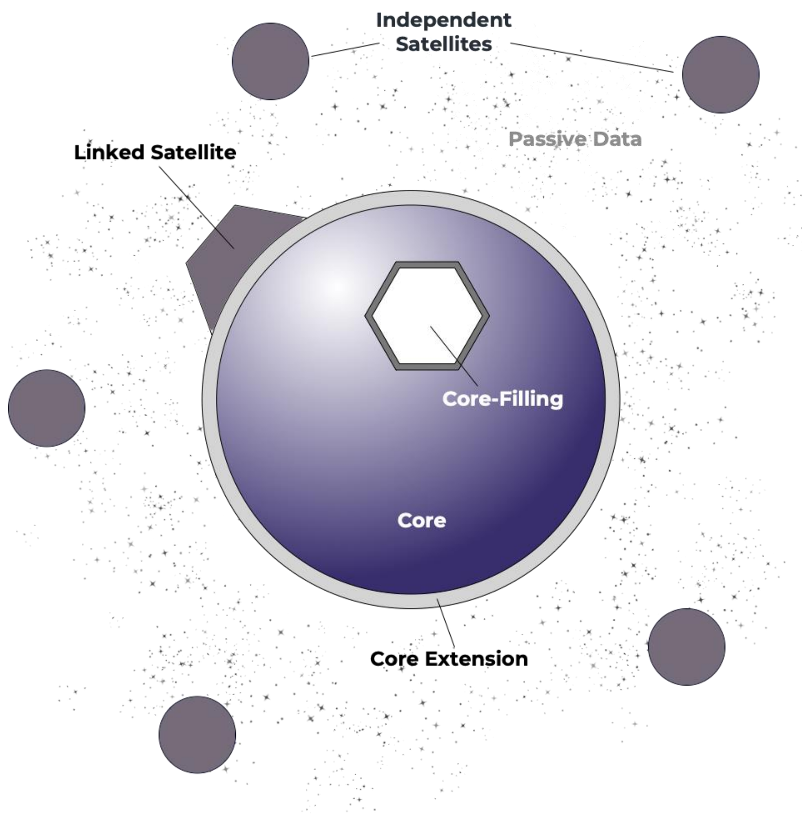


Figure 5: The Expanded Core-Satellite Framework (Srikukenthiran, et al., 2018)

Within the CSDCP, core surveys are used to obtain information from a representative sample of the population in order to support fundamental planning and policy needs and to facilitate the development of an activity-based travel demand model. In the expanded core-satellite framework, three types of core surveys (the main core, core-filling, and core-extension) are defined based on their unique objectives. The main core survey is analogous to the core survey in the original CSDCP or standalone travel surveys in the traditional travel survey framework. The purpose of the core-filling survey is to address issues of demographic or geographic under-representation in the core survey data. On the other hand, the goal of a core-extension survey is to obtain additional information from a sub-sample of the core respondents. Similarly, two types of satellite surveys are defined in the expanded core-satellite framework: linked satellites and independent satellites. The differences between the two types of satellite surveys are based on the ability to link the satellite data to the core data. The different types of core and satellite surveys are discussed in more detail in the following sections. Given the increase in the number of types of core and satellite surveys, a flow chart is provided in Figure 6. To help summarize the purpose and context of each type of survey.

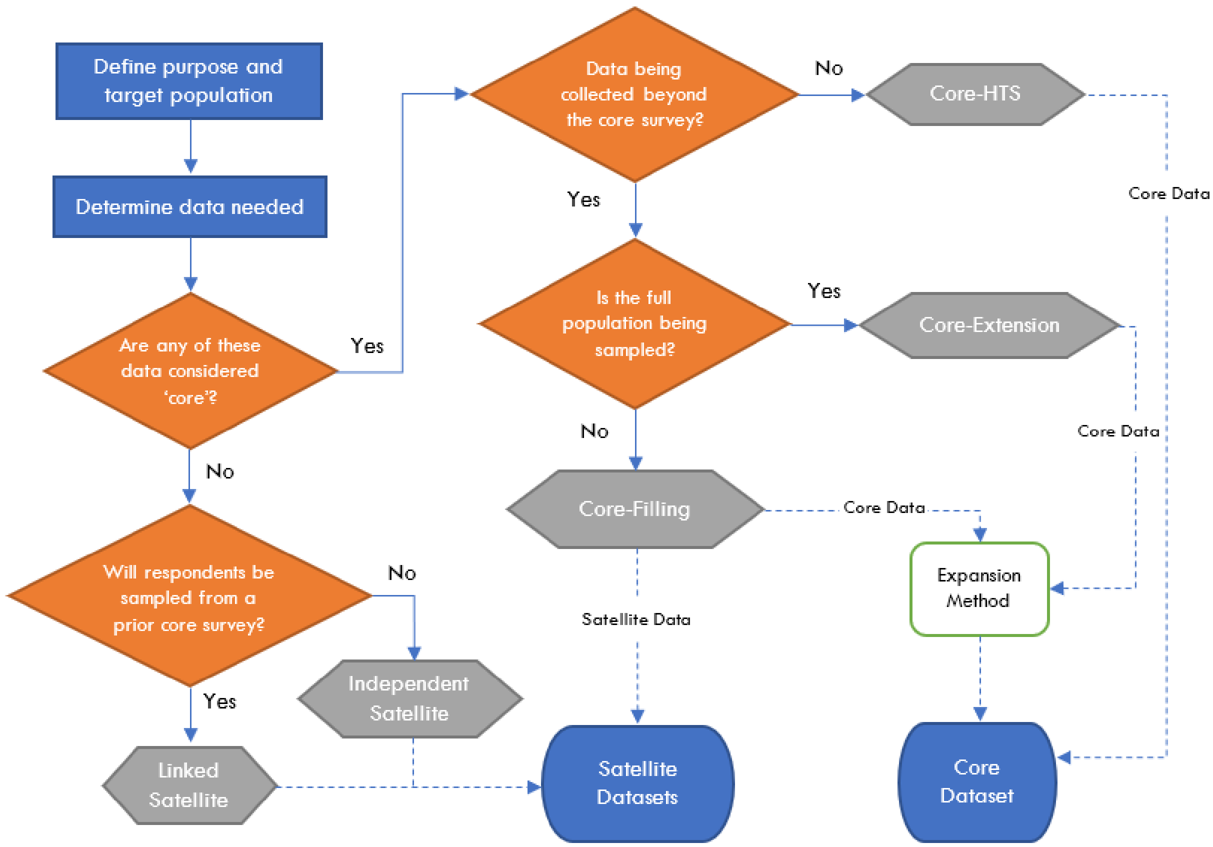


Figure 6: The Purpose and Context of Each Core and Satellite Survey (Srikukenthiran, et al., 2018)

3.2.2.1 Core Surveys

Within the expanded core-satellite framework, the information obtained through each of the three types of core surveys should still reflect the key behaviours of the target population that the stakeholders wish to capture. The expanded framework contains three surveys that are categorized as "core" surveys – the main core survey, core-filling surveys, and core-extension surveys. The main core survey in the expanded framework is analogous to the core survey in the CSDCP and is often the first attempt to obtain core data. The purpose of the main core survey is to obtain information that can provide a general understanding of the travel behaviour of the target population. This often involves the collection of information that can support travel demand modelling and planning needs, as well as information that can support the application of the core-satellite framework. The basic set of information to support these goals includes (Srikukenthiran, et al., 2018):

- Demographic information and other relevant information to facilitate the expansion of the survey data and linkages between the core and satellite surveys;
- Permission to follow-up with respondents to invite them to participate in satellite surveys (including a method of contact); and
- A travel diary to obtain information on the travel patterns of the population, with an emphasis on commuting trips to support infrastructure planning

A key consideration in the design of the core survey is whether certain information needs to be collected from the entire sample, or if obtaining the information from a random sub-sample would be sufficient. In cases where obtaining the information from a random sub-sample will suffice, a core-extension survey can be used.

The main benefit of a core-extension survey is their ability to obtain additional information from core respondents without increasing the survey length for all respondents. Key information can still be obtained from a sub-sample that is large enough to allow statistically valid inferences to be made, but without subjecting all core respondents to increased burden. Core-extension surveys are conducted by appending additional questions to the main core survey for a random sub-sample of respondents. Examples of core-extension surveys include long-distance travel surveys and attitudinal surveys. The design of a core-extension survey requires careful consideration, as the number and complexity of questions can increase response burden and jeopardize completion rates. The data obtained through core-extension surveys can be appended to the records obtained through the main core survey.

Within the expanded framework, the third type of core survey is the core-filling survey. The goal of this type of survey is to address representation issues within the main core survey. As discussed earlier, most contemporary survey modes do not provide a representative sample of the population on their own, and large-scale travel surveys tend to experience age-related representation issues. These issues can call the reliability of the survey results into question, and often cannot be addressed by simple data expansion methods. There are two key differences in the conduct of core-filling and core-extension surveys. In contrast to core-extension surveys, core-filling surveys aim to collect information that is similar to that of the main core survey. Additionally, core-filling surveys target the members of under-represented demographic groups, whereas core-extension surveys are administered to a random sub-sample of core survey respondents.

A core-filling survey should take place within a reasonable amount of time following the main core survey. Conducting this type of survey in conjunction with the main core survey, or shortly thereafter, can help to minimize temporal biases. The main issue with this approach is the need to predict under-representation issues beforehand. Conversely, a core-filling survey can be conducted after the completion of the main core survey, where representation issues can be explicitly identified. While the concept of a core-filling survey is relatively novel, similar types of surveys exist in practice, although without the explicit intent to fill gaps in a core survey. The most common type of core-filling survey is the post-secondary student survey. The StudentMoveTO (SMTO) survey, conducted among four Toronto universities in 2015, is one example of this type of survey. SMTO obtained information on travel and activity participation and included a one-day travel diary (StudentMoveTO, 2015). The conduct of this survey was motivated by the under-representation of post-secondary students and persons between the ages of 19 and 30 in the TTS.

3.2.2.2 Satellite Surveys

One of the key benefits of the core-satellite paradigm is the ability to create and exploit the synergistic relationship between the core and satellite surveys. While satellite surveys are administered to members of the target population, the design of a satellite survey will also be informed by whether records in the satellite dataset can be directly linked to the core survey. This question leads to the definition of two types of satellite surveys – linked satellites and independent satellites.

Linked satellite surveys are characterized by their ability to directly link responses to records in the core dataset. The difference between a linked satellite and a core-extension survey is the approach that is used to select participants. While core-extension surveys are administered to a random sub-sample of core respondents, the recruitment for a linked satellite is informed by the objective of the survey. Examples include the targeting of core respondents who are cyclists or post-secondary students. A linked satellite survey can be administered by appending additional questions to the main core survey, or as a follow-up survey. The data obtained through linked satellites can be appended to the corresponding record in the core dataset or kept as an independent dataset with a unique identifier that can link the two records together.

Independent satellites are characterized by the lack of a direct link to the core dataset. Despite the lack of a direct link, the data obtained through independent satellites are obtained from members of the target population and can provide valuable insights. The lack of information that can be used to link independent satellite data to the core dataset, however, means that data fusion methods may need to be applied to harmonize the two datasets. This is discussed in further detail in Section 3.7. Examples of independent satellite surveys include intercept surveys of transit users and cyclists and HOV usage surveys (Srikukenthiran, et al., 2018).

3.3 Principles for the Design of Satellite Surveys

Broadly speaking, both the CSDCP and the expanded core-satellite framework emphasize the content of the survey, rather than the use of any one survey mode (Miller E. , et al., 2011). Thus, the design of a satellite survey should be based on the desire to facilitate an approach to travel demand analysis that the core survey cannot or to address a gap in the core dataset. Because satellite surveys exist to supplement or augment the data collected in the core survey, they should be designed to ensure that the data are compatible with the core dataset. As satellite surveys take a more targeted approach to data collection, their design should be informed by the characteristics of the target population and the conditions under which respondents will participate in the survey.

The choice of survey instrument requires somewhat more consideration when designing a satellite survey than a core survey. While most core surveys utilize a questionnaire to obtain information, the targeted and purpose-specific nature of satellite surveys allow other survey instruments to be used, such as passive tracking via smartphones. Overall, the choice of a survey instrument should be informed by how the respondents will be recruited and how they will interact with the instrument. As with the design of any survey, the design of the instrument should aim to minimize response burden. Although questionnaire length is often regarded as positively correlated with response burden, respondents have been shown to tolerate longer questionnaires if they find its contents particularly relevant (Rolstad, Adler, & Ryden, 2011). In addition to the pertinence of the topic, factors such as education, technical aptitude, question characteristics, and the number of response options also affect the perception of burden and survey quality (Yan & Tourangeau, 2008).

The choice of survey mode is also an important design decision, as it can affect non-response bias and measurement error. In addition, the design of the survey mode should take the characteristics of the target population into consideration. For example, questions should be worded to ensure that they are clear and can be understood by the respondents, in order to reduce the effort required to comprehend the questions; this can help to reduce measurement errors that arise due to the misinterpretation of questions. Similarly, the survey instrument should be designed with usability in mind, and should consider the technical aptitude, literacy, and possible time constraints of the respondents. Given the number of different populations that can be the target of a satellite survey, a number of design conventions exist for satellite surveys.

3.4 Potential Satellite Surveys

Although the core-satellite framework is still a relatively novel concept, there are a number of existing purpose-specific travel surveys that can potentially fit into this framework as satellite surveys. In this section, four types of potential satellite surveys are described: transit on-board surveys, active mode user surveys, post-secondary student surveys, and employee surveys. For each of these four surveys, the typical sampling techniques and survey instruments are discussed and the application of the data obtained through each survey are summarized.

3.4.1 Transit On-Board Surveys

Transit on-board surveys are conducted by transit agencies in order to collect information about their customers, including demographics, travel patterns, and their perceptions of the service being provided by the agency. Transit on-board surveys are often conducted onboard transit vehicles, at transit stops, and within transit stations, either through the use of in-person interviews or self-administered questionnaires. This approach to data collection tends to produce information that is more detailed, accurate, and reliable than the data obtained through other survey modes.

A key difference between transit on-board surveys and traditional travel surveys is the use of transit trips as the sampling unit, rather than households or individuals. The choice of sampling units is influenced by the inability to ensure that any given customer is not invited to participate in the survey more than once. As a result, sampling for these surveys often involves the selection of a subset of route trips, in which trips along a route may be stratified based on attributes such as route direction, time-of-day, and route number (Bernardin, Lochmueller and Associates, 2010). The approach to sampling is also influenced by the choice of survey mode. The use of computer-assisted personal interviews creates the need to invite a random sample of passengers to participate in the survey.

For transit on-board surveys, the choice of survey mode should be informed by the length of the questionnaire and the context in which passengers are invited to participate in the survey. When paper questionnaires are used, a key consideration is the manner in which they will be returned to the survey administrators. Some surveys, such as the 2009 IndyGO On-Board Survey included a business reply mail permit to allow respondents to complete the questionnaire at their leisure (Bernardin, Lochmueller and Associates, 2010). On the other hand, surveys such as the 2015 On-Board survey conducted by Tri-Met in Portland gave respondents the option of depositing completed questionnaires into designated boxes or returning them to the surveyors (McHugh, Dong, Recker, & Shank, 2017). Giving respondents the option to mail completed questionnaires back to the survey administrators can help address the issues associated with obtaining data from people who are making short trips (Simas-Olivera & Casas, 2010). When respondents are asked to complete the survey before they alight from the transit vehicle, the questionnaire should be designed to be as simple as possible. This should also be the case when the survey is conducted using CAPI, as interviews must be completed before the respondents reaches their alighting stop (McHugh, Dong, Recker, & Shank, 2017).

There is a common set of demographic and travel information that appears to be collected by the majority of transit on-board surveys. Based on a review of 150 transit on-board surveys, the American Public Transit Association (APTA) found that the majority of agencies asked respondents to provide their age, gender, ethnicity, household income, household size, occupation, vehicle availability, and vehicle ownership. With regards to travel data, agencies often ask respondents to report their access and egress mode, alternative modes that they could have used for the trip, the duration and frequency of their transit usage, the frequency with which they transfer between routes, and the purpose of their current trip (American Public Transportation Association, 2007). Based on the needs of transit agencies and metropolitan planning organizations (MPOs), transit on-board surveys often collect three types of data: passenger demographics, travel behaviour, and customer satisfaction (Agrawal, Granger-Bevan, Newmark, & Nixon, 2017). In addition, some transit on-board surveys also require the surveyors to collect counts of boardings and alightings in addition to administering the survey. The data obtained through on-board surveys are typically used for travel demand modelling, long-range and area-wide planning, route planning and scheduling, service design, marketing, and customer communications (Memarian, Jeong, & Uhm, 2012).

3.4.2 Active Mode User Surveys

The term "active modes" refers to bicycling and walking. The travel information obtained through traditional household travel surveys tends to under-represent the users of these modes, which is the result of several factors. One of the main causes of under-reporting is the lack of clarity of what constitutes a "trip" (Edwards, Ivey, Lipinski, & Golias, 2012) and the tendency of respondents to neglect short and discretionary trips (Son, Khattak, Chen, & Wang, 2012). This, combined with the tendency for active mode trips to be made over relatively short distances, tends to result in the under-reporting of these trips (Edwards, Ivey, Lipinski, & Golias, 2012). Active mode user surveys can help to address this issue and provide the opportunity to gain new insights into the behaviours of active mode users.

The sampling and recruitment techniques that are used in active mode user surveys are influenced by the target population and the goals of the survey. When surveyors are interested in collecting information on the behaviour and travel patterns of cyclists, participants are often recruited through intercept surveys and snowball sampling. When surveyors are interested the decision to walk or bike, the sample frame must include both users and non-users. In these cases, participants can be selected using probability sampling, non-probability sampling, or a combination of the two. The sampling and recruitment methods used in an active mode user survey should also be influenced by the desired applications of the data and whether inferences about the population as a whole will be made.

The choice of survey mode and survey instrument should be guided by the goals of the survey, the data to be collected, and the characteristics of the target population. To address the causes of the under-reporting of trips made by active modes, the definition of a trip should be clearly described, and should cater to the characteristics of the respondents. Issues related to incomplete recall and memory decay can, to a certain extent, be addressed through the design and choice of survey instrument. The survey instrument must also account for the contexts in which the members of the sample are invited to participate in the survey. For example, long questionnaires may not be suitable for intercept surveys. Active mode user surveys tend to be administered using questionnaires, both web- and paper-based, and through passive means, using smartphones or GPS readers.

Active mode user surveys are used to obtain a wide variety of data. Surveys that are conducted using passive means are most frequently used to study the factors that influence the route choices of cyclists; however, most applications have limited themselves to the study of utilitarian cyclists. This limitation tends to stem from the application of traditional econometric approaches to explain route choice behaviours. GPS data have also been used to study the effects of the built environment on the use of active modes, such as the work presented in Broach & Dill (2016). Surveys of active mode users that aim to understand the factors that influence the decision to walk or use a bicycle tend to collect stated preference (SP) data and/ or attitudinal information. A common application of SP data is to understand the factors that influence the use of different types of cycling facilities. On the other hand, attitudinal information is often used to identify factors that motivate or dissuade people from walking or cycling. In these types of surveys, respondents are either asked to identify these factors, or are asked to indicate the extent to which various factors motivate or dissuade their use of active modes.

3.4.3 Post-Secondary Student Surveys

The primary motivation for conducting travel surveys that target post-secondary students is the desire to better understand the impact that the travel of students and staff have on the areas surrounding post-secondary institutions. Although universities and colleges can have a significant impact on travel demand in a region, post-secondary students and staff members tend to be under-represented in household travel surveys (Garikapati, et al., 2016). This under-representation makes it difficult to quantify and understand the impacts of post-secondary students and staff on the travel demand of the surrounding areas. The lack of both data and frameworks that can adequately represent student travel behaviour are issues faced by MPOs when trying to incorporate post-secondary institution sub-models into their existing modelling frameworks (Garikapati, et al., 2016). A common approach to addressing the under-representation of post-secondary students is to treat them as either a member of general the population or as a low-income, one-person household. This is often inadequate, as students often have a set of mandatory trips and access to subsidized services (Huegy, et al., 2014).

Post-secondary student surveys benefit from access to one of the most comprehensive sample frames that exists in survey practice – a register of student email addresses. The increasing prevalence of internet access, particularly on university and college campuses, makes email an effective and efficient means of reaching the entirety of the student population. Furthermore, the availability of information on the demographics of each student, such as age or student status, allows for the arrangement of students into strata. The use of email list servers to recruit students to participate in travel surveys is a fairly common practice in the literature (Akar & Clifton, 2009; Verreault & Morency, 2016). Taking this approach to sampling can help to mitigate the role that the transient nature of student populations has on their under-representation in traditional travel surveys. When stratified sampling techniques are applied, students tend to be categorized based on student status (i.e. graduate vs. undergraduate) and by residential location (on-campus vs. off-campus).

The availability of email list servers as a sample frame, combined with the relatively high level of technical aptitude of post-secondary students, leads most surveys of this nature to be conducted using web-based tools. In addition, the use of web-based interfaces to conduct these surveys can partially be attributed to the desire to exploit their ability to reduce burden, improve usability, and ensure that the survey is accessibility-compliant (Volosin, et al., 2014). The choice of survey mode, however, should be informed by the methods through which students are recruited. Recruitment that is done through non-electronic means must account for the burden that would be incurred by prospective respondents if they must access a webpage to participate in the survey. In addition, good advertising and offering incentives can have a positive impact on response rates (Huegy, et al., 2014).

Because post-secondary students have traditionally been under-represented in household travel surveys, the most common application of this type of survey is to understand the travel of post-secondary students. This is often done by using the survey data to understand the travel characteristics of the students (e.g. modal shares, trip rates, trip purpose) and to estimate travel demand models. Integrating models of post-secondary student travel into existing travel demand models has a variety of both short- and long-term benefits. The creation of models of post-secondary student travel provides the opportunity to gain insights into the factors that affect various aspect of their preferences and behaviours and creates the opportunity to forecast and analyze the impacts of policies geared towards post-secondary students.

3.4.4 Establishment Surveys

The use of establishment surveys began in the mid-1980s and coincided with the development of regional travel demand models. Establishment surveys are a specific category of special generator surveys. Also known as workplace surveys, their goal is to collect information related to trips that are made to workplaces and similar establishments. Chapter 18 of the Travel Survey Manual, published by the Transportation Research Board, provides a detailed set of guidelines for the design and conduct of workplace surveys (Southwell, Zhang, & Sharp, 2014). The key points of these guidelines are presented in this section.

Southwell, Zhang, & Sharp (2014) summarize the steps that an establishment survey takes when data is collected from both employees and visitors:

1. Call each establishment in the sample to determine if it is still in business, to verify its address, and to establish a contact
2. Send a recruiting letter to each employer
3. Interview and recruit the employer, establish a contact person
4. Schedule the survey day
5. Schedule an in-person site visit
6. Remind the contact person at each business to deliver the employee questionnaire

Due to the nature of establishment surveys, two types of sampling units are used – individuals and establishments. The selection of establishments usually takes a stratified random sampling approach, with strata being defined based on factors such as location, industry sector, and the number of employees. After establishments are selected, the decision of whether to target employees, visitors, or both must be made. Similar to the choice of sampling units, the recruitment of both employers and individuals is required. In the recruitment process, non-response can be reduced by soliciting support from the local Chamber of Commerce or business associations, contacting the most senior manager possible, and focusing on recruiting larger firms (Southwell, Zhang, & Sharp, 2014). During the recruitment process, the survey team should draft a letter that is then delivered by the Chamber of Commerce or business association, in order to improve the credibility of the invitation. The survey team should also create an invitation to recruit the employees, with an emphasis on their employer's support for the survey and the contact information of the survey liaison at the company.

Establishment surveys are often administered in one of three ways – visitor and employee intercept surveys with random selection, centralized employee surveys, or a combined visitor intercept and centralized employee survey. When an establishment survey includes an intercept component, it is often conducted using self-administered questionnaires, pen-and-paper interviews, and the use of CAPI software (Southwell, Zhang, & Sharp, 2014). When conducting a centralized employee survey, the use of a web-based interface to obtain travel information from employees tends to produce results that are statistically similar to those obtained from a traditional travel diary (Petrunoff, Xu, Rissel, Wen, & Van der Ploeg, 2013).

