Exploring Pedestrian Road Safety in Public Transit Locations

Reza Riahi Samani

A Thesis
In the Department
of
Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science Civil Engineering at
Concordia University
Montreal, Quebec, Canada

March, 2023

This is to certify that the thesis prepared

By: *Reza Riahi Samani*

Entitled *Exploring Pedestrian Road Safety in Public Transit Locations*

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science
Civil Engineering

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

__________________________ Chair
Dr. Ciprian Alecsandru

__________________________ Examiner
Dr. Fuzhan Nasiri

__________________________ Examiner
Dr. Ali Nazemi

__________________________ Supervisor
Dr. Luis Amador

Approved by

____
Dr. Mazdak Nik-Bakht, Graduate Program Director

____
Dr. Mourad Debbabi, Dean of Faculty

Date: March 27th, 2023

ABSTRACT

Exploring Pedestrian Road Safety in Public Transit Locations

Reza RiahiSamani

This thesis studies the magnitude of pedestrian road collisions in public transit locations and addresses how the road and built-environment elements affect pedestrian safety at public transit access points (PTAPs). Collision count models and hotspot identification methods are utilized to address the research questions. Chapter 1 and Chapter 2 provide an introduction and literature review over pedestrian road safety in general, and specifically in public transit locations. Chapter 3 explains the methodologies that will be utilized in this research study. Chapter 4 establishes a relationship between pedestrian-vehicle collision counts and public transit services. Pedestrian collisions occur more frequently at intersections with the presence of a PTAP and with a higher bus traffic volume, a higher number of bus routes, and a higher public transit accessibility index. Hence, Chapter 5, explores how road geometry and built environment elements affect pedestrian-vehicle collision counts at PTAPs. The analysis shows that strategies such as road narrowing, sidewalk width increase, median refuges, presence of signal's walk interval, and vehicle stop signs could improve pedestrian safety at PTAPs. Moreover, pedestrians are at more risk in PTAP where there are roads with higher road grades and more two-way streets than one-way streets. Chapter 5 continues with Empirical Bayes collision hotspot identification and examines a couple of collision hotspots in PTAPs in the case study of Montreal City. The study findings point out the need to improve pedestrian road safety at PTAP locations and offer engineering countermeasures for addressing this problem.

DEDICATION

To
My Brother, Dadash Ali, and My lovely parents who I owe for all that I have.

ACKNOWLEDGEMENTS

I am deeply indebted to Dr. Luis Amador for his support and giving me this opportunity to step toward my academic goals and ambitions. I wish all the best to him and hope I could find more opportunities in the future to learn from his profound experience in transportation management. I would like to express my appreciation to my committee members, Dr. Alecsandru, Dr. Nasiri, and Dr. Nazemi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

The following table describes abbreviations and acronyms which are frequently used throughout this thesis. Nonstandard acronyms that are used in some places to abbreviate the names of certain mathematical variables or treatments are not in this list.

DAUs	Dissemination Area Units
PT	Public Transit
PTAPs	Public Transit Access Points
PTAI	Public Transit Access Index
NB	Negative Binomial
EB	Empirical Bayes

Chapter 1: Introduction

1.1 Background

Sustainable Cities heavily rely on public transit services for the daily travel of their inhabitants. Most transit riders walk to reach the public transit access points (PTAPs); in particular, walking is the basic mode of transport that a person may take for one or more portions of the trip. Based on Canada Census 2016, in the city of Montreal 34% and 8% of daily trips to work are done by public transit and walking respectively(“Census Profile, 2016 Census” 2017). Using transit leads to interaction with vehicle traffic flow at public transit access points (PTAPs). Therefore, it is required to provide a safe road environment at these locations. Moreover, provision of safe public transit influences the users’ assessment of transit quality, and consequently, their future decision to use public transit again (Amadori and Bonino 2012; Corazza and Favaretto 2019). Hence, it is essential to provide a safe environment for pedestrians at PTAPs which in turn promotes sustainable transportation.

Pedestrian collisions hinder the achievement of a sustainable urban transportation. Road collisions result in various costs, including, direct costs, human capital costs and willingness-to-pay costs, and when it comes to vulnerable road users, i.e., pedestrians and cyclists, there is a higher risk of collision injuries and consequently higher social and financial burden (Leur 2010). In the case study of Montreal city, which is one of the top ranking cities in terms of sustainable transportation (Batten 2017), the number of pedestrian collisions is noticeable. As figure 1-1 shows there is no downward promising trend in the count of pedestrian vehicle collision between 2012 and 2019, which indicates the need to further efforts in this domain. Similarly, Figure 1-2 shows no promising trend in the collision count in the proximity of public transit in the 8-year period. Between 2012 and 2019, there are 11,120 pedestrian-vehicle collision reports of which half (49%) of total counts are within 30 m proximity of public transit access points. Although there have been several efforts to decrease collision counts, further investigation and analysis is needed to improve the pedestrian safety levels.

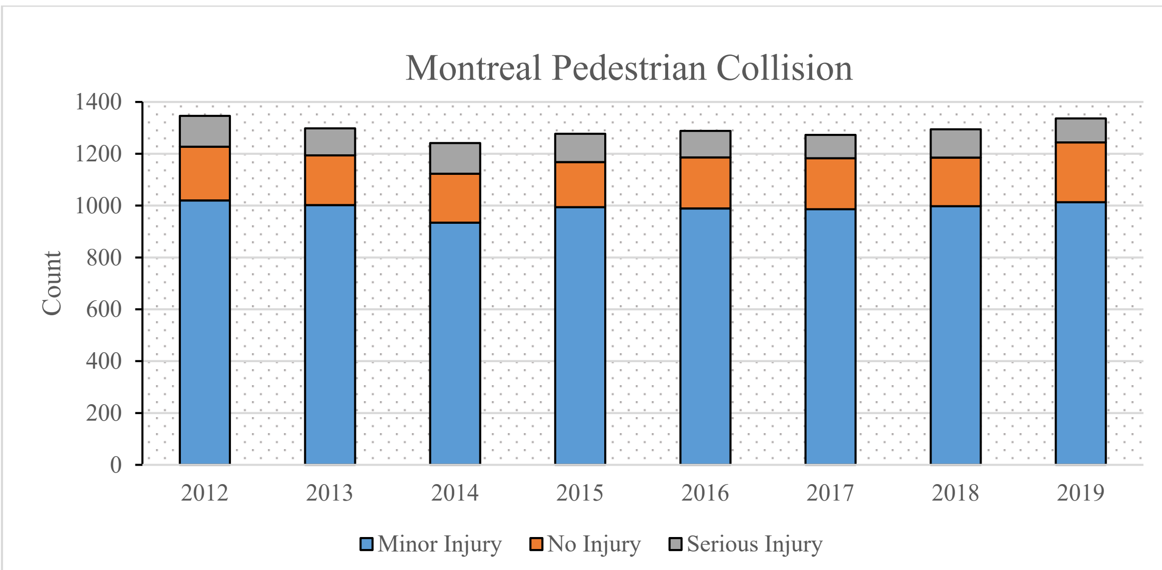

Figure 1-1Montreal Pedestrian Vehicle collision count between 2012 and 2019.

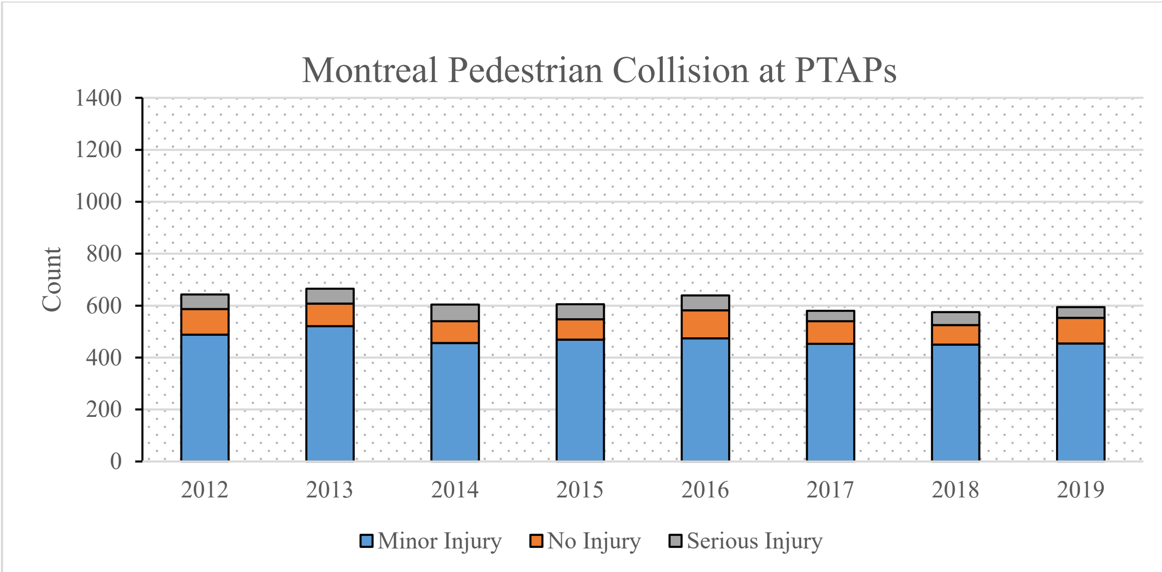

Figure 1-2Montreal Pedestrian Vehicle collision counts in the proximity to PTAPs.

Pedestrian collisions, like all road traffic crashes, should not be accepted as inevitable, because they are, in fact, both predictable and preventable (Who 2013). There is a close relationship between pedestrian safety and the walking environment. The risk of pedestrian collisions increases in an environment without pedestrian infrastructure and when more motor vehicles interact with pedestrians (Yi Zhang, Zhang, and Su 2021). In other words, inadequate or deficient engineering elements (road geometry, traffic control, and built environment) can exacerbate the pedestrian crash risk (P. Chen and Zhou 2016; Ukkusuri et al. 2012). The necessity to provide adequate safety for pedestrians has led to the development of engineering countermeasures such as the provision of pedestrian-friendly geometric designs and crossing facilities, effective traffic control, safety education programs, and implementation of pedestrian-friendly intelligent transportation systems (Zegeer and Bushell 2012; Yi Zhang, Zhang, and Su 2021).

1.2 Problem Statement

Previous studies reviewed the relation between public transit and pedestrian-vehicle collisions considering the presence of bus stops (Srinivas Subrahmanyam Pulugurtha and Penkey 2010; Shirani-bidabadi et al. 2020b; PASHKEVICH and NOWAK 2017), or the average of bus frequency at PTAPs or transit corridors (Ye et al. 2016). However, establishing a robust complementary examination of the possible relation between pedestrian road collisions and public transit locations requires evaluating the effect of exposure characteristics of public transit services, such as bus average daily traffic and public transit accessibility index. Moreover, limited studies addressed the role of road elements and built-environment factors on pedestrian safety, particularly at public transit access points. In these studies, characteristics of PTAPs are examined by utilizing qualitative categorical variables of the surveyed sites (Lakhotia et al. 2020a), and some studies are irrespective of traffic exposure and pedestrian factors (Mukherjee, Rao, and Tiwari 2022). Finally, previous studies on the quality of public transit access are typically based on a limited number of surveyed sites, e.g., 46 sites (Ye et al. 2016), 117 sites (Quistberg et al. 2015), and 70 intersections near to 15 rail transit stations (Srinivas S. Pulugurtha and Srirangam 2021); however, small sample sizes are likely to be biased in safety analysis (Afghari et al. 2019). This study attempts to address these problems through the research objective which will be shortly explained in the following.

1.3 Research Objective

1.3.1 Overall Goal

This research addresses pedestrian road safety in the proximity of public transit service locations. According to the problem statement, three research objectives lead the study to this main goal. The first objective of this study is to incorporate spatial investigation and add statistical analysis on public transit exposure variables considering transit service characteristics to establish a robust examination of the possible relation between pedestrian road collisions and public transit locations. The second objective of this thesis is to utilize quantitative road geometry characteristics and built-environment factors while controlling for pedestrian and vehicle traffic exposure to evaluate pedestrian safety at PTAPs. This research aims to establish the analysis based on a wide range of PTAPs located at midblocks and intersections and overcome the possible bias due to the small sample sizes. Finally, this thesis incorporates hotspot identification methods to nominate public transit stops with a high concentration of collisions and investigates how much applicable are suggested countermeasures to improve pedestrian safety at these hotpots.

1.3.2 First Research Task

The first research objective is to establish a robust examination on the possible relation between the pedestrian's vehicle collision counts and public transit service. Therefore, this study incorporates polygon level, intersection level, and public transit site level spatial investigation and conducts statistical analysis on public transit exposure characteristics. Public transit features include presence of public transit stop, bus traffic, number of transit routes, and public transit accessibility index at the locations. The finding of this research task could emphasise and determine whether there is a need to conduct further studies on pedestrian safety at PTAPs.

1.3.3 Second Research Task

The second objective of this study is to incorporate safety performance factors on pedestrian vehicle collisions in proximity to public transit locations to figure out the effect of engineering elements on pedestrian's safety levels at public transit access point. Engineering elements include road geometry, traffic calming, and built environment factors playing part in pedestrian road safety. The statistical analysis suggests applicable counter measures improving pedestrian safety based on the results of the models. To overcome possible bias due to the small sample size (Afghari et al. 2019), this research aims to establish the analysis based on a wide range of PTAPs and provides separate analysis for PTAPs located at midblocks and intersections, and particularly, at signalized and non-signalized intersections.

1.3.4 Third Research Task

The third research objective of this study is to utilize collision hotspot identification methods to prioritize hazardous sites and identify public transit stops with high potential for improvement. The results could indicate if the hotspots require further investigations and field specific studies. This research task addresses some hot spots example to figure out whether the findings suggested in previous analysis could be applicable to improve pedestrian safety levels at the hotspot site.

1.4 Scope and Limitations

This research study targeted pedestrian road safety in the proximity area to public transit services, considering pedestrian vehicle collisions at these locations, considering PTAPs located nearby intersections and at midblocks. This thesis takes into account aggregated collision counts in an eight-year time period between 2012 and 2019, in the defined locations. A buffer size of 30 meters

(100 foot) is suggested in the literature as an appropriate influence area for transit stops (Srinivas S. Pulugurtha and Vanapalli 2008). Therefore, 30 meters buffers around bus stops and metro access points are considered in this research study. Due to the shortage of transit ridership data, studies on the safety of individual transit users are limited (Ulak et al. 2021), however, it has been shown that pedestrian-vehicle collisions nearby public transit (PT) could be considered a surrogate measure of transit users' safety(Hess, Moudon, and Matlick 2004). Hence this research study explored reducing pedestrian vehicle collision risk in the proximity of PTAPs which in turn enhances pedestrians and PT users' road safety.

This study utilizes the short term traffic count at 1850 intersections in the city of Montreal as for traffic count data. Although the dataset provides a proper approximation of pedestrian and vehicle traffic exposure at the studied sites, the analysis could be performed based on long term traffic counts or short term traffic counts using hourly, daily, and monthly expansion factors (Stipancic et al. 2020a). The Montreal traffic count dataset provides the count of vehicles and pedestrian at the intersections, however, in this research the values are assigned to the corresponding intersections' legs which expands the traffic count to the road network (Please referee to data description section and Appendix I for further explanation).

It is assumed that pedestrian road safety at public transit access points behaves the same at bus stops and metro stations access. Therefore, PTAPs in this research study include bus stops and metro stations. The analysis in this thesis are limited in considering how safety models would vary regarding the type of PTAP, i.e., bust stop, metro station, or an multimodal interchange point.

The analysis in this thesis includes statistical inference based on Negative Binomial statistical models. Although the models show an association between the response variable, count of collisions, and the variants, such relations do not necessarily mean causality. While we can say there is an association between those variables, we cannot say that one variable caused the other. For example, the analysis shows the statistical significance of the relation between the road grades and the count of collisions; however, this statistical inference should be distinguished from a casual relation.

1.5 Research Significance

This thesis contributes to the previous studies by, first, establishing a robust relation between pedestrian collisions and public transit locations; second, by providing a comprehensive insight into the impact of quantified road geometry and built-environment characteristics on pedestrian road safety at PTAPs; and third, by conducting statistical analysis on the road configuration and geometric factors considering a wide range of PTAPs, 950 at mid-blocks, and 1700 at intersections, in order to provide more inclusive results and conclusions for city authorities and transit agencies to improve pedestrian safety at PTAPs.

1.6 Organization of the Thesis

This thesis is presented in six chapters as follows. Chapter 1 defines the problem and presents the objectives of the research and structure of the thesis. Chapter 2 contains a review of the safety performance models, particularly collision frequency models and presents the previous studies considering pedestrian safety and public transportation. This chapter ends with the literature gap in the thesis topic in which accordingly research objectives are defined. Chapter 3 presents the methodology employed in this thesis. Chapter 4 presents the research covered under research task 1. This chapter establish the need to provide safe road environment for pedestrian at public transit access points. Accordingly, in Chapter 5, the work under research task 2 is presented. This chapter suggests how road environment and traffic calming strategies could improve pedestrian safety in public transit locations. It ends with assessing the proposed safety measurements at collision hotspots in public transit locations. Chapter 6 presents the conclusions and lessons learnt from the analysis and modeling experience and, make recommendations for future research. The work described in Chapter 4 and part of the analysis in Chapter 5 could be find as publication in the Canadian Journal of Civil Engineering ("Exploring Road Safety of Pedestrians in Proximity to Public Transit Access Points (Bus Stops and Metro Stations), a Case Study of Montreal, Canada, R Riahisamani, Amador, L, Canadian Journal of Civil Engineering, Jun 2022, https://doi.org/10.1139/cjce-2022-0281").

Chapter 2: Literature review

2.1 Introduction

This chapter aims to review the previous research studies on pedestrian road safety in general and with a focus on the proximity area of public transit locations. It discusses the existing literature gap that this dissertation will address. The review includes previous methods, approaches, and results and critically discusses the studies. First, section 2.2 provides an overview of general findings and an analysis on the topic of pedestrian road safety and public transit. The section reviews the background, and possible reasons for the relation between pedestrian road collisions and the presence of public transit. It will be followed by section 2.3, which reviews collision count models and provides a comprehensive insight into statistical collision count models and it discusses the weaknesses and strengths of the methods. Section 2.4 provides an overview of collision hotspot identification methods, including GIS-based and statistical-model-based hotspot identification. Section 2.5 provides the latest updates on pedestrian road safety and road geometry and built environment elements, which enables comparing the finding of this research with previous studies. Finally, Section 2.6 discusses the literature gap which will be addressed in this dissertation.

2.2 Pedestrian Vehicle Collisions at Public Transit Locations

The safety of pedestrians and transit riders and the presence of bus stops have been studied with different approaches (Jaeyoung Lee et al. 2015a; Xie et al. 2017; Ulak et al. 2021). Some research studies have addressed the topic within the zonal level aggregated approach (Jaeyoung

Lee et al. 2015b), while others studies considered corridors with and without transit services (Hess, Moudon, and Matlick 2004). Pulugurtha and Penkey (S. Pulugurtha and Penkey 2010) compared pedestrian crashes in 30 different segments, some with public transportation corridors, and others without public transit services, with control for other variables. Pedestrians' safety was higher in segments without public transit access points (PTAP) than in those with PTAPs. It was found that segments with more PTAPs, transit ridership increases and impacts the pedestrian's road safety performance (S. Pulugurtha and Penkey 2010). In addition, the number of boarding and bus frequency were shown to increase pedestrian and vehicle collisions in the proximity of PTAPs (Ye et al. 2016). Therefore, a strong correlation between the presence of public transportation (PT) and road collisions involving pedestrians has been asserted in the literature (Quistberg et al. 2015).

Previous research discussed the probable reasons behind the relation between pedestrian safety and the presence of public transit stops. The research shows that public transit ridership is significantly associated with the location of commercial centers and education buildings and the presence of facilities such as hospitals, supermarkets, and religious facilities (Zhao et al. 2013; Ulak et al. 2018). More pedestrian traffic is expected around bus stops with higher ridership; hence bus ridership might impact pedestrian crashes due to higher pedestrian traffic. However, some interrelated factors might negatively impact the pedestrian road safety in the vicinity of PTAPs. For example, the presence of bus stops could create sight distance issues as stationary buses could be an obstacles for the moving traffic and lead to higher traffic conflicts (Shirani-bidabadi et al. 2020a; PASHKEVICH and NOWAK 2017; Chin and Quddus 2003). Also, the presence of PTAP as a pedestrian attraction might influence a pedestrian to commit a violation to avoid missing a transit connection (Cinnamon, Schuurman, and Hameed 2011). Another study asserted the poorer yielding behavior of vehicles near bus stops (Craig et al. 2019). In regards to this behavior, some psychological factors were studied, such as an increase in driver's confusion as to whether the pedestrian intends to cross or is waiting for a bus as well as driver distraction due to increased signage at public transit stops (Craig et al. 2019).

There are studies focused on pedestrian road safety at transit stops, particularly bus stops, by applying statistical analysis. Pedestrian safety at transit stops was studied by considering socio-demographic factors(Ulak et al. 2021), the quality of access to the stop(Lakhotia et al. 2020a), and geometric and traffic characteristics of the PTAP(Ye et al. 2016). Due to the shortage of transit ridership data, studies on the safety of individual transit users are limited (Ulak et al. 2021). However, pedestrian-vehicle collisions nearby PT could be considered a surrogate measure of the safety of transit users (Hess, Moudon, and Matlick 2004). Hence, addressing pedestrian-vehicle collision risks at PTAPs improves pedestrian safety in general and enhances the safety of public transit users in particular.

2.3 Collision Count Modelling

Studying collision count models requires an understanding of the primary characteristics of crash data that influence the modeling phase. The following characteristics of collision count data are described, followed by a review of the collision counts statistical models.

2.3.1 Characteristics and challenges in Collision Count Data

There are four main challenges in collison count modeling: over-dispersion, unobserved heterogeneity, the low sample mean and small sample size, and endogenous variables. The main goal of researchers has been to address these challenges through the collision count models. In the following sections, each characteristic is described separately.

2.3.1.1 Over-dispersion

Crash data are usually characterized by over-dispersion, which indicates that the variance is greater than the sample mean. When there is over-dispersion in data, the Poisson model is not recommended for crash count analysis as it leads to biased and inconsistent parameter estimations. This, in turn, leads to erroneous inferences on the models' factors that determine crash frequencies (Mannering, Shankar, and Bhat 2016). Generally, the source of over dispersion could be considered on two sides, namely overdispersion due to unobserved heterogeneity and overdispersion due to the excess count of zeros, both of which are explained specifically.

2.3.1.2 Unobserved heterogeneity

In the real world, crashes are a result of multi-interrelated factors. The crash data typically does not include all of the factors that could lead to capturing the variance of road collision counts. This could be due to the omission of important site attributes, implicit randomness of crash occurrence, inaccuracy of traffic volumes, and unmeasured variations (weather conditions, visibility, driver behavior, etc.). Hence, unobserved covariates or immeasurable variations are often the source of the heterogeneity in the data and are responsible for the over-dispersion. Moreover, in panel data (data gathered on the same spatial elements through different periods of time, e.g., months, and years), there also exists a temporal and spatial correlation in crash count data. In other words, to get a sufficient number of observations of accidents, data is collected over time and/or space which creates additional unobserved heterogeneity issues (Mannering, Shankar, and Bhat 2016).

2.3.1.3 Low sample mean and small sample size

Large costs associated with the data collection process, and the uniqueness of entities, such as rural intersections or the number of metro station, could lead to data characterized by a small sample size and low sample mean. The crash is inherently a rare event, which could lead to a high proportion of zero recordings of crash data. This problem can be observed in the large number of locations that report zero accidents for a given time period, producing an accident frequency distribution with a high proportion of zeros. With low sample means (and dominance of zeros), the distribution of crash counts will be excessively skewed to zero, leading to incorrect estimations of parameters. Hence, when the observed zero counts exceed the zero counts that the standard Poisson model can produce, it is known as over-dispersion due to the excess number of zeros (Mannering, Shankar, and Bhat 2016)(Lord, Washington, and Ivan 2005).

2.3.1.4 Endogenous variables

An endogenous variable refers to an explanatory variable whose value is influenced or determined by one or more variables in the model. For example, in exploring the effectiveness of ice warning signs in reducing the frequency of ice-related crashes, models may show a higher collision rate at locations where ice warning signs are installed. However, the endogeneity problem should be considered, as in the example, ice-warning signs are more likely to be installed at locations with high numbers of ice-related crashes. Ignoring the endogeneity may lead to an erroneous conclusion on the endogenous variables (Mannering, Shankar, and Bhat 2016).

2.3.2 Collision count statistical models

2.3.2.1 Background

Collision counts are nonnegative and discrete road incidents. They need specific analytical tools and techniques to study the road safety of a target zone. Crash frequency models are known as safety performance functions (SPFs) which can be used for different applications. For example, SPFs are used for the identification of relations, screening variables, the sensitivity of variables, prediction, and causal relationships. In this research, SPFs are mainly used to conduct association analysis between the built-environment covariates and collision counts.

