NEW INSIGHTS INTO ACTIVE TRANSPORTATION SAFETY: A MACRO-LEVEL ANALYSIS FRAMEWORK

by

Ahmed Osama Amer

B.Sc., Ain Shams University, 2010
M.Sc., Ain Shams University, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in
THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

February 2018

© Ahmed Osama Amer, 2018
Abstract

City councils worldwide have shown an increasing interest in active transportation (AT) due to its health, environmental, and economical benefits. However, active commuters are vulnerable to severe crash risk, which is a deterrent to active travel. Therefore, there is a need for developing systematic approaches to improve AT safety. This dissertation introduces a comprehensive framework for identifying, diagnosing and remediating the macro-level AT safety issues. It provides original insights into AT networks, crash models (CM), crash hot zones identification (HZID), and policy recommendations.

Data were collected from 134 traffic analysis zones (TAZs) in the City of Vancouver. Cyclist and pedestrian crash data, traffic exposure and large GIS data were incorporated in the analysis. The GIS data integrated various land use, built environment, socioeconomic, and road facility features. Moreover, bike and pedestrian network indicators, developed using graph-theory and representing connectivity, continuity, and topography of the networks, were incorporated. The state of the practice empirical Bayesian (EB) method and the state of the art full Bayesian (FB) methods were adopted for the CMs’ development and HZID. Various FB model forms were investigated, and the Spatial Poisson-Lognormal model performed the best.

Cyclist and pedestrian crashes were found positively associated with various attributes of network-connectivity, socio-demographics, built environment, arterial-collector roads, and commercial areas. Conversely, the crashes were negatively associated with various attributes of network-directness, network-topography, residential areas, recreational areas, local roads,
separated paths, and actuated signals. Most of the safety correlates had similar effects for the pedestrian and cyclist crashes. Accordingly, mixed multi-response FB CMs were developed and the correlation between pedestrian and cyclist crashes was found significant. The univariate/multivariate CMs with spatial effects consistently outperformed those without, and the multivariate CMs generally outperformed the univariate ones.

AT crash hot-zones were then identified using the novel Mahalanobis distance and the conventional potential for safety improvement (PSI) methods, and consistency tests were applied to compare both. Afterwards, trigger variables were statistically identified for the crash hot and safe zones. Lastly, remedies regarding land use, traffic demand, and traffic supply management were proposed based on the trigger variables’ analysis, field studies, and literature consultation.
Lay Summary

This dissertation introduces a comprehensive framework for identifying, diagnosing and remedying active transportation (AT) safety issues. It allows policymakers to recommend feasible countermeasures for improving AT safety by providing them with a better understanding of the determinant factors. A comprehensive dataset that covered the city of Vancouver was collected and used to model the associations of cyclist and pedestrian crashes with numerous zonal characteristics such as traffic exposure, socio-demographics, road facility, land use, built environment, and bike and pedestrian networks’ indicators. Advanced statistical techniques were adopted to develop various crash models for investigating the impact of the abovementioned covariates on AT safety. Afterwards, AT crash hot and safe zones were identified using a novel methodology, and trigger variables were detected for those zones. Lastly, trigger variables’ analysis along with field studies and literature consultation were conducted to suggest various strategies for improving AT safety at the crash hot zones.
Preface

The thesis chapters are published as 5 papers in reputable journals and 2 papers in refereed conferences. In addition, 2 papers are under review in a prominent journal. These are as follows:

Journal Papers

Parts of the introductory text in Chapter 1, the literature review in Chapter 2, the data collection in chapter 3, the methodology in chapter 4, and the results in chapter 7 have been published in:


I conducted all the analysis and wrote most of the manuscript under the supervision of Dr. Sayed.

Also, parts of the introductory text in Chapter 1, the literature review in Chapter 2, the data collection in chapter 3, the methodology in chapter 4, and the results in chapter 5 have been published in:


I conducted all the analysis and wrote most of the manuscript under the supervision of Dr. Sayed.
Moreover, parts of the introductory text in Chapter 1, the literature review in Chapter 2, the data collection in Chapter 3, the methodology in Chapter 4, and the results in Chapter 6 have been published in:


I conducted all the analysis and wrote most of the manuscript under the supervision of Dr. Sayed.