The data obtained through establishment surveys often fall into one of three categories: establishment information, employee information, and visitor information. Applications of establishment survey data include the study of the traffic impacts of an establishment, congestion management, and trip reduction programs. The most common application of establishment survey data is to determine trip attraction rates, which are often used as inputs to travel demand models.

3.5 The Role of Passive Data

The increasing availability of passive data from third-party sources has driven efforts to exploit these data to better understand travel behaviour. Passive data, defined as data that are collected without explicit input from subjects (Matsuda, Rosenstein, Scovitch, & Takamura, 1998), is creating new opportunities to obtain information that could not be collected through traditional means. As a result, the growing prevalence of passive data has the potential to provide new insights into the factors that influence travel behaviour. In this section, the advantages and issues associated with the use of four types of passive data are presented.

3.5.1 GPS Data

GPS data are among the most frequently used set of passive data in transportation research. In addition to data obtained from GPS readers and sensors, the availability of shapefiles containing information on infrastructure, road networks, and land use continues to grow. The collection of GPS data usually involves the periodic recording of the latitude and longitude readings transmitted by a GPS sensor, in addition to the time that the recordings were logged. The costs associated with the collection of data using GPS units may limit the use of GPS data to satellite surveys within the core-satellite framework. This issue can be somewhat mitigated by the increased prevalence of smartphones, however smartphone-based data collection methods tend to over-represent younger members of the population.

The main advantage of GPS data is their ability to obtain spatiotemporal information that is more detailed than self-reported information. This can help to address issues related to memory loss and incomplete recall, which can reduce the under-reporting of short and discretionary trips (Dumont, Shalaby, & Roorda, 2012). In addition, the use of GPS data can help exploit the tendency for people to more accurately recall their activities than the start and end times of their trips (Cottrill, et al., 2013). On the other hand, the quality and reliability of data obtained from GPS devices tends to suffer from issues of cold starts, short-duration trips being missed, and the canyoning effect that occurs in dense urban areas (Shen & Stopher, 2014). Another common issue with the use of GPS devices to collect travel information is the need to impute key information, such as travel mode and trip purpose. The uncertainty associated with the imputation process necessitates that respondents validate the trip information, which has the potential to induce additional burden.

3.5.2 Cellular Data

The collection of locational information by cellular service providers for billing purposes is beginning to be exploited in transportation research. These so-called call detail records (CDRs) are logged each time a customer utilizes cellular services and records a timestamp. Although CDRs can provide spatiotemporal information, their place within the core-satellite framework should be limited to complementary datasets. The inability to obtain information on the demographics of the cellphone owners precludes the ability to identify the extent to which the dataset can represent the population of the survey area. In addition, the lack of demographic information means that CDR data cannot be linked to either core or satellite surveys.

The main benefit of CDR data is that they provide a wealth of spatiotemporal information. These data can be used to gain additional insights into mobility patterns and complement the data obtained through traditional surveys (Alexander, Jiang, Murga, & Gonzalez, 2015). Depending on the cell service provider from which the CDR information is obtained, the set of records has the potential to provide travel information from a larger proportion of the population. The key issue with the use of CDRs in transportation research is that these records are maintained by cellphone service providers, which ultimately represent a subset of the population as a whole. Additionally, CDR information is logged primarily when a person uses cellphone service, creating the potential that data will be collected at different rates from different cellphone owners. As with GPS data, trips, travel modes, and trip purposes must be inferred from the spatiotemporal data, however the lack of associated demographic information in CDRs means that there is likely no means of validating the imputed information (Ge & Fukuda, 2016).

3.5.3 Transit Smart Fare Card Data

The number of transit agencies that have adopted the use of smart fare cards (i.e. “smart cards”) continues to grow. With regards to the use of smart cards, agencies tend to implement two types of fare policies: tap-on only, where a customer taps their card on a reader when boarding a vehicle, or tap-on and tap-off, where a customer must tap their card when boarding and alighting. The fare policy used by an agency can affect the advantages and issues associated with using smart card data. Within the expanded core-satellite framework, smart card data fall into the category of complementary datasets. Smart card data also have the potential to serve as a satellite survey, with the availability of information pertaining to the owner of the smartcard determining whether it is a linked or independent satellite. Presently, the most practical application of smart card data is to provide an independent record of transit trips, which can be used as a reference value.

The main advantage of smart card data is the ability to analyze the behaviour of individuals, rather than using trips as the unit of analysis. A key benefit of smart card data is the ability to provide detailed spatiotemporal information pertaining to the ridership and passenger volumes of each route (Ji, Mishalani, & McCord, 2015). This information can be used to adjust the delivery of service and to provide a more precise means of measuring demand at both the route and passenger levels (Agard, Morency, & Trepanier, 2006; Morency, Trepanier, & Agard, 2007). The main issue with attempting to use smart card data is the fact that low penetration rates can lead to results that are not representative of transit users as a whole. As is the case with many sources of passive data, smart card data may not include demographic information about the owner. In addition, smart card data tends to provide an incomplete representation of transit trips. For systems with a tap-on policy, the alighting stop of each trip must be imputed using heuristics (Munizaga & Palma, 2012). Furthermore, the best-case scenario for smart card data is that information on the boarding and alighting stop is provided; neither the origin and destination nor the purpose of the trip is known.

3.5.4 Bluetooth Data

The number of mobile devices and vehicles that are equipped with Bluetooth technology has led to the viability of this type of passive data in transportation research. Bluetooth data is most often obtained through the use of Bluetooth readers, which collects the globally unique media access control (MAC) number of passing devices (Malinovskiy, Saunier, & Wang, 2012). Herein lies the main benefit of using Bluetooth data – the ability to obtain information on the travel characteristics of the users of a transportation network. This allows the start and end points of travel along a corridor or between cordons or screenlines to be identified. In addition, Bluetooth data can be used to monitor vehicular volumes, densities, and flows, and facilitate longitudinal analyses (Friesen & McLeod, 2015). Due to the lack of associated demographic information, Bluetooth data is regarded as a complementary dataset within the core-satellite framework. The main issue with using Bluetooth data is the need for devices to be set to be discoverable in order to be tracked by Bluetooth sensors, which can lead to small sample sizes.

3.6 Ensuring Compatibility

The benefits of constructing a data collection program in accordance with the core-satellite framework are fully realized when the data obtained through the satellite surveys can be used in conjunction with the core survey. In order to facilitate the linkage or fusion of datasets, a certain level of compatibility must exist between the datasets. The compatibility of two or more datasets is heavily dependent on the context in which the data were collected. Specifically, every set of data is characterized by three factors (Miller E. , et al., 2011):

- **The spatial context:** the location of the target population;
- **The temporal context:** the time period during which the data were collected; and
- **The semantic context:** the manner in which variables and categories are defined

Thus, in order to ensure compatibility, the different sets of data should refer to a similar spatial context, a similar temporal context, or a similar semantic context, and of course, greater compatibility is preferred. The growing demand for data at a more disaggregate level places additional pressure on data collection efforts and tends to produce at least one compatibility issues that must be addressed before the data fusion process (Bayart & Morency, 2008).

Based on the work of Judson (2006), Bayart and Morency (2008) present three principles for the integration of databases:

- **The latent variable principle:** recognizing that the estimand exists, but may not always be directly observed;
- **The uncertainty principle:** understanding that the data contributing to the estimate are not flawless; and
- **The modelling principle:** formalizing the relationship between the estimand and the source(s) of data

At their core, the intention of any data fusion method is to enrich survey data in order to better meet data needs (Bayart & Morency, 2008). This goal is achieved by using multiple datasets to supplement one another and to address the gaps in each dataset to produce a new, more comprehensive set of data (Miller E. , et al., 2011). A fundamental 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 that can facilitate the use of a matching method to impute the value(s) of the variable(s) of interest (Miller E. , et al., 2011). Bayart & Morency (2008) argued that the basic issue with the data fusion process is that statistical inferences are being made about the joint distribution of two variables without being able to directly observe the distribution. Miller et al. (2011) argue that a key impediment to ensuring a sufficient level of compatibility exists between multiple datasets is the need to address different levels of aggregation before data fusion can take place.

Whenever possible, the design of a data collection program should make every effort to ensure that the surveys will produce sets of data that are compatible with one another (D'Orazio, Di Zio, & Scanu, 2006). Ensuring spatial compatibility is relatively straightforward, as the survey areas are defined based on geographic boundaries. Similarly, compatibility in the semantic context can be created by maintaining a consistent definition of each key term and value. The greatest amount of uncertainty exists when considering compatibility in the temporal context. The only foolproof means of guaranteeing temporal compatibility is to conduct the core and satellite surveys concurrently. As a rule of thumb, shorter time lags between the conduct of the two surveys are preferable, although the occurrence of a paradigm-shifting event is a greater threat to temporal compatibility than time lag. The impact of the time lag between the core and satellite surveys has not been evaluated empirically.

The data fusion process requires that the datasets must be compatible in at least one context. However, compatibility may not initially exist between two disparate sets of data. In these situations, there are several approaches to help improve the compatibility between two datasets. Building on the work of Van der Laan (2000), D'Orazio, Di Zio, & Scanu (2006) identified 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 situations where the definition of a variable or category is inconsistent between two datasets, i.e. the semantic contexts of the two variables are not consistent, a few options exist. Variables can be re-coded or re-categorized, existing variables can be replaced with a new set of variables, or variables that cannot be harmonized can be identified so as to avoid their use as a common variable (D'Orazio, Di Zio, & Scanu, 2006). Once the sets of data are sufficiently compatible, a number of data fusion methods exist.

3.7 Overview of Data Fusion Methods

Aside from the three contexts in which sets of data exist (spatial, temporal, and semantic), the process of data fusion can be categorized into one of three contexts (Miller E. , et al., 2011): the mixed context, the survey mode context, and the data type context. The mixed context aims to account for cases where different biases arise due to the use of different survey modes. This context is becoming increasingly prominent due to the increased use of mixed-mode surveys. Within the context of transportation research, the most popular data fusion method is the weighting and expansion of survey data to represent the entirety of the target population. Data expansion is often carried out by using census data to obtain population control totals, often based on factors such as age, gender, and residential location. Bayart and Morency (2008) provides an excellent overview of data fusion methods; in general, two 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 source files are used to estimate a joint distribution of relevant variables

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

- **Exact matching:** records can be matched without uncertainty, typically based on the use of a unique identifier
- **Explicit models:** a model is used to connect variables of interest between two sets of data, based on the set of common variables
- **Implicit models:** an existing record in the donor file that is as similar as possible to the receptor file is found; constraints regarding the number of times that a particular record in the donor file is used can be defined

A number of different approaches can be taken to identify the common (matching) variables, depending on the variable in question. Common variables can be identified based on the calculation of correlation coefficients, the use of regression analysis for continuous variables, the determination of the Pearson chi-squared statistic for categorical variables, or the calculation of Somers' D for ordinal variables (D'Orazio, Di Zio, & Scanu, 2006).

Chapter 4
Case Study: Designing a Satellite Survey for the National Capital Region (NCR)

As interest in encouraging the use of active modes (defined as walking and cycling) grows, it is more important than ever to understand the factors that influence the choice to use (or not use) these modes of travel. Aside from the health benefits of walking and cycling, trips made using active modes are also more sustainable than motorized travel, making the promotion of these modes conducive to the achievement of sustainable development goals. In order to promote the use of active modes, there is a need to further study their use. This issue is partially due to the lack of clarity regarding the need to report short trips (Edwards, Ivey, Lipinski, & Golias, 2012) and the tendency for short and discretionary trips to be neglected when completing travel surveys (Son, Khattak, Chen, & Wang, 2012). Furthermore, in areas where the share of trips made by bicycle and on foot is relatively low, it is difficult to obtain a sample that is large enough to support data analysis and modelling efforts when probability (random) sampling is applied. In a cost-constrained operating environment, the ability to understand the usage (or lack thereof) of active modes can be used to inform investment and policy, which will hopefully lead to the best use of scarce funds.

In order to support the TRANS Committee in their efforts to better understand the use of active modes, it is recommended that satellite surveys focused on the use of active modes be conducted in the National Capital Region (NCR). Based on the project requirements, it is recommended that satellite survey of active mode users will be carried out using two survey instruments – a smartphone-based survey and a web-based survey. The smartphone-based survey will primarily be used to collect information on the routes and facilities used by cyclists. The web-based survey will be administered to both cyclists and non-cyclists to identify the motivators and barriers to cycling, and to solicit the attitudes of respondents towards different types of proposed network investments and modifications. Additionally, the web-based survey will ask respondents to provide the route taken for their most recent recreational trip and the route they typically take to work or school. Both instruments will include a questionnaire that will collect information to facilitate linkage to the data collected through the “core” Household Origin-Destination (OD) Survey.

This chapter summarizes the design of the web-based survey, including the factors that influenced its design and its key components. The choice to use a web-based tool to administer the survey is based on the ability to integrate a map-based interface into the survey, the relatively low marginal costs, and the potential to reduce response burden. Because this satellite survey will, among other things, attempt to identify the barriers that deter people from using a bicycle to travel, both cyclists and non-cyclists must be targeted. A sample questionnaire is provided in the appendix. The chapter is organized as follows: first the motivation and requirements of the project are summarized. Next, pertinent background information is presented. Finally, the data requirements of the survey are discussed.

4.1 Motivation and Project Requirements

Founded in 1979, the TRANS Committee was established with the directive of coordinating efforts between the major transportation planning agencies in the National Capital Region (NCR) (TRANS Committee, 2014). Comprised of six agencies, the current initiatives of the TRANS Committee include (TRANS Committee, 2014):

- The ongoing development and operation of long-term transportation forecasting models
- The collection and management of data for transportation planning
- The management of transportation studies

The collection of travel data from active mode users assists in the ongoing development of travel forecasting models, as these travellers are typically under-represented in household travel survey data.

A practical example of the utilization of satellite survey data is the development of the bicycle assignment module in the 2014 iteration of the TRANS model. The current module utilizes an approach similar to that of a typical traffic assignment model. Automobiles and bicycles are assigned iteratively to the road network, with a bicycle-specific volume-delay function being applied, and bicycles being regarded in terms of passenger car equivalents (PCE) (MMM Group Limited, 2014). The key issue is that the input data used for the model need to have sufficiently large sub-samples of all road users in order to fully capture travel demand and behaviour. The other issue, especially in context of road users that represent relatively small sample shares, is the emphasis that it placed on the utilitarian aspect of travel. While this approach is relatively standard when assigning automobiles to the road network, it may be inadequate when considering the behaviour of cyclists.

Attitudes and perceptions tend to affect the facility choices made by cyclists (Maldonado-Hinarejos, Sivakumar, & Polak, 2014). In addition, cyclists rarely set out with the sole objective of minimizing their travel time, particularly in the case of recreational cyclists. Furthermore, a number of studies regarding the route choices of cyclists have found that the chosen paths tend to be longer than the shortest-distance path, and that cyclists tend to use a mix of different facility types (Broach & Dill, 2016). Even a large-scale travel survey may not contain a sufficient number of observations to facilitate the development of route choice models. Therefore, any comprehensive approach to the modelling of these road users requires detailed behavioural data.

The collection of data regarding active mode users can also support the goals laid out in the long-term plans of the Ville de Gatineau and the City of Ottawa. In its *Strategic Plan*, the Ville de Gatineau states that sustainable development is a key theme. The *Strategic Plan* also expresses a desire to promote sustainable transportation (Ville de Gatineau, 2017). The *Strategic Plan* argues that because transportation-related emissions are a public health issue, urban development plans should discourage the use of single-passenger automobiles in favour of active and public transportation. The data collected through the active mode satellite survey can be used to develop a better understanding of the travel patterns of cyclists and pedestrians, as well as the factors that affect the decision to walk and cycle. This understanding can be used to develop strategies to encourage walking and cycling. Similarly, in the Ottawa Transportation Master Plan, "*Building a Liveable Ottawa 2031"*, the advancement of strategies to improve walking and cycling is stated as one of the plan's areas of focus (The City of Ottawa, 2013). Throughout the plan, the desire to promote modes of sustainable transportation, i.e. walking, cycling, and transit, is evident, particularly in the vision of the plan, which is to reduce automobile dependence.

The utilization of the core-satellite survey framework is a viable means of obtaining data on active mode users in the NCR. The administration of an active mode user satellite survey is meant to address the under-representation of these travellers in traditional household travel surveys. Furthermore, the conduct of said survey will allow information to be obtained from respondents in a greater level of detail than would be feasible in the core survey, i.e. the Household OD Survey. These data will contribute to the further development of the understanding of the travel patterns and behaviours of active mode users. The collection of these data can also contribute to the achievement of a number of objectives outlined in the long-term plans of the Ville de Gatineau and the City of Ottawa.

4.2 Design Considerations

This section will discuss the considerations that influenced the design of the web-based satellite survey. The web-based survey will, at minimum, include an attitudinal survey, which aims to identify barriers to cycling and the factors that motivate people to cycle. This information can be used to develop a framework to prioritize different infrastructure improvements and investments, as well as policy changes. Furthermore, the results of the survey can contribute to an improved understanding of latent cycling demand, including latent demand that may be realized due to changes in infrastructure or policies. In cases where the survey is not conducted in conjunction with the core OD Survey, or when the respondents have not participated in the core survey, additional questions can be added to create a travel diary portion of the survey. The information collected through the travel diary-style questions can be expanded to represent the target population, or potentially combined with the data from the OD Survey. On the other hand, the attitudinal information collected through the survey should not be expanded.

The choice to use a web-based tool to administer the survey is based on the ability to integrate a map-based interface into the survey, the relatively low marginal costs, and the potential to reduce response burden. Because this satellite survey will, among other things, attempt to identify the barriers that deter people from using a bicycle to travel, both cyclists and non-cyclists must be targeted.

4.2.1 Approaches Taken in Similar Surveys

In order to better understand the current state of practice, a number of recent surveys pertaining to cyclists were reviewed. A total of eight surveys were reviewed, with four taking place in Canada (Toronto, Ottawa, Calgary, and Vancouver), two in New Zealand, and one in each of London (England) and New South Wales (Australia). Each of these surveys classified its respondents based on their cycling behaviour. Respondents were typically categorized based on the frequency with which they used a bicycle, as was done in the *Cycling in Cities* survey conducted in Metro Vancouver in 2006 (Winters, Davidson, Kao, & Teschke, 2011) and the TransLink *End-of-Trip Facilities* study conducted in 2009 (ENRG Research Group, 2009).

The categorization of respondents is often based on a desire to understand the attitudes and perceptions of different types of respondents, as well as how these factors vary across the groups. In the majority of the reviewed surveys, the respondents are stratified based on their categorization prior to the analysis of the responses. This trend supports the argument that cyclists should not be considered as a single, homogenous group. Furthermore, the approach to analyzing the perceptions and responses of the members of different respondent sub-groups creates the potential to identify and target segments of the population who may be more willing to change their behaviours. This can allow agencies to allocate their resources in a manner that maximizes benefit or that can fulfill as many of its objectives as possible given their budget (Piatkowski & Marshall, 2015). It is also common to ask respondents to express their beliefs about and views on the extent to which various hypothetical or proposed policy interventions or infrastructure investments will encourage people to cycle.

In practice, there appears to be a dichotomy in the manner in which revealed preference (RP) and stated preference (SP) data are collected. Population surveys that are conducted on behalf of public agencies tend to focus on the collection of RP and attitudinal data, whereas the collection of SP data is much more prevalent in academia. This trend raises the question of whether this satellite survey should include stated preference questions. Because one of the goals of the survey is to enable the ability to predict latent demand for cycling trips, the inclusion of SP questions would be worthwhile. This, however, may create an additional requirement for the platform on which the survey will be hosted. Including a question that asks non-cyclists to indicate the facilities that they would use for a hypothetical trip can be used to supplement the facility choice data obtained from cyclists in the smartphone-based survey. The key issues with this approach are the inability to immerse the respondent in the environment in which they would be making the decision, and the tendency for the perceptions of the cyclist to change as one actually begins to use a bicycle (Akar & Clifton, 2009).

4.2.2 Sampling, Sample Size Targets, and Recruitment

4.2.2.1 Sampling Methods

A fundamental consideration in the survey design process is the choice of how and from where to sample potential respondents. As with any survey, this choice should be informed by the characteristics of the target population and the intended applications of the data. Broadly speaking, two types of sampling exist – probability and non-probability sampling, with each approach having its advantages and disadvantages. For the web-based satellite survey, probability sampling should be used; in particular, a stratified random sampling approach should be used. The use of probability sampling helps to ensure that different types of cyclists and non-cyclists will be included in the sample, although the representation of each group will be affected by self-selection bias.

The use of probabilistic sampling inherently necessitates that all members of the target population have a known, non-zero probability of selection. This requires that the sampling frame be capable of providing an adequate representation of the population of the NCR. While this consideration is somewhat more important when the survey will be used to make conclusions about the population as a whole, the choice of sampling frame can be a significant source of sampling bias (Habib K. , 2014). The use of an address-based sampling approach is recommended, as it provides the best opportunity to reach the entirety of the population. Ideally, sampling would take place in at least two stages, with the second sample aiming to address the under-representation of certain demographics in the first sample. The issue with this approach is it would likely be time-consuming and costly.

In the event that the attitudinal satellite survey is conducted in conjunction with the core survey, it is recommended that respondents to the core Household OD Survey be invited to participate in the web-based satellite survey. To facilitate this, core survey respondents would be asked to provide some form of contact information, such as an email address or phone number, if they are interested in participating in future data collection efforts. Willing participants would then be stratified based on their household characteristics before being selected through stratified random sampling. Core survey respondents who also participate in the satellite survey should be given a unique household key, so as to facilitate linkage between the core and satellite surveys. This information can also be used to pull household information from the core survey, reducing the number of questions that this type of respondent would have to answer. On the other hand, if an attitudinal satellite survey will be conducted independently from the Household OD Survey, address-based sampling should be used.

For the web-based attitudinal survey in the NCR, there is an additional issue that must be addressed. Attitudinal and perceptual information are inherently individual characteristics, however, both recommended sampling strategies are household-based. Although this discrepancy needs to be addressed, the potential for an address-based sampling frame to cover the entirety of the survey area should outweigh this issue. Furthermore, the use of ABS can help to ensure that the spatial coverage of the survey is adequate, capturing areas with different built forms. Sampling households while asking individuals to provide attitudinal information reduces the potential for complications associated with the construction of the sample frame.

4.2.2.2 Sample Size Targets

The determination of a target sample size is influenced by a number of factors. As a rule of thumb, the sample size target should be based on the approach used to categorize respondents during the analysis stage of the survey. This categorization may be based on the approach used to expand the sample data, or the type of modelling exercises for which the data will be used. The standard sample size requirement is 30 or 40 samples per category, defined by the variables used to sub-divide the respondents. This standard, combined with a scaling factor that accounts for anticipated completion rates, would generally be used to identify the number of sampling units that should be invited to participate in the survey.

When dealing with travel surveys, the issue of representation takes on an additional dimension. In a typical survey, its ability to represent the target population is often determined by comparing the distribution of various household and personal attributes to those in the census. While the census provides a comprehensive set of demographic, socio-economic, and household attributes, it provides little information on travel behaviour. The lack of an objective “ground truth” of travel behaviour means that survey administrators cannot identify the extent to which their survey results represent travel behaviour. Thus, a larger sample size may not be more valuable than a smaller value if it cannot provide an adequate reflection of the travel behaviour of the target population. Despite the absence of a universally accepted standard sample size, the method proposed by Smith (1979) has had a significant impact on subsequent attempts to identify the target sample size for travel surveys (Habib, El-Assi, & Lin, 2017).

The goal of the heuristic presented in Smith (1979) was to create a means of identifying sample size targets “by statistical means” for the purpose of calibrating travel demand models (Smith, 1979). This distinction is crucial, as the implication is that the sample data are meant to augment or supplement an existing dataset, not replace it (Habib, El-Assi, & Lin, 2017). The procedure developed by Smith (1979) was meant to be applied to each of the stages of the traditional four-stage model, with Smith arguing that the trip distribution stage has the greatest influence on the sample size requirement (Habib, El-Assi, & Lin, 2017). Although the data collected through this survey are not meant to be applied in this context, Smith’s methodology was used due to its explicit consideration of confidence levels and margins of error. This methodology was applied using data collected in the 2011 Household OD Survey, in order to identify sample size targets.