The traditional accident modeling analysis is based on the accidents reported during a time period of observation and various site-specific attributes such as traffic volumes, geometry features, warning devices, weather conditions, etc. The input data for collision count modeling can be represented as the following:

Input data at site $i = \{y_i, F_{i1}, F_{i2}, x_i, T_i\}$

where $i = 1, 2, \ldots, n$ is the total number of locations under analysis, y_i is the observed number of accidents at site i, and F_{i1}, F_{i2} are traffic volumes in interactive directions for each site, commonly defined as the average annual daily traffic (AADT). It could also be defined as traffic volume of different road users, such as pedestrians average daily count, bicycles average daily count, or AADT of special vehicles such buses and emergency vehicles. x_i is a vector of site-specific attributes such as geometry features, road assets, speed limits, location type, land use characteristics, and built-environment elements. T_i is the time period of observation for site i which typically is considered constant between all entities. For example, this research considers pedestrian-vehicle collision counts between 2012 and 2019.

Based on the research purpose, accidents can be modeled by accident frequency, severity, or severity-frequency models. This research follows collision count models (frequency approach). As there is a lack of annual-based traffic data, geometry modifications, and other possible changes in built-environment, models are usually limited to cross-sectional data analysis. In the following section, different collision frequency models are discussed.

2.3.2.2 Progress in Statistical Collision Count Models

The developments in statistical count models lead to the application of these techniques to model the occurrence of road accidents and develop safety performance functions. Theoretically, an accident is a binary process of whether a road user experiences an accident or not. That is a Bernoulli trial in which each time a road user enters an intersection, road segment, highway segment, bridge, tunnel, or any other type of entity, it is assumed as a trial either crash (successes) or not crash (failure) with unequal crash probability (Barbour, Holst, and Janson 1992; Olkin, Gleser, and Derman 1980). It was shown that when there are many trials and a low probability of success (collision occurrence), the Bernoulli process can be well approximated as the Poisson trial (Paul P Jovanis 1986; Lord, Washington, and Ivan 2005). Poisson models perform well under nearly homogenous conditions. This is when there is a total number of crashes occurring on a given set of roads over a given time period rather than crashes from different sets of roads or from the same set of roads over different time periods. The Poisson process works well when the mean and variance of the sample are equal in the crash data, while in practice, the variance of crash data is by far larger than the mean. Moreover, the Poisson model assumes the same crash risk in all sites with the same characteristics (same covariates), which is theoretically impossible.

As one of the most popular crash count models, Negative Binomial (NB) has been applied in many research and practical projects (Shaik and Hossain 2020). NB is one of the most popular extentions of the standard Poisson model by assuming the Poisson parameter follows a gamma distribution. This assumption strengthens the NB to overcome the over-dispersion issue in Poisson models (Hilbe 2011). Therefore, in addition to the mean parameter of the standard Poisson model, there is an overdispersion parameter of α following the Gamma distribution. It has a closed-form equation, and the mathematics to control the relationship between the mean and the variance structures are relatively simple. Since overdispersion is a common characteristic in crash data, and NB properly deals with overdispersion, several studies have shown the significance of overdispersion parameter in NB. Therefore, the superiority of NB to the Poisson model in crash data analysis has been studied in several research studies (T.T 2019; Arévalo-Támara, Orozco-Fontalvo, and Cantillo 2020; Khattak et al. 2021). NB is recognized as the most frequently used model in crash count modeling in recent years among practitioners and researchers (Lord, Qin, and Geedipally 2021).

The Poisson-lognormal (PLN) model is another extension of the Poisson model, which provides a similar parameterization as NB, but the error term follows lognormal distribution rather than a gamma distribution. Compared to the Poisson-gamma model (NB), the PLN model is more flexible for observations located at the tail end of the distribution, and on the other side, the NB tends to fit the data better near the zero counts (Khazraee, Johnson, and Lord 2018). Following the introduction of the NB model, several other Poisson extension models have been studied in the context of crash data analysis, for example, PoissoneWeibull(Cheng, Geedipally, and Lord 2013), and Poisson-Inverse Gaussian(Zha, Lord, and Zou 2016).

Moreover, finite mixture models were introduced to capture the unobserved heterogeneity more efficiently. They assume overall data are generated from several distributions that are mixed together. In other words, individual observations are generated from an unknown number of distributions. Examples of finite mixture models are Zero-inflated models such as Zero Inflated Negative Binomial and Negative Binomial-Lindley. In these models one of the two latent classes has a long-term mean equal to zero which theoretically is questionable. It is obviously not possible to expect sites with no crashes while there is road traffic on the facilities (Lord, Washington, and Ivan 2007). However, there are recent studies that addressed such limitations and suggested the advantages of developed modes of zero inflated models (Islam, Shirazi, and Lord 2022).

Random-effects or multilevel models were introduced to address the assumption of Poisson mixtures that the effect size of the variables is fixed. In other words, in the above-discussed models, one true effect works the same for all the observations, and the unobserved heterogeneity between observations is a purely random error. However, practically, the unobserved heterogeneity may not be solely a purely random error but could also be partly explained by the differences in the observations themselves and between groups or levels of observations (Gelman and Hill 2006). Random effect (RE) models allow the variance that may exist within different levels of the data to be better depicted. It adds one random effect or random intercept term to capture the between observations variance. In fact, the RE model modifies the mean of the observation, i.e. the mean of collision count for one site, by manipulating the intercept value. It was found that the Random Effect Negative Binomial model was the best fit for the data and outperformed the other models according to goodness-of-fit measures (Yan et al. 2020; Hosseinpour, Yahaya, and Sadullah 2014). As an extention of Random Effect models, Random Parameters (RP) models were introduced. In addition to only incorporating randomness to the models' intercept, estimated coefficients of the RP model vary across each observation to better capture unobserved heterogeneity across all observations (Anastasopoulos and Mannering 2009).

Moreover, some research mixed more than two distributions, such as the Negative Binomial-Lindley model, which is the mixture of NB and the Lindley distribution (Lord and Geedipally 2011). As mentioned previously, the NB is the mixture of Poisson-Gamma distribution, and the Lindley is the mixture of exponential and Gamma distribution. NB-L has a thick tail which can be used when there is a large number of zeros in the data. It has been studied to define a threshold for the skewness that the NB-L models outperform the conventional NB mode(Shirazi et al. 2017).

Most recently, models with multi sources of risk factors have been proposed. These models support the idea that the crash count could be generated by multiple simultaneous and inter-dependent sources of risks. This is compared to the previous frequency models which are based on a single linear equation enhanced by a variety of distributional assumptions (Afghari et al. 2018). For example, a crash occurrence is well approximated using three sources of risk: engineering, unobserved spatial, and driver behavioral factors, in which, at each site, the hierarchy of risk sources could change (Afghari et al. 2018). As the results of these models, instead of a single crash frequency outcome, there could be source-specific crash count models which is based on each crash source (Afghari et al. 2018).

Table 2-1 Summary of Collision Frequency Models

Model	Main Characteristics	Initial Example of Research
Standard Poisson Distribution	• Start point for Poisson family Collision Frequency Models • Restricted to equal mean and variance • Incapable of dealing with over dispersion	(Olkin, Gleser, and Derman 1980)
Negative Binomial	• Variance to mean ratio larger than one • Capable of dealing with overdispersal data	(Paul P Jovanis 1986)
Zero Inflated Negative Binomial	• Dual state data generating process • Deal with excess count of zeroes through a separate distribution • Assuming perfectly safe and unsafe entities (theoretically questionable)	(V. Shankar, Milton, and Mannering 1997)
Poisson Lognormal (PLN)	• Poisson Parameter follows a lognormal distribution • Flexible for observations located at the tail end of the distribution	(Miaou, Bligh, and Lord 2005)
Negative Binomial-Lindley	• Multi-distribution statistical model • Analyzing data characterized by a large number of zeroes • Maintaining similar characteristics as the traditional negative binomial	(Lord and Geedipally 2011)
Random Effect	• Argue a not purely random error • Adding one or more random intercept terms to deal with spatial-temporal unobserved heterogeneity • Capturing within observations variance	(V. N. Shankar et al. 1998)
Random Parameter	• Extension of randomness from intercept only to all model's coefficients • Improve capturing unobserved heterogeneity	(Anastasopoulos and Mannering 2009)
Multiple Sources of Risk	• Multi sources of risk such as engineering, unobserved spatial, and driver behavioural • Source-specific crash count modelling	(Afghari et al. 2018)

2.4 Network Screening

The process to identify hazardous points, known as network screening, is the primary task in transportation safety management (AASHTO 2010). Hazardous sites or hotspot sites are crash-prone locations and refer to high-risk sites that are experiencing more crashes compared to the expected counts in the other sites with similar characteristics. Network screening primarily assesses the conditions of an existing safety level of infrastructure and provides transportation agencies with decision support clues for short-term and long-term countermeasures. Moreover, the identification of hazardous sites leads to the efficient dedication of limited agency resources by implementing safety improvements with the objective of reducing the number of crashes in selected hotspot sites (Meng et al. 2020). In the following section, two primary approaches, namely, statistical models-based and GIS-based hotspot identification, are explained.

2.4.1 Observed Crash Methods

The first edition of the Highway Safety Manual (HSM) provides several types of network screening methods (AASHTO 2010). The basic methods are based on the raw data of collision observations. For example, the popular crash observation-based methods are average Crash Frequency, Crash Rate, Equivalent Property Damage Only (EPDO), and relative severity index (RSI) (AASHTO 2010). The simplest of all network screening methods is the crash frequency which considers the count of observed collisions in a period of time. It does not take into account the underlying impact of traffic volumes which leads to a bias toward locations with higher traffic volume. However, Crash Rate overcomes the mentioned problem by considering the traffic volumes and returns crash frequency in proportionate to traffic volumes. Therefore, the Crash Rate method enables identifying low-volume sites that may not necessarily experience a high number of crashes but are high-risk sites. However, the Crash Rate method is biased towards lower traffic volume sites, while sites with low traffic count inflate the ratio between crash count and volume. Finally, the Severity Index method is the standardized form of EPDO method, which takes into account both the frequency and severity level of collisions.

2.4.2 Predicted Crash Method

A crash occurrence is naturally a random process, and it changes over time at any given site. The observed crash method misses the randomness and crash fluctuation over time. Therefore, the expected crash frequency is used, which is the outcome of collision frequency models and safety performance functions (SPFs). SPFs take into account all factors such as traffic exposure and built-environment elements that influence the site safety rather than the raw information of observed crashes. When the observed safety at the entity is worse than the expected safety obtained from the SPFs, the site is considered a hotspot or high risk of collision. Thus, the Potential for Improvement (PI) is defined as the difference between the expected and observed crash count at the entity. The sites with larger PI could be considered as hazardous sites. Although SPFs address the randomness of collisions by providing the expected crash count, the PI in this method does not provide information on the random nature of PI. Moreover, this method does not provide a specific performance threshold (AASHTO 2010).

2.4.3 Bayesian Methods

The basic observed crash frequency and crash-rate-based hotspot identification (HSID) methods are not reliable due to the missing regression-to-the-mean bias (Carriquiry and Pawlovich

2004). The regression to mean (RTM) bias poses a challenge to the identification of hotspots. It is when the count of collisions is extreme, and the next sampling of the random variable of count is likely to be closer to its mean. It stems from the fact that collision frequency on a roadway is stochastic in nature, which means that it may experience fluctuations at a particular site in a short period of time without reflecting any changes in factors that affect the true underlying safety (J Lee et al. 2016). To deal with the RTM bias associated with empirical crash data, some researchers proposed and implemented the Empirical Bayesian (EB) framework (Wu et al. 2014). The EB method incorporates the observed crash counts at sites with the expected collision count result of an SPF that relates crash frequency as a function of variables such as road characteristics (Hauer 1997). The EB method significantly outperforms in the estimation precision and correction for the regression to mean bias (AASHTO 2010). In the EB approach, information on past safety performance is combined with information concerning existing safety conditions, therefore, entities' estimations are combined to determine an improved site's long-term safety (A. S. Lee et al. 2018).

The EB method solves one of the problems with the crash frequency and crash rate by considering the random fluctuation in crash counts, however, it requires an abundance of samples to develop the safety analysis model (X. Wang et al. 2021). Full Bayes (FB) method was developed to estimate the posterior distribution of crashes and then taking the estimate as an index to identify hotspots. A comparison between EB and FB showed similar results when the number of samples was sufficient, however, when the sample size is small, the FB could outperform EB method (Miranda-Moreno and Fu 2007). Moreover, the complexity of the FB's model enables using a hierarchical model in a Bayesian framework considering spatial and temporal correlation, and results in a better hotspot identification than EB (Huang et al. 2009).

2.4.4 GIS based Collision Hotspot Analysis

Geographic Information System technologies (GIST) have been commonly used for safety analysis and are considered the most popular tools for the visualization of crash data and hotspot analysis. Geospatial methods consider the effects of unmeasured variables by accounting for spatial autocorrelation between the crash events over geographic space. The GIS-based hotspot identifications use various methods and techniques, such as the traditional kernel density estimations, nearest neighbor analysis, K-functions, Moran's I index, and Getis-Ord (Truong and Somenahalli 2011)(Rahman, Jamal, and Al-Ahmadi 2020). Ulak et al compared four network-based hotspot detection methods namely, Getis-Ord , Local Moran's I, KLINCS (K-function local indicators of network-constrained clusters), and KLINCS-IC (Inverse Cost) to give insight into the similarities and differences between the methods using different spatial weights (Ulak et al. 2019). Using the crash prediction accuracy index, it was shown that, the Getis-Ord Gi* and Local Moran's I with distance-based spatial weights perform better than other spatial weights such as free flow travel time and congested flow travel time (Ulak et al. 2019). The Gi* method was shown to be the most precise and reliable method among Average Nearest Neighbor, Global Moran's I, kernel density estimation (KDE), mean centre, and Getis-Ord Gi* (Amiri et al. 2021). Getis-Ord Gi * was used to identify vehicle road accidents at expressways (Borhan, Razuhanafi, and Yazid 2019) and to identify bus stops hotspots in terms of pedestrian vehicles collisions severity (Truong and Somenahalli 2011).

2.5 Traffic Calming and Road Geometry Strategies

Traffic Calming Strategies are a set of measurements for roads that try to slow down the speed of motor vehicles and restore roads for pedestrians and cyclists' users. Traffic calming deliberately slows down the traffic speed, typically in areas with a high volume of non-motorized users. It is mostly done by applying physical measures including: speed bumps, roundabouts, raised pavement at intersections, reducing road areas allotted for motor traffic, traffic diversion, signalization, and so on (Litman 1999).

Various research studies have explored the effect of different traffic calming strategies on pedestrian safety levels. Speed reduction as one of the main objectives of traffic calming strategies was addressed in many research studies (Distefano and Leonardi 2019). Previous studies show that the higher the speed limit or the operating speed is, the higher the risk of collision occurrence and severity is for pedestrians (Ewing and Dumbaugh 2009; Fridman et al. 2020). Lowering speed limits would bring better conditions for looking out, for avoidance actions, and for communication between different users.fa Although collision occurrence could also be reduced by traffic speed reduction, in the case of collision, lower speed mitigates the severity of the accident (Distefano and Leonardi 2019; Ziakopoulos and Yannis 2020).

Road median type providing safe refuge for pedestrians crossing is shown as an effective improvement for pedestrian safety (Vignali et al. 2019; Lakhotia et al. 2020b). Some researchers argue that the driver's perception of a built-environment as an urban district is due to the presence of certain road medians. Medians could significantly increase the proportion of pedestrians who look for vehicles before crossing the road (S. Pulugurtha et al. 2012). It also increases the proportion of drivers who yield to pedestrians (S. Pulugurtha et al. 2012).

Road narrowing is one common traffic calming measurement that is applied to reduce speed profile. A before-after study showed a 33% reduction in pedestrian collision injuries by road narrowing (Distefano and Leonardi 2019). Some studies showed a correlation between wider lane width and higher pedestrian road safety (Tulu et al. 2015; T. Chen et al. 2020; Lakhotia et al. 2020a). They assert that the extra space would work as a buffer between pedestrian and vehicular traffic. While some other researchers advocate to increase pedestrian safety through speed reduction achieved by the drivers' perception of narrower lanes (Gomaa Mohamed et al. 2012).

The effect of traffic control measurement, specifically the signal's walk intervals, rarely have been studied in collision count models (Stipancic et al. 2020). Previous research shows that traffic signal walking intervals decreases the probability of pedestrian conflicts and collisions (Yaohua Zhang et al. 2015).

Characteristics of road networks and geometric features such as the road length and the number of intersections have also been argued in the literature (Ukkusuri, Hasan, and Aziz 2011; Y. Wang and Kockelman 2013; Jaeyoung Lee et al. 2015a). Pedestrian collision severity at intersections will decrease with curb extensions, raised medians, and exclusive left turn lanes and will increase with the total number of lanes and the number of commercial entrances (Stipancic et al. 2020).

2.6 Literature gap and Contributions

The literature review showed that studies analyzed the relation between public transit and pedestrian-vehicle collisions considering the presence of bus stops (Srinivas Subrahmanyam Pulugurtha and Penkey 2010; Shirani-bidabadi et al. 2020b; PASHKEVICH and NOWAK 2017), or the average of bus frequency at bus stop locations or corridors with PT services (Ye et al. 2016). While, establishing a robust complementary examination of the possible relation between pedestrian road collisions and public transit locations requires evaluating the effect of other characteristics of public transit services, such as average daily traffic volume, public transit accessibility index, public transit routes characteristics, and presence of the services. Hence, the first objective of this paper is to incorporate spatial investigation and add statistical analysis on public transit exposure variables, considering transit service characteristics. Moreover, limited studies addressed the role of road design and built-environment factors on pedestrian safety, particularly at public transit access points. In these studies, characteristics of PTAPs are examined by utilizing qualitative categorical variables of the surveyed sites (Lakhotia et al. 2020a; 2020b), and some studies do not consider traffic exposure and pedestrian factors (Mukherjee, Rao, and Tiwari 2022). Therefore, the second objective of this paper is to utilize continuous road geometry and built-environment factors, taking into account pedestrian and vehicle traffic exposure to evaluate pedestrian safety at PTAPs. Finally, previous studies on the quality of public transit access are typically based on a limited number of surveyed sites, e.g., 46 sites (Ye et al. 2016), 117 sites (Quistberg et al. 2015), and 70 intersections near 15 rail transit stations (Srinivas S. Pulugurtha and Srirangam 2021). As small sample sizes are likely to be biased (Afghari et al. 2019), this research aims to establish the analysis based on a wide range of PTAPs located at midblocks and intersections.

This paper contributes to the literature by, first, establishing a robust relation between pedestrian collisions and the presence of public transit; second, providing a comprehensive insight into the impact of quantitative road geometry and built-environment characteristics on pedestrian road safety at PTAPs; and third, conducting a statistical analysis on these elements, considering a wide range of PTAPs, 950 mid-blocks, and 1700 intersections, in order to provide more inclusive and reliable conclusions. Therefore, the contribution of this thesis is to provide information for city authorities and transit agencies to improve pedestrian safety at public transit access points (PTAP).

Chapter 3: Methodology

3.1 Introduction

This chapter presents the methodology employed in this dissertation. The chapter is divided into two sections. First the statistical model, namely, Negative Binomial (NB) applied for developing safety performance functions, is discussed in detail. The aim of using this methodology is to identify the key traffic, transit, road, and built-environment factors that influence pedestrian-vehicle crash frequency at PTAPs. Secondly, two hotspot analysis approaches, namely, Empirical Bayes (EB) method and Getis Ord spatial hotspot identificatoin, are explained which are used to address network screening in this dissertation.

3.2 Collision frequency model

3.2.1 Negative Binomial (NB)

In the transportation safety analysis, the Poisson-gamma regression model (Negative Binomial) is the popular extension of the Poisson model developed to overcome the limitations of standard Poisson regression. As discussed in chapter 2, crash occurrence initially was considered as a series of Bernoulli trials, known as the binomial distribution under the assumption of equal probability of events, which the formulation is given as follows:

$$P\ (Y = n) = \binom{N}{n} P^{n}\ (1 - p)^{N-n} \qquad \textit{3-1}$$

Where N is the count of trials (count of road users), and n= 0,1,2,..,N refers to the number of collisions (success cases) and P is the probability of success (collision occurrence). The mean and variance of the binomial distribution are respectively $E(Y) = Np$ and $VAR(Y) = Np(1-p)$. It was shown that the Bernoulli process approaches Poisson distribution in crash count analysis (Olkin, Gleser, and Derman 1994). Therefore, the mean and variance of a collision count for each entity follows a Poisson distribution as the following formulation:

$$Y_i \mid \mu_i \sim Poisson\,(\mu_i)\,,\ \mu_i = f(F_{i1}, F_{i2}, x_i; \beta) \qquad 3\text{-}2$$

Where Y_i is the random variable represents estimated collision counts, μ_i is the mean of crash counts which follows Poisson distribution, μ_i commonly specified as an exponential function of site-specific attributes or covariates, and $\beta = (\beta_0, \ldots, \beta_k)$ is the vector of regression parameters to be estimated from the data and F_{i1}, F_{i2} are traffic volumes in interactive directions. In addition, x_i is a vector of covariates representing site-specific attributes. $Y = (Y1, \ldots, Yn)$ is the number of accidents at site i $(i = 1, \ldots, n)$ represents random variables corresponding to the n sites under analysis.

The illustrated Poisson distribution is restricted to the assumption of equal variance and mean, while it is not valid in the real world. Therefore, Negative Binomial provides random variations in the mean of the Poisson model through a two-stage mixed Poisson model such as the Poisson/Gamma:

$$Y_i \mid \theta_i \sim Poisson\,(\theta_i) \sim Poisson\,(\mu_i \,.\, e^{\varepsilon_i})$$

$$e^{\varepsilon_i} \sim Gamma\,(\phi, \delta)$$

$$e^{\varepsilon_i} \mid \phi \sim Gamma\,(\phi, \phi)$$

where $Gamma()$ denotes a Gamma probability density function with parameters ϕ>0 and δ >0, which ensures $\theta_i > 0$ and e^{ε_i}>0, and specifying $\phi = \delta$ the classical NB model is obtained. e^{ε_i} follows a Gamma distribution with $E[e^{\varepsilon_i}] = 1$ and $Var[e^{\varepsilon_i}] = {}^1/_\phi$. The term ϕ is defined as the inverse dispersion parameter of NB model. The α = ${}^1/_\phi$ is defined as the dispersion parameter of the over-dispersion and is considered constant. The mean and variance of collision counts are respectively as the following:

$$E(y_i|\mu_i, \alpha) = \mu_i \qquad 3\text{-}3$$

$$Var(y_i|\mu_i, \alpha) = \mu_i + \alpha\mu_i^2 \qquad 3\text{-}4$$

And The Negative Binomial regression can be formulated as follows:

$$\mu_i = F_{1i}^{\beta_1}.F_{2i}^{\beta_2}.\exp(\beta_0 + \beta_3 X_{3i} + \cdots + \beta_k X_{ki}) \qquad \text{or}$$

$$\mu_i = exp(\beta_0 + \beta_1 ln(F_{1i}) + \beta_2 ln(F_{2i}) + \beta_3 X_{3i} + \cdots + \beta_k X_{ki}) \qquad 3\text{-}5$$

Where μ_i is the mean number of accidents at site i as a function of site specific variables of X_i and traffic volume of vehicles (F_1) and pedestrians (F_2) in each site i, and β is the vector of regression coefficients estimated in max likelihood process.

The elasticity of variables in the regression models is defined as the percentage change in the response variable due to one percentage change in the models' variables. The value of measuring in percentage terms is that the units of measurement do not play a role in the value of the measurement and thus it allows direct comparison between elasticities. We can write the calculation for elasticity as:

$$E_{X_{ki}}^{\mu_i} = \frac{d\mu_i}{dX_{ki}}\left(\frac{X_{ki}}{\mu_i}\right) \qquad 3\text{-}6$$

Considering the equation 3-5 and 3-6, the elasticity of variables, for example F_1 and X_3 are equal to β_1, and $\beta_3 X_3$ where β_k is the regression coefficients and X_k is the mean of each variable. Therefore, it is possible to calculate the elasticity of variables whether they are traffic count variables (log format) or they are site specific variables (integer and float format).

3.2.2 Model performance and evaluation metrics

In the Negative Binomial (NB) model, the variables' coefficients are estimated using the maximum likelihood procedure. There are three steps to this procedure. Estimates from each step are used as initial values for the next parameter estimate iteration. First, a standard Poisson model is applied to the data. Second, the maximum likelihood is estimated for the mean and dispersion parameter of the response variable, using the constant-only model. The dispersion parameter is plugged in as the starting value for the dispersion parameter for the next iterations. Third, using the starting values obtained in the previous steps, the negative binomial model iterates until the algorithm converges. The final log-likelihood of the applied model is used in the calculation of the AIC measure, which will be discussed shortly.

The NB model estimates the variables' coefficients. The coefficients could be interpreted as follows: for a one-unit change in the predictor variable, the difference in the logs of expected counts of the response variable is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant. The intercept term is the negative binomial regression estimate when all variables in the model are evaluated at zero. The model shows the estimate of the dispersion parameter. If the dispersion parameter equals zero, the model reduces to the simpler Poisson model. If the dispersion parameter is significantly greater than zero, then the data is over dispersed and the negative binomial model outperforms the Poisson model.