**Conference Papers**

Parts of the introductory text in Chapter 1, the literature review in Chapter 2, the data collection in Chapter 3, the methodology in Chapter 4, and the results in Chapter 8 have been presented as follows:


I conducted all the analysis and wrote most of the manuscript under the supervision of Dr. Sayed and Dr. Sacchi.
Additionally, parts of the introductory text in Chapter 1, the literature review in Chapter 2, the data collection in chapter 3, portions of the methodology in chapter 4, and the results in chapter 5 and 6 have been presented as follows:


I conducted all the analysis and wrote most of the manuscript under the supervision of Dr. Sayed.

**Papers Being Reviewed**

Lastly, parts of the literature review in Chapter 2, the data collection in chapter 3, the methodology in chapter 4, and the results in chapters 7, 8, and 9 are in two submitted papers that are under review in a prominent journal as follows:

- “A Cross-Comparison of Different Techniques for Modeling Macro-Level Cyclist Crashes.”

- “A Framework for Identifying, Diagnosing and Treating Active Transportation Crash-Prone Zones; City of Vancouver as a Case Study.”
# Table of Contents

Abstract .............................................................................................................................................. ii

Lay Summary ..................................................................................................................................... iv

Preface ................................................................................................................................................ v

Table of Contents ............................................................................................................................... viii

List of Figures ..................................................................................................................................... xv

List of Symbols ................................................................................................................................... xvii

List of Abbreviations ......................................................................................................................... xviii

Acknowledgements ............................................................................................................................. xix

Chapter 1: Introduction ..................................................................................................................... 1

1.1 Motivation ................................................................................................................................... 1

1.2 Problem Statement ....................................................................................................................... 5

1.3 Thesis Structure ............................................................................................................................ 9

Chapter 2: Literature Review .......................................................................................................... 10

2.1 Crash Models ............................................................................................................................... 10

2.1.1 Statistical Inference Approaches ............................................................................................ 10

2.1.2 Modeling Techniques .............................................................................................................. 13

2.1.3 Unobserved Heterogeneity .................................................................................................... 15

2.1.3.1 Random Effects Models .................................................................................................. 15

2.1.3.2 Random Parameters Models .......................................................................................... 18

2.1.3.3 Multivariate (Multi-response) Models ............................................................................ 19

2.1.4 Macro-level Analysis .............................................................................................................. 24
4.2.3 FB Estimation ........................................................................................................ 71
4.2.4 Models Comparison ................................................................................................. 72

Chapter 5: Evaluating the Impact of Network Configuration on Cyclist and Pedestrian Safety

5.1 Pedestrian Safety Models ................................................................................................. 75
5.1.1 Traffic Exposure Indicators ......................................................................................... 75
5.1.2 Network Graph Indicators ......................................................................................... 76
5.1.2.1 Network Connectivity ......................................................................................... 76
5.1.2.2 Network Directness ............................................................................................ 77
5.1.2.3 Network Topography .......................................................................................... 78
5.1.2.4 Combined FB CMs with Spatial Effects ............................................................... 79
5.2 Cyclist Safety Models ................................................................................................. 80
5.2.1 Traffic Exposure Indicators ......................................................................................... 81
5.2.2 Network Graph Indicators ......................................................................................... 81
5.2.2.1 Connectivity ...................................................................................................... 82
5.2.2.2 Directness ......................................................................................................... 83
5.2.2.3 Topography ....................................................................................................... 83
5.2.2.4 Combined FB CMs with Spatial Effects ............................................................... 84

Chapter 6: Evaluating the Impact of Zonal Characteristics on Cyclist and Pedestrian Safety

6.1 Pedestrian Safety Models ................................................................................................. 88
6.1.1 Traffic Exposure Indicators ......................................................................................... 88
6.1.2 Non-Graph Zonal Characteristics .............................................................................. 89
6.1.2.1 Socio-Economic Model ................................................................. 89
6.1.2.2 Land Use Models ........................................................................ 89
6.1.2.3 Built Environment Models.......................................................... 90
6.1.2.4 Road Facility Models .................................................................. 91
6.1.2.5 Combined FB CMs with Spatial Effects ....................................... 91

6.2 Cyclist Safety Models ....................................................................... 94
6.2.1 Traffic Exposure Indicators ............................................................. 94
6.2.2 Non-Graph Zonal Characteristics .................................................. 95
  6.2.2.1 Socio-Economic Models ............................................................. 95
  6.2.2.2 Land Use Models ...................................................................... 95
  6.2.2.3 Built Environment Models ......................................................... 96
  6.2.2.4 Road Facility Models ................................................................. 97
  6.2.2.5 Combined Models FB CMs with Spatial Effects ...................... 98