The stratification of households was based on four variables – home location (either sampling district or district), income level, household size, and dwelling type. Because of the impact that income tends to have on the propensity to cycle, this variable was deemed to be fundamental to the determination of a target sample size. The location of the household was also deemed to be fundamental, due to the potential for differences in the propensity to cycle and the availability of facilities across the NCR. The impact of including the household size and dwelling type variables was examined, with the results being summarized in Table 2. Two sets of sample sizes were calculated, one using general trip rates and another using bicycle trip rates. As shown in Table 2, the sample sizes that were calculated based on the bicycle trip rate vary from 0.38% of households to over 300% of households. Consequently, the sample size should be determined based on the general trip rate. Based on the results shown below, a minimum of 1,622 households should be invited to participate in the attitudinal satellite survey, with the sample frame being stratified based on the location of the household, dwelling type, and household size. If respondents are not required to state their household income in the core Household OD Survey, income should not be used to stratify survey respondents. It is important to note that these sample sizes must be scaled up to account for anticipated response rates; this, along with the choice of target sample size, is left to the discretion of the TRANS Committee.

Table 2: Potential Sample Sizes for the Attitudinal Satellite Survey

Individual or Household Socio-demographic Attributes	Design Sample Size (considering general trip rate)	% of total households in the NCR*	Design Sample Size (considering only bicycle trip rate)	% of total households in the NCR*
Household Income; Sampling District; Household Size	1,599.92	0.299	245,031.45	45.758
Household Income; Sampling District; Dwelling Type	1,898.63	0.355	643,479.64	120.165
Household Income; Sampling District; Dwelling Type; Household Size	1,720.04	0.321	2,034.44	0.380
Household Income; Home District; Household Size	1,993.71	0.372	1,146,724.40	214.143
Household Income; Home District; Dwelling Type	2,132.69	0.398	1,636,787.63	305.659
Household Income; Home District; Dwelling Type; Household Size	1,621.27	0.303	37,859.16	7.070
* According to the 2016 census, there are 535,495 households in the NCR (source: Statistics Canada)				

4.2.2.3 Sample Recruitment

The design of the survey invitation is a key consideration when recruiting potential survey participants, as generic or impersonal invitations are more likely to be ignored. To address this issue, survey invitations should be addressed to a specific member of the household whenever possible. When recruiting core survey respondents to participate in the satellite survey, this can be accommodated by asking respondents to provide both a contact name and email address.

Regardless, including an indication that the survey is sponsored by the TRANS Committee and its member agencies can also help to reduce the likelihood that the survey is ignored. The topic and sponsor of a survey tend to carry considerable weight on the decision to participate in the survey, particularly when it is being sponsored by a government agency (Chung, Srikukenthiran, Habib, & Miller, 2016). Although invitations should be addressed to a specific member of the household, the recipient should also be encouraged to invite other members of the household to participate, as a means of obtaining as many responses as possible. Proxy responses should not be allowed, due to the inherently personal nature of attitudinal information.

In addition to recruitment through stratified random sampling, consideration should be given to the recruitment of participants through media and advertising. This open approach to recruitment was used by Bachand-Marleau, Larsen and El-Geneidy (2011) in their study of the behaviours and preferences pertaining to bicycle-transit integration in the Montreal region. The choice to use this approach to recruitment stemmed from the authors' concerns about the tendency for web-based surveys to over-represent certain subsets of the population, and wanted to ensure that "a broad cross-section of the public was reached" (Bachand-Marleau, Larsen, & El-Geneidy, 2011). The survey was advertised through email newsletters, mailing lists, newspaper articles in both English and French, radio interviews, social media, and by distributing flyers at major transit stations (Bachand-Marleau, Larsen, & El-Geneidy, 2011). The adoption of this approach in the NCR would help to increase the number of survey responses, although it introduces the potential for self-selection bias. Because responses to attitudinal questions should not be expanded, recruiting participants in this manner would not affect the applicability of the attitudinal information.

4.2.3 Designing Attitudinal Questions

Collecting attitudinal data is of particular importance when studying the choice to travel by bicycle. This importance is magnified when studying the decision to cycle on a regular basis, because it tends to be influenced by a variety of perceptual and attitudinal factors (Maldonado-Hinarejos, Sivakumar, & Polak, 2014). Response options for attitudinal questions tend to take one of three forms: open responses, Likert scales, or checkboxes. Using textboxes allows respondents to identify a motivator or deterrent to cycling that would have otherwise been overlooked. However, this method would only provide insights into the frequency with which a factor is identified as a motivator or deterrent, not the extent to which it affects behaviour. The potential for a similar issue also exists when checkboxes are used. Regardless of the response option that is used, it is important to identify the extent to which each potential motivator or deterrent would affect the behaviour of the respondent.

The use of a Likert scale to solicit responses to attitudinal questions facilitates the use of principal component analysis (PCA), also known as factor analysis. This approach to analysis, which can be used to identify the underlying factors that affect a person's attitudes and perceptions (Akar & Clifton, 2009). With the caveat that the response options should be both mutually exclusive and exhaustive, the use of a Likert scale to collect information from respondents can provide the same utility as textboxes and checkboxes. Furthermore, the use of a Likert scale provides a means of using attitudinal and perception data in more quantitative applications, such as model estimation. Thus, a five-point Likert scale (e.g. strongly agree/ disagree, agree/ disagree, and no opinion) should be used when posing attitudinal and perceptual questions to respondents.

4.2.4 Designing Stated Preference Questions

When administering stated preference questions, it is important to try to mimic the conditions under which the respondent will be making the decision. Compared to when a person is faced with a choice in their daily life, survey respondents tend to have access to a relatively more comprehensive set of information. When presented with a completely hypothetical choice, access to information at this level of detail, when combined with their past experiences, may cause them to make a decision that does not accurately reflect the choice that they would make in their daily lives. In order to help ensure that a respondent's choice is reflective of his or her behaviour, the premise of a stated preference question should aim to replicate the context in which the choice would be made in their daily life, where possible. By designing SP questions in this manner, the effects of past experiences can be more accurately captured, because the context of the trip is consistent with at least some of the contexts in which the experiences have been gained. Furthermore, the choice set presented to respondents should be characterized using only the information to which they would have access, including the availability of facilities and travel times, but excluding specific roadway grades or the proportion of the trip that would be made using a given facility.

An interesting example of an attempt to provide respondents with an idea of the conditions associated with each response option is outlined in Tilahun, Levinson and Krizek (2007). In their study of the trade-off between facility choice and travel time, the authors used a web-based adaptive SP survey that presented respondents with 10-second videos that reflected the nature of each facility that respondents were given the option of choosing. Additionally, the conditions shown in the video were reflective of the outdoors over the period during which the survey was administered (summer and winter) (Tilahun, Levinson, & Krizek, 2007). The ability to immerse respondents in the environment in which the decision will be made is important, as one's perceptions can affect the choice process. This is particularly important when it comes to cycling, as the negative aspects can diminish, and the positive aspects can be enhanced, when one engages in the activity (AMR Interactive Consultants, 2009).

4.3 Survey Data Requirements

This section will present recommendations on the data that must be collected in order to develop the capabilities outlined in the project proposal. Due to the fact that the survey will be carried out as a satellite, the ability to link the data collected through this survey to the data collected through the Household OD Survey is imperative. Based on the trends identified in recent surveys of cyclists, the categorization of respondents is recommended, in part to facilitate a more targeted approach to attempts to encourage cycling.

4.3.1.1 Facilitating Linkage to the Core Survey

In cases where information from the travel diary portion of the survey are to be expanded, it is important for the satellite survey to collect information that can facilitate linkage to the core survey. In order to identify the information needed to link the core and satellite survey data together, the conduct of the core Household OD Survey was used as a reference point. At minimum, the information used to expand the results of the core survey should also be collected in the satellite survey, in order to facilitate the most basic form of data fusion – weighting (or expansion) (Bayart & Morency, 2008). Consequently, the age and gender of survey respondents should be collected regardless of how they were recruited. The age and gender of each member of the respondent's household should also be collected, in order to facilitate the expansion of non-attitudinal information (Verreault & Morency, 2016). In addition, participants who have completed the core Household OD survey should have information on their household location, dwelling type, and household size carried over from the core survey dataset. Participants who were recruited through other means should be asked to provide this information (R.A. Malatest & Associates Ltd., 2013). This process is summarized in Figure 7.

Because the sampling will be carried out in a probabilistic manner, the collection of these data will allow the data collected in the non-attitudinal portions of the satellite survey to be expanded in the same manner as the core survey, i.e. through the calculation of both household- and person-level weights. While this approach can be used to address the under-representation of cyclists in the core OD Survey, it is imperative that the attitudinal and SP data not be expanded to represent the entirety of the target population.

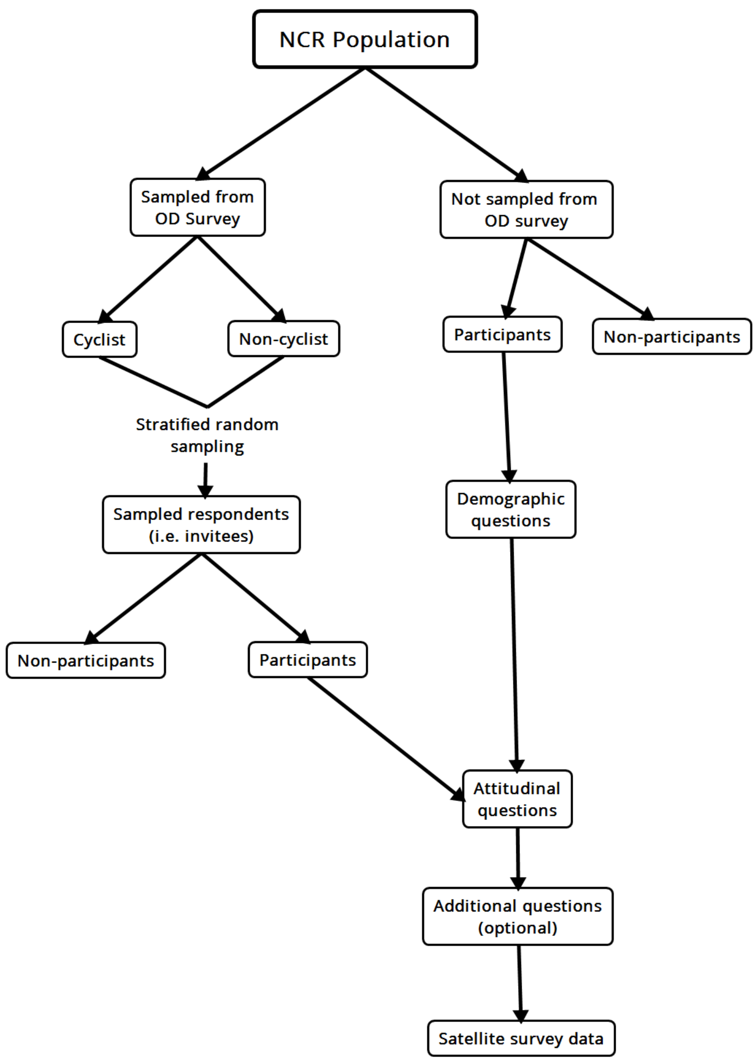


Figure 7: Data Requirements, by Sample Frame

4.3.1.2 Categorizing Respondents

The categorization of survey respondents has two significant benefits. First, it allows surveyors to analyze the behaviour of different types of respondents. Second, it facilitates the use of a more targeted approach to the encouragement of cycling. Treating cyclists as the aggregation of several sub-groups, rather than one homogenous group, allows for the identification of groups that are more prone to behavioural changes, which can lead to a more efficient use of resources (Piatkowski & Marshall, 2015). Generally speaking, respondents in surveys that focus on cycling have been categorized in two ways – the frequency with which they cycle, or the purpose for which they cycle (if at all). In the literature, both factors have been shown to affect a person's cycling behaviour.

The extent to which internal and external factors affect the decision to cycle has been found to vary based on the frequency with which a person uses a bicycle. The behaviour of regular cyclists tends to differ from that of casual users (Khatri, Cherry, Nambisan, & Han, 2016); for example, regular cyclists tend to be less sensitive to variations in weather (Godefroy & Morency, 2012). The purpose for which a person cycles is also an important consideration, as the behaviour of utilitarian cyclists tends to differ from that of recreational cyclists. Although recreational cycling tends to be more prevalent than utilitarian cyclists in parts of the world with low cycling rates (Olafsson, Nielsen, & Carstensen, 2016), the modelling of cyclist behaviour tends to focus on utilitarian cyclists. Thus, the inclusion of questions that facilitate the categorization of cyclists should be included in the survey, in order to assist those who will analyze the survey data, and to identify segments of respondents who could be targeted.

Chapter 5
Application of Satellite Survey Data

Satellite surveys offer the opportunity to obtain data that may not be feasible or cost-effective to obtain through traditional household travel surveys. For reasons discussed earlier, satellite surveys offer the ability to collect very detailed and specific information that may introduce excessive amounts of burden if included in the main core survey. Within the core-satellite framework, satellite surveys serve to help augment and enrich the core data. However, a number of surveys that could be implemented as satellite surveys have existed prior to the creation of the core-satellite framework. Aside from their ability to enrich the data collected through core surveys, satellite survey data can provide a wealth of information about a specific subpopulation. This chapter presents an empirical study of the location choice behaviour of university students in Toronto that makes use of the data collected by a travel and activity survey of university students, named StudentMoveTO (StudentMoveTO, 2015).

The decision to use StudentMoveTO data in lieu of TTS data was based on the potential of the former to reach students who live on campus or with roommates, as well as those who live at home. In addition, the structure of the StudentMoveTO dataset is more conducive to an activity-based approach to the study of travel behaviour. Compared to the structure of the TTS data, which includes person, household, trip, and transit trip files, the StudentMoveTO dataset includes a record of the locations visited by each respondent. This eliminates the need to construct activity diaries using the TTS trip file. Future analyses should attempt to combine the StudentMoveTO data with observations of post-secondary students in the TTS.

The remainder of the chapter is structured as follows. First, the motivation for this empirical study is presented. Then, a literature review that describes prior studies that have focused on the travel behaviour of post-secondary students is presented. An overview of location choice models is provided, and the various approaches that have been used to measure accessibility are discussed. Next, the description of the dataset that was used for the empirical investigation is provided. Then, the theoretical background of the empirical model is discussed, and the process used to impute the choice sets of decision-makers is described. Following this discussion, the results of the empirical are presented and discussed. Finally, the empirical model is used to calculate utility-based measures of accessibility, which are compared to count-based measures of accessibility.

5.1 Motivation

The travel behaviour of post-secondary students, which has traditionally been poorly understood, has received increased attention in recent years. In operational travel demand models, university campuses are often modelled as special generators (Eom, Stone, & Ghosh, 2009). This approach can account for the impacts that a university campus can have on travel demand in the surrounding area, but provides little insight into the travel behaviours of students. Furthermore, this approach places primary emphasis on non-discretionary trips, such as trips to school or to return home. Also, operational travel demand models tend to be developed based on the data obtained through household travel surveys. With regards to understanding travel behaviour, this is problematic because post-secondary students tend to be under-represented in household travel surveys (Wang, Khattak, & Son, 2012). Despite many efforts in the literature to better understand the travel behaviour of post-secondary students, more work needs to be done to incorporate this understanding into travel demand models. Overall, developing a better understanding of the factors that influence the travel behaviour of students presents another step towards being able to provide a more accurate basis for the evidence-based approach to policy development.

The increased focus on student travel behaviour is warranted, as it tends to differ from that of the general population, even when considering persons in the same age group. Compared to the general population, post-secondary students tend to make more trips on a given day (Wang, Khattak, & Son, 2012) and tend to utilize alternative modes, namely public transit and active modes, at a greater frequency (Whalen, Paez, & Carrasco, 2013). The difference in travel behaviour may also stem from the relatively flexible and irregular nature of the daily activity schedules of post-secondary students, which tend to vary from one day to the next. The effects of activity schedules cannot be analyzed using the traditional trip-based approach to travel demand modelling due to the assumption that trips are made independently of one another (Castiglione, Bradley, & Gliebe, 2015). Consequently, the effects of activity schedules on travel behaviour must be analyzed through the lens of activity-based models.

Travel demand is often regarded as the by-product of a desire to participate in out-of-home activities; a critical component of this is the decision of where to participate in said activity. The choice of activity location can have a significant impact on the utilization of transportation infrastructure and services, particularly when public transit is used. Students' accessibility by transit to different locations plays a critical role in defining their well-being and mobility needs, specifically for a large city like Toronto. Toronto is well-served by transit. However, does it mean that Toronto transit serves its student population well? To answer this question, this analysis uses a discrete choice model-based measurement of accessibility of post-secondary students by transit in Toronto. It presents an investigation into the factors that influence the location choice process of post-secondary students when using public transit for discretionary trips in Toronto. The emphasis placed on discretionary trips stems from the fact that school and work location choice are often modelled explicitly in the literature and the role that discretionary travel plays in travel demand. The location choice model is used to derive measures of choice model-based accessibility by transit, a concept that is seldom studied in the literature. It is then compared against the traditional count-based measure to highlight how accessibility by transit can be over-estimated and misrepresented.

5.2 Background

The increased efforts to understand the factors influencing post-secondary student travel behaviour are necessary to improve the ability of travel demand models to produce accurate forecasts. The continued need to better understand the travel behaviour of students is necessary as household travel surveys and regional travel demand models often under-represent this segment of the population (Wang, Khattak, & Son, 2012). This can result in student travel behaviour being represented in a simplified manner. While these simplifications can produce relatively accurate results at the aggregate level, they may not provide an adequate representation of preferences or behaviours at the individual level. The under-representation of student travel behaviour and the resulting simplifications are problematic for two reasons. First, university campuses can have a significant impact on the travel demand generated in a region, which is why they tend to be treated as special generators in some travel demand models (Garikapati, et al., 2016). Secondly, the representation of university campuses as special generators provides no insights into the travel behaviour and preferences of individual students.

Furthermore, this approach predominantly considers the school-based trips made by students, which is only a subset of their overall travel. As argued by Eom et al. (2009), this approach can neither reflect nor account for the role that factors such as class schedules, the land use patterns of university campuses, and access to subsidized services play in the travel patterns of students. The contribution of the participation of students in non-school-related activities to the overall travel demand of a region necessitates that this type of travel is explicitly considered in travel demand models.

The inadequate representation of post-secondary student travel behaviour in travel demand models can primarily be attributed to the availability, or lack thereof, of the requisite data. Many operational travel demand models are built using data from regional household travel surveys, which for a variety of reasons, tend to do a poor job of obtaining responses from post-secondary students. Volosin et al. (2014) theorize that the under-representation of post-secondary students may be due to their lack of engagement in civic processes, the omission of student residences in sampling frames, and the tendency for students to change residences on a fairly frequent basis. The once-common utilization of lists of landline phone numbers as survey sample frames may also have contributed to said under-representation (Wang, Khattak, & Son, 2012). Moreover, large-scale travel surveys rarely aim to investigate the travel behaviours of students, allowing post-secondary students to be neglected (Khattak, Wang, Son, & Agnello, 2011). As a result, studies in the literature that investigate the travel behaviour of students often rely on purpose-built surveys to obtain travel diaries – see (Volosin, et al., 2014; Wang, Khattak, & Son, 2012; Whalen, Paez, & Carrasco, 2013; Hasnine, Lin, Weiss, & Habib, 2018; Searcy, et al., 2018) for examples. Data from such surveys is facilitating a broader range of applications than has previously been seen in the literature.

Until recently, much of the literature about post-secondary students consisted of investigations into fundamental aspects of travel behaviour, such as trip frequency and mode choice. In the literature, the factors that influence trip frequencies are often analyzed through the development of regression models, such as the fractional polynomial model described in (Searcy, et al., 2018) or the Poisson and negative binomial regression models outlined in (Wang, Khattak, & Son, 2012). In terms of mode choice, many early studies limited themselves to the use of the multinomial logit (MNL) models. Recently, more sophisticated model specifications, such as the nested logit, cross-nested logit, and generalized extreme value models have been applied to study commuting mode choice and mobility tool ownership (Hasnine, Lin, Weiss, & Habib, 2017; Habib, Weiss, & Hasnine, 2018). Also, travel-related data collected from post-secondary students have been analyzed through data visualization in geographic information system (GIS) software and has also been used to study attitudes towards safety and driving, the enjoyment of different modes, and activity patterns (Whalen, Paez, & Carrasco, 2013).

Although the overall understanding of student travel behaviour continues to grow, most of this understanding is in the context of the trip-based approach to analyzing travel demand. While this improvement is promising, it is happening at a time where the paradigm of travel demand analysis is moving towards an activity-based approach. This shift creates the potential for the understanding of student travel behaviour to once again lag behind the state-of-the-art. Some studies in the literature argue that activity-based models have the potential to more accurately represent travel behaviours related to special generators (Eom, Stone, & Ghosh, 2009). With regards to post-secondary students, one aspect of the activity-based approach to travel demand modelling that has been overlooked is activity location choice. To the authors' knowledge, Garikapati et al. (2016) is the only study in the literature that has developed location choice models for post-secondary students. Using data obtained from students at Arizona State University in the spring of 2012, the authors estimated a location choice model to identify the non-university end of university-based trips. The MNL model of location choice was implemented as part of a broader activity-based framework that was used to model university travel demand in the Greater Phoenix Area and the Albuquerque Metropolitan Area.

Simply put, location choice models are used to analyze the factors that influence the choice of a location for a specific purpose. Location choice models are often used to analyze the choice of where to live, where to establish a firm, or where to participate in an activity. In the context of activity participation, location choice is often analyzed using the MNL model, such as models described in (Yang, Du, Sun, & Zhao, 2009; Kim & Lee, 2017). Commonly used explanatory variables include measures of inter-zonal impedance, measures of zonal size, non-size measures, and indicator variables (Kim & Lee, 2017). In addition to modelling the choice of activity location, location choice models can also be used to derive measures of accessibility.

Since the seminal work on the topic of accessibility was published by (Hansen, 1959), the concept of accessibility has taken on a variety of forms. Each definition that is offered in the literature is consistent with two fundamental principles: 1) accessibility is a product of the interaction between land use and transportation systems, and 2) accessibility is related to the ability to reach goods, opportunities, or services (Albacete, Olaru, Paul, & Biermann, 2017). In the literature, measures of accessibility are often grouped into one of three categories: count-based, gravity-based, or utility-based measures. Count-based measures of accessibility are reflective of the total number of opportunities that can be reached within a given time or cost threshold (Dong, Ben-Akiva, Bowman, & Walker, 2006). While this measure is easy to interpret, there are two critical issues with this approach. First, the results will be influenced by the decision of the threshold value (Cascetta, Carteni, & Montanino, 2012). Second, each opportunity within the threshold is regarded as equally attractive (El-Geneidy, et al., 2016). To address this shortcoming, gravity-based measures of accessibility discount the attractiveness of opportunities based on the cost associated with reaching said opportunity (El-Geneidy, et al., 2016). One of the critical issues with gravity-based accessibility measures is the inherent assumption that the impedance associated with reaching an opportunity will be the same for all travellers (Hasnine, Graovac, Camargo, & Habib, 2019). Because they overlook the context in which trips are made, both count- and gravity-based measures have the potential to over-estimate accessibility (Hasnine, Graovac, Camargo, & Habib, 2019). This issue can be addressed through the use of utility-based accessibility measures, which are predominantly defined by the log-sum value of a destination choice model (Paez, Scott, & Morency, 2012). The ability to include socio-economic and demographic variables in the discrete choice model allows this measure to account for the perceptions of individual travellers. A critical shortcoming of this approach is that the results are relatively difficult to interpret and to compare across the spatial and temporal dimension. In spite of this issue, the utility-based approach has the potential to provide a more representative and context-sensitive measure of accessibility. For a more detailed review of accessibility, and the measures used to quantify it, see (Albacete, Olaru, Paul, & Biermann, 2017; Cascetta, Carteni, & Montanino, 2012; Curtis & Scheurer, 2010; Geurs & and van Wee, 2004).

The contribution of this work to the existing literature is two-fold. First, the activity-based approach to studying the travel behaviour of post-secondary students addresses a gap in the literature about post-secondary students. Specifically, there is a dearth of studies that investigate the location choice processes of post-secondary students. Also, the accessibility of post-secondary students is rarely examined. Second, the analysis presented in this study focuses on the location choice process and accessibility for trips made using public transit. While accessibility by public transit has been studied in past works, it has rarely been studied specifically for post-secondary students.