As mentioned before, safety performance functions (collision count models) can be used for different applications, including screening variables, the sensitivity of variables, estimation and prediction, identification of relations and causal relationships. In this thesis, SPFs are developed mainly to conduct association analysis between the traffic volume and built-environment covariates and collision counts to figure out the relation between these variables. Thus, statistics metrics, namely, hypothesis testing (p-value), z-test statistics, and Akaike Information Criterion (AIC) are used to evaluate the models. The Figure 3-1 shows an example of NB model estimated in R programing language using the MASS library (Venables and Ripley 2002). The evaluation metrics and parameters are discussed below. Appendix II is provided for further investigation on how the table of coefficients and models' performance metrics were generated for NB in RStudio script.

```
Call:
MASS::glm.nb(formula = Count_4 ~ ., data = db, link = "log",
    init.theta = 1.396278975)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1516  -1.0886  -0.4917   0.3832   4.6800

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.003799   0.358953 -11.154  < 2e-16 ***
Ln_V_TrfSum30  0.109538   0.033775   3.243  0.00118 **
Ln_P_VolSum30  0.343803   0.023349  14.725  < 2e-16 ***
Avg_Road_W     0.073489   0.012541   5.860 4.64e-09 ***
SIgnalzed      0.615094   0.076541   8.036 9.27e-16 ***
Ave_SideWa    -0.061691   0.015481  -3.985 6.75e-05 ***
Green_Med_    -0.150158   0.079089  -1.899  0.05762 .
Sum_grad_a     0.019722   0.008148   2.421  0.01550 *
Avg_Num_Di     0.326072   0.110342   2.955  0.00313 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(1.3963) family taken to be 1)

    Null deviance: 2606.7  on 1708  degrees of freedom
Residual deviance: 1848.7  on 1700  degrees of freedom
AIC: 6142.8

Number of Fisher Scoring iterations: 1


              Theta:  1.3963
          Std. Err.:  0.0997

 2 x log-likelihood:  -6122.8270
[1] "Performance Metrics of  NB-SPF0"
 Degree of Freedom:        1699
 Pearson Chi Square:       2319.9342
 Over Dispersion Factor: 1.3655
 Log Likelihood:           -3061.4134
 AIC:                      6142.8268
```

Figure 3-1 An Example of Negative Binomial output in R Software

In figure 3-1, the columns z and P>|z| are the z-test and p-value, respectively. These are the measurement for the test on the null hypothesis which is if an individual predictor’s regression coefficient is zero, given that the rest of the predictors are in the model. The test statistic z is the ratio of the coefficient to the standard error of the respective predictor. The z value follows a standard normal distribution which is used to test against a two-sided alternative hypothesis that the coefficient is not equal to zero. The probability that a particular z test statistic is as extreme as, or more so than what has been observed under the null hypothesis is defined by P>|z|.

Hypothesis testing determines the truth of a hypothesis using statistics based on the sample data. There are four steps in the hypothesis testing test. First, null and alternative hypotheses are explained in a mutually exclusive manner, which means, if the null is false, then the alternative is true or vice versa. In the chapter 5 analysis, the null hypothesis (H_0) is that a single variable is no different than its mean, and the alternative hypothesis (H_a) is the variation of the variable is different that its mean. Secondly, a confidence level should be considered which many studies select a confidence level of 95% leading to a significance level of 0.05. Third, the two-sided test method is selected and the probability (P-value) that provides sufficient evidence whether to accept or reject the null hypothesis is considered. Finally, the P-value is compared with the selected threshold for the statistical significance of the hypothesis testing. If the P-value is greater than $1 - c,$ confidence level, then the null hypothesis cannot be rejected. However, when the P-value is less than the threshold, there is evidence to reject the null hypothesis and consider the alternative hypothesis as statistically significant. When the P-value becomes smaller, there is stronger evidence against the null hypothesis. As can be seen in figure 3-1, there are codes like three stars or two stars, respectively for p-values equal to zero or less than 0.001, which in our analysis, p-values less than 0.05 is regarded as statistically significant.

Akaike information criterion (AIC) is presented in figure 3-1. When several non-nested models are calibrated based on the same data using maximum likelihood, we can compare their performance using the Akaike Information Criterion (Cameron and Trivedi 1998). The Akaike information criterion (AIC) is one of the most commonly used measures for the comparison of non-nested models based on maximum likelihood. AIC is defined as the following formula, where k is the number of parameters in the model, and L is the maximum value of the likelihood function for the model. The regression model with the lowest AIC is preferred.

$$AIC = -2ln(L) + 2k \qquad 3\text{-}7$$

3.3 Hotspot Identification

3.3.1 Empirical Bayes Method

Consider Y_i as the count of crashes at entity i, which can be estimated with the probability density function of $f(Y_i|\theta_i)$. In the Negative Binomial model, the mean crash count θ_i is assumed fixed and it is estimated through the Maximum Likelihood method. However, the Bayesian approach assumes a prior distribution to θ_i which enables making inferences on the variable crash means (Lord, Qin, and Geedipally 2021). Therefore, Bayesian approaches can provide posterior information on the mean count of collisions (θ_i) as the following Bayes' theorem:

$$p(\theta_i|Y_i,\eta) = \frac{f(Y_i|\theta_i) \,.\, \pi(\theta_i|\eta)}{m(Y_i|\eta)} \qquad 3\text{-}8$$

Where, $f(Y_i|\theta_i)$ is the likelihood of accident count at the entity, $\pi(\theta_i|\eta)$ is the prior information on the parameter θ_i, where η is the vector of prior parameters. $m(Y_i|\eta)$ is referred to as the marginal distribution of Y_i, which is constant among all sites, hence the Bayes equation can be simplified as:

$$p(\theta_i|Y_i,\eta) \sim f(Y_i|\theta_i) \,.\, \pi(\theta_i|\eta)$$

Therefore, performing the Bayes approach requires two primary steps. First, select a prior distribution, $\pi(.\,|\eta)$, second, regular Bayes update with the historical accident data to achieve the posterior distribution of crash counts. Selection of prior distribution can be challenging, however, commonly conjugate priors are selected the context of road safety. In Bayesian theory, when the posterior distribution is in the same distribution family as the prior distribution, the prior and posterior are conjugate distributions, and the prior is called a conjugate prior for the likelihood function. For example, when the likelihood function is a Poisson-Gamma distribution (NB), and the prior is Gamma, the posterior follows the same distribution as the prior, Gamma distribution, in which NB and Gamma distribution are called conjugate models. Therefore, the formulation could be as the fowling:

$$Y_i\,|\mu_i, e^{\varepsilon} \sim Poisson\,(\mu_i\, e^{\varepsilon}),$$

$$e^{\varepsilon} \sim Gamma(\phi, \phi)$$

We know that when $X \sim Gamma(\phi, \phi)$, then kX~Gamma($\phi, \frac{\phi}{k}$), therefore:

$$Y_i\,|\,\theta_i \sim Poisson\,(\theta_i),$$

$$Y_i\,|\,\theta_i \sim \text{Gamma}(\phi, \frac{\phi}{\mu_i})$$

Now, applying Bayes Theorem:

$$p(\theta_i|y_i) \sim f(Y_i|\theta_i)\,.\,\pi(\theta_i)$$

$$\sim \frac{e^{-\theta_i}.\,\theta_i^{\,y_i}}{y_i!}\,.\frac{\left(\phi/_{\mu_i}\right)^{\phi}}{\Gamma(\phi)}.\,\theta_i^{\,\phi-1}e^{-\left(\frac{\phi}{t\mu_i}\right)\theta_i}$$

$$\sim e^{\theta_i^{\,[1+\phi/_{\mu_i}]}}.\,\theta_i^{\,[y_i+\phi-1]}$$

Knowing that the posterior is Gamma distribution (conjugate prior):

$$\theta_i \mid y_i \sim Gamma\ \left(y_i+\phi, 1+{}^{\phi}/_{\mu_i}\right) \quad 3\text{-}9$$

Where the posterior mean follows the Gamma distribution with the shape parameter of $y_i + \phi$ and scale parameter $1 + {}^{\phi}/_{\mu_i}$. In fact, the Empirical Bayes approach implicitly uses the data twice, once in providing the prior distribution in which the μ_i and ϕ are the parameters of the Negative Binomial model, and EB again uses the number of the observed crash (y_i) to make an inference on the posterior mean of collision (Lord, Qin, and Geedipally 2021). Commonly, $w_i = \frac{\phi}{(\mu_i+\phi)}$ and, $1 - w_i = \frac{\mu_i}{(\mu_i+\phi)}$ are defined, where the posterior mean of θ_i is as:

$$E(\theta_i|y_i) = \frac{\mu_i(y_i+\phi)}{\mu_i+\phi} = \frac{\mu_i y_i}{\mu_i+\phi} + \frac{\mu_i\phi}{\mu_i+\phi} = (1-w_i)y_i + w_i\mu_i \quad 3\text{-}10$$

3.3.2 Getis Ord (G^*) Spatial Hotspot Identification

Getis and Ord defined a family of G statistics is applied to investigate spatial patterns (Songchitruksa and Zeng 2010a; Ord and Getis 1995; Getis and Ord 1992). The G_i^* (pronounced

G-i-star) index is highly useful as it could identify hotspots as points, lines or regions on a global scale and makes clusters of high or low concentration among observations.

Consider that there are 1,2,3,…,i identified spatial elements, that a random variable of incidents (X), i.e., pedestrian crashes count or public transit accessibility index, is associated to each spatial element i as x_i. The goal is to figure out the existence of a spatial pattern among spatial elements i which are weighted by the random variable x_i (explanatory feature in each of the regions). It could be asserted that there is a spatial autocorrelation of variable X over the region i when x_i exhibit similarities between contiguous elements. The following G_i^* equation could be mentioned:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij}(d)\, x_j}{\sum_{j=1}^{n} x_j} \qquad \text{3-11}$$

Where G_i^* is the statistic that describes the spatial dependency of incident X in region i, n is the number of regions, x_j is a random variable X in the entity j (j may equal i), and w_{ij} is the weight value between region i and j that represents their spatial interrelationship. w_{ij} is usually associated with the conceptualized spatial relationship and is dependent on the distance d. The value of d could be defined as a user-specific threshold to study the incident in the threshold distance. In this study, two thresholds are used. The conceptualized spatial relationship based on zone of indifference in which features within the specified critical distance (distance band or threshold distance) of a target feature receives a weight of one and influence computations for that feature. Once the critical distance is exceeded, weights (and the influence a neighboring feature has on target feature computations) diminish with distance. Moreover, the second threshold is the conceptualized spatial relationship based on Contiguity Edges Corners Polygon features that share a boundary, share a node, or overlap will influence computations for the target polygon feature. The w_{ij} factor could be defined as a non-binary value as well. The sum of weights for each region i is defined as equation 3-11 and the expectation (E) of G_i^* and the variance of G_i^* is defined as equation 3-12 and 3-13 respectively.

$$W_i = \sum_{j=1}^{n} w_{ij}\, x_j \qquad \textit{3-12}$$

$$E(G_i^*) = \frac{W_i}{n} \qquad \textit{3-13}$$

$$Var(G_i^*) = \frac{s^2}{\bar{x}} = \frac{W_i(n-W_i)}{n-1} \ \ where\ s^2 = \frac{\sum_{j=1}^{n} x_j^2}{n} - \bar{x}^2 \ \ and\ \bar{x} = \frac{\sum_{j=1}^{n} x_j}{n} \qquad \textit{3-14}$$

When the distribution of variable X is normal the G_i^* distribution is similarly normal, however, if the underlying distribution is nonnormal (e.g. heavily skewed), the G_i^* statistic becomes nonnormal correspondingly. In such cases, an increase in the number of spatial units in the cluster analyzed will help the distribution of the G_i^* statistic approach normality. One common method is to raise the value of d to include more x_j (Songchitruksa and Zeng 2010b).

Under the exact or asymptotical normal conditions, G_i^* is usually standardized based on its sample mean and variance:

$$Z(G_i^*) = \frac{\sum_{j=1}^{n} w_{ij}\, x_j - \bar{x}\sum_{j=1}^{n} w_{ij}^2}{s \ * \sqrt{\dfrac{n\sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij}\right)^2}{n-1}}} \qquad 3\text{-}15$$

The standardized G_i^* is essentially a Z score and can be associated to being statistical significance. A close-to-zero G_i^* implies random distribution of the observations (incidents), however, negative and positive G_i^* statistics with high absolute values implies for low and high statistically significant of observations (incidents) occurrence.

Chapter 4: Pedestrian-Vehicle Collisions and Public Transit

Abstract: Public transit is an underlying component of sustainable urban transportation. Urban residents walk or cycle to reach public transit for their daily trips. PTAPs are primary elements of public transit where ridership interacts with vehicle traffic flow and could result in road conflicts and collisions. Examining the magnitude of pedestrian vehicle collision risk at public transit locations would show whether there is a potential pedestrian vehicle collision risk at these locations. It would also show if there is a need to conduct further safety analysis focusing on these locations. Chapter 4 studies the potential relationship between pedestrian vehicle collision counts and public transit exposure and attempts to figure out if there is a spatial relation between pedestrian vehicle collisions and public transit service locations. The analysis results in this chapter prove that there is a higher risk of pedestrian collision at locations where there are public transit services. Moreover, pedestrian collision risk increases when public transit exposure characteristics, such as bus daily traffic volume, public transit accessibility index, and the number of transit routes increase. Therefore, it is proposed to conduct further safety analysis to study pedestrian vehicle collisions at PTAPs.

4.1 Introduction

Public transportation a dominant part of a sustainable urban transportation. It enables urban dwellers to commute to their work, education institutes, or recreational centres through a sustainable form of transportation. Transit users primarily walk or cycle to reach the public transportation access points, i.e., transit stops. Hence, public transit promotes walking and cycling which in turn leads to sustainable transportation. Public transit stops are typically located in proximity to urban spots with a high concentration of urban activities and consequently a higher pedestrian traffic volume. Transit users may interact with vehicle traffic flow, particularly in the proximity of public transit stops, which could result in pedestrian-vehicle collisions. Therefore, it is hypothesized that, there would be a higher pedestrian vehicle collision risk where there are public transit services.

4.2 Objective

Chapter 4 research objective is to establish a robust examination on the possible relation between the pedestrian-vehicle collision counts and public transit services. This study incorporates spatial investigations with polygon levels, intersection levels, and public transit sites and adds statistical analysis on public transit exposure characteristics. Public transit exposure features include the presence of public transit stops, bus traffic volume, the number of transit routes, and the public transit accessibility index. The finding of this research task emphasizes the need for further consideration on pedestrian safety at public transit locations.

4.3 Methodology

To address the research objective of this chapter, two methodologies are applied in the analysis. G* hotspot analysis is primarily used to identify spatial hotspots within a zonal level. Moreover, Negative Binomial statistical analysis is utilized to find whether there is a statistical association between public transit exposure variables and the counts of pedestrian-vehicle collisions. In the following sections, the two research methodologies applied in this chapter will be reviewed. Please refer to chapter 3, Methodology, for further details and explanations.

4.3.1 Spatial Hotspot Analysis, Getis-Ord Method

Getis-Ord spatial hotspot identification method, known as G* (G-star) hotspot analysis, returns hotspots where features are either high or low value clusters spatially. It considers each feature within the context of neighboring features, and therefore, the G* returns the spatial hotspot of the feature considering the spatial relation with the neighborhood.

Consider that there are 1,2,3,…,i identified spatial elements, such as the urban region, dissemination area units, or defined influential zone of intersections or public transit access points. There is a random variable of incident (X), is associated to each spatial element i as x_i. In this chapter, you will find pedestrian crash counts and public transit accessibility index (PTAI) as the features under the focus of the study. The existence of a spatial pattern among spatial elements i needs to be investigated given the value of the random variable x_i. It could be asserted that there is a spatial autocorrelation of pedestrian collision counts or the PTAI over the region i when the variable exhibits similarities between contiguous elements. The G_i^* metric can be defied as the formula 3-1 explained in the chapter 3.

To determine the neighbor of the spatial elements and the corresponding spatial weights, two kinds of thresholds are utilized, namely, the conceptualized spatial relationship based on the zone of indifference and the conceptualized spatial relationship based on Contiguity Edges Corners. The former considers a specified critical distance (distance band or threshold distance) and gives a weight of one to the features in the threshold and once the critical distance is exceeded, weights diminish with distance. The latter, the conceptualized spatial relationship based on Contiguity Edges Corners Polygon considers the spatial elements that share a boundary, or overlap with the target spatial element. The spatial analysis in figures 4-4 and 4-5 are utilizing the above methods.

4.3.2 Negative Binomial regression

Negative Binomial model is one the most applied statistical models in safety studies which performs well both in practical and scientific objectives (Lord, Qin, and Geedipally 2021). In this chapter, the Negative Binomial model is used to build up the statistical analysis on the counts of pedestrian-vehicle collisions at the intersection level, taking into account public transit exposure variables. Hence, the Negative Binomial establishes the possible association between the higher pedestrian road risk at the locations with public transit exposure. There could be a formulation as the following:

$$\mu_i = exp(\beta_0 + \beta_1 ln(F_{1i}) + \beta_2 ln(F_{2i}) + \beta_2 ln(F_{3i}) + \beta_4 X_{4i} + \cdots + \beta_k X_{ki}) \quad 4\text{-}1$$

Where μ_i is the mean number of pedestrian vehicle collisions at intersection i as a function of site specific variables which includes public transit exposure variables including PTAI, number of bus routes, and presence of PT stop nearby the intersection. The logarithm of pedestrian traffic volume (F_1) and vehicle traffic volume (F_2) and bus traffic volume (F_3) in each site i, and β is the vector of regression coefficients estimated in the max likelihood process.

The Negative Binomial distribution takes into account an additional parameter to the standard Poisson distribution, namely, over dispersion parameter, to address the over dispersion in collision data. This leads to the privilege of Negative Binomial to standard poison model, as collision data is typically over dispersed. The formulation of the Negative Binomial could be reviewed as the following:

$$Y_i \mid \theta_i \sim Poisson\ (\mu_i \, . \, e^{\varepsilon_i}) \quad \text{4-2}$$

$$e^{\varepsilon_i} \mid \phi \sim Gamma\ (\phi, \phi) \quad \text{4-3}$$

where $Gamma()$ denotes a Gamma probability density function with parameters φ >0 and δ >0, which ensures $\theta_i > 0$ and e^{ε_i}>0, and e^{ε_i} follows a Gamma distribution with $E[e^{\varepsilon_i}] = 1$ and $Var[e^{\varepsilon_i}] = {}^{1}/_{\phi}$. The term ϕ is defined as the inverse dispersion parameter of NB model. The α = ${}^{1}/_{\phi}$ is defined as dispersion parameter of over-dispersion parameter and is considered constant. Numeric optimizations could likelihood function and estimate the parameters of each variable and the possible association between the response, pedestrian vehicle collision count, and the dependent variables public transit exposure variables.

4.4 Data Description

This case study utilizes data from various online open sources datasets from the island of Montreal. The data sources are available online and the links are provided in the references. The following sections provide a detailed description of the data collection procedure of this research.

4.4.1 Sites

There are two spatial levels in this chapter: (1) spatial analysis on the zonal level in which Dissemination Area Units (DAUs) are incorporated for the analysis. DAU is the smallest Canada census area, and there are 3200 dissemination area units (DAU) in the city of Montreal. (2) an intersection level analysis is utilized to address the research question in this chapter. Figure 4-1 demonstrates the Montreal base map with DAUs and the 1850 intersections where traffic data is available for these sites.

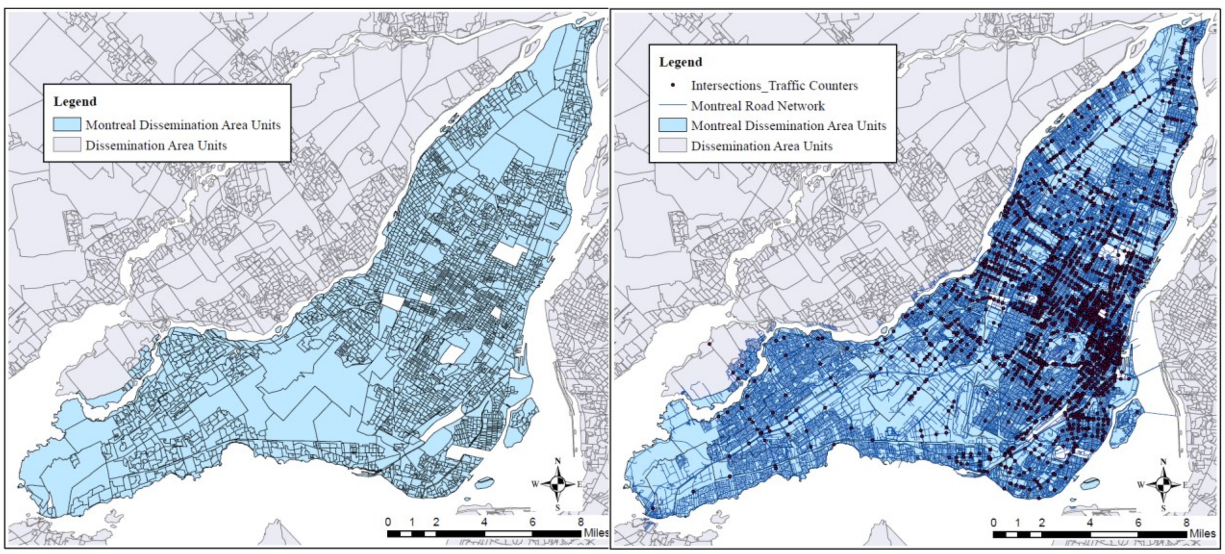

Figure 4-1Montreal Dissemination Area Units, Road Network, and Intersection with traffic counters. Maps were created using ArcGIS® software by Esri ("GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri" 2022).

4.4.2 Traffic Count

Vehicle daily traffic count, bus daily traffic count, and pedestrian daily traffic volume are available online for 1850 intersections in the city of Montreal ("Traffic Lights — Vehicle and Pedestrian Counts at Intersections with Lights - Open Government Portal" 2021a). Figure 4-1 (right-hand side) shows the location of the intersections that have been equipped with traffic count facilities.

4.4.3 Count of collisions

Montreal Road Collision Dataset is available online("Road Collisions - Open Government Portal" 2021). It provides the location and type of collisions including pedestrian-vehicle collisions. Figure 4-2 shows 11,120 pedestrian-vehicle collisions between 2012 and 2019 in the city of Montreal.

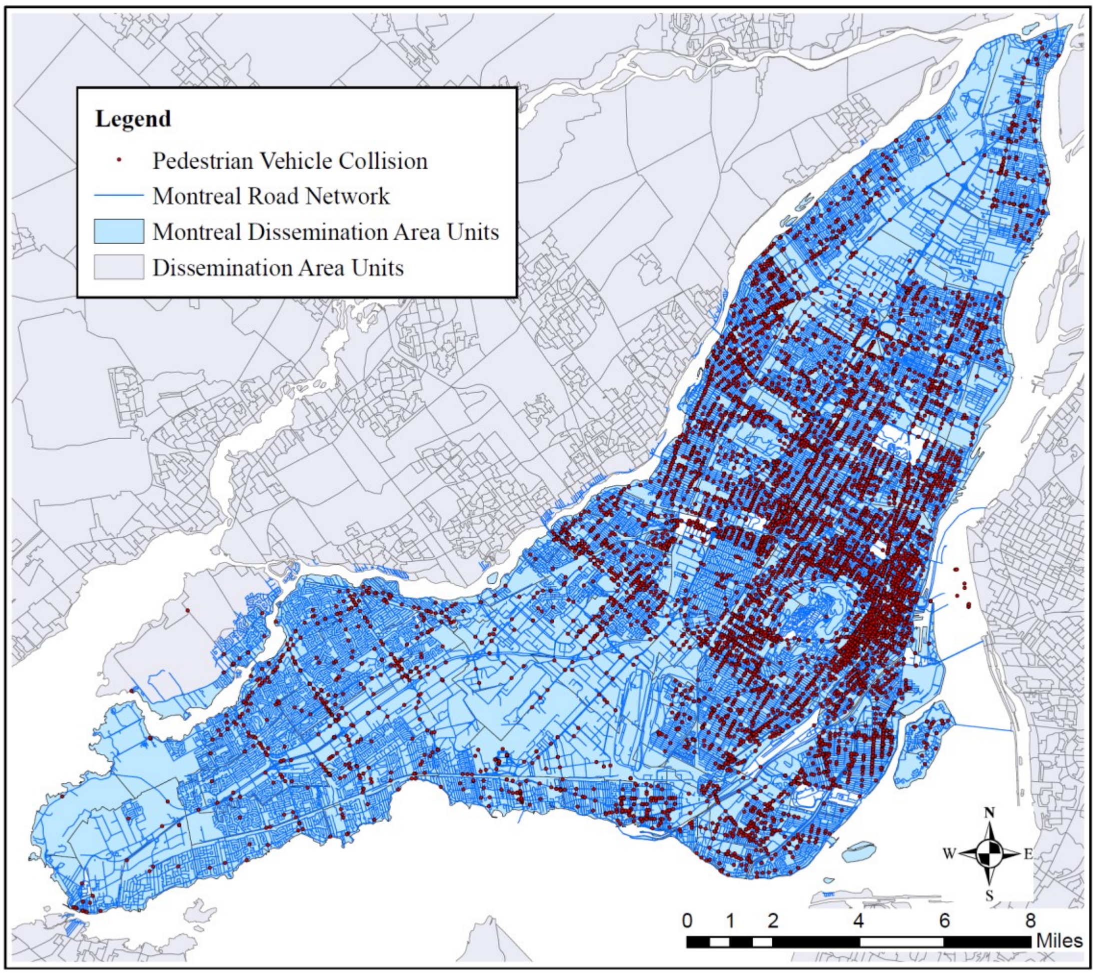

Figure 4-2Pedestrian Vehicle Collisions in the City of Montreal. Maps were created using ArcGIS® software by Esri ("GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri" 2022).