5.3 Data Description and Descriptive Statistics

The results presented in this investigation were obtained by analyzing data collected through the StudentMoveTO survey. The survey, which was conducted in 2015, collected detailed travel and activity information from students attending Toronto's four universities (OCAD University, Ryerson University, York University, and the University of Toronto) (StudentMoveTO, 2018). This survey was administered across seven university campuses and obtained information from a total of 15,226 students (StudentMoveTO, 2018). As Figure 8 shows, six of the campuses are located within the city of Toronto, while the other campus is located in Mississauga. Respondents of the survey were asked to, among other things, provide a one-day travel diary that summarized the locations that they visited and the mode of transportation that they used to go from one location to the next.

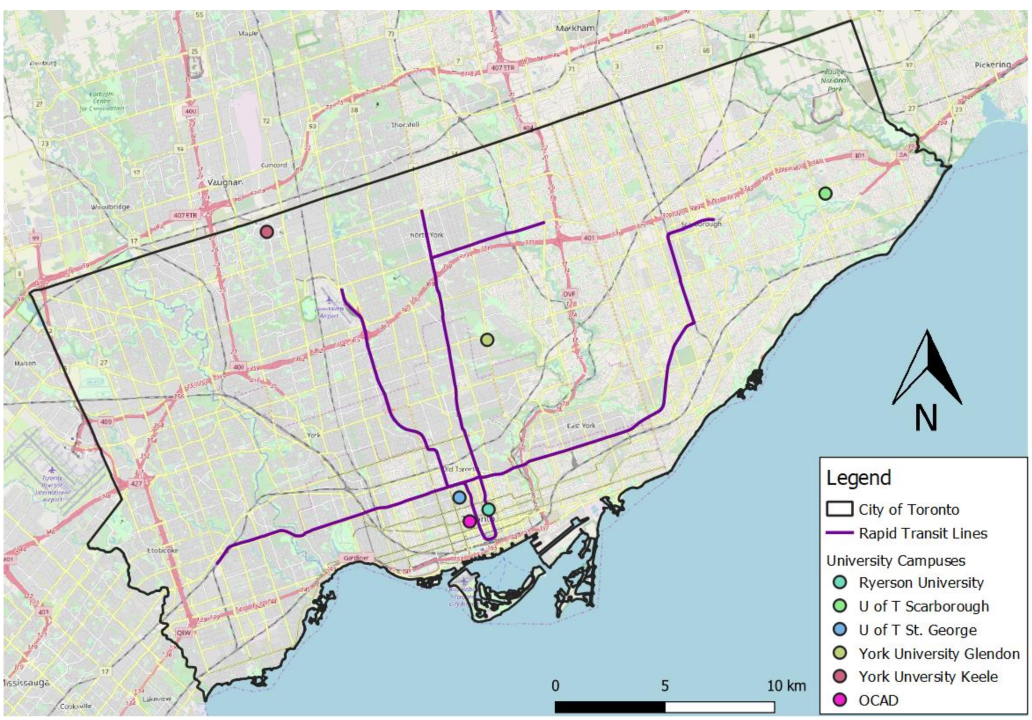


Figure 8: Locations of participating universities in Toronto

To compare and contrast the travel behaviors of post-secondary students and the general population, the trip rates, modal shares, and the distribution of trip distances were examined. Information on the travel behaviour of the residents of the City of Toronto was obtained through the 2016 iteration of the Transportation Tomorrow Survey (TTS). Three groups were identified for this comparison: residents of Toronto ("TTS"), post-secondary students who live in Toronto and were included in the TTS ("TTS Student"), and respondents to the StudentMoveTO survey ("SMTO").

Overall, it appears that the travel behaviour of students in Toronto differs from that of the rest of the population. For example, as shown in Figure 9, both the *TTS Student* and *SMTO* groups had a higher average trip rate than the *TTS* group. Furthermore, the two student groups displayed a higher propensity for using public transit than the general population. The public transit trip rate for the *TTS Student* group being higher than that of the *SMTO* group may be due to the ability of the TTS to capture students who live with their families and the ability of the StudentMoveTO survey to capture students who live in dormitories or with roommates. As a result, a smaller proportion of the members of the *TTS Student* may live within walking distance from campus, meaning that they may be more reliant on private automobiles or public transit. It also appears that students in Toronto are more likely to use public transit and active modes than the population as a whole, as shown in Figure 10. Roughly half of the trips reported by the *TTS* group were made by driving, while almost half of the trips reported by the *TTS Student* group were made by public transit. Additionally, roughly 70% of the trips reported by the *SMTO* group were made either on foot or by public transit. The relatively high share of trips made by walking among the *SMTO* group may be because the Euclidean distance of roughly one-third of the trips reported by this group is less than 1 km, as shown in Figure 11.

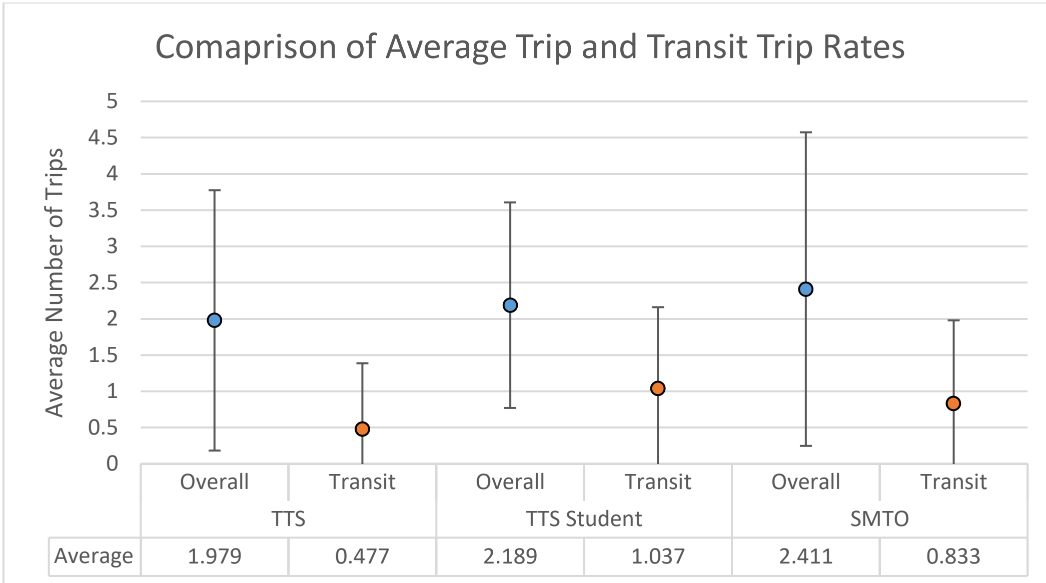


Figure 9: Average and Standard Deviation of Trip and Transit Trip Rates

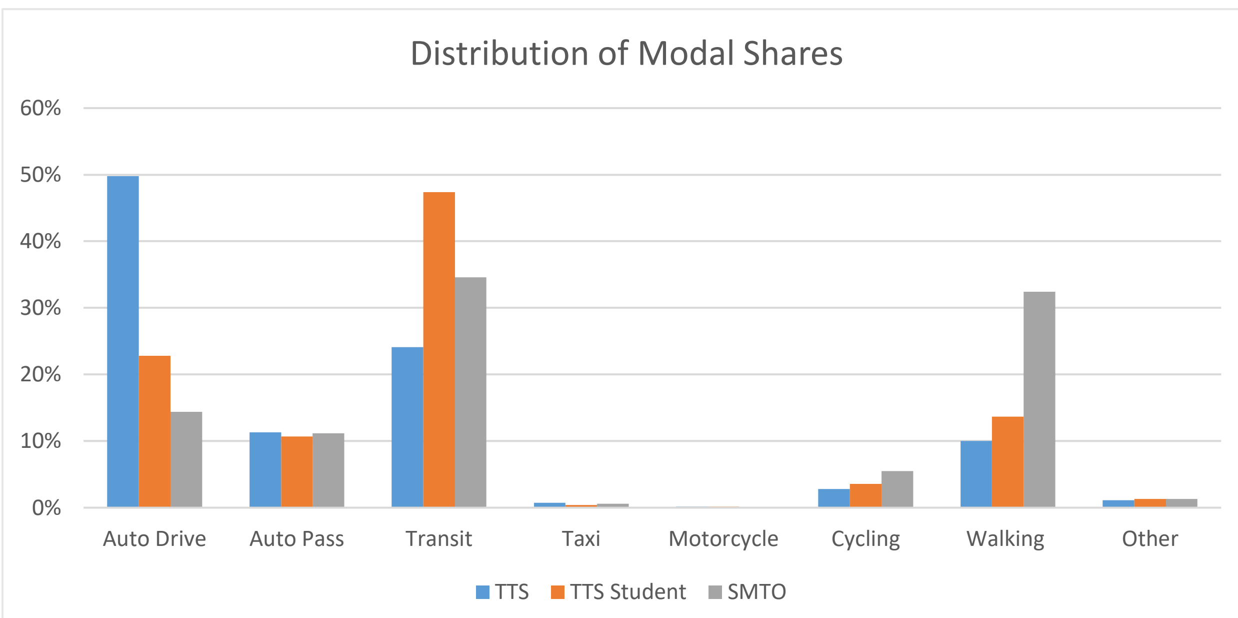


Figure 10: Modal Shares of Reported Trips

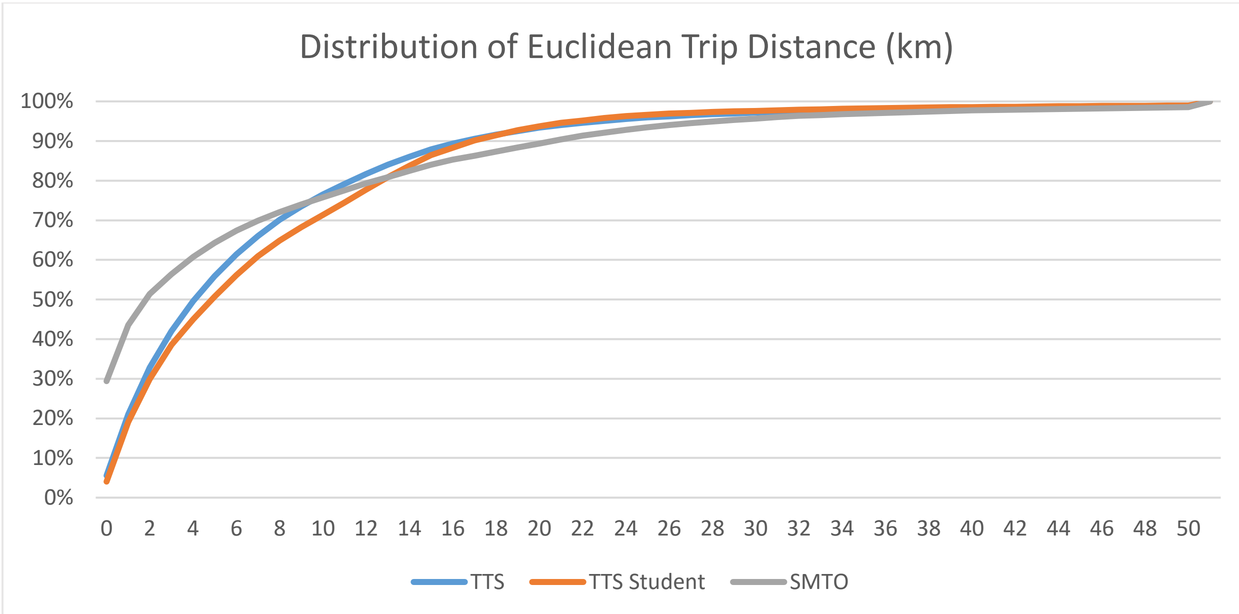


Figure 11: Distribution of the Euclidean Distances of the Reported Trips

5.4 Activity Location Choice Model and Accessibility

The empirical investigation of the location choice of post-secondary students was carried out using the multinomial logit (MNL) model of discrete choice. In this investigation, the MNL model is used to model the choice of traffic analysis zone (TAZ) in which a university student will participate in a non-work and non-school activity. The MNL model predicts the probability that a given alternative will be chosen by a decision-maker as a function of variables related to the alternatives and the decision-maker. The decision-maker is assumed to obtain a certain level of utility from each alternative, with the choice being assumed to be driven by the level of utility that would be obtained from the chosen alternative (Train, 2009). The utility that is derived from choosing alternative *i* (denoted as U_i) can be divided into two additive components – the deterministic component (V_i) and the random component (ε_i) (Ben-Akiva & Lerman, 1985). The probability that person *n* chooses alternative *i* is given by (McFadden, 1974):

$$P_{ni} = P(U_i > U_j) = P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}\ \forall\ j \neq i) = P(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \forall\ j \neq i)$$

The deterministic component of utility can be modelled by the researcher as a function of observable characteristics. The deterministic component is often represented as the product of a matrix of variables that are believed to influence the utility obtained from each alternative (denoted as X_{ni}) and a vector of parameters (denoted as β):

$$V_{ni} = \beta' X_{ni}$$

The exact value of the random components of utility cannot be determined by the researcher; rather, the distribution of these components must be specified prior to the estimation of the model. The distribution of the random components of utility will define the exact specification of the discrete choice model. For the MNL model, the random components are assumed to be independent and identically distributed according to the Type I extreme value (Gumbel) distribution, with location parameter η and scale parameter $\mu > 0$ (Ben-Akiva & Lerman, 1985). This assumption produces the following:

$$P_{ni} = \frac{exp(\mu V_{ni})}{\sum_j exp(\mu V_{nj})} = \frac{exp(\mu \beta' X_{ni})}{\sum_j exp(\mu \beta' X_{nj})}$$

The value of μ in above equation cannot be identified separately from the value of β (Train, 2009). When estimating an MNL model, it is common to normalize the value of the scale parameter to 1. This closed-form expression can be estimated using maximum likelihood estimation techniques. The model is estimated by using the MAXLIk algorithm of GAUSS (Aptech Systems, Inc., 2013).

A key consideration in the estimation of a discrete choice model, mainly when location choice is being studied, is the process through which the choice set of each decision-maker will be determined. For location choice models, where there are potentially hundreds or thousands of alternatives, it is common to estimate the model using a subset of the universal choice set (Lee & Waddell, 2010). This subset often must be imputed, because most revealed preference surveys collect information about the alternative that was chosen, not the alternatives that were considered. In the literature, the formation of choice sets is approached in one of three ways: 1) the random selection of alternatives from the universal choice set 2) the consideration of all plausible alternatives, or 3) the application of rules to reduce the size of the feasible choice set prior to the use of the full feasible choice set (Rashidi, Auld, & Mohammadian, 2012). This investigation takes the third of the above approaches, applying the time-space prism concept first proposed by Hägerstrand (1970), as shown in Figure 12. The application of this concept is consistent with prior studies, which typically focus on the participation of a "flexible" activity that takes place between two fixed locations (Chen & Kwan, 2012).

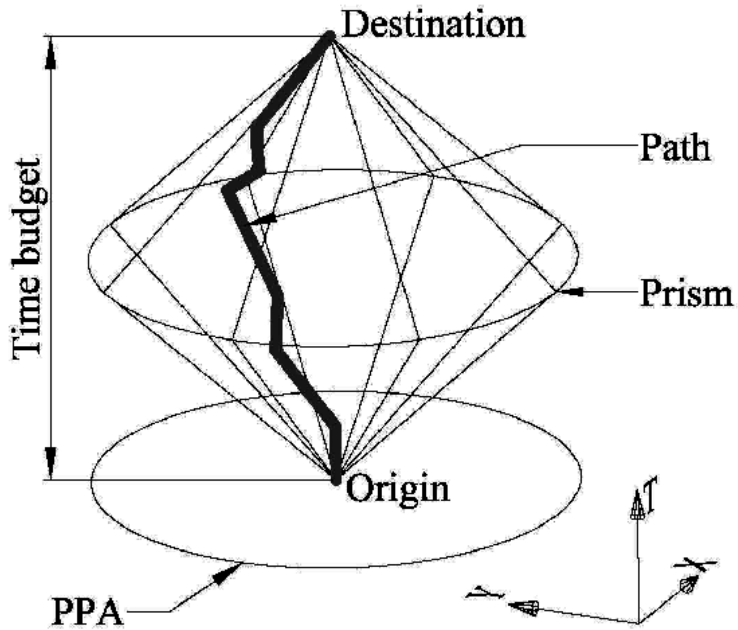


Figure 12: A Visualization of the Time-Space Prism (Neutens, Witlox, Van De Weghe, & De Maeyer, 2007)

For this study, the choice set formation process began with the classification of the locations that each survey respondent visited. These locations were classified as either mandatory or discretionary locations based on the reported activity purpose. The mandatory locations were defined as locations where non-discretionary activities take place, i.e., school, work, and home. Locations for which the activity purpose was reported as shopping, restaurant, recreation, services, health, visiting, and others were classified as discretionary locations. Next, the characteristics of the trips made to the discretionary locations were examined. The discretionary locations that were accessed using public transit on a weekday between 6 AM and 7 PM were kept for analysis. The feasible choice set was imputed for the trips that met these criteria. Because the level-of-service (LOS) attributes of the unchosen alternatives cannot be obtained from survey data, these values had to be imputed. These attributes were imputed using the Tool for Incorporating Level of Service attributes (TILOS) tool described in (Hasnine, Kamel, & Habib, 2017). Using a Google application programming interface (API)-based framework, TILOS was used to obtain the in-vehicle travel time, access time, egress time, total travel time, walking distance, and the number of transfers required to travel from the centroid of a given TAZ to that of every other TAZ. Due to cost constraints associated with the use of the Google API, LOS information was only obtained for three periods – 6 AM to 10 AM, 10 AM to 3 PM, and 3 PM to 7 PM. These periods are consistent with the periods that are used by the Toronto Transit Commission to plan service.

For each trip under investigation, the set of feasible alternatives was determined based on the time budget that was available for each activity. The time budget is defined as the time that has elapsed between the arrival at the next mandatory location and the departure from the previous mandatory location, minus the time that was spent at the location in question. Consider the activity schedule shown in Figure 13. For activity 3, denoted as A3, the time budget for this activity is determined by subtracting TD2 and D3 from TA6. For each observation in the dataset, a TAZ i was considered to be a feasible alternative if a person could travel from their previous mandatory to the zone and from the zone to their next mandatory location within the time budget (TB), i.e.:

$$TT(Previous \rightarrow Zone\ i) + TT(Zone\ i \rightarrow Next) \leq TB$$

The total travel time imputed using TILOS was used to determine the time needed to travel between zones.

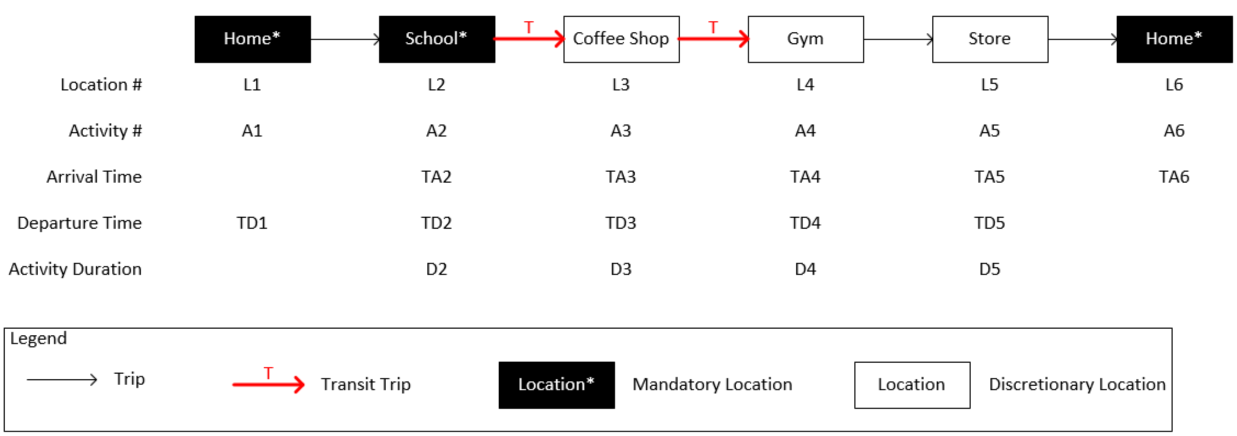


Figure 13: Example of an Activity Schedule

Through this process, each feasible choice set was imputed (denoted as C_n for each person n). Consequently, the probability that a particular TAZ will be chosen can be expressed as**:**

$$P_{ni} = \frac{exp(\mu V_{ni})}{\sum_{j \in C_n} exp(\mu V_{nj})} = \frac{exp(\mu \beta' X_{ni})}{\sum_{j \in C_n} exp(\mu \beta' X_{nj})}$$

The distribution of choice set sizes is shown in Figure 14. The relationship between the time budget for a given activity and the size of the choice set is shown in Figure 15.

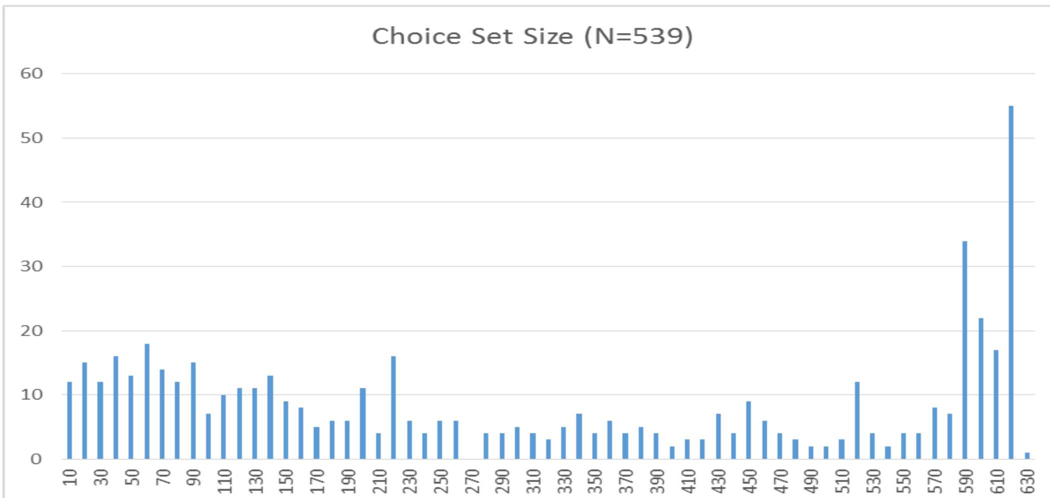


Figure 14: Distribution of Imputed Choice Set Size

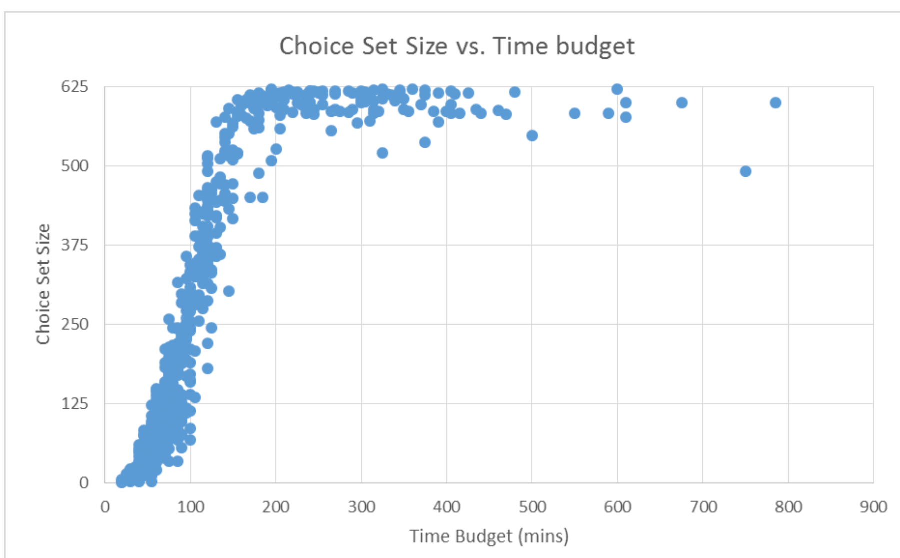


Figure 15: The Relationship between Time Budget and Choice Set Size

As expected, the size of the feasible choice set increases as the length of the time budget increases. This relationship, however, does not hold after the time budget exceeds roughly 150 minutes, which is understandable given the range of inter-zonal travel times. Once the feasible choice set was imputed for each trip, the accessibility by public transit experienced by a student in zone i can be computed as:

$$A_i = \frac{1}{\mu} ln\left[\sum_{j=1}^{L} exp\left(\mu V_{nj}\right)\right] = \frac{1}{\mu} ln\left[\sum_{j=1}^{L} exp\left(\mu \beta' X_{nj}\right)\right]$$

$$for \;\; i \neq j, L \; is \; the \; total \; number \; of \; feasible \; locations$$

5.5 Empirical Model

Five different sources of data were used to investigate the factors that influence the location choice behaviours of university students in Toronto. Two of the sources – the StudentMoveTO survey and the LOS information obtained using TILOS have already been discussed. Two of the remaining datasets were incorporated to analyze the impacts of land use variables on the location choice process. The first is the Enhanced Points of Interest (EPOI) dataset produced by DMTI Spatial Inc., which contains the location (i.e., latitude and longitude) and classification of each so-called point-of-interest (POI) in Canada (DMTI Spatial Inc., 2016). Examples of points of interest include accommodation and food services, retail trade, educational services, and manufacturing. The second dataset was the land use information contained in the CanMap RouteLogistics file, also produced by DMTI Spatial Inc. (DMTI Spatial Inc., 2014). These are used to count the number of points of interest in each zone and the proportion of each zone that corresponds to particular land use. Finally, data collected through the 2016 Canadian census was also projected onto each TAZ.

Variables related to socio-economic characteristics and trip purpose were interacted with land use and LOS variables in the generic utility function of location choice model and included in the final specification. The variables that were included in the final model specification are described in Table 3. The decision of whether to include a variable in the final model was based on the significance of the parameter and the sign of the estimated coefficient. There were a few cases where parameters that are not significant were retained in the model to account for the impact of various attributes on the choice of location, and ultimately, accessibility.

As reported in Table 4, two final models are estimated. One with socio-economic variables that allow investigation of location choice behaviour at an individual level. The other is without socio-economic variables that allows for the calculation of aggregate zone-to-zone accessibility of students making discretionary trips by transit in Toronto. The latter allows the performance of utility-based accessibility measures to be compared against that of commonly used count-based measures. Both empirical models were estimated using a dataset containing 539 observations.

The decision to estimate two versions of the model stems from the inherent issues with applying an individual-based model to an aggregate zonal system. Other studies that have applied discrete choice models, such as Hasnine, Graovac, Camargo, & Habib (2019) benefit from having larger sample sizes, which allows them to apply the model to individuals and aggregate the model outputs based on the origin zone of the trip. Due to the small sample size used for this study, the decision was made to base the accessibility calculations on a version of the final model that only included zone-level attributes.

Table 3: Summary of Explanatory Variables in Final Models

Variable Name	Description
nTrans	The number of times that a person must transfer between transit routes to reach their destination
Sub_Campus	A dummy variable that equals 1 if a person is attending York University, U of T Mississauga, or U of T Scarborough, 0 otherwise
IVTT/km	The time spent travelling in a transit vehicle minutes), divided by the length of the trip (km)
DrivLic	A dummy variable that equals one if a person possesses a driver's license, 0 otherwise
Auto_Access	A dummy variable that equals one if a person reports owning a car or a membership in a car-sharing service, 0 otherwise
Inc>$120K	A dummy variable that equals one if a person's household income was more than $120,000 in the past year, 0 otherwise
AT	The time spent by a person traveling from the origin of their trip to the access stop of their trip (minutes)
ET	The time spent by a person traveling from the egress stop of their trip to the destination of their trip (minutes)
CBD	A dummy variable that equals one if a zone is in the central business district, 0 otherwise
Shopping	A dummy variable if the purpose of the trip was reported as 'shopping,' otherwise
Man/sqkm	The number of points of interest classified as 'manufacturing' in the zone divided by the area of said zone [sq. km]
Edu/sqkm	The number of points of interest classified as 'educational services' in the zone divided by the area of said zone [sq. km]
FA/sqkm	The number of points of interest classified as 'food and accommodation' in the zone divided by the area of said zone [sq. km]
RT/sqkm	The number of points of interest classified as 'retail trade' in the zone divided by the area of the said zone [sq. km]
FTG	A dummy variable that equals one if the person is a full-time graduate student, 0 otherwise
FEMALE	A dummy variable that equals one if a person is female, 0 otherwise
nVeh_1	A dummy variable that equals one if a person reports that their household owns one vehicle, 0 otherwise
Age>22	A dummy variable that equals one if a person is over the age of 22, 0 otherwise
Zone_Area	The area of the TAZ, in square kilometers
%Res	The proportion of a TAZ that is classified as 'residential.'
%Comm	The proportion of a TAZ that is classified as 'commercial.'

Table 4: Model Parameters

Variable Description	Final Model		Model to Calculate Aggregate Accessibility	
	Parameter	t-stat	Parameter	t-stat
Level-of-Service and Interactions				
nTrans – mean effect	-0.5844	-6.102	-0.6228	-8.624
nTrans x Sub_Campus = 1	-0.1415	-1.003	-	-
IVTT/km – mean effect	-0.0132	-1.002	-0.0174	-1.643
IVTT/km x DrivLic = 1 x Auto_Access = 1	-0.0160	-0.670	-	-
IVT/km x Inc>$120K = 1	-0.0472	-0.979	-	-
AT – mean effect	-0.0476	-1.986	-0.0461	-1.937
ET x CBD = 0 x Shopping = 1	-0.0317	-1.480	-	-
Points of Interest and Interactions				
ln(Man/sqkm + 1) – mean effect	-0.5017	-7.385	-0.5181	-6.373
ln(Edu/sqkm + 1) – mean effect	-0.0122	-0.187	0.1233	1.987
ln(Edu/sqkm + 1) x FTG = 1	0.2001	2.309	-	-
ln(FA/sqkm + 1) – mean effect	0.7007	6.745	0.564	5.867
ln(FA/sqkm + 1) x FEMALE = 1	-0.1568	-2.576	-	-
ln(FA/sqkm + 1) x nVeh_1 = 1	-0.1060	-1.770	-	-
ln(RT/sqkm + 1) – mean effect	0.5378	5.220	0.3877	3.536
ln(RT/sqkm + 1) x Age>22	-0.1259	-2.002	-	-
Zonal Size Measures				
Zone_Area – mean effect	0.5655	6.446	0.5391	5.815
%Res – mean effect	-0.8953	-4.591	-1.0763	-5.346
%Comm – mean effect	0.149	0.287	0.6592	1.385
Adjusted rho-squared	0.10		0.06	
Number of observations	539		539	

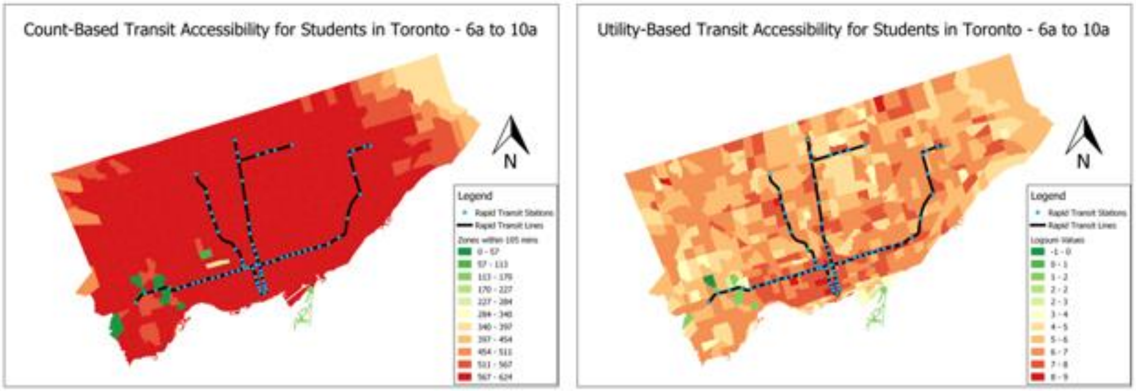
Overall, many of the variables are significant, and the parameter estimates for the land use and LOS attributes have the expected signs. For example, the number of transfers and access time reduces the probability that a zone is chosen. Interestingly, neither in-vehicle travel time nor egress time proved to have the same effect when included in the model on their own. This result may stem from the fact that mode choice was not explicitly modelled in this research, combined with the fact that the Google API tries to find the best route possible. This result implies that, to choose a location to travel to via public transit to participate in a discretionary activity, university students are sensitive to the number of times they must change routes and the time required to reach a public transit service. Also, the negative parameter associated with the IVTT/km variable may imply a preference for transit routes with shorter travel times, such as those with a dedicated right-of-way. The interaction of the nTrans and Sub_Campus variables implies that students who attend classes at university campuses located outside of downtown Toronto are more sensitive to the number of transfers required to reach their destination. Similarly, it appears that persons who could have completed their trip by car (i.e., those with both a driver's license and access to an automobile) and persons from households who reported earning more than $120,000 are both less likely to choose a location that requires them to spend more time in a transit vehicle relative to the distance travelled.

The density of points of interest (i.e., number per zonal area) was used instead of the number of points to account for the fact that larger zones have the potential to include more of a given classification of point of interest. The densities were first log-transformed before they were included in the model, due to the wide range of values that were observed for these attributes. A value of 1 was added to the density values before their transformation to ensure that all of the resulting values were positive, enabling the parameter values to be interpreted more easily. Based on the parameter estimates, a higher density of retail trade and food and accommodation locations is positively associated with the propensity to travel to a zone for a discretionary trip. Similarly, the density of manufacturing and educational services locations is associated with a decreased probability that the zone will be selected. This result is reasonable, given the trip purposes that were defined as discretionary, and the tendency for manufacturers to locate their facilities in more industrialized areas.

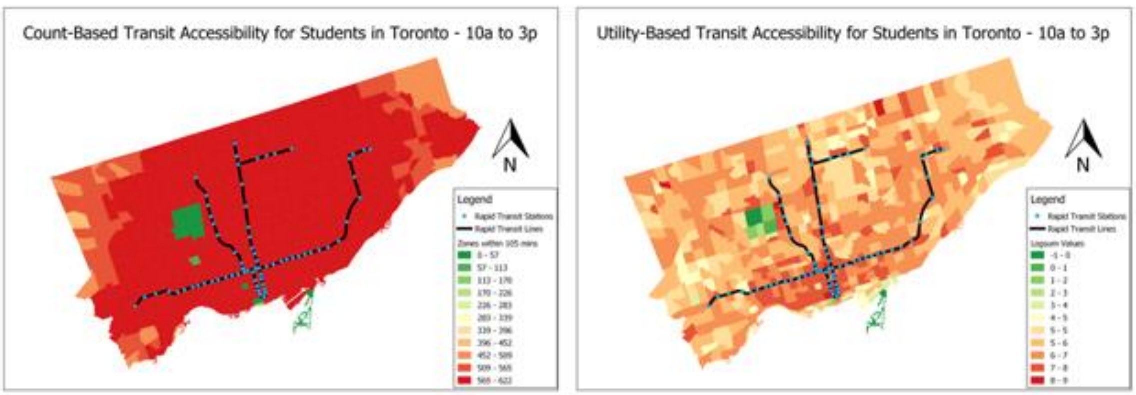
Full-time graduate students are more likely to choose a location if it has a higher density of educational services, while the mean effect indicates the opposite is true for university students overall. It appears that female students and students from households that own one vehicle are less likely to choose a zone with a higher density of food and accommodation locations in comparison to university students as a whole. It should be noted that even for female students with one household vehicle, the density of these locations is still associated with a higher probability that the corresponding zone will be chosen. Finally, it appears that the area of a zone is positively associated with the likelihood that it will be chosen.

The second location choice model containing only the LOS and land use attributes was estimated to derive utility-based measurements of accessibility. The specification of this model is also shown in Table 4. Aside from the %Comm parameter, each of the estimated parameters is statistically significant at a 95% confidence level. Because the set of feasible zones that can be reached from a given TAZ cannot be inferred using activity schedules, a zone was considered feasible if it could be reached within the maximum travel time reported in the StudentMoveTO survey – 105 minutes. This criterion also formed the basis for how the count-based accessibility measures were derived. For each TAZ, the location choice model was applied to every other zone that could be reached by public transit within 105 minutes; this entailed the calculation of the utility that could be obtained by choosing each zone. Using this approach, the utility-based accessibility of university students in Toronto by public transit for discretionary activities was calculated for each TAZ. The count-based accessibility of university students in Toronto by public transit was calculated by counting the number of TAZs that could be reached within 105 minutes by transit. These calculations were performed for each of the three periods for which LOS attributes were imputed.

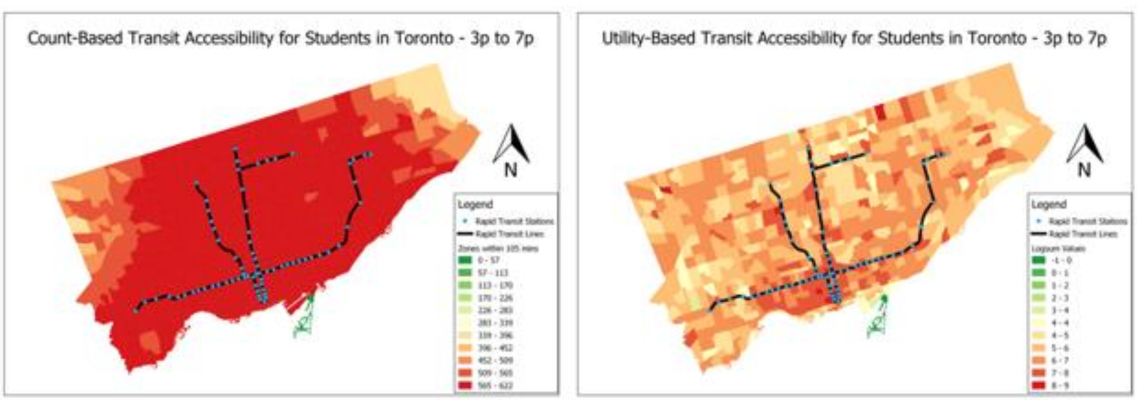
The values of the count- and utility-based accessibility measures for each TAZ in Toronto were plotted using GIS software; the results are shown in Figure 16. It should be noted that count-based and utility-based accessibility measures cannot be compared in quantitative terms, due to the differences in the approach used to compute these values. To facilitate a qualitative comparison between the two measures, each value was assigned to one of 11 equally spaced categories. The spacing of the categories was determined by dividing the difference between the highest and lowest values by 11. Examining the maps shown in Figure 16, two trends are immediately clear. First, accessibility by public transit for university students is highest in the downtown core of Toronto. Second, in comparison to utility-based accessibility measures, count-based measures appear to over-estimate the accessibility of a given TAZ. Particularly in the north-east and north-west of the city, where land use patterns tend to be auto-centric, it appears that the count-based approach overstates accessibility by public transit. This may be because count-based measures combine different aspects of transit travel time, whereas the utility-based measures explicitly account for the impact of access time. Also, utility-based accessibility measures can account for the role that land use plays in the attractiveness of the zones that can be reached within a set amount of time. Generally speaking, the utility-based measures of accessibility tend to be higher in zones that include or are adjacent to subway stations. A few exceptions are seen in the western part of the city, and upon further examination, this appears to be due to the inclusion of parks and ravines within the TAZ and the proximity of the zonal centroid to these types of areas.



(a) Count-based (left) and utility-based (right) accessibility, 6 AM to 10 AM



(b) Count-based (left) and utility-based (right) accessibility, 10 AM to 3 PM



(c) Count-based (left) and utility-based (right) accessibility, 3 PM to 7 PM

Figure 16: A Comparison of Count- and Utility-Based Accessibility Measures for University Students in the City of Toronto by Public Transit

5.6 Conclusions and Future Work

This chapter presents the results of an investigation into the factors that influence the location choice behaviour of university students when using public transit to participate in discretionary activities. This work aims to address several gaps in the literature and to contribute to an improved understanding of the travel behaviour of post-secondary students. In many operational travel demand models, post-secondary student travel behaviour is accounted for by the representation of post-secondary institutions as special generators. This approach is insufficient because it neglects the impacts of post-secondary students making non-school trips on the overall travel demand of a region. Also, this application of a trip-based approach to analyzing travel demand is becoming outdated with the rise of the activity-based approach.

The results of the empirical model indicate that the location choice process for university students is influenced by the time required to access transit, the number of transfers required to complete the trip, and the land use patterns of a zone. Besides, it appears that the socio-economic characteristics of the decision-maker affect the extent to which the aforementioned variables affect the location choice process. The location choice model was used to derive a utility-based measure of accessibility, which reflects the accessibility experienced by a university student in Toronto when travelling by public transit. This measure was compared to a count-based accessibility measure, which was computed by counting the number of TAZs that can be reached from a given TAZ within 105 minutes, which was the longest reported travel time. These values were computed for three different times of the day and were mapped to facilitate a qualitative analysis between the two measures. Overall, it appears that count-based measures over-estimate accessibility, in part due to the treatment of the access, egress, and in-vehicle travel time as single travel time. From a policy standpoint, the accessibility of university students can be improved by reducing the amount of time it takes to reach a transit stop and providing more direct services, i.e., reducing the number of transfers required. Also, accessibility could be improved by encouraging the inclusion of retail trade and food services in new multi-family residential developments.

The results presented in this paper aim to contribute to an improved understanding of post-secondary student travel behaviour within an activity-based framework. Future work includes the imputation of transit level-of-service attributes throughout the day and the utilization of a dataset that includes both university and college students. Also, estimating the model using the entirety of the feasible choice set may produce biased results if the decision-makers only consider a small number of alternatives. This issue can be assuaged by developing a joint model of choice set generation and location choice to provide a more realistic basis for model estimation.

Chapter 6
Combining Passive and Core Survey Data

The growing availability of data from passive sources creates the potential to gain insights into behaviours that are overlooked by or under-reported in traditional household travel surveys. Passive data often contain a greater number of records than household travel surveys and tend to represent greater proportion of the members of the target population. One key issue with the use of passive data is the lack of associated demographic information, and in some cases, trip information (such as mode and purpose). In spite of these issues, passive data can supplement data obtained through a main core survey, particularly when used for aggregate approaches to analysis.

This chapter presents an empirical study of the relationship between the use of ride-hailing services, such as those offered by Uber and Lyft, and public transit. The analysis presented in this chapter is based on the use of data obtained through the 2016 TTS and data on ride-hailing trips provided by the City of Toronto. The data provided by the City was used to quantify the utilization of ride-hailing services in the study area. Although the 2016 TTS included “paid rideshare” as a primary mode of travel, there is a large discrepancy in the number of relevant observations in each dataset. The 2016 TTS included 1,357 reported trips that were made using a paid ridesharing service, which when expanded, represents a total of 31,417 trips originating in Toronto. By comparison, the data provided by the City contains 4,469,025 trip records for the period corresponding to the TTS survey period, September 7 to December 16, 2016. The data from the City provides a near-comprehensive record of ride-hailing usage in Toronto over this period of time.

The rest of the chapter is organized as follows: first, the motivation for this study is summarized. Then, contemporary studies on ride-hailing usage and transit ridership are reviewed. Next, a description of the study area and data used for the study is presented. Afterward, the structure of the empirical model is described, and the results of the model are presented and discussed.

6.1 Motivation

Ride-hailing services, such as those offered by Uber and Lyft, have had a transformative impact on how people travel and are challenging traditional perceptions of mobility. The growing adoption and popularity of these services have inevitably impacted the use of other modes of transportation. In the North American context, the relationship between ride-hailing and public transit services has received increased attention from academicians, policymakers, and journalists alike. This focus partially stems from the fact that the growth of ride-hailing services has come at a time where transit ridership has stagnated or declined in North American cities. This decline should be of concern to transit agencies, as larger agencies tend to rely on fare revenues to help fund operations. Besides, the increased use of ride-hailing has the potential to worsen congestion, which, in turn, could degrade the quality of the services offered by transit agencies. Consequently, the nature of the relationship between the use of ride-hailing and public transit services, and the factors that influence it, should be of interest to both transit agencies and policymakers.

Despite the growing body of work on the topic, the nature of the relationship between ride-hailing and public transit tends to vary from one context to another. In the literature, studies on the relationship between the two services tend to find that ride-hailing tends to complement public transit in some cases while acting as a substitute in others. Overall, the understanding of the impact of ride-hailing services on travel behaviour is still in the developmental stages, which can partially be attributed to the lack of publicly available data on the use of ride-hailing services. Consequently, many studies on the use of ride-hailing services tend to be based on data obtained through surveys or a sample of data on ride-hailing trips. Notable exceptions include (Gerte, Konduri, & Eluru, 2018) who used data on Uber trips made in New York City, and (Dias, Lavieri, Kim, Bhat, & Pendyala, 2019), who used data on ride-hailing trips provided by RideAustin. Both of these studies utilized comprehensive sets of information on ride-hailing trips, collected over several months, to understand the factors that influence the extent to which ride-hailing services are used. The increasing availability of this type of data creates new opportunities to develop a more comprehensive understanding of the travel behaviour of ride-hailing users and its impact on overall travel demand.

Trip generation is a fundamental component of the traditional four-stage model of travel demand analysis. Although this stage traditionally precedes mode choice in this framework, taking a mode-specific approach to analysis can provide insights into the influence of zonal and other aggregate attributes on the generation of trips made by a given mode. This chapter presents the results of an investigation into the factors that affect the magnitude of the ride-hailing and transit trips that are generated by each Dissemination Area (DA) in the City of Toronto. The investigation combines regional household travel survey data with a comprehensive set of data on ride-hailing trips made in Toronto between September 2016 and September 2018. The results of this study aim to provide insights into the role that built environment and zonal attributes play in the extent to which public transit and ride-hailing services are utilized and the influence of one service on the other. These insights can help agencies take a more proactive approach to identify areas where ridership may be cannibalized by ride-hailing services, while also helping to inform policies that aim to encourage a complementary relationship between the two.

6.2 Background

The adoption and growing usage of ride-hailing services have led to numerous attempts to understand the factors that influence the use of ride-hailing services. In terms of adoption, the common finding in the literature is that the use of ride-hailing services tends to be more prevalent among younger people, especially those with higher incomes or higher levels of education (Clewlow & Mishra, 2017). Alemi, et al. (2018) investigated the determinants of the adoption of ride-hailing services using data obtained from the California Millennials dataset. Using this dataset, the authors estimated a binary logit model to identify the factors that influence whether a person has used ride-hailing services before. Overall, it was found that living in a metropolitan area, land use mix, and regional automobile accessibility have a positive impact on the probability that a person has used a ride-hailing service (Alemi, Circella, Handy, & Mokhtarian, 2018). Besides, the authors also found that older millennials (aged 25 to 34) with higher incomes and education levels were more likely to adopt ride-hailing services. Using the same dataset, Alemi, Circella, and Sperling (2017) found that cost and the preference to use one's vehicle were crucial factors that prevented respondents from using ride-hailing services (Alemi, Circella, & Sperling, 2017).

Aside from the adoption of ride-hailing services, the extent to which ride-hailing services are used has received significant attention. Attempts to understand the use of ride-hailing services are typically based on the use of descriptive statistics or the estimation of empirical models. Studies that fall into the first category typically rely on data obtained through surveys to understand the tendencies of ride-hailing users. Due to the relative novelty of ride-hailing services, a common goal of these studies is to understand why ride-hailing services are used.

Compared to other travel surveys, surveys of ride-hailing users are distinguished by the inclusion of questions about why the respondents chose to use ride-hailing services for a particular trip, and the mode that they would have used had said services not been available. One example of a study that uses this approach is described in (Rayle, Dai, Chan, Cervero, & Shaheen, 2016), who conducted an intercept survey of persons in San Francisco who had either just completed a ride-hailing trip or had completed one in the past two weeks. Using this information, the authors found that the availability of ride-hailing services appears to induce travel; 8% of respondents reported that they would not have made their trip if ride-hailing had not been available (Rayle, Dai, Chan, Cervero, & Shaheen, 2016). Also, the authors found that the desire to avoid drinking and driving, as well as speed and convenience, were essential motivators of ride-hailing use.