4.4.4

4.4.5 Public Transit

Montreal public transit network datasets provide the public transit links and transit stop locations in the city of Montreal. There are 8950 stops through 220 transit service routes. Figure 4-3 shows the Montreal public transit network.

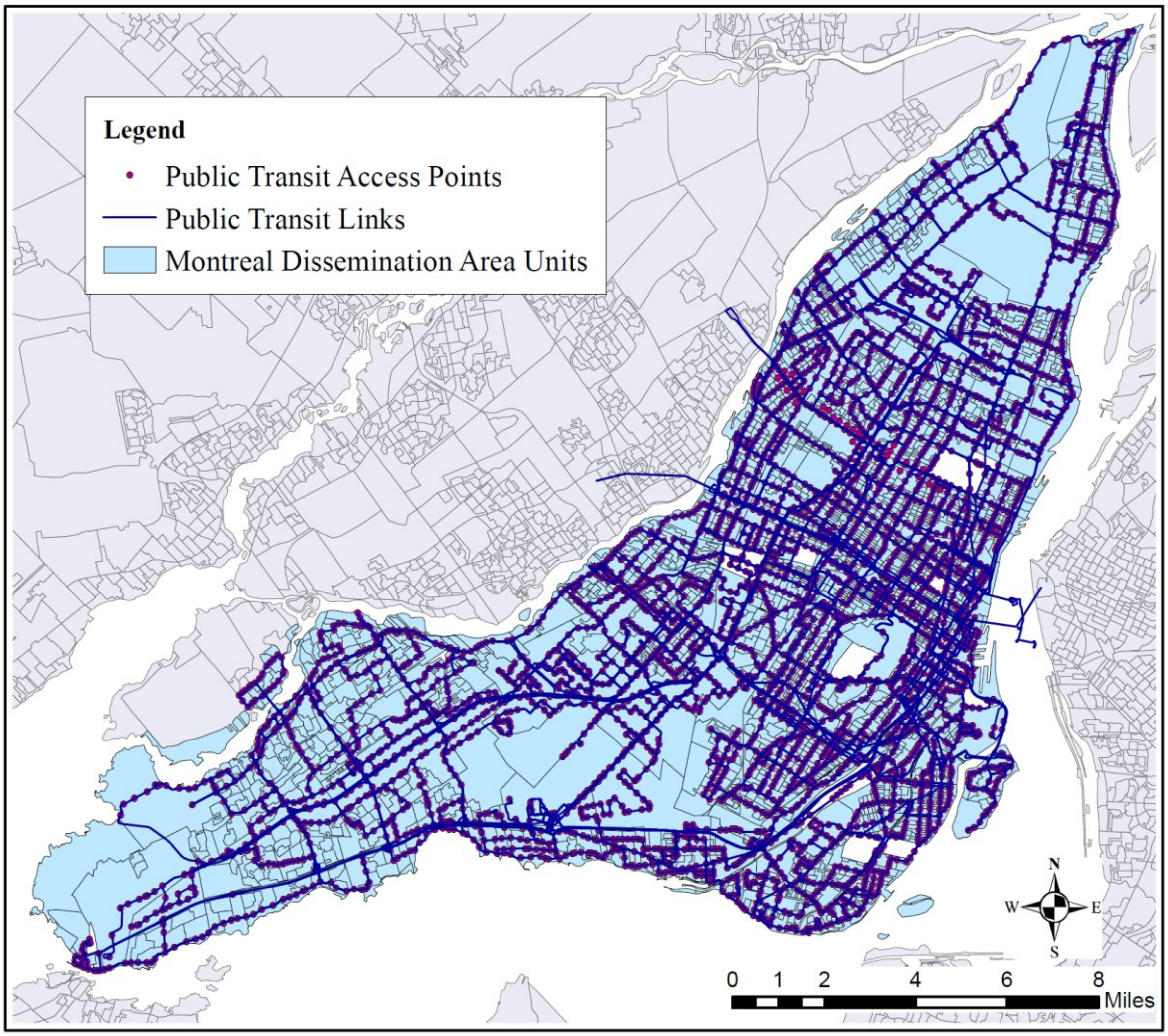

Figure 4-3Montreal Public Transit Network. Maps were created using ArcGIS® software by Esri ("GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri" 2022).

Public Transit Accessibility Index (PTAI) is used to quantify the level of accessibility of public transit services in a target region. PTAI is a measure of connectivity showing how well the zone is connected to the public transit network. It takes into account the walking distance to the nearest station, waiting time at the transit stop, the frequency of services and the presence of rail transit, and returns access index as a quantitative measure (Transport for London 2014). The first step is to calculate walking time to PTAPs. Next, the scheduled waiting time for each access points is calculated and average waiting time for each route will be calculated for each access point. A reliability factor of two minutes and 0.75 minutes is used respectively for buses and metro stations. The total access time at each access point is the summation of walking time and the average waiting time at the access point. Finally, the equivalent doorstep frequency is calculated at each PTAP which could be aggregated for a neighborhood. The calculation process is done in ArcGIS software by defining a central point as the representative of the polygon (DAUs), and then following the above steps to calculate the PTAI for each polygons. Also, intersections located inside the polygons get the PTAI of their region.

Aggregating the mentioned datasets provides us with the dataset addressing the research objective of this chapter. Table 4-1 provides the descriptive statistics of the variables of the dataset used through the analysis in this chapter.

Table 4-1. Descriptive Statistics of Variables in the Analysis of the relation between public transit and Pedestrian collisions.

Public Transit Presence and Accessibility	**Type**	**Mean**	**SD**	**Min**	**Max**
Counts of Collisions in Proximity of Intersections	Integer	2.27	0.07	0	25
Vehicle Daily Traffic Logarithm	Float	9.39	0.02	5.7	11.25
Pedestrians Volume Logarithm	Float	6.86	0.04	0.0	10.95
Bus Daily Traffic Logarithm	Float	4.86	0.03	0.0	7.28

Public Transit Presence and Accessibility	Type	Mean	SD	Min	Max
Public Transit Accessibility Level at the Intersection	Float	10.4	0.29	0.0	7.42
Presence of Public Transit Access Points Within 30 meters of the Intersection, (Yes or No)	Binary	0.62	0.01	0.0	1.00
Number of Bus Routes at the Intersection	Integer	2.09	0.06	0	12

4.5 Analysis and Results

The analysis of this chapter follows two approaches: firstly, the spatial comparison and the spatial analysis on pedestrian-vehicle collisions and PTAI. Secondly, a statistical approach is incorporated to cover the previous analysis. Getis-Ord G* spatial hot spot analysis is mainly used to study the spatial relationship between pedestrian-vehicle collision occurrence and PTAI. Moreover, the Negative Binomial model is utilized to study a statistical approach and cover the previous spatial hotspot analysis regarding the relation between pedestrian collisions and public transit services locations.

In the city of Montreal, there are almost 3200 dissemination area units (DAUs), which are the smallest standard geographic unit in Canada. Figure 4-3(left side) shows the distribution of the PTAI through DAUs. The index is ranged between 0 and 70 and the darker in color the DAU is, the higher PTAI it has. Besides, Figure 4-3 (right side) demonstrates the Pedestrian-Vehicle Collision Counts (2012-2019) per area unit (km^2).

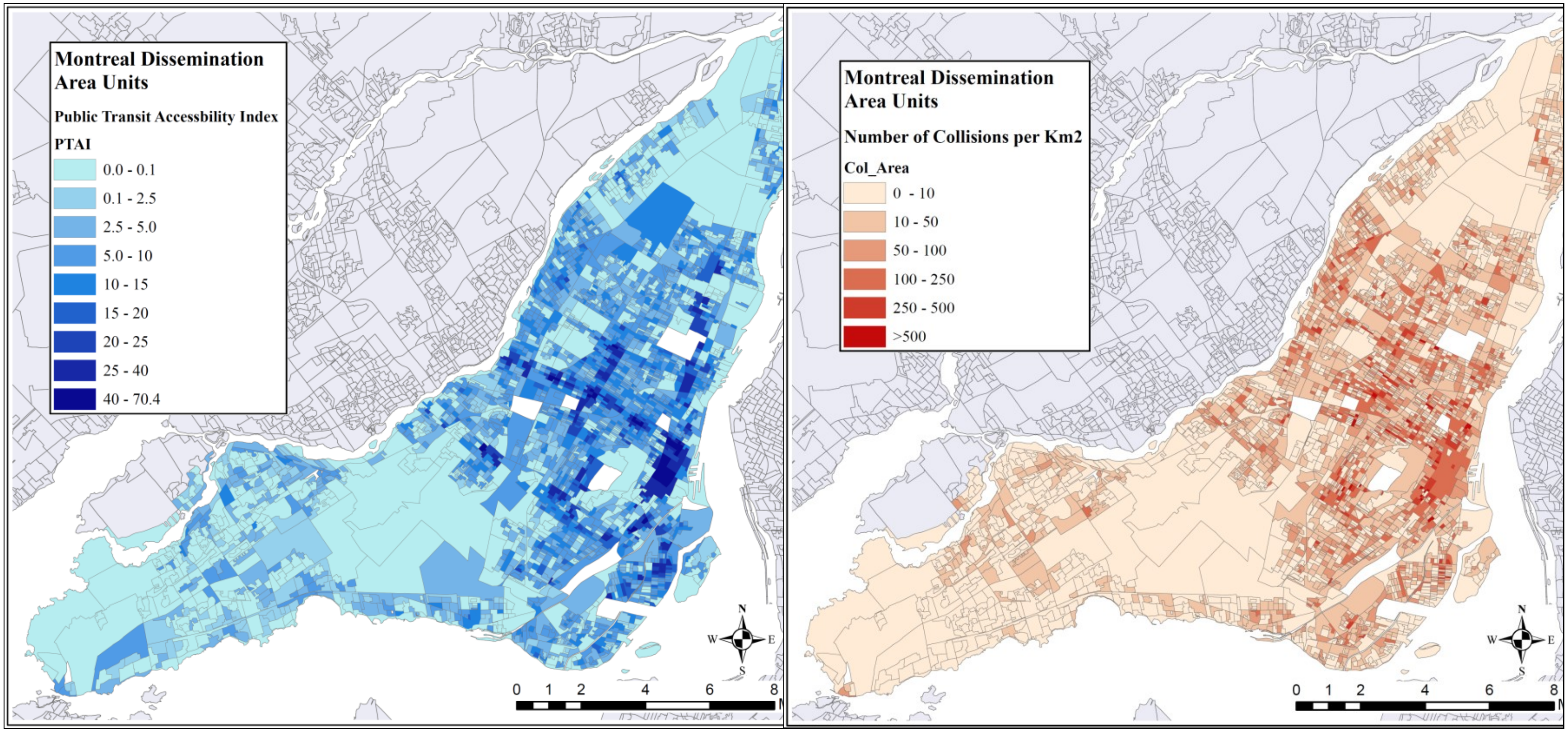

Figure 4-4Public Transit Accessibility level and Collision Counts per Area in the Montreal Dissemination Area Units. Maps were created using ArcGIS® software by Esri (“GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri” 2022)

In the following, Getis-Ord (G*) spatial hotspot analysis is used for analysis on DAUs (polygons) and PTAPs (points). Figure 4-5 shows the hotspot analysis on the dissemination area unit to find DAUs which are spatially statistical significant concerning PTAI (left side) and collision counts (right side). Moreover, Figure 4-6 shows G* applied on the PTAPs pedestrians-vehicles collision counts with the zone of indifference criterion. The dark red crosses (⌖) show PTAPs that are statistically significant hotspots with regard to the number of collisions.

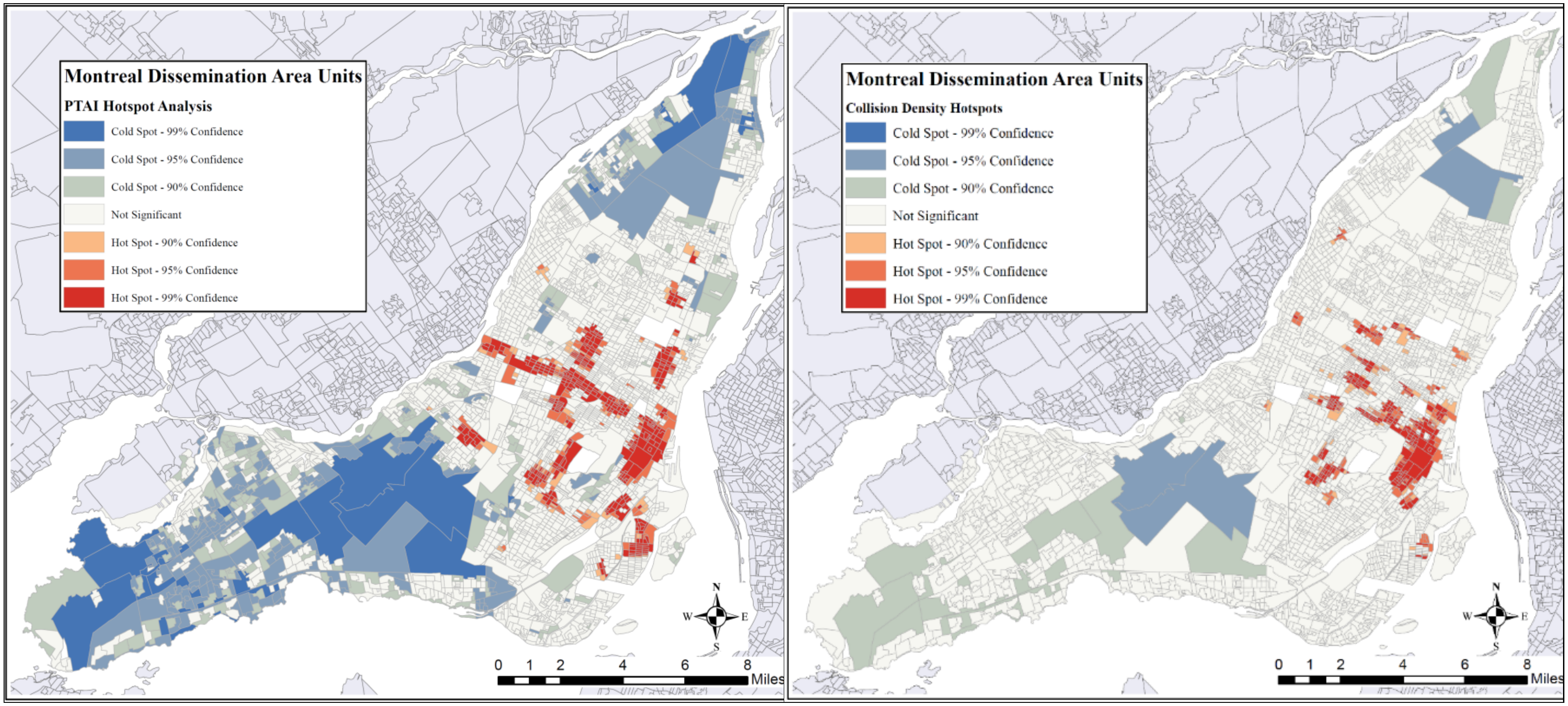

Figure 4-5 Hotspot Analysis on Public Transit Accessibility level and Collision Counts per Area in the Montreal Dissemination Area Units, hotspot analysis is based on Contiguity Edges Corners. Maps were created using ArcGIS® software by Esri ("GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri" 2022).

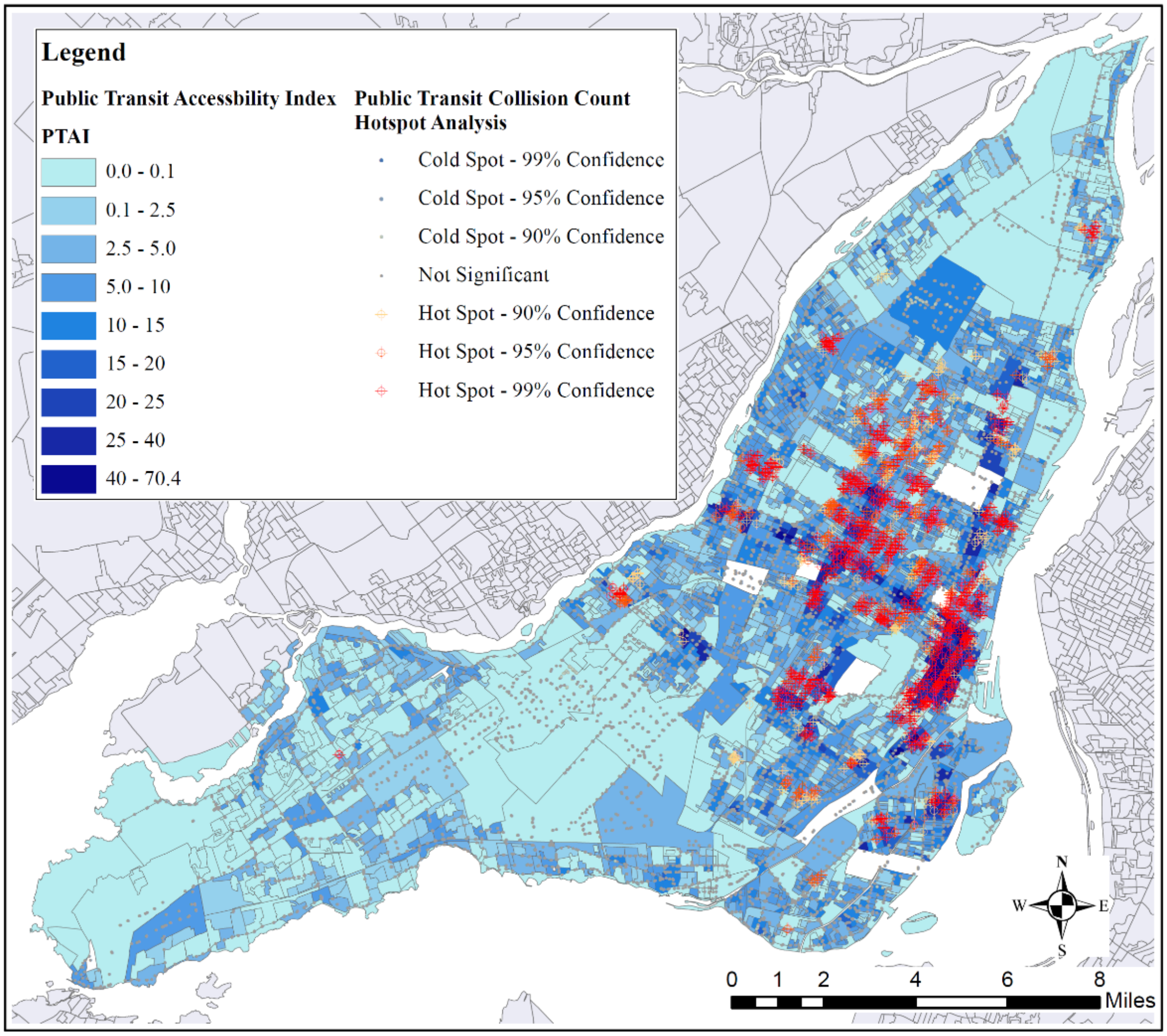

Figure 4-6Public Transit Access Points with statistically significant high collision counts are located inside Dissemination Area Units with high PTAI, hotspot analysis is based on zone of indifference criterion. Maps were created using ArcGIS® software by Esri ("GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri" 2022).

Table 2 is the Negative Binomial model on 1638 intersections of the city of Montreal. Among the 1638 intersections, 1018 have at least one PTAP in proximity (30 meters). The model found the logarithm of daily bus traffic count at intersections, the presence of PTAP at the intersection, the PTAI at the intersection, and the number of bus routes as statistically significant variables with a positive coefficient. In the following section, the discussion on the analysis is provided.

Table 4-2 Negative Binomial Model, Statistical Analysis on intersections to study the relation between Public Transit Access Points and Pedestrian Collision Counts, Number of observations = 1637 intersections, Log-likelihood= -2900.3

Collision Count	Coef	Std. Err.	z	P>z
Vehicle Daily Traffic Logarithm	0.55	0.04	12.850	0.000
Pedestrians Volume Logarithm	0.35	0.02	15.880	0.000
Bus Daily Traffic Logarithm	0.07	0.03	2.400	0.016
Public Transit Accessibility Level at the Intersection	0.01	0.002	2.960	0.003
Presence of PublicTransit Access Points Within 30 meters of Intersectoin, (Yes or No)	0.20	0.08	2.650	0.008
Number of Bus Routes at the Intersection	0.05	0.01	3.610	0.000
Constant	-7.36	0.37	-19.720	0.000

4.6 Discussion

An initial investigation on PTAI and collision counts on DAUs, demonstrated in figure 4-4 (left side) and figure 4-4 (right side), shows the substantial similarities between the distribution pattern of PTAI and collision counts per area unit. This is accompanied by figure 4-5 (left side) and figure 4-5 (right side), which demonstrate similar patterns through G* spatial hotspot analysis. Figure 4-6 shows the collision count hotspots of PTAPs located within the DAUs with statistically significant PTAI (dark blue regions). Aligned with the previous analysis, Table 2 shows more pedestrian collisions occur at intersections with a higher volume and number of bus routes (which together contribute towards higher PTAI) and with the presence of a PTAP within 30 meters of the intersection. The abovementioned analysis supports the hypothesis that PTAPs as a daily origin-destination of pedestrians suffers from higher counts of pedestrian collisions. Providing a safe environment for pedestrians in the proximity of PTAPs implies the necessity of further analysis to enhance the safety index of the pedestrians in these areas. Therefore, urban transportation planners should take into account pedestrian safety at the current PTAP locations and when developing public transit.

4.7 Conclusions

Chapter 4 argued the potential relationship between the increase in pedestrian collision risk at the locations with public transit services. The spatial investigation and statistical analysis revealed the significance of pedestrian-vehicle collisions at locations where there are public transit services. The statistical approach in this chapter showed an association between pedestrian-vehicle collision counts and public transit exposure characteristics such as bus average daily traffic, public transit accessibility index, and the number of public transit routes. Pedestrians are at higher collision risk at intersections where there is PTAP rather than the intersections with no PTAPs. Hence, it is needed to conduct further research to examine pedestrian road safety at PTAPs. The next chapter analyses and discusses the role of road geometry and built environment elements on pedestrian-vehicle collision counts at PTAPs and proposes applicable countermeasures to address this problem.

Chapter 5: Road Geometry, and Built-Environment Effect on Public Transit Pedestrian Safety

Abstract: Public Transit Access Points (PTAP) are the origin and destination of daily trips of urban residents in urban regions. Public transit promotes sustainable transportation and it is necessary to provide safe transit access points for transit users as they interact with vehicle traffic at PTAPs. Therefore, it is needed to study pedestrian road safety at these locations. The role of engineering elements, namely, traffic calming, road geometry and built-environment are studied in this chapter. It was found that shows that traffic calming strategies such as road width reduction, sidewalk width increase, median refuges, separate pedestrian crossing (walk) intervals, and vehicle stop signs could improve pedestrian safety of PTAP. Moreover, the statistical models in this chapter found that pedestrians are at more risk in PTAP at locations where high road grades and in proximity to intersections with more two way streets than one way streets. The Empirical Bayes collision hotspot identification method is applied to the PTAPs and it studied how proposed engineering elements would improve pedestrian road safety in public transit hotspots.

5.1 Introduction

As discussed in chapter 4, public transit locations are likely to face a higher risk of pedestrian vehicle collisions. Through two geographical study areas namely, intersection level and polygon level (dissemination area units), pedestrian vehicle collision was found to be associated with public transit exposure and accessibility factors. Hence, pedestrian road safety should be studied at public transit access points.

In recent years, numerous studies investigated the role of road geometry and built environment factors on pedestrian safety. Those studies showed that such engineering factors could influence pedestrian safety levels in the intersection level (Stipancic et al. 2020a) and road segments (T. Chen et al. 2020). Hence, in the analysis performed in this chapter, it is expected that road geometry and built environment factors affect pedestrian road safety at the proximity of public transit access points and reduce the risk of pedestrian vehicle collision in these locations.

This chapter first studies the role of traffic calming, road geometry, and built-environment factors on pedestrian vehicle collisions in the proximity of public transit access points located nearby intersections and at midblock. Furthermore, top ranking pedestrian collision hotspots at public transit locations will be studied to figure out the potential of traffic calming and built environment elements, such as, road narrowing, sidewalk curb extension, road median, and pedestrian crossing facilities at intersections on improving pedestrian safety at the public transit hotspots.

5.2 Objective

Chapter 5 research objective is to incorporate pedestrian collision models on pedestrian vehicle collisions in proximity to public transit locations to figure out the effect of engineering elements on pedestrian safety levels at public transit access points. Engineering elements include continuous road geometry, traffic calming, and built environment factors playing part in pedestrian road safety. The statistical analysis proposed in this chapter suggests countermeasures improving pedestrian safety. To overcome possible bias due to the small sample size (Afghari et al. 2019), this research aims to establish the analysis based on a wide range of PTAPs and provide separate analysis for PTAPs located at midblock and intersections, particularly, at signalized and non-signalized intersections. Moreover, this chapter aims to utilize collision hotspot identification methods to prioritize hazardous sites, identify public transit stops with high potential for improvement. The hotspot identification suggest the sites need to have further investigations and field studies specific for each point. Moreover, the research task addresses the question of whether the findings suggested in the previous analysis could be applicable to improve pedestrian safety levels at the hotspot site.