Apart from survey data, some studies have been able to utilize travel information provided by ride-hailing companies to analyze the usage of ride-hailing services quantitatively. One of the earliest examples of a study of this nature was (Gerte, Konduri, & Eluru, 2018), who used the information on the pick-up location of Uber trips obtained by the website FiveThirtyEight. This study utilized weekly pick-up volumes in New York City, aggregated at the taxi zone level, from April through September 2014 and January through June 2015 to quantify temporal trends in ride-hailing use. Using a panel-based random-effects model, the authors identified the influence of several variables related to demographics and the built environment. Specifically, the authors found that the total built area and the proportion of a zone (by floor area) dedicated to retail and residential use had a positive impact on the number of weekly Uber pick-ups in a zone. Additionally, the authors found that the percentage of residents under the age of 19 in a zone had a positive impact on the number of pick-ups, while the average household size displayed a negative association (Gerte, Konduri, & Eluru, 2018). Another example of a comprehensive set of information on ride-hailing trips is the data provided by RideAustin, a ride-hailing company that began operations in Austin in 2016 (Dias, Lavieri, Kim, Bhat, & Pendyala, 2019).

Containing information on trips that took place between June 2016 and April 2017, this dataset has been used to investigate a variety of different aspects of travel behaviour. For example, Lavieri, et al. (2018) used this dataset to model and analyze the demand for ride-hailing services. In this study, the authors modelled the generation of ride-hailing trips in traffic analysis zones located in central Austin on weekdays and weekends and developed a fractional split model to examine the factors that influence the attraction of ride-hailing trips. In terms of trip generation, the authors found that the average frequency of buses at the average bus stop within a zone was negatively associated with the generation of ride-hailing trips on weekdays. On the other hand, population density, retail employment density, and the median annual household income all displayed a positive relationship (Lavieri, Dias, Juri, Kuhr, & Bhat, 2018). With regards to the distribution of ride-hailing trips, the authors found that retail employment density and median household income, among other factors, were positively associated with the attraction of ride-hailing trips. Finally, Dias et al. (2019) used the same data set to investigate the influence of socio-economic and demographic factors on the number of ride-hailing trips made to a set of six purposes (work, airport, visit CBD, education, commercial, and recreation) (Dias, Lavieri, Kim, Bhat, & Pendyala, 2019). The authors utilized a multivariate ordered probit model to analyze the number of trips that "frequent" users (i.e., those who visited the same location more than ten times during the study period) made for each of the aforementioned purposes. This modelling framework allowed for the correlation of the error terms corresponding to the latent propensity for each destination purpose. The authors found that the influence of factors such as gender, age, and income vary based on the destination purpose and that some correlations that exist between the error terms of different trip purposes were significant.

Many contemporary studies on the utilization of ride-hailing services include an analysis of the impacts of these services on the use of public transit. For example, in their synthesis of prior studies that investigate the effects of ride-hailing on travel behaviour, (Rodier, 2018) found evidence that ride-hailing acts as both a complement to and substitute for transit service. Specifically, the complementary effect is observed when providing access to heavy rail and transit stations, although ride-hailing is also frequently used as a substitute for transit. This finding was echoed by (Clewlow & Mishra, 2017), who observed that ride-hailing services tend to draw ridership away from surface transit while complementing commuter rail services. Overall, it appears that the substitution effect is stronger than the complementary effect, in part due to the relatively shorter travel times that can be offered by ride-hailing services (Rodier, 2018). In an attempt to understand the motivators for using hailing ride services instead of public transit, some studies have compared the characteristics of ride-hailing trips to those of the trip if it were made by transit. For example, (Rayle, Dai, Chan, Cervero, & Shaheen, 2016) found that roughly two-thirds of the ride-hailing trips reported in their survey would take twice as long if they were completed by transit, with 86% taking at least 50% longer. Interest in the influence of ride-hailing services on transit ridership has reached the point where the former is explicitly considered in models of the latter.

Some studies of transit ridership have found that ride-hailing services have a positive influence on transit ridership. For example, in their study of transit ridership in 25 North American cities over 14 years, (Boisjoly, et al., 2018) found that the presence of Uber has a positive but insignificant effect on ridership. Similarly, (Hoffmann, Ipcirotis, & Sundararajan, 2016) found a positive correlation between the volume of subway turnstile entries and Uber pick-ups in their study of New York City. On the other hand, other studies have found that the impacts of ride-hailing depend on the size of the agency in question. For example, in their study of the transit ridership of 103 Canadian agencies between 2002 and 2016, Diab, Kasraian, Miller, & Shalaby (2018) found that the presence of Uber had a small but positive impact on ridership. However, for agencies whose ridership was less than 1.2 million linked trips, the presence of ride-hailing services was typically associated with a decrease in ridership. A similar result was reported in (Hall, Palsson, & Price, 2018), in their study on the net effect of Uber on transit ridership in Metropolitan Statistical Areas (MSA) in the United States. The authors found that transit ridership slowly increased for the first 24 months following the entry of Uber; however, the presence of Uber tends to decrease ridership in smaller MSAs (Hall, Palsson, & Price, 2018). The work outlined in Sadowsky and Nelson (2017) takes an interesting approach to this question by analyzing the impacts of the first and second entry of a ride-hailing company into an urbanized area. It was found that all else being equal, the first entry of a ride-hailing company was associated with an increase in public transit ridership, while the entry of a second company completely reverses this increase. The authors speculate that competition created by the entrance of a second company reduces the cost of ride-hailing trips to the point that it is more cost-effective to use these services for the entirety of a trip, rather than solely as a means to access transit (Sadowsky & Nelson, 2017). This study aims to contribute to the literature by investigating the role that zonal and built environment attributes play in the generation of ride-hailing and transit trips in Toronto. In particular, this study involves the joint estimation of a model of the usage of ride-hailing and public transit services in Toronto using the bivariate ordered probit model, which also allows for the identification of the correlation between the factors that drive the demands for the two modes.

6.3 Data Description and Study Area

The empirical work presented in the chapter focuses on the City of Toronto. As shown in Figure 17, the number of ride-hailing trips made on the average weekday has more than doubled between September 2016 and September 2018, while the average number of weekday transit trips has declined during the same period. Two agencies provide transit services in Toronto – the Toronto Transit Commission (TTC), which provides local transit services, and GO Transit, which provides regional transit service. In terms of ride-hailing companies, Uber began operating in Toronto in 2014, while Lyft began its operations in 2017 (City of Toronto - Big Data Innovation Team, 2019). Per the 2016 Canadian Census, Toronto is home to 2.73 million residents, with the area of the city being divided into 3,702 Dissemination Areas (DAs) (Statistics Canada, 2017), as shown in Figure 18. These dissemination areas form the basis of the analysis presented in this chapter.

Figure 17: The average number of weekday ride-hailing trips (left) (City of Toronto - Big Data Innovation Team, 2019)and transit trips (right) since September 2016

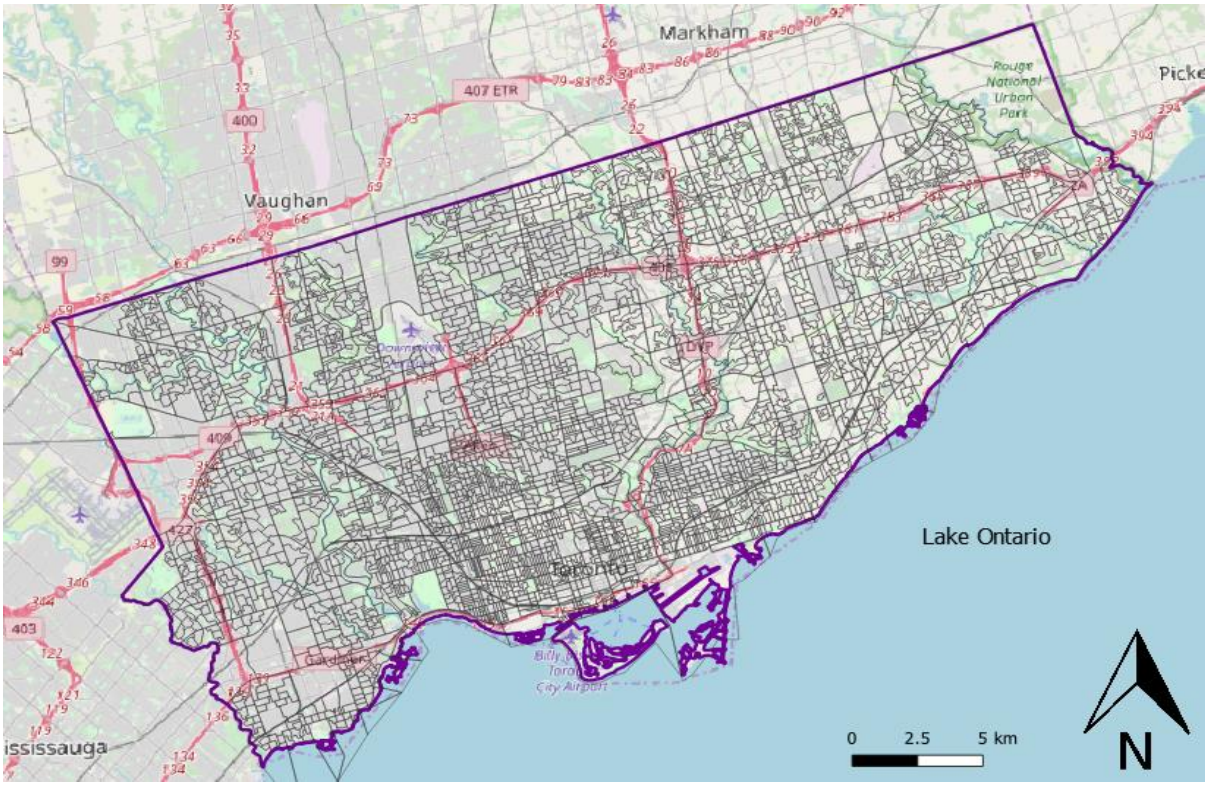


Figure 18: Dissemination Area (DA) Boundaries in Toronto

The travel information used in the analysis was obtained from two sources. The number of transit trip generated by each zone was obtained from the Transportation Tomorrow Survey (TTS). The 2016 version of the TTS was conducted from September 7 to December 16, 2016 and contains information on 395,885 persons and 798,093 trips (R.A. Malatest & Associated Ltd., 2018). Information on the use of ride-hailing in Toronto was provided by Uber and Lyft, via the City of Toronto, as part of the requirements of the by-law governing the operations of ride-hailing companies in Toronto (City of Toronto - Big Data Innovation Team, 2019). The data correspond to the period between September 2016 and September 2018 and contains roughly 23 million time-stamped records of the origin and destination of each trip, mapped to the nearest intersection. These records were mapped to the corresponding dissemination area.

One of the issues that had to be addressed before using the two sets of data is that the data obtained from the TTS is meant to represent a typical fall weekday, while the data provided by Uber and Lyft contains trip information collected throughout two years. Additionally, the number of ride-hailing trips made on an average weekday has increased throughout the study period, meaning that choosing any single date could produce biased results. The process used to create a dataset of zonal ride-hailing usage that was compatible with the TTS data is summarized in the equation below. First, the records of all ride-hailing trips made during the survey period were extracted. These records reflect the number of ride-hailing trips that began in each DA on each day that the survey took place (denoted as D_{ij} for DA j and survey day i). These records were weighted based on the proportion of total survey responses that were received from households in Toronto on day i of the survey (denoted as P(i) for survey day i). This process produced the number of ride-hailing trips generated by the j^{th} DA on a typical day during the survey period (denoted as RH_j).

$$RH_j = \sum_{i=1}^{N} P(i) * D_{ij}$$

Summary statistics for the DA-level generation of ride-hailing and transit trips are shown in Table 5. Of the 3,702 DAs in Toronto, 3,557 included information on transit trip generation, while 3,610 had information about ride-hailing trips. In total, 3,473 dissemination areas had records of both ride-hailing and transit trip generations; these DAs were used for the analysis presented in this study in order to avoid the need to make assumptions in cases where records did not exist.

Table 5: Summary Statistics of Ride-Hailing and Public Transit Trip Generation

Statistic	Ride-Hailing	Public Transit
Minimum	0.00	3.55
First Quartile	3.12	74.20
Median	6.10	161.36
Third Quartile	13.31	362.08
Maximum	2095.74	75760.88
Mean	17.98	418.84
Standard Deviation	61.41	1660.59

In addition to travel data, information on zonal attributes and built environment characteristics were obtained. Two datasets that are produced by DMTI Spatial Inc., specifically the CanMap RouteLogistics and Enhanced Points of Interest (EPOI) datasets were included in the analysis. The former includes information on land use classifications (DMTI Spatial Inc., 2014), while the latter assigns a classification to each so-called point-of-interest (POIs) in Canada (DMTI Spatial Inc., 2016). Classifications in the EPOI dataset include food and accommodation, education, and retail. Zone-level demographic characteristics were obtained from the 2016 Canadian Census.

Finally, zone-level transit network attributes were obtained from the General Transit Feed Specification (GTFS). Because the boundaries of dissemination areas typically coincide with major roads, transit network attributes were assigned to DAs based on the definition of buffers centred on the centroid of each zone. Specifically, the length of the bus, streetcar, and subway routes, as well as the number of transit stops, within 400 and 800 m of each centroid were calculated for use as explanatory variables. The inclusion of these variables follows the work presented by studies such as (Boisjoly, et al., 2018; Diab, Kasraian, Miller, & Shalaby, 2018) who have found that the vehicle revenue kilometers provided by transit agencies influence ridership.

6.4 Empirical Model

The empirical analysis presented in this chapter utilizes the bivariate ordered probit model, which is an extension of the univariate probit model first proposed by (McKelvey & Zavoina, 1975). This model analyzes the influence of various factors on the outcomes of an ordinal nature and can capture interactions between the two response variables. Similar to the univariate case, the explanatory variables are assumed to influence the value of a latent variable, whose value is reflected as a discrete, ordinal outcome. For the bivariate case, this is formalized as (Greene & Hensher, 2009):

$$y_{i,1}^{*} = \beta' x_{i,1} + \varepsilon_{i,1} \rightarrow y_{i,1} = j \; if \; \mu_{j-1} < y_{i,1}^{*} < \mu_j \; for \; j = 0, \dots, J_1$$

$$y_{i,2}^{*} = \beta' x_{i,2} + \varepsilon_{i,2} \rightarrow y_{i,2} = j \; if \; \delta_{j-1} < y_{i,2}^{*} < \delta_j \; for \; j = 0, \dots, J_2$$

The unobserved factors that influence the latent variables (ε_1 and ε_2) are assumed to follow the standard bivariate normal distribution, i.e.:

$$\begin{pmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]$$

Consequently, the probability of a given pair of outcomes, $y_{i,1} = j$ and $y_{i,2} = k$ is given by:

$$P\left(y_{i,1} = j, y_{i,2} = k \mid x_{i,1}, x_{i,2}\right) =$$

$$\left[\Phi\left[\left(\mu_j - \beta_1' x_{i,1}\right), \left(\delta_k - \beta_2' x_{i,2}\right), \rho\,\right] - \Phi\left[\left(\mu_{j-1} - \beta_1' x_{i,1}\right), \left(\delta_k - \beta_2' x_{i,2}\right), \rho\,\right]\right]$$

$$- \left[\Phi\left[\left(\mu_j - \beta_1' x_{i,1}\right), \left(\delta_{k-1} - \beta_2' x_{i,2}\right), \rho\,\right] - \Phi\left[\left(\mu_{j-1} - \beta_1' x_{i,1}\right), \left(\delta_{k-1} - \beta_2' x_{i,2}\right), \rho\,\right]\right]$$

One of the more common applications of this model in transportation research is to understand the factors that influence both the severity and number of automobile accidents. An example of the former is described in (Li, Hasnine, Habib, Persaud, & Shalaby, 2017), who used a binary ordered probit model using crash data from Toronto to investigate the factors that influence higher levels of injury severity. Multivariate versions of this model have been used to jointly model the number of times that each outcome, among a given set of outcomes, is observed. While these numbers are cardinal, they are effectively treated as ordered in these applications, particularly when less frequent outcomes are grouped. An example of this application is described in (Ferdous, Eluru, Bhat, & Meloni, 2010), who jointly modelled the number of non-work activity episodes generated by the respondents of the 2007 American Time Use Survey. Other applications of the bivariate ordered probit model in transportation research is the modelling of household mobility tool ownership, as is described in (Scott & Axhausen, 2006).

In order to use the model mentioned above, the trip generation values had to be converted from a continuous to an ordered outcome. Due to the exploratory nature of this analysis, a relatively simplistic approach to categorization was taken, wherein each level of transit and ride-hailing trip generation contained the same number of observations. Specifically, each DA was assigned two levels, one related to the number of transit trips generated and another based on the number of ride-hailing trips generated. Each DA was assigned these levels based on the segmentation of the two sets of data into halves, thirds, and quarters. The inclusion of various categorization strategies was based on a desire to ensure that each pair of categories included at least 100 observations. Using this approach, four different models were defined and estimated. The categorization approaches for the generation of transit and ride-hailing trips are summarized in Table 6

Table 6: Definitions of Empirical Models

Model #	Ride-Hailing Classification	Public Transit Classification	Total Number of Categories
1	Halves	Thirds	6
2	Thirds	Halves	6
3	Thirds	Thirds	9
4	Quarters	Quarters	16

6.5 Results and Discussion

Each of the four models was estimated using ‘mvord’ package that is available in the statistical computing software R. The package allows for the estimation of multivariate ordered regression models (Hirk, Hornik, Vana, & Genz, 2019). The final specification of each model contained three types of explanatory variables: transit network attributes, land use density indicators, and zonal attributes. The land use density indicators (LUDI) were derived using information from the EPOI dataset described earlier; for classification k, the density in zone i is given by:

$$LUDI(purpose\ k) = ln\left[\frac{Number\ of\ POIs\ of\ type\ k\ in\ zone\ i}{Area\ of\ zone\ i\ (km^2)} + 1\right]$$

For each of the models, the same explanatory variables were included in both equations in order to facilitate a direct comparison of their impacts on the usage of the two modes. During the model estimation process, parameters were retained in the model based on their significance and their sign; however, some insignificant parameters were retained to allow for comparisons with other parameters. Variables about land use, zonal demographics, and transit network attributes were tested. Most notably, variables regarding land use classifications were omitted from the final specification of the models instead of variables derived from the EPOI dataset. The decision was made based on the ability of the latter to capture the number of different types of opportunities that are present within a DA, rather than just the proportion of the zonal area devoted to each type of opportunity. Also, the EPOI dataset provides a broader range of classifications. In terms of zonal demographics, several different approaches to the aggregation of age and income groups were tested. The explanatory variables that were included in the final specifications of the four models are summarized in Table 7, with the model results being summarized in Table 8.

Table 7: Summary and Description of Explanatory Variables

Parameter	Definition
bus_400	The length of bus routes (km) within a 400 m buffer
streetcar_400	The length of streetcar routes (km) within a 400 m buffer
subway_800	The length of subway lines (km) within an 800 m buffer
stops_400	The number of stops within a 400 m buffer
ln_acco_food_density	The density of 'accommodation and food' locations in the DA
ln_art_ent_rec_density	The density of 'art, entertainment, and recreational' locations in the DA
ln_edu_density	The density of 'education services' locations in the DA
ln_health_density	The density of 'health services' locations in the DA
ln_priv_dwelling_density	The density of private dwellings in the DA
ln_retail_density	The density of 'retail trade' locations in the DA
under_25	The number of persons under the age of 25 that live in the DA
area_sq_km	The area of the DA, in square kilometers
avg_income ($1000's)	The average income of the residents of the DAs, in $1000s

Table 8: Summary of Model Results

Parameter	Model 1		Model 2		Model 3		Model 4	
	RH	PT	RH	PT	RH	PT	RH	PT
Transit Attributes								
Bus_400	0.014	0.001	0.013	0.045 *	0.012	0.002	0.016	0.001
Streetcar_400	0.498 #	0.228 #	0.527 #	0.173 #	0.529 #	0.227 #	0.565 #	0.189 #
Subway_800	0.296 #	0.371 #	0.310 #	0.324 #	0.312 #	0.369 #	0.360 #	0.377 #
Stops_400	0.027 #	0.028 #	0.026 #	0.030 #	0.026 #	0.028 #	0.026 #	0.027 #
Land Use Density Indicators								
ln_acco_food_density	0.188 #	0.069 ^	0.154 #	0.118*	0.156 #	0.069 ^	0.216 #	0.065%
ln_art_ent_rec_density	0.365 #	0.207 #	0.289 #	0.210 #	0.291 #	0.210 #	0.351 #	0.238 #
ln_edu_density	0.091 *	0.211 #	0.099%	0.171 #	0.099%	0.210 #	0.108%	0.211 #
ln_health_density	0.182 #	0.254 #	0.175 #	0.232 #	0.175 #	0.255 #	0.148 #	0.265 #
ln_priv_dwelling_density	0.271 #	0.253 #	0.262 #	0.334 #	0.261 #	0.252 #	0.245 #	0.222 #
ln_retail_density	0.157 #	0.001	0.198 #	0.067	0.198 #	0.001	0.144 #	0.001
Zonal Attributes								
under_25	0.001 #	0.004 #	0.001 #	0.004 #	0.001 #	0.004 #	0.001 #	0.003 #
area_sq_km	1.228 #	1.933 #	1.265 #	1.868 #	1.267 #	1.934 #	1.351 #	1.806#
avg_income ($1000's)	0.003 #	-0.001 #	0.001 #	-0.001 #	0.001 #	-0.001 #	0.001 #	-0.001 #
Correlation Parameter	0.001 #	0.004 #	0.001 #	0.004 #	0.001 #	0.004 #	0.001 #	0.003 #
Threshold Parameters								
1|2	2.441 #	2.057 #	1.797 #	2.900 #	1.794 #	2.055 #	1.437 #	1.533 #
2|3	-	3.276 #	2.959 #	-	2.955 #	3.273 #	2.312 #	2.424 #
3|4	-	-	-	-	-	-	3.278 #	3.428 #
Model Fit Statistics								
Number of Observations	3473		3473		3473		3473	
McFadden's Pseudo R2	0.251		0.2388		0.2296		0.2022	
Loglikelihood of Final Model	-4661		-4737		-5879		-7682	

Note: Significance levels: #: >99.9%, %: 99.9%, *: 99%, ^: 95%

The results summarized in Table 8 appear to indicate that the transit network and land use density indicators have the same type of effect on the use of ride-hailing services as they do on the use of public transit. The parameters pertaining to the length of streetcar and subway lines, as well as the parameter corresponding to the number of transit stops, were positive for both transit and ride-hailing. This result implies that increases in these attributes are associated with increased levels of transit and ride-hailing usage. Similar results were found for each of the land use density indicators, implying that increases in the density of private dwelling and locations that provide services such as ‘food and accommodation,’ ‘health,’ and ‘art, entertainment, and recreation’ are associated with the increased generation of ride-hailing and transit trips. The only case where the signs were not the same for the two modes was the average income parameter, which was positive for ride-hailing and negative for public transit. This result may be due to the design of the road network in wealthier residential areas, which is often meant to prioritize private vehicle usage and reduce through traffic. Consequently, the residents of these areas tend to lead more auto-centric lifestyles, and public transit services tend to be provided on a relatively infrequent basis. Each of the four correlation parameters have a positive sign, which is reasonable given that the majority of the explanatory variables have the same sign in the two models. This result implies that the unobserved factors that influence the generation of ride-hailing trips are positively correlated with the unobserved factors that influence the generation of transit trips. This may be due to the fact that the variables included in the models influence overall trip generation, not just the generation of trips by any one mode. The positive sign of the correlation parameter may be indicative of the fact that mode choice was not considered in this analysis, rather than evidence of a complementary relationship between ride-hailing and public transit services.

Because the regression reveals the relationship between the explanatory variables and a latent variable that is reflected in the ordered outcome, it is difficult to infer the impact that each explanatory variable has on the ordered outcome. In order to better understand the influence of the explanatory variables on the ordered outcomes, the partial effects for each of the two services were calculated according to the procedure outlined in (Greene & Hensher, 2009):

$$\frac{\partial Prob(y = j|\overline{\boldsymbol{x}})}{\partial A} = \left[f\left(\mu_{j-1} - \boldsymbol{\beta}'\overline{\boldsymbol{x}}\right) - f\left(\mu_j - \boldsymbol{\beta}'\overline{\boldsymbol{x}}\right)\right]\beta_A$$

The partial effects were computed for each of the explanatory variables in each model; the results are summarized in Figure 19**,** Figure 20**,** and Figure 21. In terms of magnitude, it appears that the zonal area has the most significant impact on the likelihood that a particular outcome will be observed. This is understandable, as the size of a zone can influence both the number of residents and the number of activity locations that are present in a zone. With regards to transit network attributes, it appears that the length of streetcar routes within a 400 m radius of the centroid of the DA and the length of subway lines within 800 m of the same had the most significant influence on the outcome. When considering the land use density indicators, it appears that the density of locations classified as ‘art, entertainment, and recreation’ and the density of private dwellings were the most impactful variables. Overall, these results appear to indicate that the generation of both transit and ride-hailing trips is higher in dissemination areas with higher densities of both commercial and recreational buildings, as well as private dwellings. Besides, increases in both the number of stops and the coverage of transit routes are positively associated with the generation of both transit and ride-hailing trips. These findings are likely related, as areas with a more significant number of activities tend to attract more people, which in turn can increase the number of trips generated by the area. Besides, areas with a higher number and diversity of activities tend to be better served by transit, in part because the travel demand justifies the higher levels of service.