5.3 Methodology

This chapter builds up statistical safety analyses to assess the effect of road geometry and built-environment elements on the count of pedestrian vehicle collisions in the vicinity of public transit locations. Negative Binomial statistical models are mainly incorporated to conduct the pedestrian safety analysis at these locations. The analysis will follow by Empirical Bayes collision hotspot identification which sorts the sites with higher concentration of pedestrian vehicle collisions. Firstly, Empirical Bayes model is reviewed in the following section and the reader is referred to the methodology in chapters 3 and 4 for the Negative Binomial (NB) model.

5.3.1 Empirical Bayes Method

Estimating the mean expected count of collisions through the conventional NB model, the crash count mean is assumed fix and there is no statistical inference on the parameter, however, applying the Empirical Bayes approach, in addition to the expected mean count of collisions, there could be a statistical inference on the value leading to valuable information on the network screening phase and hotspot identification.

The Empirical Bayes approach assumes having prior knowledge (often called "prior information") of the mean count of collisions, which will be updated by the collision data leading to the posterior value on the crash counts. In other words, thanks to Bayesian statistics, the likelihood function is calculated by the fitting of an NB distribution on the collision data, and is merged with the "prior information" leading to the "posterior information" on the mean crash count. In this study, the prior information on the mean count of collisions is assumed to follow a Gamma distribution. The Gamma function is a conjugate prior to the NB likelihood function $f(Y_i|\theta_i)$ which could result in a Gamma distribution in the posterior probability ($p(\theta_i|Y_i,\eta)$ (for more explanations please refer to the Chapter 3). The following formulation shows the Empirical Bayes framework:

$$p(\theta_i|Y_i,\eta) \sim f(Y_i|\theta_i)\ .\ \pi(\theta_i|\eta) \qquad 5\text{-}1$$

Where, $f(Y_i|\theta_i)$ is the likelihood of the accident count at the entity, $\pi(\theta_i|\eta)$ is the prior information on the accident count θ_i, where η is the vector of prior parameters. Commonly, $w_i = \frac{\phi}{(\mu_i+\phi)}$ and, $1 - w_i = \frac{\mu_i}{(\mu_i+\phi)}$ are defined, to calculate the expected posterior mean of θ_i, $E(\theta_i|y_i)$, as the following:

$$E(\theta_i|y_i) = (1 - w_i)y_i + w_i\mu_i \qquad 5\text{-}2$$

Where μ_i and ϕ are respectfully the estimated collision count mean at the site i and the over dispersion parameter of the Negative Binomial model.

5.4 Data Description

The analysis in this chapter utilizes data from various online open sources and a closed source dataset from the island of Montreal. The majority of data sources are available online and the links are provided in the references. The following sections provide a detailed description of the data gathering procedure of this research.

5.4.1 Site Selection

Sites were selected across the Public Transit Access Points, where bus stops and metro stations were found, in the city of Montreal. A buffer size of 30 meters (100 foot) is suggested in the literature as an appropriate influence area for transit stops (Srinivas S. Pulugurtha and Vanapalli 2008). Bus stops and metros access points, in the vicinity to each other, were clustered with a Euclidean distance of 60 meters which ensures 30 meter buffers are left as separate PTAP instances. Then, PTAPs within 20 meters of an intersection are considered as part of a single PTAPs instance at such intersection, and PTAPs located more than 20 meters further from intersections are considered as midblock stops. Among the access points, daily traffic count of pedestrians and vehicles is available for 2650 sites, out of which 1709 access points are classified as “nearby intersections”, and 941 access points are classified as “midblock locations”. Figure 5-1 illustrates the public transit access points which are the targets of safety performance models in this chapter.

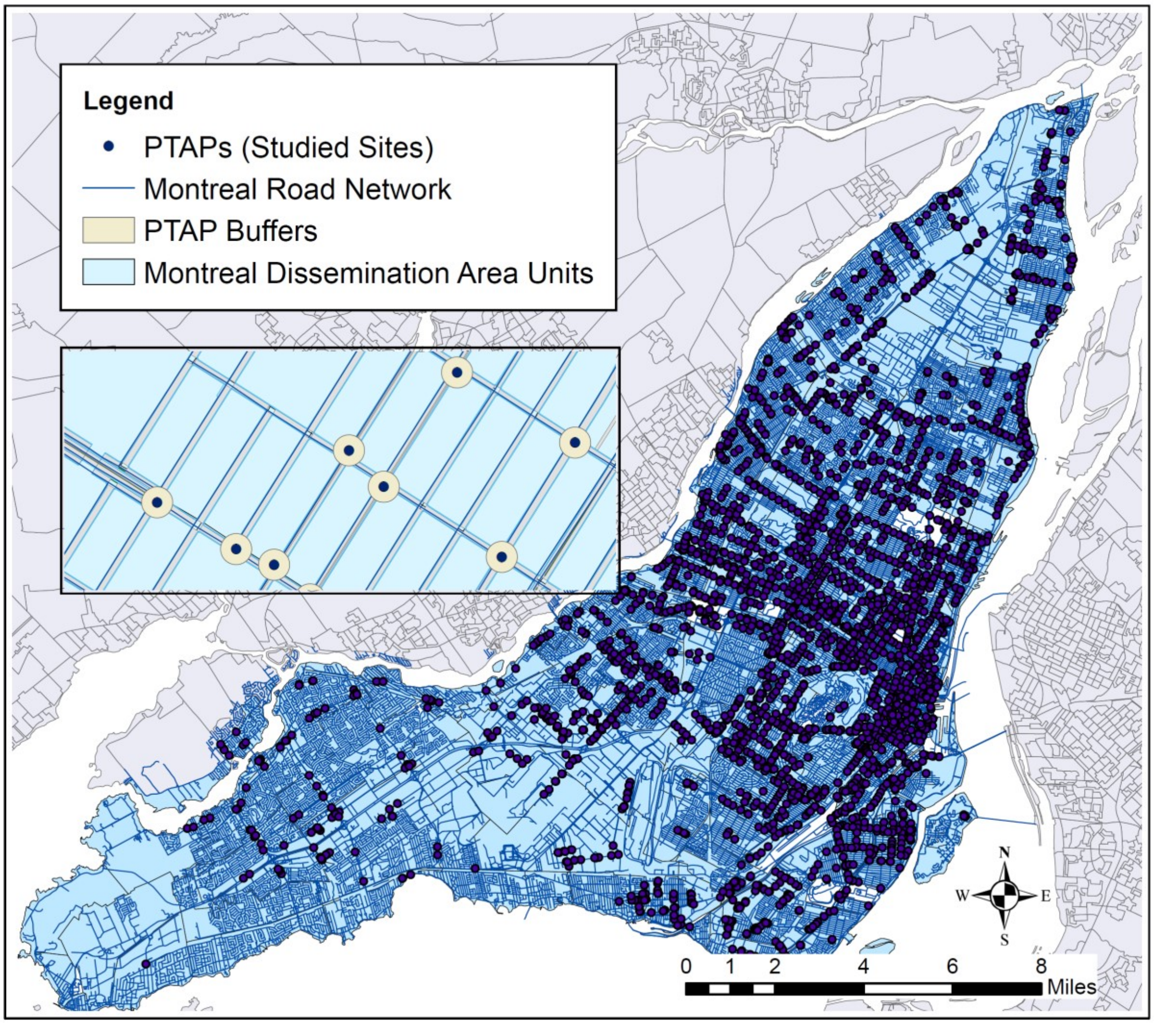

Figure 5-1Selected sites, 2650 Public Transit Access Points (PTAPs) in the Montreal Island. *The map was created using ArcGIS® software by Esri ("GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri" 2022)*

5.4.2 Count of collisions

Montreal road collisions data set ("Road Collisions - Open Government Portal", 2020) provides the location of pedestrian vehicle collisions. For the PTAPs located nearby intersections, the counts of pedestrian collisions are aggregated in the 30 meters' buffer around access points (bus stops and metro station entrances). For the PTAPs further from intersections (20 meters), the count of Pedestrian vehicle collisions at the road segments is associated to the PTAPs, as pedestrians might be influenced to commit jaywalking where there is public transit access at midblock (Zheng et al. 2015; Cinnamon, Schuurman, and Hameed 2011). The collisions between 2012 and 2019 are aggregated on the sites.

5.4.3 Traffic Count

Vehicle Traffic Count and Pedestrians' volume are calculated for 6100 road segments based on the available traffic count for intersections in Montreal city which is available online ("Traffic Lights — Vehicle and Pedestrian Counts at Intersections with Lights - Open Government Portal" 2021b). The available dataset provides the count of vehicle traffic count at intersections with the detail of vehicles' trajectories. That is, vehicle traffic count is available at intersections with respect to the 4 cardinal coordinates of intersections legs. Therefore, the azimuth (angle to the north) of intersections was measured to find the cardinal direction and correspondingly to assign the traffic count to the road segment. The figure below demonstrates Montreal road networks with the measured directions which are used in the intersections traffic assignment. The procedure of assigning intersections traffic count to the legs of the intersection is done using python codes which are available in appendix 1 of this dissertation.

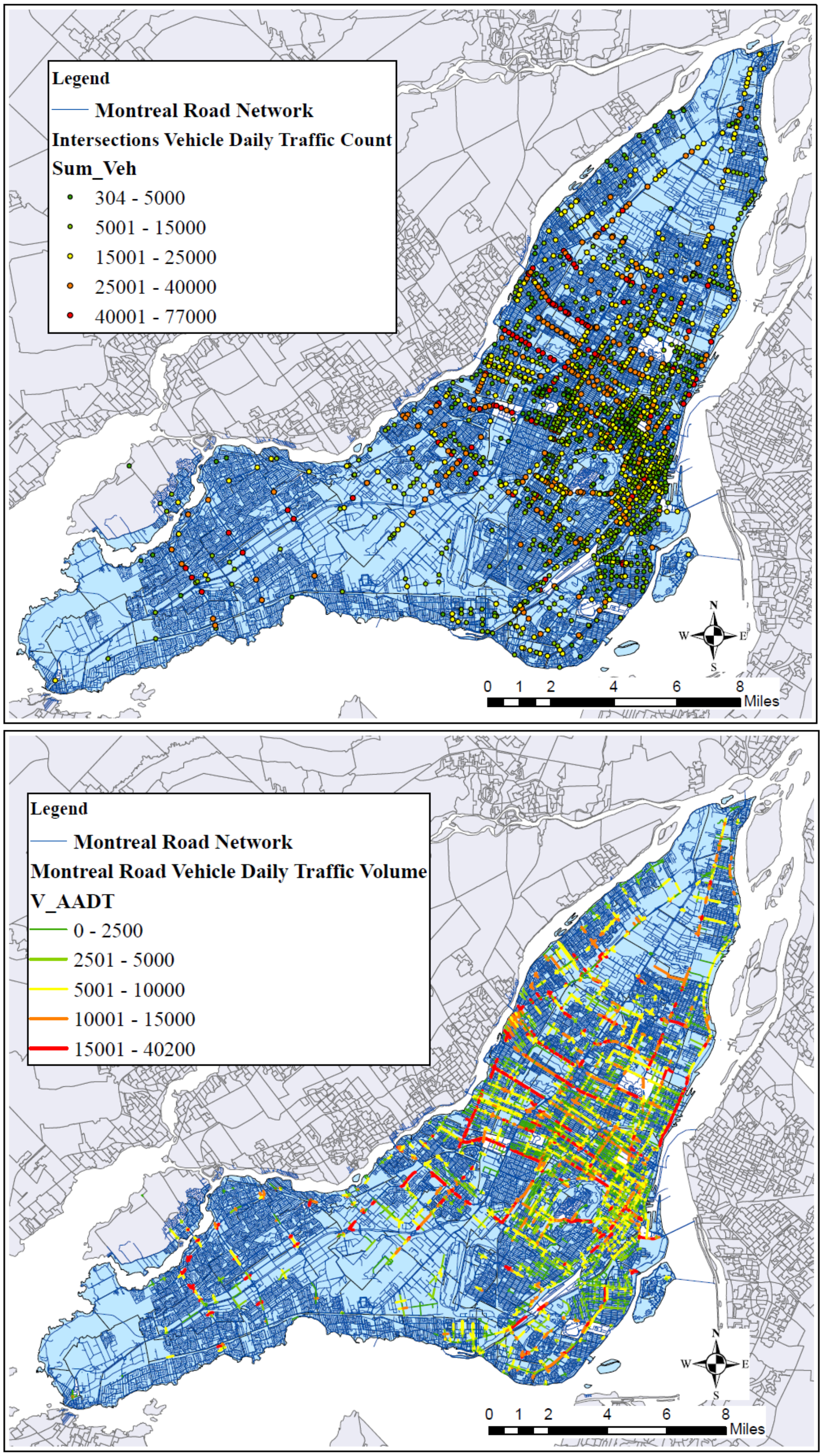

Figure 5-2Color map for Montreal daily vehicle traffic volume at intersections (top) and intersections legs (bottom). The maps were created using ArcGIS® software by Esri ("GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri" 2022)

Expanding the single values of traffic counts at intersections to the corresponding intersections' legs, and road segments, expands the available traffic data on PTAPs, which in turn, increase the count of PTAPs studied in the analysis. The analysis in this chapter is not limited to PTAP nearby intersections with traffic counters, but it also includes the PTAPs nearby other intersections or at mid-block locations.

5.4.4 Road Geometrics and Road Assets

The open source dataset of Montreal road assets (Complete database) ("Road Assets (Complete Database - Pavement, Island, Intersection, Sidewalk, Area) - Open Government Portal" , 2020) gives the geometric detail of the roads segments, intersections, road medians and sidewalks. Besides, the Canadian Geospatial data from DMIT provides access to the characteristics of road way such as the number of directions (one-way or two-way). Based on the sources, the Average Road and Sidewalk Width, Presence of Certain Road Median, average number of vehicle flow directions (an indicator of one-way and two-way roads), presence of signals' walk interval, and stop signs are gathered for the entities ("Geoindex - Public" n.d.; "Traffic Lights — Pedestrian Lights - Open Government Portal" n.d.; "Road Signs (Excluding Parking) - Open Government Portal" n.d.). Figure 5-3 shows the overview of road geometry and built-environment datasets in the case study of Montreal.

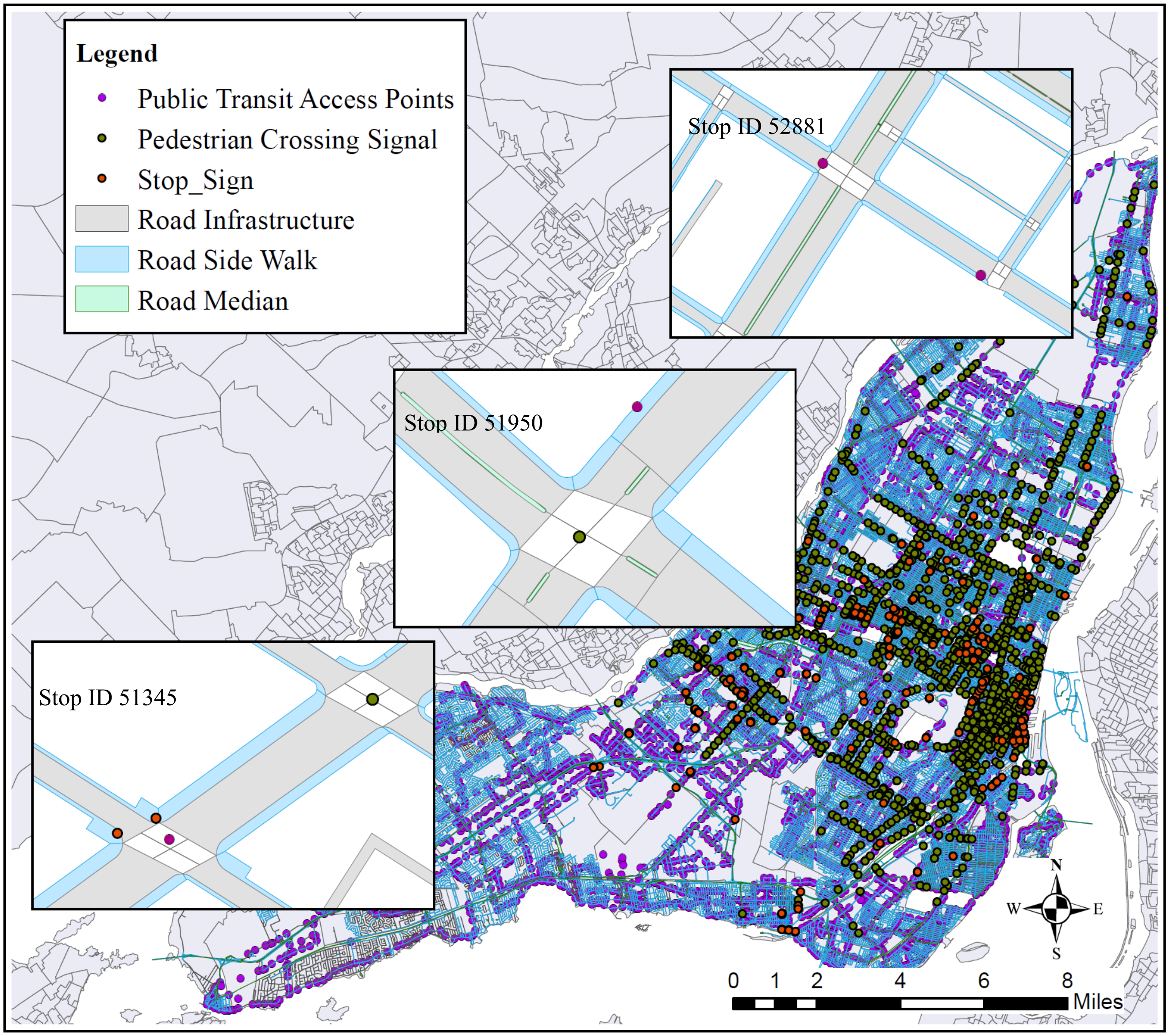

Figure 5-3Montreal Road Infrastructure, with a close view of three example of sites. The map was created using ArcGIS® software by Esri ("GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri" 2022)

Table 4-1 summarizes the descriptive statistics of the variables used in the analyses in this chapter.

Table 5-1 Descriptive Statistics of Variables in the Analysis

Variables	Type	Mean	S.D.	Min	Max
Counts of collision in proximity of PTAPs	Integer	1.55	0.05	0	25
Logarithm of vehicle average daily traffic in the proximity of PTAPs	Float	9.06	0.02	7.02	11.22
Logarithm of pedestrian daily volume	Float	5.86	0.03	0.69	10.94
Average number of traffic flow directions per road in the proximity of PTAPs, (one-way vs two way, indicator)	Integer	1.54	0.01	1	2
Average road width	Float (m)	9.32	0.09	1.96	20.55
Average sidewalk width	Float (m)	5.55	0.09	0.09	8.87
Presence of island (median refuge)	Binary	0.18	0.01	0.00	1.00
PTAP's road grade	Float	0.55	0.75	0.00	7.83
Signalized Intersection	Binary	0.66	0.01	0.00	1.00
Presence of signal's walk interval in proximity of PTAPs	Binary	0.55	0.02	0.00	1.00
Presence of stop sign in proximity of PTAPs	Binary	0.50	0.02	0.00	1.00

5.5 Analysis and Results

5.5.1 Safety Performance Functions

Safety performance functions are developed using Negative Binomial statistical models to study pedestrian safety at PTAPs. Sites are divided based on their position regarding the nearest intersection. That is, the sites through the study are PTAPs at midblock and the PTAPs in the proximity to the intersection. PTAPs at intersections are studied more specifically by developing further analysis whether the PTAP is near to signalized intersections or nearby a non-signalized intersection. Figure 5-4 shows an overview of analysis in chapter 4 which is accompanied by the analysis in this chapter.

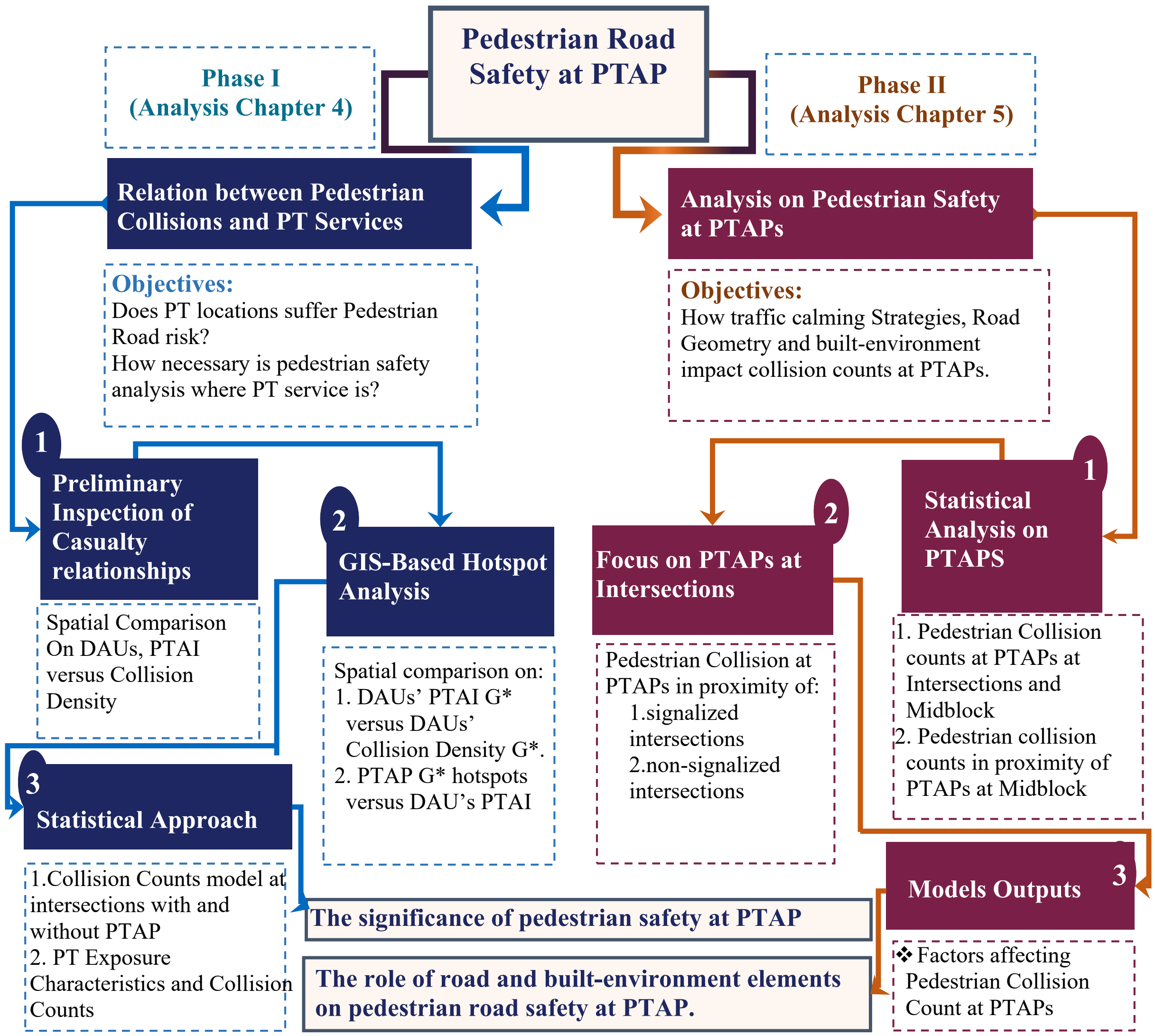

Figure 5-4High level overview of the analysis in Chapter 4 and 5.

Negative Binomial models confirm the statistical significance of vehicle traffic and pedestrian volume with positive coefficients in all the analyses in Table 5-2:5. The analysis on the geometrics and built environment characteristics of PTAPs at intersections and mid-blocks is shown in Table 5-2 and Table 5-3 respectively. The model found that both: (1) average road width and (2) sidewalk width in the proximity of PTAP as statistically significant (95% confidence level) factors with positive and negative coefficients respectively. Moreover, road medians and road grades are significant variables. In the PTAP nearby intersections, the average number of directions of vehicle traffic flow (an indicator of one-way vs two way) at the intersection, with coefficient +0.29, and the PTAP's distance to the intersections found statistically significant with coefficient +0.01, in the PTAPs at road segments.

Table 5-2 Negative Binomial Regression on Pedestrians Vehicle Collision Counts in the proximity of Public Transit Access Points nearby intersections- Road Geometrics, Number of observations = 1709*, Log Likelihood = -714.7*

Pedestrian Vehicle Collision Counts	Coef	Std. Err.	z	P>z
Logarithm of vehicle average daily traffic	0.12	0.03	3.56	0.00
Logarithm of pedestrian daily volume	0.34	0.02	14.45	0.00
Average road width in proximity of PTAP	0.08	0.01	6.05	0.00
Average sidewalk width in proximity of PTAP	-0.06	0.02	-4.06	0.00
Presence of median in proximity of PTAP (Yes or No)	-0.14	0.08	-1.82	0.07
Signalized intersections (Yes or No)	0.61	0.08	8.07	0.00
Average number of traffic flow directions in proximity of PTAP, (an indicator of one-way vs two way)	0.29	0.11	2.61	0.01
Average road grade in proximity of PTAP	0.06	0.04	1.66	0.10
Constant	-4.00	0.35	-11.39	0.00
Alpha (NB extra parameter)	0.72	0.05		

* *Public Transit Access Point (PTAP)*

Table 5-3 Negative Binomial Regression on Pedestrians Vehicle Collision Counts in the proximity of Public Transit Access Points at Road Segments (Mid-blocks), number of observations = 941, *Log Likelihood = -3061.3*

Pedestrian Vehicle Collision Counts	Coef	Std. Err.	z	P>z
Logarithm of vehicle average daily traffic	0.17	0.08	2.21	0.03
Logarithm of pedestrian daily volume	0.30	0.06	5.35	0.00
Average road width in proximity of PTAP	0.06	0.03	2.25	0.02
Average sidewalk width in proximity of PTAP	-0.04	0.03	-1.48	0.14
Presence of median in proximity of PTAP (Yes or No)	-0.50	0.24	-2.04	0.04
Average road grade of PTAP	0.29	0.10	2.79	0.01
Distance to the nearest intersection	0.01	0.00	4.23	0.00
Constant	-4.85	0.74	-6.56	0.00
Alpha	2.69	0.38		

** Public Transit Access Point (PTAP)*

Table 5-2 shows PTAPs at signalized intersections are more at risk of pedestrian-vehicle collisions compared to PTAPs at non-signalized intersections, with a coefficient of +0.61. This motivated further analysis on PTAP in the vicinity of intersections. Hence, PTAP at signalized and non-signalized intersections were studied respectively in Table 5-4 and Table 5-5. Table 5-4 shows pedestrian crossing light in Proximity to PTAPs is statistically significant with a negative coefficient (-0.13). Besides, Table 5-5 shows the statistical significance of stop signs in the proximity of a PTAP located at non-signalized intersection with a negative coefficient (-0.33).

Table 5-4 Negative Binomial Regression on Pedestrians Vehicle Collision Counts, PTAP *nearby Signalized intersections- Crossing Facilities, number of observations = 926, Log Likelihood = -1847.4*

Pedestrian Vehicle Collision Count	Coef	Std. Err.	z	P>z
Logarithm of vehicle average daily traffic	0.50	0.06	8.38	0.00
Logarithm of pedestrian daily volume	0.51	0.03	17.24	0.00
Presence of signal's walk interval in proximity of PTAPs (Yes or No)	-0.13	0.07	-2.00	0.05

Presence of median in the proximity of PTAP (Yes or No)	-0.17	0.08	-2.08	0.04
Average number of traffic flow directions in proximity of PTAP, (an indicator of one-way vs two way)	,0.09	0.01	6.69	0.00
Constant	-8.03	0.54	-14.77	0.00
Alpha	0.41	0.04		

* *Public Transit Access Point (PTAP)*

Table 5-5 Negative Binomial Regression on Pedestrians Vehicle Collision Counts, PTAP *nearby Non-signalized intersections-Crossing Facilities, number of observations = 638, Log Likelihood = -714.7*

Pedestrian Vehicle Collision Count	Coef	Std.Err.	z	P>z
Logarithm of vehicle average daily traffic	0.12	0.05	2.19	0.03
Logarithm of pedestrian daily volume	0.35	0.06	5.89	0.00
Presence of stop sign (Yes or No)	-0.33	0.15	-2.25	0.03
Presence of median in the proximity of PTAP (Yes or No)	-0.34	0.19	-1.78	0.08
Constant	-2.76	0.51	-5.44	0.00
Alpha	1.96	0.26		

* *Public Transit Access Point (PTAP)*

5.5.2 Hotspot Identification and Site Selection

As discussed before, site selection and hotspot identification is one of the significant steps in the transportation safety management process. In this section, hotspots of pedestrian vehicle collisions are identified based on Empirical Bayes estimation. Table 5-6 and Table 5-7 show the potential improvements and the hotspots ranking based on the mentioned methods and provide the collision count, the estimated count, and the Empirical Bayes estimation at these sites. Public Transit hotspots are ranked based on the EB potential improvement.

Table 5-6 Site Hotspot Identification, top ranking based on Empirical Bayes method, public transit access points nearby intersections.

Site	Close intersection	Collision Count	Estimated Count	EB Potential Improvement (PI)	EB Method Ranking
51392	Saint Michel/ Jarry	25	5.28	18.70	1
53010	de Lorimier/ Rue Logan	20	4.42	14.09	2
52129	Pie-IX / Beaubien	20	2.66	11.86	3
9	du Mont-Royal / Berri St.	16	5.79	11.84	4
52329	Guy / René-Lévesque	14	7.55	10.82	5
34	Guy / Saint Catherin	13	8.84	10.26	6
50556	Sauvé / Papineau	16	3.35	10.10	7
15	Gauchetière / Cathedrale	13	7.97	10.07	8
34	Guy / De Maisonneuve	12	11.31	9.75	9
51	Vanhorn/ Victoria	13	5.67	9.37	10

Table 5-7 Site Hotspot Identification, top ranking based on Empirical Bayes method, public transit access points at midblock.

Site	Closest intersection	Collision Count	Estimated Count	EB Potential Improvement (PI)	EB Method Ranking
51950	Sherbrook / Atwater	11	0.79	7.29	1
61896	Saint Denis Mont-Royal Ave	8	0.70	5.04	2
50820	Vanhorn/Victoria	7	1.01	4.97	3
51049	Côte-sainte-catherine/ CDN	6	1.38	4.61	4
52238	Guy / Saint Catherin	5	2.34	4.23	5
52330	Guy / René-Lévesque	5	2.23	4.19	6
57930	Sources Blvd/ Salaberry	8	0.40	3.89	7
51050	Queen-Mary / Circle	6	0.76	3.85	8
57947	Sauvé / Papineau	8	0.38	3.78	9
54019	Vanhorn/ Victoria	6	0.47	3.15	10

In the following section, first, the results of NB models built in this research study are interpreted. The discussion explains how road elements could improve the pedestrian safety level of PTAPs. Moreover, the usefulness of the safety analysis findings is discussed in a couple of hotspot examples in Table 5-6 and Table 5-7.

5.6 Discussion

5.6.1 Road Elements and Pedestrians Collision count at PTAPs

Statistical analysis on pedestrian vehicle collisions in the proximity of public transit locations shows road geometry and built environment factors play part in the count of collisions in these areas. Engineering elements such as road and sidewalk presence, presence of certain road medians, road configuration, and pedestrian crossing facilities were found statistically significant variables in the analysis which are discussed shortly in the following. Moreover, Empirical Bayes collision hotspot identification methods sorted PTAPs regarding the EB collision risk. In the following, the discussion of the analysis is provided.

All the negative binomial models show the significance of traffic calming strategies in PTAP's pedestrian safety while controlling for pedestrian vehicle exposure variables. Average daily vehicle traffic and pedestrian daily volume are statistically significant with positive coefficients in all models. That is higher pedestrian vehicle collisions where there is higher traffic of vehicles and higher pedestrian volume.

The negative binomial models show the effectiveness of traffic calming strategies in PTAP's pedestrian safety. Table 5-2 shows as the width of the sidewalk increases and the road width decreases, a lower count of pedestrian collisions are expected for the PTAPs. This is consistent with the objectives of traffic calming strategies reviewed in the literature (Distefano and Leonardi 2019; Gomaa Mohamed et al. 2012). The NB models in Table 3 proved the positive role of road medians in pedestrian safety. Certain road medians could channelize vehicle traffic flow and provide refuges for pedestrians to cross the road safely. Besides, the presence of a road median could increase the distance drivers stop or yield for pedestrians before the crosswalk, and can significantly increase the share of pedestrians who look for vehicles before crossing the street (S. Pulugurtha et al. 2012).

Table 3 confirms the statistical significance of road grade for the safety of PTAP at intersections and mid-blocks. Analysis shows that more pedestrian-vehicle collisions are expected in the proximity of PTAPs on roads with higher grade levels. High road grades would lead to higher speed on the down flow of vehicles, reduced visibility if vehicles are coming from a crest and then going downslope, or if vehicles are coming from a sag at night time and then going upslope. Besides, Table 3, and Table 4 confirm the positive effect of two-way to one-way conversion or reduction of traffic directions to improve the safety of pedestrians in the proximity of public transit. When the number of traffic flow directions increases, the expected number of pedestrian-vehicle collisions increases.

Table 3 suggests locating PTAP close to the intersections and decreasing the distance between the PTAP and the nearest intersection. As PTAP distance to intersections increases, the expected count of collisions increases, possibly because of more pedestrians attempting to cross the street at locations where formal crossings do not exist. This implies the priority to locate PTAPs near intersections rather than road segments and possibly the need to add pedestrian fencing in certain locations.

PTAPs at signalized intersections are more at risk of pedestrian-vehicle collisions than PTAPs at non-signalized intersections (Table 3). This suggested further analysis of PTAP in the vicinity of intersections. Table 4 shows a pedestrian crossing signal as a traffic calming device that provides temporal separation for the vehicle traffic flow and pedestrian crossing in proximity to PTAPs. It is highly recommended to equip PTAPs nearby signalized intersections with a separate pedestrian crossing interval to address pedestrian collision risk at these locations, as found in Table 3. More, for the PTAP nearby non-signalized intersections, Table 4 confirms the positive role of stop signs since it implies vehicles' full stop at the intersections leading to providing a safe environment around PTAP nearby non-signalized intersections.

5.6.2 PTAP Hotspots and Countermeasures

The safety analysis in Table 5-2 and Table 5-3 shows the effect of the built environment and road elements on the count of collisions in the vicinity of public transit locations. In the network screening phase, the hotspot sites are identified to conduct a further field study and examine how engineering metrics could improve the safety level at the fields. Hence, in the following a couple of examples of the hotspots sites in the Table 5-6 and Table 5-7 will be presented.

Figure 5-5 shows the stop ID 53010, located nearby the intersection Lorimier/ Rue Logan in the case study of Montreal city. As can be seen in the top aerial photo, one of the intersection legs stems from an arterial road, the Jacques-Cartier bridge, which the data shows the average daily traffic of 9900 vehicles and 800 pedestrians at this location. The GIS database and the Google Map source show this intersection as a non signalized intersection in which the bus stop is located in the proximity of the intersection. Vehicle Traffic flow is properly separate due to the presence of certain road median, however, a pedestrian crossing is not secured with proper pedestrian crossing facilities. Since the analysis in Table 5-2 and Table 5-3 shows the positive effect of stop signs and signalized intersections on pedestrian safety, the feasibility of such crossing facilities could be studied for this site. Further site specific investigations could be conducted to improve pedestrian road safety at this site.

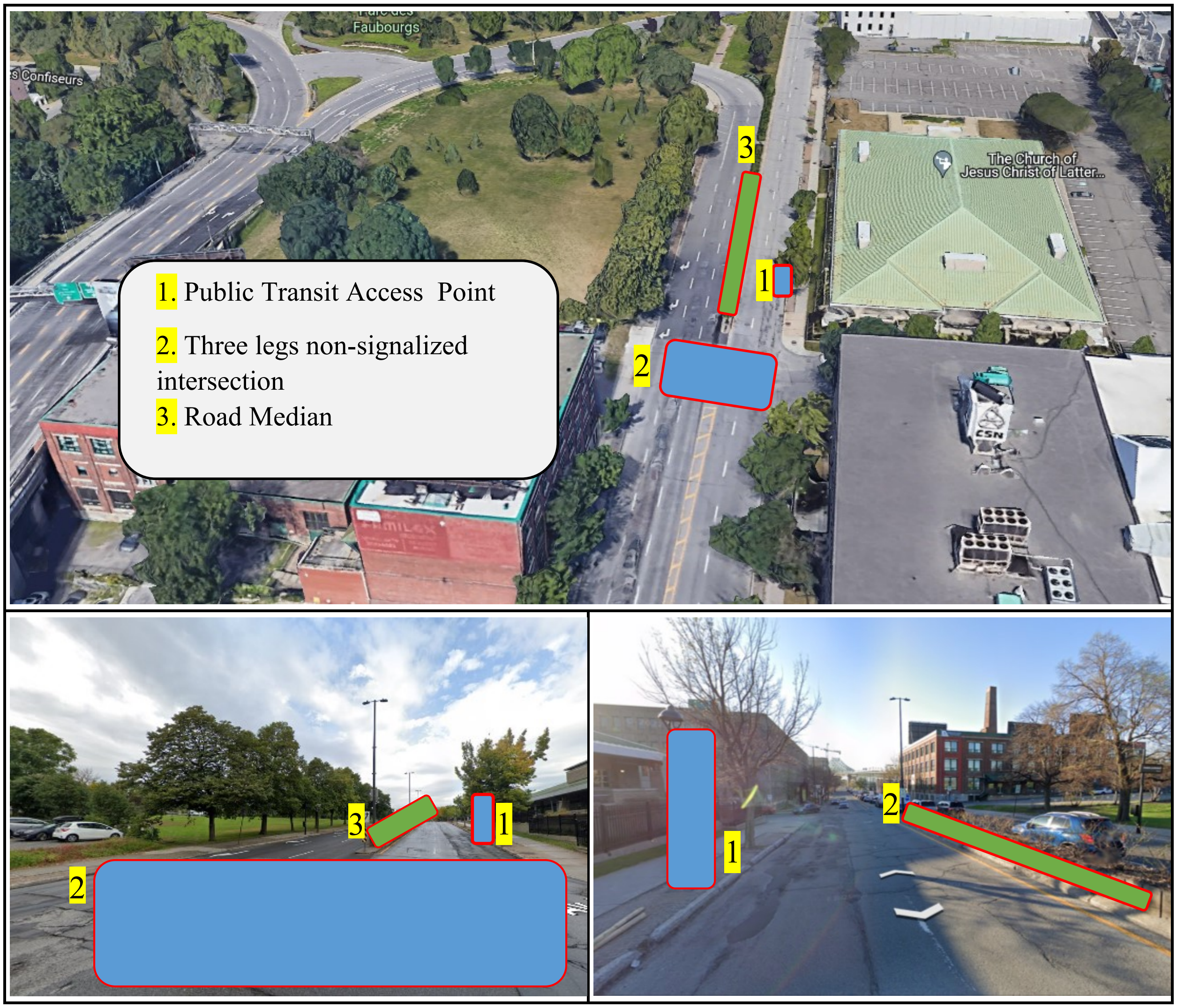

Figure 5-5 Transit stop de Lorimier/ Rue Logan Stop, ID 53010, hotspot of pedestrian vehicle collisions among transit stops. The top picture is the aerial photo of the site, and the two photos at bottoms are the field photos. Photo Source: Google Earth.

The hotspot analysis in Table 5-6 found the transit stop Gauchetière / Cathedrale, stop ID 15, as a pedestrian vehicle collision hotspot. Figure 5-6 shows an initial field investigation by providing the Google Maps 3D photos of the site. The noticeable point at this site is the recent change that has been applied to the road geometry and design at this location. As can be seen in Figure 5-6, the top picture demonstrates the site before the recent changes in which the intersections used to be a four leg intersections, however, in the new road design, one of the intersection's leg is removed and pedestrian pavement was installed instead. The new design, reduce the interaction between pedestrian and vehicles which in turn reduce the pedestrian collision risk at this site. This site instance shows the positive and negative sides of the hotspot analysis through this study. On the positive side, the analysis is capable of finding and introducing sites that lack proper road design and the safety level could be enhanced by a field study and site specific engineering countermeasures. However, on the other hand, this example shows the weakness of analysis due to the aggregated approach of the analysis which could miss the built environment changes through the analysis period.

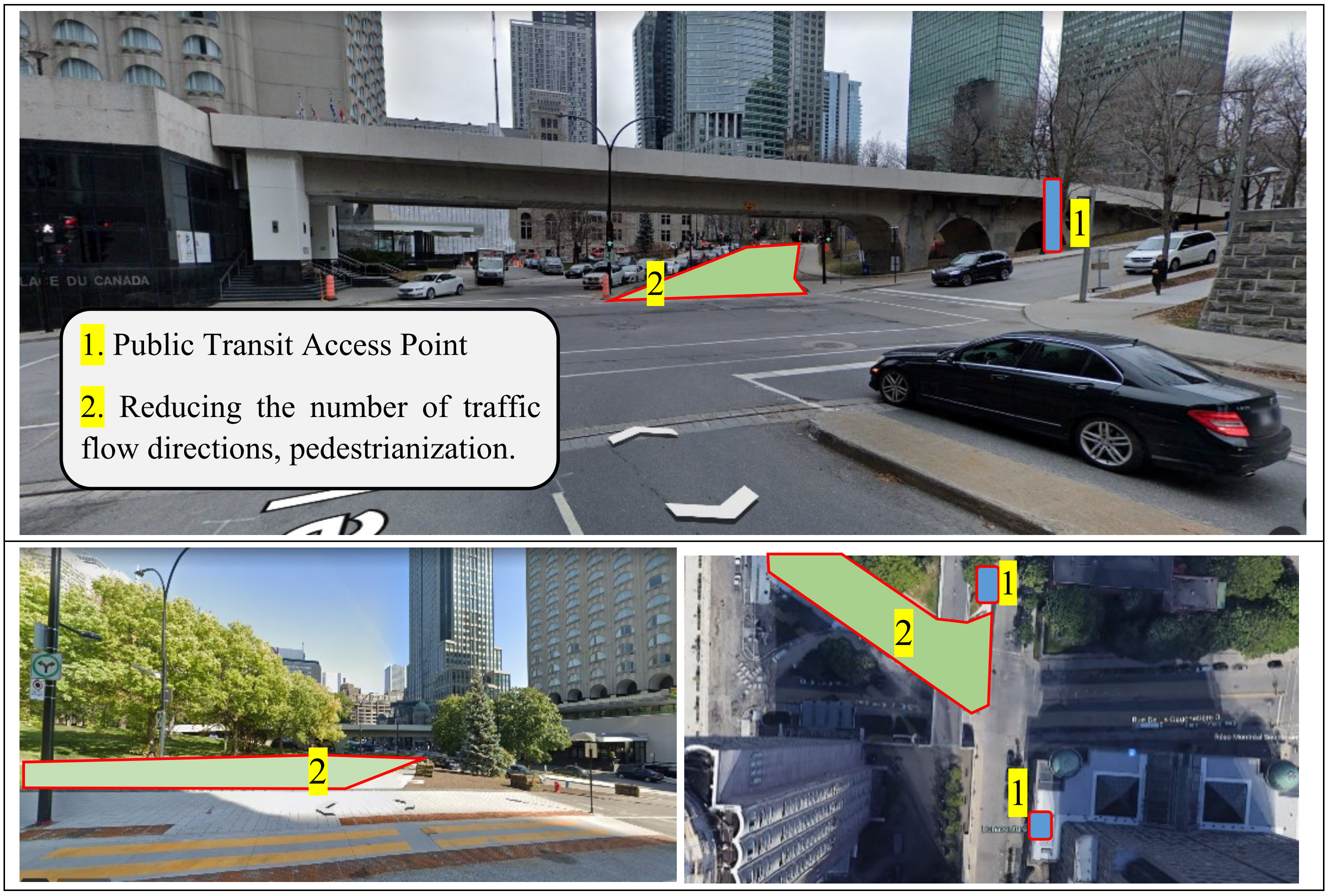

Figure 5-6 Transit stop Gauchetière / Cathedrale, Stop ID 15, hotspot of pedestrian vehicle collisions among transit stops. The top picture refers to the site before the road redesign. The two bottom picture shows the contemporary situation which road elements has been redesigned and one of the roads changed to pedestrian pavement. Photo Source: Google Earth.

Figure 5-7 shows the transit stop nearby the intersection of Sources Blvd/ Salaberry. The transit is located at the side of an arterial road with 4 lanes including 3 main lanes and one left turn lane. The bottom picture in figure 5-7 shows a proper demonstration of the built-environment around this stop. Although the stop is located at a location further from the intersection, the road median at this site can be improved to lead pedestrians to the intersection area to cross the road rather than jaywalk in the proximity of the transit stop. Moreover, it would be helpful to relocate the stop to a closer location to the intersection crossing. This would tempt the transit users to not jaywalk in the proximity of the PTAP.

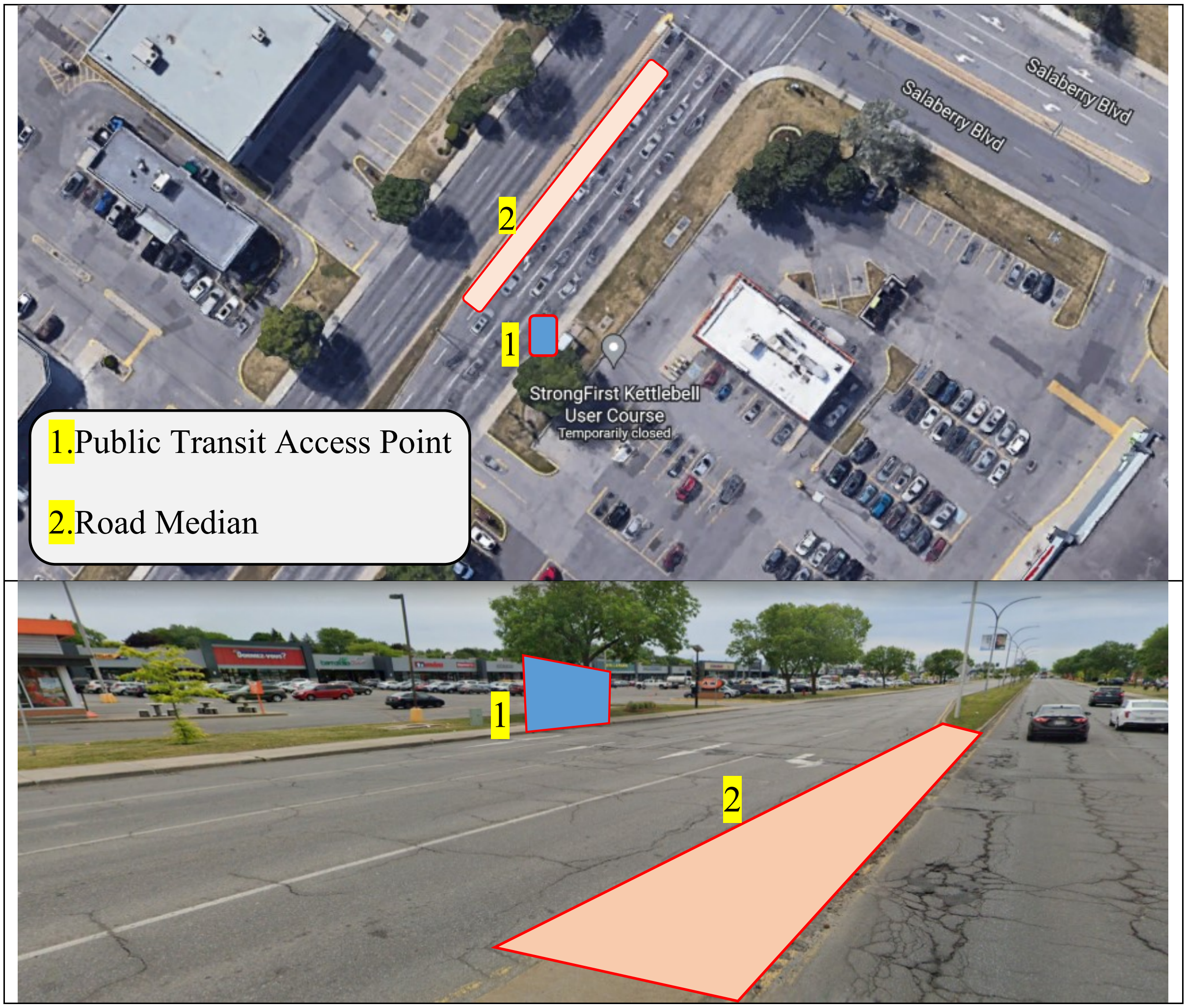

Figure 5-7 Transit stop Sources Blvd/ Salaberry, Stop ID 57930, hotspot of pedestrian vehicle collisions among transit stops. The top picture is the aerial photo of the site, and the photo at bottoms shows the field. Photo Source: Google Earth.

5.7 Conclusions

The safety analysis presented in this chapter showed the effectiveness of road and built environment elements on pedestrian road safety in the proximity of PTAPs. The Negative Binomial collision count models suggested the strategies like road narrowing, sidewalk curb extension (increasing sidewalk width), and installing road medians to improve pedestrian road safety at PTAPs. Moreover, pedestrian crossing facilities like the walking intervals in signalized intersections and the vehicle stop sign at non-signalized intersections were found effective to improve pedestrian road safety at the PTAPs. The number of vehicle traffic flow directions could influence the risk of traffic conflicts between pedestrians and vehicles, hence, strategies like two ways to one-way conversion were found statistically significant to provide a safer built environment for pedestrians to access public transit.

Collision hotspot identification is one of the most important steps in road safety management. Hotspot identification suggests the sites with a high concentration of pedestrian-vehicle collisions and sorts out the sites with a higher potential for improvement. Based on the EB hotspot analysis conducted in this chapter, there are PTAPs with a high pedestrian-vehicle collision risk, so further site-specific investigations will be necessary to determine how to implement appropriate countermeasures. The road element and built environment strategies proposed by this chapter's safety analysis could also be considered as an effective countermeasure for the sites.