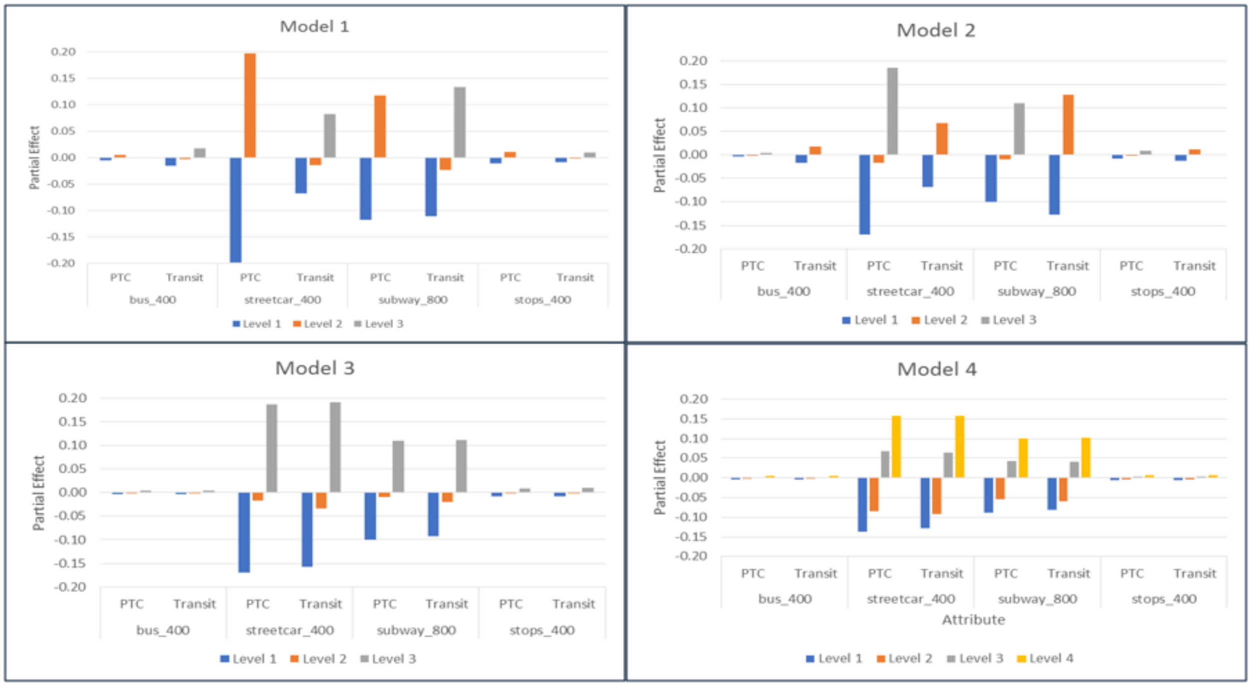


Figure 19: Partial Effects of Transit Attributes

Figure 20: Partial Effects of Land Use Density Indicators

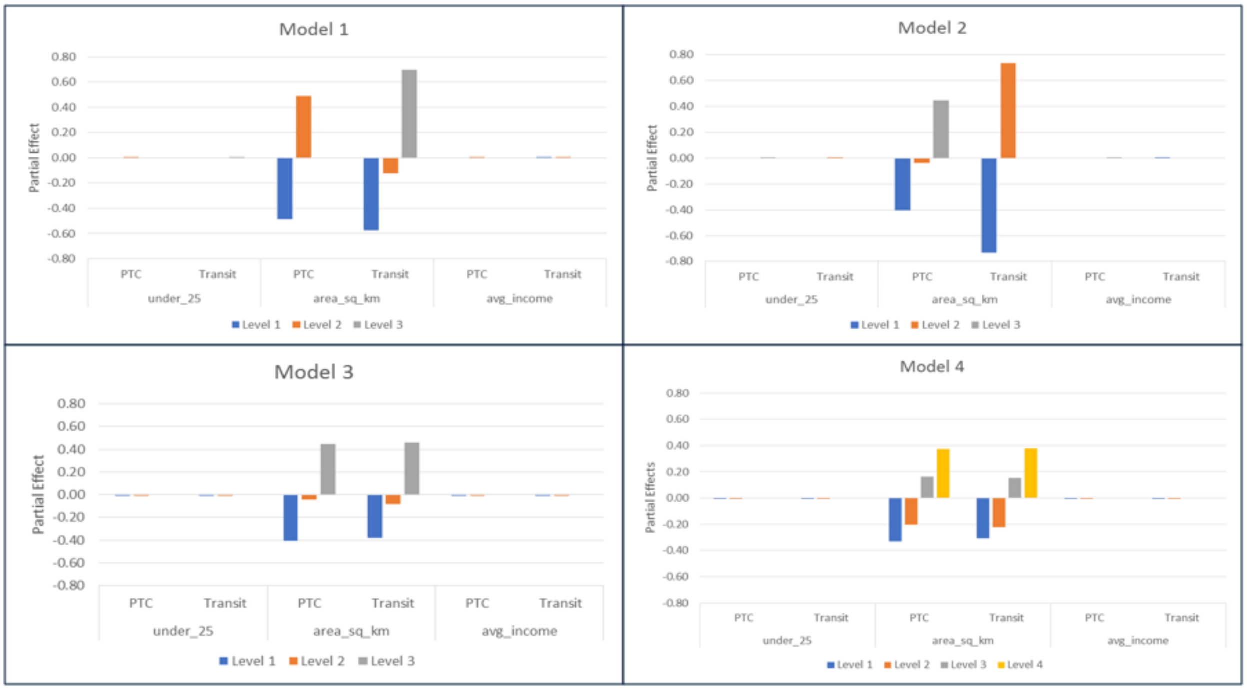


Figure 21: Partial Effects of Zonal Attributes

6.6 Conclusions and Future Work

This chapter presents an analysis of the influence of zonal attributes and built environment characteristics on the generation of ride-hailing and public transit trips. This work aims to contribute to the literature by jointly modelling the generation of public transit and ride-hailing trips using the bivariate ordered probit model. The use of this model allows for the estimation of the correlation between the factors that influence the generation of ride-hailing and transit trips. The analysis presented in this study utilized ride-hailing trip information provided by Uber and Lyft, which contains spatiotemporal information about each ride-hailing trip made in Toronto for over two years. To determine the impact of the strategy that was used to define the ordered response variables, four different models were defined and estimated.

The influence of the explanatory variables was qualitatively similar across the different models, in terms of the sign and significance of the parameters, implying that the classification strategy did not significantly impact the results. The results themselves indicate that the factors that influence the generation of transit trips had similar impacts on the generation of ride-hailing trips. This result was somewhat surprising for the transit network attributes, which showed that increases in the number of stops and the coverage of the network were also associated with an increased level of ride-hailing trip generation. Also, the density of private dwellings and the density of commercial and recreational locations in a zone have a positive influence on the generation of ride-hailing and transit trips. The correlation parameters estimated in the four models all imply that a positive correlation exists between the unobserved factors that influence the generation of ride-hailing and transit trips.

The results presented in the chapter aim to contribute to the improved understanding of the factors that influence the number of transit and ride-hailing trips that originate in each zone. Future work includes taking an approach to model development that allows explanatory variables only to be included in a single equation, rather than in both equations. Furthermore, this analysis should be extended to consider the trip generated by each mode of transit, rather than treating transit service as a single homogenous mode. Similarly, the explicit consideration of different ride-hailing services, i.e., shared and exclusive services should also be incorporated into the model.

Chapter 7
Conclusions and Future Work

Although traditional household travel survey methods are becoming increasingly obsolete, they still play a critical role in the planning, operation, and maintenance of transportation networks. Household travel surveys are the primary source of urban passenger travel data, which are used to understand the use of existing facilities and forecast the future demand and utilization of transportation infrastructure. The obsolescence of traditional household travel surveys is primarily being driven by declining rates of household landline ownership, particularly among younger households. The traditional reliance on landline telephones as both a survey mode and sample frame is the main reason why this decline has been detrimental to household travel surveys. Aside from technological trends, the obsolescence of traditional household travel surveys can also be attributed to changes in the data required to facilitate contemporary travel demand analyses.

Travel demand analysis has typically focused on weekday commuting trips, however other aspects of travel are beginning to receive increased attention. In addition, there has been a growing desire to analyze and model travel at a more disaggregate level. The most prominent of the emerging approaches to travel demand analysis is the activity-based approach to modelling travel behaviour. Other aspects of travel behaviour that are receiving increased attention from analysts include the study of seasonal variations in travel behaviour, long distance travel, the influence of attitudinal factors of travel behaviour, and attempts to understand the impacts of new technologies (such as ride-hailing). The desire to study these aspects of travel behaviour creates the need for data that is typically not collected through large-scale household travel surveys and the need for data at a greater level of detail than can be accommodated by a traditional household travel survey.

In order to address the issues faced by the Transportation Tomorrow Survey (TTS), the TTS 2.0 research and development program was created. The primary goal of TTS 2.0 was to modernize the conduct of the TTS, a regional household travel survey that has been conducted in the Greater Golden Horseshoe Area once every five years since 1986. One of the focuses of the TTS 2.0 project was to further build on the core-satellite data collection paradigm outlined in

(Goulias, Pendyala, & Bhat, 2011), and to determine how this framework can be applied in the Greater Golden Horseshoe Area.

The original core-satellite data collection paradigm was comprised of three components – the core survey, satellite surveys, and complementary datasets. The core survey is analogous to a standalone household travel survey, however, in this framework, core surveys serve to collect fundamental information of travel behaviour, as well as personal and household characteristics. The data collected through core surveys is meant to support foundational planning and policy applications, as defined by the survey administrators. Satellite surveys, on the other hand, are primarily meant to enrich or supplement the data obtained through the core survey by collecting additional data from a specific subset of the target population. This subset is commonly defined based on socio-economic attributes or behavioural attributes. The third component of this framework, the complementary dataset, is meant to augment the data obtained through the core and satellite surveys.

The expanded core-satellite framework proposed in this thesis aims to address two key shortcomings of the original paradigm proposed by (Goulias, Pendyala, & Bhat, 2011). The first issue with the original paradigm is that it assumes that the core survey can provide an accurate representation of the target population, which is rarely the case for standalone household travel surveys. The second issue is that the original paradigm does not differentiate between different types of satellite surveys. The expanded core-satellite framework aims to address these issues by distinguishing between different types of core and satellite surveys. In the expanded framework, three types of core surveys are defined. The main core survey functions as a typical household travel survey. Core-filling surveys are meant to address the first issue with the original paradigm; they serve to address issues of representation in the main core survey. On the other hand, core-extension surveys are meant to collect additional "core" information from a random sub-sample of core respondents. In addition, two types of satellite surveys are defined – linked and independent satellite surveys. The main distinction between these two types of surveys is whether they can be directly linked to responses obtained through the core survey.

Two empirical studies are presented in this thesis, one that makes use of data obtained through a satellite survey, and another that combines core survey data with passive data. The satellite survey data was obtained through StudentMoveTO, a travel and activity survey of University students in Toronto conducted in 2015. Compared to the TTS, StudentMoveTO directly recruits students, providing the opportunity to reach students who live on campus or with roommates. The StudentMoveTO dataset was used to investigate the factors that influence the location choice behaviour of students when they use public transit for discretionary trips. The travel behaviour of post-secondary students is still not well understood in the literature and is often represented in a simplified manner when it is included in operation travel demand models. The location choice model was estimated using the multinomial logit specification, with the choice set of the students being imputed using the time-space prism concept. The results of the location choice model were used to determine the utility-based accessibility of university students in Toronto; these measures were qualitatively compared to count-based measures. The results of the location choice model indicate that the time required to access transit and the number of transfers required for a trip influence the location choice process. In addition, the socio-economic attributes of the student also affect their location choice process. The comparison of the count- and accessibility-based measures indicate that the former can over-estimate accessibility, particularly in the north-west and north-east parts of Toronto.

The second empirical study presented in this thesis combined data obtained through the 2016 TTS with passive data on trips made using Uber and Lyft from September 2016 to September 2018 provided by the City of Toronto. The data on ride-hailing trips originating in Toronto provides the ability to replace the records obtained through the TTS, which appear to under-represent trips made using this mode. The goal of the study was to determine whether zonal and built environment attributes affect the relationship between ride-hailing services and public transit. In order to make the ride-hailing data compatible with the TTS data, a weighted average of the number of ride-hailing trips originating from each zone over the duration of the TTS was used. The weights were taken from the TTS and represented the proportion of survey responses that were obtained on each day of the survey. To analyze the impacts of transit, land use, and zonal attributes on the use of ride-hailing and public transit, as well as the relationship between the two modes, the multivariate ordered probit model structure was used. The results of the empirical model showed that the factors that lead to higher transit usage also tend to lead to higher ride-hailing usage. It is unclear whether these results speak to the general nature of trip generation, or it there is a complementary relationship between the two services.

Future work should focus on further operationalizing the expanded core-satellite framework for use in the Greater golden Horseshoe Area. Based on the nature of the core-satellite framework, it is imperative that the harmonization of different dataset be studied in further detail. Being able to identify, quantify, and address the inherent biases that are captured in an individual dataset would help to significantly reduce the uncertainty associated with utilizing pooled datasets. In addition, methods to make better use of the vast amounts of passive data that are now available should be investigated. Passive data has the potential to provide insights into travel behaviour that cannot be obtained through traditional dataset, in part due to the volume and variety of information that can be obtained through passive means. In terms of empirical work, the proposed framework should be put into practice in order to identify and address its shortcomings and weaknesses and to build on its strengths.

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.

Albacete, X., Olaru, D., Paul, V., & Biermann, S. (2017). Measuring the Accessibility of Public Transport – A Critical Comparison Between Methods in Helsinki. *Applied Spatial Analysis*, *10*, 161 - 188.

Alemi, F., Circella, G., & Sperling, D. (2017). Adoption of Uber and Lyft, Factors Limiting and/or Encouraging Their Use and Impacts on Other Travel Modes among Millennials and Gen Xers in California. Washington, D.C.: 97th Annual Meeting of the Transportation Research Board.

Alemi, F., Circella, G., Handy, S., & Mokhtarian, P. (2018). What Influences Travelers to use Uber? Exploring the Factors Affecting the Adoption of On-Demand Ride Services in California. *Travel Behaviour and Society*, *13*, 88 - 104.

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.

Alsnih, R. (2006). Characteristics of Web Based Surveys and Applications in Travel Research. *Travel Survey Methods. Quality and Future Directions*, 569-592.

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.

AMR Interactive Consultants. (2009). *Research into Barriers to Cycling in NSW.* St. Leonards: AMR Interactive Consultants.

Andrews, D., Nonnecke, B., & Preece, J. (2003). Conducting Research on the Internet: Online Survey Design, Development, and Implementation Guidelines. *International Journal of Human-Computer Interaction*, 185-210.

Aptech Systems, Inc. (2013). *GAUSS 14 User Guide.* Higley: Aptech Systems, Inc.

Bachand-Marleau, J., Larsen, J., & El-Geneidy, A. (2011). The Much Anticipated Marriage of Cycling and Transit - But How Will It Work? *Transportation Research Record - Journal of the Transportation Research Board*, 109 - 117.

Bavdaz, M., Giesen, D., Cerne, S., Lofgren, T., & Raymond-Blaess, V. (2015). Response Burden in Official Business Surveys: Measurement and Reduction Practices of National Statistical Institutes. *Journal of Official Statistics*, *31*(4), 559 - 588.

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

Beebe, T., McAlpine, D., Ziegenfuss, J., Jenkins, S., Haas, L., & Davern, M. (2012). Deployment of a Mixed-Mode Data Collection Strategy Does Not Reduce Nonresponse Bias in a General Population Health Survey. *Health Services Research*, 1739 - 1754.

Ben-Akiva, M., & Lerman, S. (1985). *Discrete Choice Analysis - Theory and Applications to Travel Demand.* Cambridge: The MIT Press.

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

Boisjoly, G., Grise, E., Maguire, M., Veillette, M., Deboosere, R., Berebi, E., & El-Geneidy, A. (2018). Invest in the Ride – A 14 Year Longitudinal Analysis of the Determinants of Public Trnasport Ridership in 25 North American Cities. *Transportation Research Part A*, *116*, 434 - 445.

Bradley, M., Bergman, A., Lee, M., Greene, E., & Childress, S. (2015). Predicting and Applying Differential Response Rates in Address-Based Sampling for a Household Travel Survey . *Transportation Research Record: Journal of the Transportation Research Board, 2526*(13), 119 - 125.

Broach, J., & Dill, J. (2016). Using Predicted Bicyclist and Pedestrian Route Choice to Enhance Mode Choice Models. *Transportation Research Record: Journal of the Transportation Research Board, 2564*(6), 52 - 59.

Canada Post. (2015). *Canada Complete Mailing Lists*. Retrieved from Canada Post: https://www.canadapost.ca/web/en/products/details.page?article=rent_canadas_most_c

Cascetta, E., Carteni, A., & Montanino, M. (2012). A New Measure of Accessibility Based on Perceived Opportunities. *Procedia - Social and Behavioural Sciences, 87*, 117 - 132.

Castiglione, J., Bradley, M., & Gliebe, J. (2015). *Activity-Based Travel Demand Models: A Primer.* Washington, D.C.: Transportation Research Board.

Chen, X., & Kwan, M. (2012). Choice Set Formation with Multiple Flexible Activities Under Space-Time Constraints. *International Journal of Geographical Information Science, 26*(5), 941 - 961.

Chung, B., Srikukenthiran, S., Habib, K., & Miller, E. (2016). *Web Survey Design for Housheold Travel Surveys.* Toronto: University of Toronto Transportation Research Institute.

City of Toronto - Big Data Innovation Team. (2019). *The Transportation Impacts of Vehicle-for-Hire in the City of Toronto.* Toronto: City of Toronto - Big Data Innovation Team.

Clewlow, R., & Mishra, G. (2017). *Disruptive Transportation – The Adoption, Utilization, and Impacts of Ride-Hailing in the United States.* Davis: UC Davis Institute of Transportation Studies.

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

Couper, M., Tourangeau, R., Conrad, F., & Crawford, S. (2004). What They See is What We Get: Response Options for Web Surveys. *Social Science Computer Review, 22*(1), 111-127.

Curtis, C., & Scheurer, J. (2010). Planning for Sustainable Accessibility – Developing Tools to Aid Discussion and Decision-Making. *Progress in Planning, 74*, 53 - 106.

Data Management Group. (2018). *Design and Conduct of the Survey.* Toronto: R.A. Malatest & Associates Ltd.

de Bruijne, M., & Wijnant, A. (2013). Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment with a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. *Social Science Computer Review*, 482-504.

Diab, E., Kasraian, D., Miller, E., & Shalaby, A. (2018). The Rise and Fall of Transit Ridership Across Canada – Understanding the Determinants. Washington, D.C.: Prepared for Presentation at the Transportation Research Board 98th Annual Meeting.

Dias, F., Lavieri, P., Kim, T., Bhat, C., & Pendyala, R. (2019). Fusing Multiple Sources of Data to Understnad Ride-Hailing Use. *Transportation Research Record*, 1 - 11. doi:0361198119841031

Dillman, D., & Smyth, J. (2007). Design Effects in the Transition to Web-based Surveys. *American Journal of Preventive Medicine*, S90-S96.

DMTI Spatial Inc. (2014). *CanMap RouteLogistics User Manual.* Richmond Hill: DMTI Spatial Inc.

DMTI Spatial Inc. (2016). *CanMap Content Suite Data Dictionary.* Richmond Hill: DMTI Spatial Inc.

Dong, X., Ben-Akiva, M., Bowman, J., & Walker, J. (2006). Moving from Trip-Based to Activity-Based Measures of Accessibility. *Transportation Research Part A, 40*, 163 - 180 .

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 and Pedestrian Studies Based on Data from National Household Travel Survey. *Journal of the Transportation Research Board*, 150–156.

El-Geneidy, A., Levinson, D., Diab, E., Boisjoly, G., Verbich, D., & Loong, C. (2016). The Cost of Equity – Assessing Transit Accessibility and Social Disparity Using Travel Cost. *Transportation Research Part A, 91*, 302 - 316.

ENRG Research Group. (2009). *Cycling End-of-Trip Facilities Survey.* Vancouver: ENRG Research Group.

Eom, J., Stone, J., & Ghosh, S. (2009). Daily Activity Patterns of University Students. *Journal of Urban Planning and Development, 135*(4), 141 - 149.

Ferdous, N., Eluru, N., Bhat, C., & Meloni, I. (2010). A Multivariate Ordered-Response Model System for Adults' Weekday Activity Episode Generation by Activity Purpose and Social Context. *Transportation Research Part B, 44*, 922 - 943.

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.

Gerte, R., Konduri, K., & Eluru, N. (2018). Is There a Limit to Adoption of Dynamic Ridesharing Systems? Evidence from Analysis of Uber Demand Data from New York City. *Transportation Research Record, 2672*, 127 - 136.

Geurs, K., & and van Wee, B. (2004). Accessibility Evaluation of Land-Use and Transport Studies - Review and Research Directions. *Journal of Transport Geography, 12*, 127 - 140.

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

Greene, W., & Hensher, D. (2009, January). *Modeling Ordered Choices - A Primer.* Cambridge: Cabridge University Press. Retrieved June 11, 2019, from http://pages.stern.nyu.edu/~wgreene/DiscreteChoice/Readings/OrderedChoiceSurvey.pdf

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

Habib, K., El-Assi, W., & Lin, T. (2017). *How Large is too Large? The Issue of Sample Size Requirements of Regional Household Travel Surveys, the Case of the Transportation Tomorrow Survey in the Greater Toronto and Hamilton Area.* Toronto: University of Toronto Transportation Research Institute.

Habib, K., Weiss, A., & Hasnine, S. (2018). On the Heterogeneity and Substitution Patterns in Mobility Tool Ownership Choices of Post-Secondary Students. *Transportation Research Part A, 116*, 650 - 665.

Hall, J., Palsson, C., & Price, J. (2018). Is Uber a Substitute or a Complement for Public Transit? *Journal of Urban Economics, 108*, 36 - 50.

Hansen, W. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners, 25*(2), 73 - 76.

Har Group Management Consultants. (2011). *Cycling Strategy Research - Online Survey.* Calgary: Har Group Management Consultants.

Hasnine, M., Graovac, A., Camargo, F., & Habib, K. (2019). A Random Utility Maximization (RUM) based Measure of Accessibility to Transit: Accurate Capturing of the First-Mile Issue in Urban Transit. *Journal of Transport Geography, 74,* 313 - 320. doi:January 7–11, 2018

Hasnine, M., Kamel, I., & Habib, K. (2017). Using Google Map to Impute Transportation Level-of-Service Attributes - Application in Mode and Departure Time Choice Modelling. *11th International Conference on Transport Survey Methods*, (pp. 24 - 29). Esterel, Quebec, Canada.

Hasnine, M., Lin, T., Weiss, A., & Habib, K. (2017). Is It the Person or the Urban Context? Role of Urban Travel Context in Defining Mode Choice for School Trips of Postsecondary Students in Toronto, Canada. *Presented at the 96th Annual Meeting of Trans. Res. Board*. doi:January 8–12, 2017

Hasnine, M., Lin, T., Weiss, A., & Habib, K. (2018). Determinants of Travel Mode Choices of Post-Secondary Students in a Large Metropolitan Area – The Case of the City of Toronto. *Journal of Transport Geography, 70,* 161 - 171.