Chapter 6: Conclusions and Recommendations

6.1 Conclusions

This master thesis addressed pedestrian road safety in the proximity of public transit access points. First, this research study emphasizes the significance of pedestrian road safety at locations where there are public transit services. In a zonal level approach, higher pedestrian-vehicle collision density (count of collisions per area units) was found associated with the polygons with higher public transit accessibility index. Moreover, at an intersection level, higher pedestrian-vehicle collision risk is expected where there is higher bus transit daily traffic, higher number of public transit routes, and higher public transit accessibility index. Intersections where there is a PTAP suffer higher pedestrian collision risk than the intersections with no PTAPs. Hence, it is concluded that there is a potential relation between the pedestrian vehicle collision counts and public transit services, and public transit access points suffer pedestrian vehicle collision risk; therefore, these locations require further pedestrian safety measures to reduce the collision risk.

Furthermore, this thesis studied the effect of road and built-environment elements to improve pedestrian safety at PTAPs. Certain road geometry and built-environment characteristics could provide safer pedestrian environment at PTAPs. City authorities should incorporate the findings of this study to improve the pedestrian safety at PT locations. In particular, road geometry and traffic calming strategies, namely road narrowing and side walk extension, are highly recommended for both PTAPs located nearby intersections or at midblock. In the case of PTAPs nearby intersections, curb side extension could meet this suggestion as it simultaneously could decrease road width and extend pedestrian side walk, which in turn, would reduce the chance of pedestrian vehicle exposure and collision risk. Traffic signals equipped with a walk interval improve pedestrian safety at public transit access points nearby signalized intersection. Therefore, it is suggested to add certain walk intervals to traffic signals nearby public transit stops. Moreover, if a public transit stop is located nearby a non-signalized intersection, it is recommended to improve pedestrian road safety by adding vehicle stop signs at the non-signalized intersection near by the PTAPs.

It is recommended to located PTAPs nearby intersection compared to the locations in a further distance to the nearest intersection. if the PTAP is located at midblock locations, it is highly recommended to utilize certain road medians as a refuge providing proper pedestrian crossing facilities at the location, or as for channeling pedestrian traffic to cross the road using the nearest intersection or crossing facilities. PTAPs located in roads with high grades and high number of vehicle traffic directions (two way roads) are exposed to higher pedestrian vehicle collision risk. Hence, it is recommended to convert two way streets to one way streets, particularly where there are streets with high grades descending from a hill nearby the PTAP.

Incorporating Empirical Bayes hotspot identification suggests the PTAPs which are in a high risk of pedestrian-vehicle collision. Suggested hotspot sites require further field and site specific studies to utilize the applicable countermeasures to improve pedestrian safety at the hotspots of PTAPs. The finding of safety analysis in chapter 5 of this thesis showed the potential to be considered for alleviating the collision risk at the sites.

6.2 Lessons Learned and Recommendations for Future Research

This study established the potential relationship between the pedestrian collision counts and public transit service locations by taking into account public transit exposure characteristics. The analysis utilized bus traffic count and pedestrian volume derived from short term traffic counters at intersections, however, future research could explore more how the safety analysis vary by utilizing passenger count and vehicle location datasets from public transit agencies.

This thesis studies the role of road geometry on pedestrian road safety at PTAPs, and argued decreasing the road width and increasing the sidewalk width would lead to the lower pedestrian collision risk is at PTAPs. This would provide a general idea behind the relation between road width and pedestrian road safety at the locations, however, the analysis could be improved if the dataset is potential for considering lane width and the number of lane in addition to the road width parameter. Therefore, it is recommended that future research study how lane width change would affect pedestrian road safety at PTAPs, and simultaneously, if there could be further details on the lane width providing a safer PTAP.

This research study showed and discussed the effectiveness of presence of walk interval in the traffic signal in proximity of PTAPs on improving pedestrian safety at these locations, however, the analysis is incapable of discussing walk interval specifications like walk interval and clearance duration. Therefore, future research could study how the pedestrian walk interval specifications affect pedestrian road safety at PTAPs near by the intersection. Moreover, this study discussed the effectiveness of presence of stop signs where PTAPs is nearby a non-signalized intersection. However, this question was not answered that whether these stop signs are for the streets being crossed by pedestrians or stop signs on the side streets. Hence, it is recommended that future studies focus on the specifications of these two traffic control devices on the pedestrian road safety at PTAPs.

The current study found the presence of road median as an effective element to improve pedestrian road safety at both PTAPs nearby intersection and located at midblock. It was argued that certain road medians would channelize pedestrian traffic to cross the road at a proper crossing facility, and it may provide pedestrian a refuge when they are crossing the road. Therefore, future research could study how different types of road median, with different configurations, would play part in providing a safer road environment at the PTAPs.

Moreover, although the current study addresses the role of road geometry and built environment elements on pedestrian road safety, it does not take into account the effect of the factors on traffic operation and network performance. Therefore, future research could establish safety analysis considering both road safety geometry and built-environment measurements and traffic operation through an integrated research study.

Future research could explore and compare other collision count models and hot spot identifications in investigating pedestrian road safety at PTAPs. This research study utilized NB collision count model and EB collision hotspot identification method to address the research tasks. Although Negative Binomial is widely used in both research and practice, future research could assess and compare the performance of other collision count models and hotspot identification methods to study the pedestrian road safety at these locations.

This research study incorporated an aggregated temporal and spatial approach in the safety analysis. The analysis considers an eight year aggregated collision count between 2012 and 2019, which could miss the built environment changes during this time. Therefore, it is recommended that future research studies consider shorter time periods, for example, four or six years of aggregated collision counts, and compare how the analysis performance would change. Moreover, the analysis in this thesis does not distinguish the sites whether are bus stops only, metro station access points or they are multi-modal interchange points. Hence, for future research, it is recommended to determine how the pedestrian safety models could vary considering the type of PTAP. In the case of data availability, applying before-after safety analysis is also recommended as it could provide specific knowledge on the effectiveness of the suggested countermeasures leading to improve road safety management at PTAPs.

It is highly recommended that future research studies take into account the type of pedestrian-vehicle collision, namely, midblock dart/dash, intersection turning/ through vehicle, bus-related, failure to yield at non-signalized location, walking along the roadway, working in the road, crossing expressway. This would provide crash-type-specified inferences and accordingly pave the way to address the count of collisions based on the collision type and enhance the safety levels at the zones.

It is recommended that future research studies take into account the temporal variation in collision counts. That is, collision risk would vary based on the time and date of crashes. For example, whether the collisions occurred in the morning or evening rush hours compared to the other time steps during weekdays or weekends. This also could include monthly and seasonal timing and the potential countermeasures accordingly. Therefore, city authorities would have a timeline to take the proposed counter measurements.

Appendix I

In this section, the codes for assigning intersections' traffic count values to the intersections' legs are provided. The following is the Python script, when the cardinal Azimuths of intersections' legs were calculated in ArcGIS direction tool and it is used to assign the values to the corresponding legs.

```
import pandas as pd
import numpy
import math
df = pd.read_csv("G:/Master's of Applied Science in Civil
Engineering/Research/repository/Export_Output_105_TableToExcel.csv")
print(df.columns)

df['Ab_Vec'] = (df['Deltha_X'] ** 2 + df['Deltha_Y'] ** 2) ** 0.5
df['Costheta'] = df['Deltha_Y'] / df['Ab_Vec']
df['ArCosthet'] = numpy.arccos(df['Costheta']) * (180 / math.pi)
df['ArcTan'] = numpy.arctan(df['Deltha_Y'] / df['Deltha_X']) * (180 /
math.pi)
df['appcomp'] = 0
df['miss']=0
df['appcomp_MidPoint']=0
df['difference']=0

'''refine3'''
for i in range(len(df)):
    if ((df['CompassA'][i] < 45)
          and ((df['ArCosthet'][i] > 90))):
        df['appcomp'][i] = df['CompassA'][i] + 180
    elif ((df['CompassA'][i] > 315)
          and ((df['ArCosthet'][i] > 90))):
        df['appcomp'][i] = df['CompassA'][i] - 180
    elif ((135<df['CompassA'][i] < 180)
          and ((df['ArCosthet'][i] < 90))):
        df['appcomp'][i] = df['CompassA'][i] + 180
    elif ((180<df['CompassA'][i] <225)
          and ((df['ArCosthet'][i] < 90))):
        df['appcomp'][i] = df['CompassA'][i] - 180
    elif ((45 < df['CompassA'][i] < 135)
          and (((df['ArCosthet'][i] < 90) and (df['ArcTan'][i] < 0))or
((df['ArCosthet'][i] > 90) and (df['ArcTan'][i] > 0)))):
        df['appcomp'][i] = df['CompassA'][i] + 180
    elif ((225 < df['CompassA'][i] < 315)
          and (((df['ArCosthet'][i] < 90) and (df['ArcTan'][i] > 0)) or
```

```
((df['ArCosthet'][i] > 90) and (df['ArcTan'][i] < 0)))):
        df['appcomp'][i] = df['CompassA'][i] - 180
    else:
        df['appcomp'][i] = df['CompassA'][i]

# refined 2 for i in range(len(df)):
#     if ((df['CompassA'][i] < 90)
#            and ((df['ArCosthet'][i] > 90) and (df['ArcTan'][i] > 0) or
((df['ArCosthet'][i] < 90) and (df['ArcTan'][i] < 0)))):
#         df['appcomp'][i] = df['CompassA'][i] + 180
#     elif ((90 < df['CompassA'][i] < 180)
#            and (((df['ArCosthet'][i] < 90) and (df['ArcTan'][i] < 0))or
((df['ArCosthet'][i] > 90) and (df['ArcTan'][i] > 0)))):
#         df['appcomp'][i] = df['CompassA'][i] + 180
#     elif ((180 < df['CompassA'][i] < 270) and
#              (((df['ArCosthet'][i] < 90) and (df['ArcTan'][i] > 0)) or (
#                     (df['ArCosthet'][i] > 90) and (df['ArcTan'][i] < 0)))):
#         df['appcomp'][i] = df['CompassA'][i] - 180
#     elif ((270 < df['CompassA'][i] < 360)
#            and ((df['ArCosthet'][i] > 90) and (df['ArcTan'][i] < 0) or (
#                     (df['ArCosthet'][i] < 90) and (df['ArcTan'][i] > 0)))):
#         df['appcomp'][i] = df['CompassA'][i] - 180
#     else:
#         df['appcomp'][i] = df['CompassA'][i]

for i in range(len(df)):
    if df['ArCosthet'][i] < 90 and df['ArcTan'][i] > 0:
        df['appcomp_MidPoint'][i] = df['ArCosthet'][i]
    elif df['ArCosthet'][i] > 90 and df['ArcTan'][i] < 0:
            df['appcomp_MidPoint'][i] = df['ArCosthet'][i]
    elif (df['ArCosthet'][i] > 90) and (df['ArcTan'][i] > 0):
            df['appcomp_MidPoint'][i] = 360-df['ArCosthet'][i]
    elif (df['ArCosthet'][i] < 90) and (df['ArcTan'][i] < 0):
            df['appcomp_MidPoint'][i] = 360-df['ArCosthet'][i]

for j in range(len(df)):
    df['difference'][j] = abs(df['appcomp'][j] - df['appcomp_MidPoint'][j])
    if 300>df['difference'][j]>50:
                df=df.drop([j])



import queue as Q
print('start')
angel = 60
dir = {}
for i in range(len(df)):
    if df['Id_Interse'][i] == 0:
        continue
    if df['Id_Interse'][i] not in dir:
        dir[df['Id_Interse'][i]] = {}
        dir[df['Id_Interse'][i]]['To_N1'] = Q.PriorityQueue()
        dir[df['Id_Interse'][i]]['To_E'] = Q.PriorityQueue()
        dir[df['Id_Interse'][i]]['To_S'] = Q.PriorityQueue()
        dir[df['Id_Interse'][i]]['To_W'] = Q.PriorityQueue()
```

```
    if abs(df['appcomp'][i] - 0) < angel:
        dir[df['Id_Interse'][i]]['To_N1'].put((abs(df['appcomp'][i] - 0), i,
'N', df['appcomp'][i]))
    if abs(df['appcomp'][i] - 360) < angel:
        dir[df['Id_Interse'][i]]['To_N1'].put((abs(df['appcomp'][i] - 360),
i, 'N', df['appcomp'][i]))
    if abs(df['appcomp'][i] - 90) < angel:
        dir[df['Id_Interse'][i]]['To_E'].put((abs(df['appcomp'][i] - 90), i,
'E', df['appcomp'][i]))
    if abs(df['appcomp'][i] - 180) < angel:
        dir[df['Id_Interse'][i]]['To_S'].put((abs(df['appcomp'][i] - 180),
i,'S', df['appcomp'][i]))
    if abs(df['appcomp'][i] - 270) < angel:
        dir[df['Id_Interse'][i]]['To_W'].put((abs(df['appcomp'][i] - 270), i,
'W', df['appcomp'][i]))

print('next step')
finaldir = {}
import sys
for dot in dir.keys():
    if dot == 0:
        continue

    finaldir[dot] = Q.PriorityQueue()
    if not dir[dot]['To_N1'].empty():
        finaldir[dot].put(dir[dot]['To_N1'].get())
    if not dir[dot]['To_E'].empty():
        finaldir[dot].put(dir[dot]['To_E'].get())
    if not dir[dot]['To_S'].empty():
        finaldir[dot].put(dir[dot]['To_S'].get())
    if not dir[dot]['To_W'].empty():
        finaldir[dot].put(dir[dot]['To_W'].get())

final = {}

df['dir'] = 0
for dot in finaldir.keys():
    while not finaldir[dot].empty():
        street = finaldir[dot].get()
        if dot not in final:
            final[dot] = {}
        if street[2] not in final[dot].keys() and street[1] not in
final[dot].values():
            final[dot][street[2]] = street[1]
            df['dir'][street[1]] = street[2]
df.to_csv("G:/Master's of Applied Science in Civil
Engineering/Research/repository/Export_Output_105_TableToExcel_Refined7.csv")
```

Appendix II

In this section the codes for developing safety performance functions (collision count models) are provided. The following is the R script which usese statistical libraries to build up Negative Binomial models and in the next phase, the Empirical Bayes hotspot analysis.

```
# Crash Count Statistical Models, Negatice Binomial Aapproach on Pedestrain
Vehicle Collisons at PTAPs


#Loading Libraries

library("MASS")
library("stats")
library("sjPlot")
library("ggplot2")

library("GGally")
library("lmtest")


rm(list=ls())

#the current directory as the active direcory of R-studio
current_directory = dirname(rstudioapi::getSourceEditorContext()$path)
setwd(current_directory)

dir.create("results", showWarnings = TRUE, recursive = FALSE, mode = "0777")
dir.create("results/general", showWarnings = TRUE, recursive = FALSE, mode =
"0777")
dir.create("results/negative_binomial", showWarnings = TRUE, recursive =
FALSE, mode = "0777")


#*** loading dataset


db1 <- read.csv("Join_Output_246_20mINsct.csv", header = TRUE)

View(db1)
#check for NA and Missing
db1[which(is.na(db1)),]
db1 <- na.omit(db1)
```

```r
#library(readxl)
# db_x <- read_excel("Crash Frequency.xlsx", sheet = "Crash_Dataset")
 View(db1)


# column names and check for model input variables
colnames_db <- colnames(db1)
colnames_db

db <- subset(db1, select = c(Ln_V_TrfSum30,Ln_P_VolSum30,
Avg_Road_W,SIgnalzed,Count_4,Ave_SideWa, _Med_,Sum_grad_a,Avg_Num_Di ))


#Negateive Binomial Regression, using MASS library
glm_nb_0 <-  MASS::glm.nb(Count_4 ~ ., data = db, link = "log")
# glm_nb_0$

beta_nb_0 = glm_nb_0[["coefficients"]][["(Intercept)"]]

#~ Dispersion parameter PHI
phi_nb_0 = glm_nb_0$theta

#~ In thesis we saw alpha, which is the inverse of PHI:
alpha_nb_0 = 1/phi_nb_0
# we will use phi in most of the equation.

#~ Mean and Variance of NB is equal to:
mu_nb_0 = exp(beta_nb_0)
variance_nb_0 = mu_nb_0 + (1/phi_nb_0)*mu_nb_0^2

#~ Probability of crashes using NB distribution:
# Pr_Y_NB_0 <- dnbinom(crash_count, mu=mu_nb_0, size=phi_nb_0, log = FALSE)


source("Customized_Writing_Functions.R")


generating_NegativeBinomial_results(glm_nb_0, "NB-SPF0")

 #Hotspot Identification_ based on Potential Improvement in expected count of
collisions


db_hotspot <- data.frame(ID=db1[["stop_id"]])


db_hotspot$ACC <- db[["Count_4"]]


db_hotspot$mu <- glm_nb_0$fitted.values

mu_rp <- mean(db_hotspot$mu)
```

```
db_hotspot$PI_mu <- db_hotspot$mu - mu_rp
db_hotspot$rank_mu <- rank(-db_hotspot[["PI_mu"]], ties.method="min")



##Hotspot Identification_ based on EB Method:

phi = glm_nb_0[["theta"]]
db_hotspot$w <- phi / (phi + db_hotspot[["mu"]])

db_hotspot$EB <- ((1 - db_hotspot[["w"]]) * db_hotspot[["ACC"]])
+(db_hotspot[["w"]] * db_hotspot[["mu"]])

EB_rp <- mean(db_hotspot$EB)

db_hotspot$PI_EB <- db_hotspot$EB - EB_rp

db_hotspot$rank_EB <- rank(-db_hotspot[["PI_EB"]], ties.method="min")

#Method: Posterior Probability of Excess ====
alpha = db_hotspot[["ACC"]] + phi          # shape
beta = 1/(1 + phi/db_hotspot[["mu"]])      # scale

db_hotspot$Pr_Theta_1_n <- 1 - pgamma(8, shape=alpha, scale = beta, log =
FALSE)

db_hotspot$PI_Pr_Theta <- db_hotspot[["Pr_Theta_1_n"]] - 0.95

db_hotspot$rank_Pr_Theta <- rank(-db_hotspot[["PI_Pr_Theta"]],
ties.method="min")


write.csv(db_hotspot, "results/Hotspot_Identification_Results.csv")
View(db_hotspot)

###################################################

db2 <- read.csv("Join_Output_251_noninsectt.csv", header = TRUE)

db2[which(is.na(db2)),]
db1 <- na.omit(db1)




#Extracting column names
colnames_db <- colnames(db2)
colnames_db

db <- subset(db2, select = c(LN_V100, LN_P100,Dis_INscti ,Avg_grad_a,
Grn_MedB30, CC_at_Segs, AV_SidWE2, AV_RWPD  ))

#Negateive Binomial Regression, using MASS library
glm_nb_Non <-  MASS::glm.nb(CC_at_Segs ~ ., data = db, link = "log")
```

```
# glm_nb_0$
generating_NegativeBinomial_results(glm_nb_Non, "glm_nb_Non")

beta_nb_Non = glm_nb_Non[["coefficients"]][["(Intercept)"]]

#~ Dispersion parameter PHI
phi_nb_Non = glm_nb_Non$theta

#~ In thesis we saw alpha, which is the inverse of PHI:
alpha_nb_Non = 1/phi_nb_Non
# we will use phi in most of the equation.

#~ Mean and Variance of NB is equal to:
mu_nb_Non = exp(beta_nb_Non)
variance_nb_Non = mu_nb_Non + (1/phi_nb_Non)*mu_nb_Non^2

#~ Probability of crashes using NB distribution:
# Pr_Y_NB_0 <- dnbinom(crash_count, mu=mu_nb_0, size=phi_nb_0, log = FALSE)


source("Customized_Writing_Functions.R")


generating_NegativeBinomial_results(glm_nb_Non, "NB-SPFNon")

#Hotspot Identification_ based on Potential Improvement in expected count of
collisions


db_hotspot2 <- data.frame(ID=db2[["Stop_ID_1"]])


db_hotspot2$ACC <- db2[["CC_at_Segs"]]


db_hotspot2$mu <- glm_nb_Non$fitted.values

mu_rp <- mean(db_hotspot2$mu)

db_hotspot2$PI_mu <- db_hotspot2$mu - mu_rp
db_hotspot2$rank_mu <- rank(-db_hotspot2[["PI_mu"]], ties.method="min")



##Hotspot Identification_ based on EB Method:

phi = glm_nb_Non[["theta"]]
db_hotspot2$w <- phi / (phi + db_hotspot2[["mu"]])

db_hotspot2$EB <- ((1 - db_hotspot2[["w"]]) * db_hotspot2[["ACC"]])
+(db_hotspot2[["w"]] * db_hotspot2[["mu"]])

EB_rp <- mean(db_hotspot2$EB)

db_hotspot2$PI_EB <- db_hotspot2$EB - EB_rp
```

```
db_hotspot2$rank_EB <- rank(-db_hotspot2[["PI_EB"]], ties.method="min")

#Method: Posterior Probability of Excess ====
alpha = db_hotspot2[["ACC"]] + phi          # shape
beta = 1/(1 + phi/db_hotspot2[["mu"]])      # scale

db_hotspot2$Pr_Theta_1_n <- 1 - pgamma(8, shape=alpha, scale = beta, log =
FALSE)

db_hotspot2$PI_Pr_Theta <- db_hotspot2[["Pr_Theta_1_n"]] - 0.95

db_hotspot2$rank_Pr_Theta <- rank(-db_hotspot2[["PI_Pr_Theta"]],
ties.method="min")


write.csv(db_hotspot2, "results/Hotspot_Identification_Results2.csv")
View(db_hotspot2)
```

References

AASHTO. 2010. *Highway Safety Manual. National Research Council (U.S.). Transportation Research Board. Task Force on Development of the Highway Safety Manual. American Association of State Highway and Transportation Officials. Joint Task Force on the Highway Safety Manual. National Cooperativ*. American Association of State Highway and Transportation Officials.

Afghari, Amir Pooyan, Md Mazharul Haque, Simon Washington, and Tanya Smyth. 2019. "Effects of Globally Obtained Informative Priors on Bayesian Safety Performance Functions Developed for Australian Crash Data." *Accident Analysis & Prevention* 129 (August): 55–65. https://doi.org/10.1016/J.AAP.2019.04.023.

Afghari, Amir Pooyan, Simon Washington, Md Mazharul Haque, and Zili Li. 2018. "A Comprehensive Joint Econometric Model of Motor Vehicle Crashes Arising from Multiple Sources of Risk." *Analytic Methods in Accident Research* 18 (June): 1–14. https://doi.org/10.1016/J.AMAR.2018.03.002.

Amadori, Marco, and Tommaso Bonino. 2012. "A Methodology to Define the Level of Safety of Public Transport Bus Stops, Based on the Concept of Risk." *Procedia - Social and Behavioral Sciences* 48 (January): 653–62. https://doi.org/10.1016/J.SBSPRO.2012.06.1043.

Amiri, Amir Mohammadian, Navid Nadimi, Vahid Khalifeh, and Moe Shams. 2021. "GIS-Based Crash Hotspot Identification: A Comparison among Mapping Clusters and Spatial Analysis Techniques." *Https://Doi.Org/10.1080/17457300.2021.1925924* 28 (3): 325–38. https://doi.org/10.1080/17457300.2021.1925924.

Anastasopoulos, Panagiotis Ch, and Fred L. Mannering. 2009. "A Note on Modeling Vehicle Accident Frequencies with Random-Parameters Count Models." *Undefined* 41 (1): 153–59. https://doi.org/10.1016/J.AAP.2008.10.005.

Arévalo-Támara, Andrea, Mauricio Orozco-Fontalvo, and Victor Cantillo. 2020. "Factors Influencing Crash Frequency on Colombian Rural Roads." *Promet - Traffic&Transportation* 32 (4): 449–60. https://doi.org/10.7307/PTT.V32I4.3385.

Barbour, A. D., Lars. Holst, and Svante. Janson. 1992. *Poisson Approximation*. Oxford [England] New York: Clarendon Press ; Oxford University Press.

Batten, John. 2017. "SUSTAINABLE CITIES MOBILITY INDEX 2017."

Borhan, Muhamad Nazri, Muhamad Razuhanafi, and Mat Yazid. 2019. "Determining Spatial Patterns of Road Accidents at Expressway by Applying Getis-Ord Gi* Spatial Statistic Assessing and Modeling Adverse Impacts of Roadside Service Facilities (RSFs) on Crash

Frequency and Severity Along Malaysian's Expressway View Project EXPERT SYSTEM FOR GREEN TRANSPORTATION SYSTEM View Project." *International Journal of Recent Technology and Engineering*, 2277–3878. https://doi.org/10.35940/ijrte.C1004.1183S319.

Cameron, A. Colin, and Pravin K. Trivedi. 1998. "Regression Analysis of Count Data." *Regression Analysis of Count Data*. https://doi.org/10.1017/CBO9780511814365.

Carriquiry, Alicia L, and Michael Pawlovich. 2004. "From Empirical Bayes to Full Bayes: Methods for Analyzing Traffic Safety Data." https://rosap.ntl.bts.gov/view/dot/24410.

"Census Profile, 2016 Census." 2017. 2017. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E.

Chen, Peng, and Jiangping Zhou. 2016. "Effects of the Built Environment on Automobile-Involved Pedestrian Crash Frequency and Risk." *Journal of Transport & Health* 3 (4): 448–56. https://doi.org/10.1016/J.JTH.2016.06.008.

Chen, Tiantian, N. N. Sze, Sikai Chen, and Samuel Labi. 2020. "Urban Road Space Allocation Incorporating the Safety and Construction Cost Impacts of Lane and Footpath Widths." *Journal of Safety Research* 75 (December): 222–32. https://doi.org/10.1016/J.JSR.2020.09.014.

Cheng, Lingzi, Srinivas Reddy Geedipally, and Dominique Lord. 2013. "The Poisson–Weibull Generalized Linear Model for Analyzing Motor Vehicle Crash Data." *Safety Science* 54 (April): 38–42. https://doi.org/10.1016/J.SSCI.2012.11.002.

Chin, Hoong Chor, and Mohammed Abdul Quddus. 2003. "Applying the Random Effect Negative Binomial Model to Examine Traffic Accident Occurrence at Signalized Intersections." *Accident Analysis & Prevention* 35 (2): 253–59. https://doi.org/10.1016/S0001-4575(02)00003-9.

Cinnamon, Jonathan, Nadine Schuurman, and S. Morad Hameed. 2011. "Pedestrian Injury and Human Behaviour: Observing Road-Rule Violations at High-Incident Intersections." *PLOS ONE* 6 (6): e21063. https://doi.org/10.1371/JOURNAL.PONE.0021063.

Corazza, Maria Vittoria, and Nicola Favaretto. 2019. "A Methodology to Evaluate Accessibility to Bus Stops as a Contribution to Improve Sustainability in Urban Mobility." *Sustainability 2019, Vol. 11, Page 803* 11 (3): 803. https://doi.org/10.3390/SU11030803.

Craig, Curtis M, Nichole L Morris, Ron Van Houten, and David Mayou. 2019. "Pedestrian Safety and Driver Yielding Near Public Transit Stops." *Transportation Research Record* 2673 (1): 514–23. https://doi.org/10.1177/0361198118822313.

Distefano, Natalia, and Salvatore Leonardi. 2019. "Evaluation of the Benefits of Traffic Calming on Vehicle Speed Reduction.” *Civil Engineering and Architecture* 7 (4): 200–214. https://doi.org/10.13189/cea.2019.070403.

Ewing, Reid, and Eric Dumbaugh. 2009. “The Built Environment and Traffic Safety: A Review of Empirical Evidence.” *Http://Dx.Doi.Org/10.1177/0885412209335553* 23 (4): 347–67. https://doi.org/10.1177/0885412209335553.

Fridman, Liraz, Rebecca Ling, Linda Rothman, Marie Soleil Cloutier, Colin Macarthur, Brent Hagel, and Andrew Howard. 2020. “Effect of Reducing the Posted Speed Limit to 30 Km per Hour on Pedestrian Motor Vehicle Collisions in Toronto, Canada - A Quasi Experimental, Pre-Post Study.” *BMC Public Health* 20 (1): 1–8. https://doi.org/10.1186/S12889-019-8139-5/TABLES/2.

Gelman, A, and J Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. https://books.google.ca/books?hl=en&lr=&id=c9xLKzZWoZ4C&oi=fnd&pg=PR17&dq=+Gelman+2007.+Data+analysis+using+regression+and+multilevel/hierarchical+models&ots=bcRcPXSouf&sig=wdw6KZLpp1jhBpungDJQPB4PfDo.

“Geoindex - Public.” n.d. Accessed May 8, 2021. https://geoapp.bibl.ulaval.ca/login.

Getis, Arthur, and J. K. Ord. 1992. “The Analysis of Spatial Association by Use of Distance Statistics.” *Geographical Analysis* 24 (3): 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Gomaa Mohamed, Mohamed, Nicolas Saunier, Luis F Miranda-Moreno, and Satish V Ukkusuri. 2012. “A Clustering Regression Approach: A Comprehensive Injury Severity Analysis of Pedestrian-Vehicle Crashes in New York, US and Montreal, Canada.” https://doi.org/10.1016/j.ssci.2012.11.001.

Hauer, Ezra. 1997. “OBSERVATIONAL BEFORE-AFTER STUDIES IN ROAD SAFETY -- ESTIMATING THE EFFECT OF HIGHWAY AND TRAFFIC ENGINEERING MEASURES ON ROAD SAFETY.” *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*, 175–221.

Hess, Paul, Anne Moudon, and Julie Matlick. 2004. “Pedestrian Safety and Transit Corridors.” *Journal of Public Transportation* 7 (2): 73–93. https://doi.org/10.5038/2375-0901.7.2.5.

Hilbe, Joseph M. 2011. “Negative Binomial Regression, Second Edition.” *Negative Binomial Regression, Second Edition*, January, 1–553. https://doi.org/10.1017/CBO9780511973420.

Hosseinpour, Mehdi, Ahmad Shukri Yahaya, and Ahmad Farhan Sadullah. 2014. “Exploring the Effects of Roadway Characteristics on the Frequency and Severity of Head-on Crashes: Case Studies from Malaysian Federal Roads.” *Accident Analysis and Prevention* 62: 209–22. https://doi.org/10.1016/j.aap.2013.10.001.

Huang, H, HC Chin, MM Haque - Transportation Research, and undefined 2009. 2009. "Empirical Evaluation of Alternative Approaches in Identifying Crash Hot Spots: Naive Ranking, Empirical Bayes, Full Bayes Methods." *Journals.Sagepub.Com*, no. 2103: 32–41. https://doi.org/10.3141/2103-05.

Islam, A.S.M. Mohaiminul, Mohammadali Shirazi, and Dominique Lord. 2022. "Finite Mixture Negative Binomial-Lindley for Modeling Heterogeneous Crash Data with Many Zero Observations." *Accident Analysis & Prevention* 175 (September): 106765. https://doi.org/10.1016/J.AAP.2022.106765.

Khattak, Muhammad Wisal, Ali Pirdavani, Pieter De Winne, Tom Brijs, and Hans De Backer. 2021. "Estimation of Safety Performance Functions for Urban Intersections Using Various Functional Forms of the Negative Binomial Regression Model and a Generalized Poisson Regression Model." *Accident Analysis & Prevention* 151 (March): 105964. https://doi.org/10.1016/J.AAP.2020.105964.

Khazraee, S. Hadi, Valen Johnson, and Dominique Lord. 2018. "Bayesian Poisson Hierarchical Models for Crash Data Analysis: Investigating the Impact of Model Choice on Site-Specific Predictions." *Accident; Analysis and Prevention* 117 (August): 181–95. https://doi.org/10.1016/J.AAP.2018.04.016.

Lakhotia, Sneha, Sylvain Lassarre, K. Ramachandra Rao, and Geetam Tiwari. 2020a. "Pedestrian Accessibility and Safety around Bus Stops in Delhi." *IATSS Research* 44 (1): 55–66. https://doi.org/10.1016/J.IATSSR.2019.07.001.

———. 2020b. "Pedestrian Accessibility and Safety around Bus Stops in Delhi." *IATSS Research* 44 (1): 55–66. https://doi.org/10.1016/j.iatssr.2019.07.001.

Lee, Alexander S., Wei Hua Lin, Gurdiljot Singh Gill, and Wen Cheng. 2018. "An Enhanced Empirical Bayesian Method for Identifying Road Hotspots and Predicting Number of Crashes." *Https://Doi.Org/10.1080/19439962.2018.1450314* 11 (5): 562–78. https://doi.org/10.1080/19439962.2018.1450314.

Lee, J, K Chung, S Kang - Accident Analysis & Prevention, and Undefined 2016. 2016. "Evaluating and Addressing the Effects of Regression to the Mean Phenomenon in Estimating Collision Frequencies on Urban High Collision Concentration." *Elsevier*. https://www.sciencedirect.com/science/article/pii/S0001457516303050.

Lee, Jaeyoung, Mohamed Abdel-Aty, Keechoo Choi, and Helai Huang. 2015a. "Multi-Level Hot Zone Identification for Pedestrian Safety." *Accident Analysis and Prevention* 76: 64–73. https://doi.org/10.1016/j.aap.2015.01.006.

———. 2015b. "Multi-Level Hot Zone Identification for Pedestrian Safety." *Accident Analysis and Prevention* 76: 64–73. https://doi.org/10.1016/j.aap.2015.01.006.

Leur, Paul de. 2010. “Collision Cost Study.”

Litman, Todd. 1999. “Victoria Transport Policy Institute ‘Efficiency-Equity-Clarity’ Traffic Calming Benefits, Costs and Equity Impacts.” *Phone & Fax*. www.walkable.org.

Lord, Dominique, and Srinivas Reddy Geedipally. 2011. “The Negative Binomial–Lindley Distribution as a Tool for Analyzing Crash Data Characterized by a Large Amount of Zeros.” *Accident Analysis & Prevention* 43 (5): 1738–42. https://doi.org/10.1016/J.AAP.2011.04.004.

Lord, Dominique, Xiao Qin, and Srinivas R. Geedipally. 2021. “Highway Safety Analytics and Modeling.” *Highway Safety Analytics and Modeling*, January, 1–488. https://doi.org/10.1016/B978-0-12-816818-9.01001-5.

Lord, Dominique, Simon Washington, and John N. Ivan. 2007. “Further Notes on the Application of Zero-Inflated Models in Highway Safety.” *Accident; Analysis and Prevention* 39 (1): 53–57. https://doi.org/10.1016/J.AAP.2006.06.004.

Lord, Dominique, Simon P. Washington, and John N. Ivan. 2005. “Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory.” *Accident; Analysis and Prevention* 37 (1): 35–46. https://doi.org/10.1016/J.AAP.2004.02.004.

Mannering, Fred L., Venky Shankar, and Chandra R. Bhat. 2016. “Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data.” *Analytic Methods in Accident Research* 11 (September): 1–16. https://doi.org/10.1016/J.AMAR.2016.04.001.

Meng, Yi, Lingtao Wu, Chaolun Ma, Xiaoyu Guo, and Xiubin Wang. 2020. “A Comparative Analysis of Intersection Hotspot Identification: Fixed vs. Varying Dispersion Parameters in Negative Binomial Models.” *Https://Doi.Org/10.1080/19439962.2020.1779421* 14 (2): 305–22. https://doi.org/10.1080/19439962.2020.1779421.

Miaou, Shaw-Pin, Roger P. Bligh, and Dominique Lord. 2005. “Developing Guidelines for Median Barrier Installation.” *Transportation Research Record: Journal of the Transportation Research Board* 1904 (1): 2–19. https://doi.org/10.1177/0361198105190400101.

Miranda-Moreno, LF, and L Fu. 2007. “Traffic Safety Study: Empirical Bayes or Full Bayes?” https://trid.trb.org/view/801811.

Mukherjee, Deotima, K. Ramachandra Rao, and Geetam Tiwari. 2022. “Built-Environment Risk Assessment for Pedestrians near Bus-Stops: A Case Study in Delhi.” *Https://Doi.Org/10.1080/17457300.2022.2109175*. https://doi.org/10.1080/17457300.2022.2109175.

Olkin, Ingram., Leon Jay. Gleser, and Cyrus. Derman. 1994. “Probability Models and

Applications," 715.

Olkin, Ingram., Leon Jay Gleser, and Cyrus Derman. 1980. "Probability Models and Applications," 576.

Ord, J. K., and Arthur Getis. 1995. "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application." *Geographical Analysis* 27 (4): 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x.

PASHKEVICH, Anton, and Michał NOWAK. 2017. "Road Safety Risk Assessment at Pedestrian Crossings: A Case Study from Sułkowice." *Scientific Journal of Silesian University of Technology. Series Transport* 95 (June): 159–70. https://doi.org/10.20858/sjsutst.2017.95.15.

Paul P Jovanis, Hsin-Li Chang. 1986. "Modeling the Relationship of Accident to Miles Traveled." *Transportation Research Record Journal of the Transportation Research Board 1068*. https://www.researchgate.net/publication/246851294_Modeling_the_relationship_of_accident_to_miles_traveled.

Pulugurtha, Srinivas S., and L. Prasanna Srirangam. 2021. "Pedestrian Safety at Intersections near Light Rail Transit Stations." *Public Transport*, June, 1–26. https://doi.org/10.1007/S12469-021-00276-Y/FIGURES/3.

Pulugurtha, Srinivas S., and Vinay K. Vanapalli. 2008. "Hazardous Bus Stops Identification: An Illustration Using GIS." *Journal of Public Transportation* 11 (2): 4. https://doi.org/http://doi.org/10.5038/2375-0901.11.2.4.

Pulugurtha, Srinivas Subrahmanyam, and Eshwar N. Penkey. 2010. "Assessing Use of Pedestrian Crash Data to Identify Unsafe Transit Service Segments for Safety Improvements." *Transportation Research Record*, no. 2198 (December): 93–102. https://doi.org/10.3141/2198-11.

Pulugurtha, Srinivas, Vinod Vasudevan, Shashi Nambisan, and Mukund Dangeti. 2012. "Evaluating Effectiveness of Infrastructure-Based Countermeasures for Pedestrian Safety:" *Https://Doi.Org/10.3141/2299-11* 2299 (2299): 100–109. https://doi.org/10.3141/2299-11.

Quistberg, D. Alex, Thomas D. Koepsell, Brian D. Johnston, Linda Ng Boyle, J. Jaime Miranda, and Beth E. Ebel. 2015. "Bus Stops and Pedestrian–Motor Vehicle Collisions in Lima, Peru: A Matched Case–Control Study." *Injury Prevention* 21 (e1): e15–22. https://doi.org/10.1136/INJURYPREV-2013-041023.

Rahman, Muhammad Tauhidur, Arshad Jamal, and Hassan M. Al-Ahmadi. 2020. "Examining Hotspots of Traffic Collisions and Their Spatial Relationships with Land Use: A GIS-Based Geographically Weighted Regression Approach for Dammam, Saudi Arabia." *ISPRS International Journal of Geo-Information* 9 (9): 1–22. https://doi.org/10.3390/ijgi9090540.

"Road Assets (Complete Database - Pavement, Island, Intersection, Sidewalk, Area) - Open Government Portal." n.d. Accessed February 17, 2022. https://open.canada.ca/data/en/dataset/0acbc6c8-bbfc-4aae-a0fa-ec74ba0686c6.

"Road Collisions - Open Government Portal." 2021. 2021. https://open.canada.ca/data/en/dataset/cd722e22-376b-4b89-9bc2-7c7ab317ef6b.

"Road Signs (Excluding Parking) - Open Government Portal." n.d. Accessed May 8, 2021. https://open.canada.ca/data/en/dataset/c5bf5e9c-528b-4c28-b52f-218215992e35.

Shaik, Md Ebrahim, and Quazi Sazzad Hossain. 2020. "Application of Statistical Models: Parameters Estimation of Road Accident in Bangladesh." *SN Computer Science* 1 (5): 1–10. https://doi.org/10.1007/S42979-020-00284-Y/TABLES/4.

Shankar, V., J. Milton, and F. Mannering. 1997. "Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical Inquiry." *Accident Analysis and Prevention* 29 (6): 829–37. https://doi.org/10.1016/S0001-4575(97)00052-3.

Shankar, Venkataraman N., Richard B. Albin, John C. Milton, and Fred L. Mannering. 1998. "Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model:" *Https://Doi.Org/10.3141/1635-06*, no. 1635 (January): 44–48. https://doi.org/10.3141/1635-06.

Shirani-bidabadi, Niloufar, Naveen Mallipaddi, Kirolos Haleem, and Michael Anderson. 2020a. "Developing Bicycle-Vehicle Crash-Speci Fi c Safety Performance Functions in Alabama Using Di Ff Erent Techniques." *Accident Analysis and Prevention* 146 (August): 105735. https://doi.org/10.1016/j.aap.2020.105735.

———. 2020b. "Developing Bicycle-Vehicle Crash-Specific Safety Performance Functions in Alabama Using Different Techniques." *Accident Analysis and Prevention* 146 (October): 105735. https://doi.org/10.1016/j.aap.2020.105735.

Shirazi, Mohammadali, Soma Sekhar Dhavala, Dominique Lord, and Srinivas Reddy Geedipally. 2017. "A Methodology to Design Heuristics for Model Selection Based on the Characteristics of Data: Application to Investigate When the Negative Binomial Lindley (NB-L) Is Preferred over the Negative Binomial (NB)." *Accident Analysis & Prevention* 107 (October): 186–94. https://doi.org/10.1016/J.AAP.2017.07.002.

Songchitruksa, Praprut, and Xiaosi Zeng. 2010a. "Getis-Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data." *Transportation Research Record*, no. 2165: 42–51. https://doi.org/10.3141/2165-05.

———. 2010b. "Getis-Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data." *Transportation Research Record*, no. 2165 (January): 42–51.

https://doi.org/10.3141/2165-05.

Stipancic, Joshua, Luis Miranda-Moreno, Jillian Strauss, and Aurélie Labbe. 2020a. "Pedestrian Safety at Signalized Intersections: Modelling Spatial Effects of Exposure, Geometry and Signalization on a Large Urban Network." *Accident Analysis and Prevention* 134 (January). https://doi.org/10.1016/J.AAP.2019.105265.

———. 2020b. "Pedestrian Safety at Signalized Intersections: Modelling Spatial Effects of Exposure, Geometry and Signalization on a Large Urban Network." *Accident Analysis and Prevention* 134 (January). https://doi.org/10.1016/j.aap.2019.105265.

T.T, Ahmed. 2019. "Statistical Approaches to Crash Frequency Modelling: An Overview of Poisson and Negative Binomial Models." *Journal of Advances in Civil Engineering and Management* 2 (3). http://hbrppublication.com/OJS/index.php/JACEM/article/view/943.

"Traffic Lights — Pedestrian Lights - Open Government Portal." n.d. Accessed February 17, 2022. https://open.canada.ca/data/en/dataset/9a42fb1d-7656-4839-8ac6-0f5d2b519503.

"Traffic Lights — Vehicle and Pedestrian Counts at Intersections with Lights - Open Government Portal." 2021a. 2021. https://open.canada.ca/data/en/dataset/584de76b-13b9-47ea-af12-0c37b8eb5de5.

———. "Traffic Lights — Vehicle and Pedestrian Counts at Intersections with Lights - Open Government Portal." 2021b. 2021. https://open.canada.ca/data/en/dataset/584de76b-13b9-47ea-af12-0c37b8eb5de5.

Transport for London. 2014. "Assessing Transport Connectivity in London." content.tfl.gov.uk/connectivity-assessment-guide.pdf.

Truong, Long Tien, and Sekhar V.C. Somenahalli. 2011. "Using GIS to Identify Pedestrian-Vehicle Crash Hot Spots and Unsafe Bus Stops." *Journal of Public Transportation* 14 (1): 99–114. https://doi.org/10.5038/2375-0901.14.1.6.

Tulu, Getu Segni, Simon Washington, Md Mazharul Haque, and Mark J. King. 2015. "Investigation of Pedestrian Crashes on Two-Way Two-Lane Rural Roads in Ethiopia." *Accident Analysis & Prevention* 78 (May): 118–26. https://doi.org/10.1016/J.AAP.2015.02.011.

Ukkusuri, Satish, Samiul Hasan, and H. M.Abdul Aziz. 2011. "Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency." *Transportation Research Record*, no. 2237 (December): 98–106. https://doi.org/10.3141/2237-11.

Ukkusuri, Satish, Luis F. Miranda-Moreno, Gitakrishnan Ramadurai, and Jhael Isa-Tavarez. 2012. "The Role of Built Environment on Pedestrian Crash Frequency." *Safety Science* 50 (4):

1141–51. https://doi.org/10.1016/j.ssci.2011.09.012.

Ulak, Mehmet Baran, Ayberk Kocatepe, Anil Yazici, Eren Erman Ozguven, and Ashutosh Kumar. 2021. “A Stop Safety Index to Address Pedestrian Safety around Bus Stops.” *Safety Science* 133 (January): 105017. https://doi.org/10.1016/J.SSCI.2020.105017.

Ulak, Mehmet Baran, Eren Erman Ozguven, O. Arda Vanli, and Mark W. Horner. 2019. “Exploring Alternative Spatial Weights to Detect Crash Hotspots.” *Computers, Environment and Urban Systems* 78 (August): 1–9. https://doi.org/10.1016/J.COMPENVURBSYS.2019.101398.

Ulak, Mehmet Baran, Eren Erman Ozguven, Omer Arda Vanli, Maxim A. Dulebenets, and Lisa Spainhour. 2018. “Multivariate Random Parameter Tobit Modeling of Crashes Involving Aging Drivers, Passengers, Bicyclists, and Pedestrians: Spatiotemporal Variations.” *Accident Analysis & Prevention* 121 (December): 1–13. https://doi.org/10.1016/J.AAP.2018.08.031.

Venables, W. N., and B. D. Ripley. 2002. “Modern Applied Statistics with S,” Statistics and Computing, . https://doi.org/10.1007/978-0-387-21706-2.

Vignali, Valeria, Federico Cuppi, Ennia Acerra, Arianna Bichicchi, Claudio Lantieri, Andrea Simone, and Marco Costa. 2019. “Effects of Median Refuge Island and Flashing Vertical Sign on Conspicuity and Safety of Unsignalized Crosswalks.” *Transportation Research Part F: Traffic Psychology and Behaviour* 60 (January): 427–39. https://doi.org/10.1016/J.TRF.2018.10.033.

Wang, Xuesong, Yingying Pei, Minming Yang, and Jinghui Yuan. 2021. “Meso-Level Hotspot Identification for Suburban Arterials.” *Accident Analysis & Prevention* 156 (June): 106148. https://doi.org/10.1016/J.AAP.2021.106148.

Wang, Yiyi, and Kara M. Kockelman. 2013. “A Poisson-Lognormal Conditional-Autoregressive Model for Multivariate Spatial Analysis of Pedestrian Crash Counts across Neighborhoods.” *Accident Analysis and Prevention* 60: 71–84. https://doi.org/10.1016/j.aap.2013.07.030.

Who. 2013. “Pedestrian Safety: A Road Safety Manual for Decision-Makers and Practitioners.” *World Health Organization*, 120. http://apps.who.int/iris/bitstream/10665/79753/1/9789241505352_eng.pdf?ua=1.

Wu, Lingtao, Yajie Zou, Dominique Lord, L Wu, and D Lord. 2014. “Comparison of Sichel and Negative Binomial Models in Hot Spot Identification.” *Journals.Sagepub.Com* 2460 (1): 107–16. https://doi.org/10.3141/2460-12.

Xie, Kun, Kaan Ozbay, Abdullah Kurkcu, and Hong Yang. 2017. “Analysis of Traffic Crashes Involving Pedestrians Using Big Data: Investigation of Contributing Factors and Identification of Hotspots.” *Risk Analysis* 37 (8): 1459–76. https://doi.org/10.1111/risa.12785.

Yan, Ying, Ying Zhang, Xiangli Yang, Jin Hu, Jinjun Tang, and Zhongyin Guo. 2020. "Crash Prediction Based on Random Effect Negative Binomial Model Considering Data Heterogeneity." *Physica A: Statistical Mechanics and Its Applications* 547 (June): 123858. https://doi.org/10.1016/J.PHYSA.2019.123858.

Ye, Zhirui, Chao Wang, Yongbo Yu, Xiaomeng Shi, and Wei Wang. 2016. "Modeling Level-of-Safety for Bus Stops in China." *Traffic Injury Prevention* 17 (6): 656–61. https://doi.org/10.1080/15389588.2015.1133905/SUPPL_FILE/GCPI_A_1133905_SM1023.PDF.

Zegeer, Charles V., and Max Bushell. 2012. "Pedestrian Crash Trends and Potential Countermeasures from around the World." *Accident Analysis & Prevention* 44 (1): 3–11. https://doi.org/10.1016/J.AAP.2010.12.007.

Zha, Liteng, Dominique Lord, and Yajie Zou. 2016. "The Poisson Inverse Gaussian (PIG) Generalized Linear Regression Model for Analyzing Motor Vehicle Crash Data." *Journal of Transportation Safety and Security* 8 (1): 18–35. https://doi.org/10.1080/19439962.2014.977502.

Zhang, Yaohua, Sha A. Mamun, John N. Ivan, Nalini Ravishanker, and Khademul Haque. 2015. "Safety Effects of Exclusive and Concurrent Signal Phasing for Pedestrian Crossing." *Accident; Analysis and Prevention* 83 (July): 26–36. https://doi.org/10.1016/J.AAP.2015.06.010.

Zhang, Yi, Yicheng Zhang, and Rong Su. 2021. "Pedestrian-Safety-Aware Traffic Light Control Strategy for Urban Traffic Congestion Alleviation." *IEEE Transactions on Intelligent Transportation Systems* 22 (1): 178–93. https://doi.org/10.1109/TITS.2019.2955752.

Zhao, Jinbao, Wei Deng, Yan Song, and Yueran Zhu. 2013. "What Influences Metro Station Ridership in China? Insights from Nanjing." *Cities* 35 (December): 114–24. https://doi.org/10.1016/J.CITIES.2013.07.002.

Ziakopoulos, Apostolos, and George Yannis. 2020. "A Review of Spatial Approaches in Road Safety." *Accident Analysis and Prevention* 135 (February). https://doi.org/10.1016/J.AAP.2019.105323.