Hirk, R., Hornik, K., Vana, L., & Genz, A. (2019, March 6). *Package 'mvord'.* Retrieved June 16, 2019, from https://cran.r-project.org/web/packages/mvord/mvord.pdf

Hoffmann, K., Ipcirotis, P., & Sundararajan, A. (2016). Ridesharing and the Use of Public Transportation. *ThirtySeventh International Conference on Information Systems.* Dubin.

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.

Ipsos. (2013). *Auckland Transport Cycling Research.* Auckland: Ispos.

Ipsos Reid. (2010). *City of Toronto Cycling Study.* Toronto: Ipsos Reid.

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, 2255*(15), 137 - 145.

Kim, J., & Lee, S. (2017). Comparative Analysis of Traveller Destination Choice Models by Method of Sampling Alternatives:. *Transportation Planning and Technology, 40*(4), 465 - 478.

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.

Lavieri, P., Dias, F., Juri, N., Kuhr, J., & Bhat, C. (2018). A Model of Ridesourcing Demand Generation and Distribution. *Transportation Research record, 2672*(46), 31 - 40.

Lee, B., & Waddell, P. (2010). Residential Mobility and Location Choice – A Nested Logit Model with Sampling of Alternatives. *Transportation, 37*, 587 - 601.

Li, L., Hasnine, S., Habib, K., Persaud, B., & Shalaby, A. (2017). Investigating the Interplay between the Attributes of At-fault and Not-at-fault drivers and the associated ipacts on crash injury occurrence and severity level. *Journal of Transportation Safety & Security, 9*(4), 439 - 456.

Maldonado-Hinarejos, R., Sivakumar, A., & Polak, J. (2014). Exploring the Role of Individual Attitudes and Perceptions in Predicting the Demand for Cycling: A Hybrid Choice Modelling Appraoch. *Transportation*, 1287 - 1304.

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.

Massey, D., & Tourangeau, R. (2013). Where do we Go from Here? Non-Response and Social Measurement. *The Annals of American Political and Social Science, 645*(1), 222 - 236.

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

McFadden, D. (1974). Conditional Logit Analysis of Quantitative choice Behaviour. In *Frontiers in Econometrics* (pp. 105 - 142). New York: Academic Press.

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.

McKelvey, R., & Zavoina, W. (1975). A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology, 4*, 103 - 120.

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

Messer, B., & Dillman, D. (2011). Surveying the General Public over the Internet Using Address-based Sampling and Mail Contact Procedures. *Public Opinion Quarterly*, 429-457.

Millar, M., & Dillman, D. (2011). Improving Response to Web and Mixed-mode Surveys. *Public Opinion Quarterly*, 249-269.

Miller, E., & Habib, K. (2014). *Passenger Travel Survey Methods for the Greater Golden Horseshoe: A Discussion Paper.* Toronto: U of T Data Management Group.

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.

MMM Group Limited. (2014). *TRANS Model.* Ottawa: MMM Group Limited.

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.

Neutens, T., Witlox, F., Van De Weghe, N., & De Maeyer, P. (2007). Space–time opportunities for multiple agents: a constraint-based approach. *International Journal of Geographical Information Science, 21*(10), 1061 - 1076.

Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., & Maurer, P. (2014). Supporting large-scale travel surveys with smartphones – A practical approach. *Transportation Research Part C: Emerging Technologies, 43*(2), 212 - 221.

Olafsson, A., Nielsen, T., & Carstensen, T. (2016). Cycling in Multimodal Transport Behaviours: Emploring Modality Styles in the Danish Population. *Journal of Transport Geography, 52*, 123 - 130.

Paez, A., Scott, D., & Morency, C. (2012). Measuring Accessibility – Positive and Normative Implementations of Various Accessibility Indicators. *Journal of Transport Geography, 25*, 141 - 153.

Pan, B. (2010). Online Travel Surveys and Response Patterns. *Journal of Travel Research*, 121-135.

Parsons, C. (2007). Web-based Surveys – Best Practices Based on the Research Literature. *Visitor Studies*, 13-33.

Petrunoff, N., Xu, H., Rissel, C., Wen, L., & Van der Ploeg, H. (2013). Measuring Workplace Travel Behaviour - Validity and Reliability of Survey Questions. *Journal of Environmental and Public Health, 2013,* 1 - 7.

Piatkowski, D., & Marshall, W. (2015). Not All Prospective Bicyclists are Created Equal: The Role of Attitudes, Socio-demographics, and the Built Environment in Bicycle Commuting. *Travel Behaviour and Society,* 166 - 173.

R.A. Malatest & Associated Ltd. (2018). *TTS 2016 Data Guide.* Toronto: University of Toronto Data Management Group.

R.A. Malatest & Associates Ltd. (2013). *2011 NCR Household Origin-Destination Survey.* Ottawa: R.A. Malatest & Associates Ltd.

Rashidi, T., Auld, J., & Mohammadian, A. (2012). A Behavioural Housing Search Model – Two-Stage Hazard-Based and Multinomial Logit Approach to Choice-Set Formation and Location Selection. *Transportation Research Part A, 46*, 1097 - 1107.

Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a Better Taxi? A Survey-based Comparison of Taxis, Transit, and Ridesourcing Services in San Francisco. *Transport Policy, 45,* 168 - 178 .

Resource Systems Group, Inc. (2013). *Utah Travel Study.* Hartford: Resource Systems Group, Inc.

Rodier, C. (2018). *The Effects of Ride Hailing Services on Travel and Associated Greenhouse Gas Emissions.* Davis: National Center for Sustainable Transportation.

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.

Sadowsky, N., & Nelson, E. (2017). *The Impact of Ride-Hailing Services on Public transportation Use – A Discontinuity Regression Analysis.*

Scott, D., & Axhausen, K. (2006). Household Mobility Tool Ownership – Modeling Interactions between Cars and Season Tickets. *Transportation*, *33*, 311 - 328.

Searcy, S., Findley, D., Huegy, J., Ingram, M., Mei, B., Bhadury, J., & Wang, C. (2018). Effect of Residential Proximity on University Student Trip Frequency by Mode. *Travel Behaviour and Society* , *12*, 115 - 121.

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

Sills, S., & Song, C. (2002). Innovations in Survey Research – An Application of Web-based Surveys. *Social Science Computer Review*, 22-30.

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.

Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Sciences*, 112 - 141.

Smith, M. (1979). Design of Small-Sample Home-Interview Travel Surveys. *Transportation Research Record*, 29 - 35.

Son, S., Khattak, A., & Kim, N. (2013). Non-coverage Errors in Travel Surveys Due to Mobile Phone-Only Households. *Transportation Research Record: Journal of the Transportation Research Board*, *2354*(4), 29 - 39.

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.

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

Srikukenthiran, S., Loa, P., Hossain, S., Chung, B., Habib, K., & Miller, E. (2018). *Transportation Tomorrow Survey 2.0 - Final Report.* Toronto: University of Toronto Transportation Research Institute.

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

Statistics Canada. (2017, February 8). *Census Profile, 2016 Census*. (Statistics Canada) Retrieved June 24, 2019, from https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMACA&Code1=535&Geo2=PR&Code2=35&Data=Count&SearchText=Caledon%20East&SearchType=Begins&SearchPR=01&B1=All

Statistics Canada. (2018). *Survey of Household Spending, 2017.* Ottawa: Statistics Canada.

Stern, M., Bilgen, I., & Dillman, D. (2014). The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design. *Field Methods*, 284-301.

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

StudentMoveTO. (2015, August). *About*. (StudentMoveTO) Retrieved October 30, 2017, from http://www.studentmoveto.ca/about/

StudentMoveTO. (2018, June). *Summary Findings 2015*. Retrieved June 10, 2019, from http://www.studentmoveto.ca/resources-2/2015summary/

The American Association for Public Opinion Research. (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* Lenexa: The American Association for Public Opinion Research.

The City of Ottawa. (2013). *Transportation Master Plan.* Ottawa: The City of Ottawa.

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

Toepoel, V., & Dillman, D. (2011). Words, Numbers, and Visual Heuristics in Web Surveys: Is There a Hierarchy of Importance? *Words, Numbers, and Visual Heuristics in Web Surveys: Is There a Hierarchy of Importance?*, 193-207.

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.

Train, K. (2009). Discrete Choice Methods with Simulation. In K. Train, *Discrete Choice Methods with Simulation* (pp. 34 - 75). New York City: Cambride University Press.

TRANS Committee. (2014, June 20). *About TRANS*. Retrieved September 19, 2017, from http://www.ncr-trans-rcn.ca/about-trans/

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

Ville de Gatineau. (2017). *Strategic Plan*. Retrieved September 20, 2017, from http://www.gatineau.ca/portail/default.aspx?p=la_ville/administration_municipale/plan_strategique

Volosin, S., Pendyala, R., Kerrigan, J., Greene, E., Livshits, V., & Samuelson, J. (2014). Measuring the Travel Characteristics of a University Population – Experiences from the Design and Administration of a Web-Based Travel Survey. *No. 14-0240*.

Waksberg, J. (1978). Sampling Methods for Random Digit Dialing. *Journal of the American Statistical Association*, 40 - 46.

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

Weiner, M., Puniello, O., & Noland, R. (2016). 2016. *Transportation Research Record: Journal of the Transportation Research Board, 2594*(8), 44 - 50.

Whalen, K., Paez, A., & Carrasco, J. (2013). Mode Choice of University Students Commuting to School and the Role of Active Travel. *Journal of Transport Geography, 31*, 132 - 142.

Winters, M., Davidson, G., Kao, D., & Teschke, K. (2011). Motivators and Deterrent of Bicycling – Comparing Influences on Decisions to Ride. *Transportation*, *38*(1), 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.

Yang, J., Du, Y., Sun, D., & Zhao, Y. (2009). The Effect of Sampling on Alternatives on MNL Models - An Empirical Analysis in the Context of Shopping-Destination Choice. *IEEE Journal*, 75 - 80.

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

Appendix: Sample Questionnaire

Screening Questions:

Question #	1	Were you provided with a unique household key?
Response Type		Checkbox (select one)
Response Options		Yes; No

Question #	2	Please enter your unique household key.
Ask If		Respondent was given a household key
Response Type		Textbox (numerical values only)

Demographic Questions:

Question #	1	Please state your age
Response Type		Textbox (numerical values only)

Question #	2	Please state your gender
Response Type		Checkbox (select one)
Response Options		Male; Female; Other (Please specify); Prefer not to specify

Question #	3	Please enter your home address
Response Type		Textbox with map interface
Notes		Terminate the survey if the household is outside of the NCR

Question #	4	Please select the option that best describes your home
Response Type		Checkbox (select one)
Response Options		Single-detached; Semi-detached; Row/Townhouse; Apartment or Condo (Tenant); Apartment or Condo (Owner); Other
Notes		Taken from 2011 Household OD Survey

Question #	5	How many people (including yourself) reside in your household
Response Type		Textbox (numerical values only)

Question #	6	Please enter the ages and genders of the other members of your household.
Ask If		More than one person lives in the household
Response Type		Grid of dropdown menus (gender) and textboxes (age)
Notes		Response options same as those for questions 1 and 2.

Question #	7	Please select the option that best describes your current occupation status.
Response Type		Checkbox (select one)
Response Options		Full-time worker; Part-time worker; Student; Retired; Homemaker; Not employed

Question #	8	Please select the option that best represents your household income (optional)
Response Type		Checkbox (select one)
Response Options		$0 - $29,999; $30,000 - $59,999; $60,000 - $89,999; $90,000 - $119,999; $120,000 - $149,999; $150,000 - $179,999; $180,000 - $209,999; $210,000 and above
Notes		Taken from 2011 Household OD Survey

Question #	9	Do you own or have access to a bicycle
Response Type		Checkbox (select one)
Response Options		Yes; No

Question #	10	Are you enrolled in a bike sharing service?
Response Type		Checkbox (select one)
Response Options		Yes; No

Question #	11	Do you have a driver's license?
Response Type		Checkbox (select one)
Response Options		Yes; No

Question #	12	Do you own a transit pass?
Response Type		Checkbox (select one)
Response Options		Yes; No

Cycling Questions:

Question #	1a	How frequently do you use a bicycle in the spring?
Response Type		Checkbox (select one)
Response Options		Every day; At least once per week; At least once per month; At least once per season; Do not use a bicycle during this season

Question #	1b	How frequently do you use a bicycle in the summer?
Response Type		Checkbox (select one)
Response Options		Every day; At least once per week; At least once per month; At least once per season; Do not use a bicycle during this season

Question #	1c	How frequently do you use a bicycle in the autumn?
Response Type		Checkbox (select one)
Response Options		Every day; At least once per week; At least once per month; At least once per season; Do not use a bicycle during this season

Question #	1d	How frequently do you use a bicycle in the winter?
Response Type		Checkbox (select one)
Response Options		Every day; At least once per week; At least once per month; At least once per season; Do not use a bicycle during this season

Question #	2	How do you typically travel in the winter?
Ask If		Respondent reported using a bicycle in the spring, summer, or autumn, but not in the winter
Response Type		Checkbox (select one or more)
Response Options		Car driver; Car passenger; Transit; Walking; Other (please specify)
Notes		Taken from (Ipsos, 2013)

Question #	3	What kind of travel do you use your bicycle for?
Ask If		Respondent reported using a bicycle in any season
Response Type		Checkbox (select one or more)
Response Options		For exercise; For recreation; To get to/from shops; To get to/from work; To get to/from school; To see friends and family; To get to/from public transit; For sport; As part of my job; Other (please specify)
Notes		Taken from (Ipsos, 2013)

Question #	3a	Please enter your work or school location. Please update the route shown on the map so that it reflects the route that you typically take for this trip.
Ask If		Respondent reported using a bicycle to travel to work or school
Response Type		Map-based Interface (one text box for origin, one for destination)

Question #	3b	Please enter the start and end points of the last recreational trip that you made by bicycle. Please update the route shown on the map if it does not accurately reflect the route that you took.
Ask If		Respondent reported using a bicycle to travel to work or school
Response Type		Map-based Interface (one text box for origin, one for destination)

<u>Cycling and Transit:</u>

Question #	1a	How often have you brought your bicycle onto a transit vehicle in the spring?
Ask If		Respondent reported using a bicycle in the spring
Response Type		Checkbox (select one)
Response Options		Every day; Every week; Every month; Once per season; Do not use a bicycle during this season

Question #	1b	How often have you brought your bicycle onto a transit vehicle in the summer?
Ask If		Respondent reported using a bicycle in the past summer
Response Type		Checkbox (select one)
Response Options		Every day; Every week; Every month; Once per season; Do not use a bicycle during this season

Question #	1c	How often have you brought your bicycle onto a transit vehicle in the autumn?
Ask If		Respondent reported using a bicycle in the autumn
Response Type		Checkbox (select one)
Response Options		Every day; Every week; Every month; Once per season; Do not use a bicycle during this season

Question #	1d	How often have you brought your bicycle onto a transit vehicle in the winter?
Ask If		Respondent reported using a bicycle in the winter
Response Type		Checkbox (select one)
Response Options		Every day; Every week; Every month; Once per season; Do not use a bicycle during this season

Question #	2	In the past year, have you parked your bicycle at a nearby transit stop in order to take public transit?
Ask If		Respondent reported using a bicycle at least once in any season
Response Type		Checkbox (select one)
Response Options		Yes; No
Notes		Taken from (Ipsos, 2013)

Question #	3a	Which of these factors, if any, would entice you use a combination of bicycle *and* transit to reach your destination?
Ask If		Respondent answered “No” to both questions 1 and 2
Response Type		Checkbox (select one or more)
Response Options		Secure bike parking provided at transit stop; Secure bike parking provided at destination; Possibility of bringing bike onto transit vehicle; None of the above; Other (please specify)

Question #	3b	Which of these factors, if any, would discourage you from using a combination of bicycle *and* transit to reach your destination?
Response Type		Checkbox (select one or more)
Response Options		Lack of secure bike parking provided at transit stop; Lack of secure bike parking provided at destination; Lack of space on-board transit vehicles; Destination is within walking distance of nearest transit stop; Need to carry other items; I prefer to bike the full distance to my destination; I am unsure of how to load my bike onto the bus; My bike will not fit onto the bus’s bike rack; My bike is too heavy to lift onto the bus’s bike rack; I do not want to become sweaty/ I would have to shower or change; Bike racks not available along usual route; None of the above; Other (please specify)

Question #	4a	If you were to use a bicycle to travel from home to work, which of these factors, if any, would entice you to use transit to complete part of this trip?
Ask If		Respondent reported not using a bicycle in the past year
Response Type		Checkbox (select one or more)
Response Options		Secure bike parking provided at transit stop; Secure bike parking provided at destination; Possibility of bringing bike onto transit vehicle; None of the above; Other (please specify)

Question #	4b	If you were to use a bicycle to travel from home to work, which of these factors, if any, would discourage you from use transit to complete part of this trip?
Response Type		Checkbox (select one or more)
Response Options		Lack of secure bike parking provided at transit stop; Lack of secure bike parking provided at destination; Lack of space on-board transit vehicles; Destination is within walking distance of nearest transit stop; Need to carry other items; I prefer to bike the full distance to my destination; I am unsure of how to load my bike onto the bus; My bike will not fit onto the bus's bike rack; My bike is too heavy to lift onto the bus's bike rack; I do not want to become sweaty/ I would have to shower or change; Bike racks not available along usual route; None of the above; Other (please specify)

<u>Attitudinal Questions:</u>
Questions adapted from the 13 categories of attitudinal questions in (Winters, Davidson, Kao, & Teschke, 2011)

Question Format	Please describe the extent to which you agree or disagree with each statement.
Response Options	Strongly disagree; Disagree; No opinion; Agree; Strongly agree; I don't know
Statements	1. I am/ would be comfortable sharing the road with motor vehicles. 2. I am/ would be comfortable cycling alongside parked vehicles 3. I would use a bicycle more often if there were fewer vehicles along my path. 4. I am/ would be less comfortable travelling alongside trucks and buses than I am travelling alongside cars. 5. I am/ would be comfortable riding a bicycle on a roadway if there are shared lane markings ("sharrows"). 6. I am/ would be comfortable riding a bicycle in a dedicated on-street lane. 7. I am/ would be comfortable riding a bicycle in an elevated lane. 8. My decision to bike/ not bike is affected by the availability of bike lanes. 9. My decision to bike/ not bike is affected by the availability of bike paths. 10. I am/ would be comfortable making left- or right-turns on a bicycle.

	11. I am/ would be concerned about turning vehicles when I ride a bicycle on a road. 12. I prefer/ would prefer to have a traffic signal dedicated to cyclists. 13. I am uncomfortable crossing an intersection on a bicycle. 14. I am/ would be comfortable using a bicycle to travel a long distance. 15. My decision to use/ not use a bicycle is affected by the distance that I have to travel. 16. I prefer/ would prefer to use a bicycle to make trips that I could make on foot. 17. I prefer/ would prefer to use a bicycle to make trips that I could make by car. 18. I prefer/ would prefer to use a bicycle to make trips that I could make using transit. 19. I prefer/ would prefer to take routes with fewer/ less steep hills when using a bicycle. 20. I avoid/ would avoid routes where I would have to cycle uphill. 21. I prefer/ would prefer to use routes that are continuous when I am riding a bicycle. 22. My decision to bike/ not bike is affected by the availability of a continuous bicycle facility along my route. 23. Having to bike uphill has affected my decision to bike/ not bike. 24. I prefer/ would prefer to ride a bicycle on surfaces that are smooth. 25. I prefer/ would prefer to avoid areas that are poorly maintained (e.g. because of potholes or garbage). 26. I consider/ would consider the state of the roadway when planning a route to take on a bicycle. 27. The nature of the areas surrounding a bicycle route has affected by decision to bicycle. 28. The quality of the areas around the route have/ would have an impact on my choice of route when riding a bicycle. 29. I avoid/ would avoid a route based on the areas that surround it when riding a bicycle. 30. If I used a bicycle more often, I would take transit less. 31. If I used transit more often, I would bike less. 32. I feel/ would feel comfortable taking a bicycle onto public transit. 33. I feel/ would feel comfortable parking my bicycle near a transit stop. 34. Secure bicycle parking should be provided at transit stations and major transit stops. 35. The cycling network provides adequate access to the transit network. 36. More needs to be done to integrate the cycling network with the transit network. 37. I would cycle more often if my friends or family cycled more often. 38. I feel/ would feel comfortable cycling after dark. 39. I feel/ would feel comfortable cycling in the rain. 40. I feel/ would feel comfortable cycling in the snow. 41. Current legislation adequately protects cyclists. 42. Cyclists should receive more protections under traffic laws.

	43. I would like more information about cycling. 44. I have access to information about cycling. 45. Having secure storage facilities at my destination would encourage me to cycle more. 46. Having change rooms and/ or showers at my destination would encourage me to cycle more. 47. I feel/ would feel confident securing my bicycle in a public area.

Questions adapted from (Ipsos Reid, 2010) and (Har Group Management Consultants, 2011):

Question Format	Please describe the extent to which you agree or disagree with each statement. I cycle ____
Response Options	Strongly disagree; Disagree; No opinion; Agree; Strongly agree; I don't know
Statements	1. Because it is more convenient than other forms of transportation. 2. Because it is fun and enjoyable. 3. Because it is better for the environment. 4. To get exercise. 5. To spend time with friends and family. 6. Despite having a car. 7. Because I do not have access to a car. 8. Because I enjoy being outdoors.

Question Format	Please describe the extent to which you agree or disagree with each statement. I do not cycle because _____
Response Options	Strongly disagree; Disagree; No opinion; Agree; Strongly agree; I don't know
Statements	1. Of the distance I need to travel. 2. Of unsafe traffic condition. 3. I need to carry things. 4. I would need to change my clothes. 5. I need a car for work. 6. It is less convenient that other modes. 7. It is too time consuming. 8. It is too tiring. 9. Of physical limitations. 10. I need to transport passengers. 11. Of a lack of dedicated facilities. 12. I am worried about my bike being stolen. 13. Bike routes are too far from my destination. 14. I do not know any safe routes. 15. I am concerned about inclement weather. 16. I do not own a bicycle.

<u>Stated Preference Question:</u>

Question #	1a	Currently, how do you commute to work or school?
Response Type		Checkbox (select one or more)
Response Options		Car; Transit; Walk; Bicycle; Other (please specify)

Question #	1b	Which of the following factors, if any, would motivate you commute by bicycle?
Ask If		Respondent did not select “Bicycle” in part A
Response Type		Checkbox (select one or more)
Response Options		Trip was twice as fast; Workplace or school was closer to home; Trip was safer; Facilities were in better condition; Parking were more secure; Trip was more comfortable; I did not need a car for work; I did not need to change my clothes; I did not need to transport passengers or cargo; None of the above; Other (please specify)

Question #	2a	Currently, how do you travel to stores or malls?
Response Type		Checkbox (select one or more)
Response Options		Car; Transit; Walk; Bicycle; Other (please specify)

Question #	2b	Which of the following factors, if any, would motivate you to make this trip by bicycle?
Ask If		Respondent did not select “Bicycle” in part A
Response Type		Checkbox (select one or more)
Response Options		Trip was twice as fast; Trip was more direct; Trip was safer; Facilities were in better condition; Parking were more secure; Trip was more comfortable; I did not need a car for work; I did not need to change my clothes; I did not need to transport passengers or cargo; None of the above; Other (please specify)

Question #	3a	Currently, how do you travel to engage in recreational activities?
Response Type		Checkbox (select one or more)
Response Options		Car; Transit; Walk; Bicycle; Other (please specify)

Question #	3b	Which of the following factors, if any, would motivate you to make this trip by bicycle?
Ask If		Respondent did not select “Bicycle” in part A
Response Type		Checkbox (select one or more)
Response Options		Trip was twice as fast; Trip was more direct; Trip was safer; Facilities were in better condition; Parking were more secure; Trip was more comfortable; I did not need a car for work; I did not need to change my clothes; I did not need to transport passengers or cargo; None of the above; Other (please specify)

Question #	4a	Currently, how do you travel to visit friends and family?
Response Type		Checkbox (select one or more)
Response Options		Car; Transit; Walk; Bicycle; Other (please specify)

Question #	4b	Which of the following factors, if any, would motivate you to make this trip by bicycle?
Ask If		Respondent did not select “Bicycle” in part A
Response Type		Checkbox (select one or more)
Response Options		Trip was twice as fast; Trip was more direct; Trip was safer; Facilities were in better condition; Parking were more secure; Trip was more comfortable; I did not need a car for work; I did not need to change my clothes; I did not need to transport passengers or cargo; None of the above; Other (please specify)