Chapter 7: Univariate Joint Crash Models, Multivariate Mixed Crash Models, and
Statistical Discussion ............................................................................ 101

  7.1 Univariate Joint Crash Models ......................................................... 101
  7.2 Comparison of the Cyclist and Pedestrian Safety Correlates’ Effects .... 108
  7.3 Multivariate Mixed Crash Models .................................................... 108
  7.4 Statistical Discussion on the Univariate and Multivariate Crash Models .. 109
     7.4.1 Spatial Correlation Effects ......................................................... 110
     7.4.2 Mode Correlation Effects .......................................................... 110
     7.4.3 Spatial Correlation Effects Compared to Mode Correlation Effects ... 111
     7.4.4 Omitted Variables Effects ........................................................ 112
Chapter 8: Hot and Safe zones Identification ................................................................. 117
   8.1 The EB PSI Method ............................................................................................... 117
   8.2 The Mahalanobis Distance Method ....................................................................... 118
   8.3 Hot Zones Identification Results for PSI and Mahalanobis Distance Methods ...... 121
   8.4 Consistency Tests ................................................................................................. 124
   8.5 Active Transportation Safe Zones ........................................................................ 126

Chapter 9: Policy Recommendations ........................................................................... 128
   9.1 Hot and Safe Zones’ Trigger variables ................................................................. 128
   9.2 Discussion and Remedies ...................................................................................... 132
      9.2.1 Land Use Management ................................................................................... 133
      9.2.2 Traffic Demand Management ....................................................................... 135
      9.2.3 Traffic Supply Management ......................................................................... 137
         9.2.3.1 Road Network .......................................................................................... 137
         9.2.3.2 Transit Network ....................................................................................... 142
         9.2.3.3 Active Transportation Network ............................................................. 145

Chapter 10: Conclusions ............................................................................................. 154
   10.1 Summary .............................................................................................................. 154
   10.2 Contributions ....................................................................................................... 158
   10.3 Limitations and Future Work ............................................................................... 160

Bibliography ................................................................................................................ 164
List of Tables

Table 1 Variables Definition and Data Summary (n=134 TAZs) ................................................................. 60
Table 2 Negative Binomial GLM Analysis Estimates for Pedestrian Network Graph Indicators .... 79
Table 3 FB Analysis Estimates for Pedestrian Network Graph indicators .............................................. 80
Table 4 Negative Binomial GLM Analysis Estimates for Cyclist Network Graph Indicators .... 85
Table 5 FB Analysis Estimates for Cyclist Network Graph Indicators .................................................... 86
Table 6 Negative Binomial GLM Analysis Estimates for Pedestrian Non-Graph Zonal Characteristics .................................................................................................................. 92
Table 7 FB Analysis Estimates for Pedestrian Non-Graph Zonal Characteristics ............................ 93
Table 8 Negative Binomial GLM Analysis Estimates for Cyclist Non-Graph Zonal Characteristics .................................................................................................................. 99
Table 9 FB Analysis Estimates for Cyclist Non-Graph Zonal Characteristics ................................... 100
Table 10 Negative Binomial GLM Estimates for the Pedestrian and Cyclist Crash Joint Models ................................................................. 102
Table 11 Parameter Estimates and 95% Credible Intervals for FB Cyclist Crash Joint Models 106
Table 12 Parameter Estimates and 95% Credible Intervals for FB Pedestrian Crash Joint Models .................................................................................................................. 107
Table 13 FB Multivariate Joint Crash Model .......................................................................................... 109
Table 14 FB Traffic Exposure Crash Models ......................................................................................... 114
Table 15 FB Joint Crash Models ........................................................................................................... 115
Table 16 Variance-Covariance Matrices ............................................................................................... 116
Table 17 Trigger Variables for the HZs ................................................................................................. 131
Table 18 Summary of Land Use Safety Effects and Policy Recommendations .................. 135
Table 19 Summary of Traffic Demand Safety Effects and Policy Recommendations .......... 137
Table 20 Summary of Road Network Safety Effects and Policy Recommendations .......... 142
Table 21 Summary of Transit Network Safety Effects and Policy Recommendations .......... 144
Table 22 Summary of Active Transportation Network Safety Effects and Policy Recommendations ................................................................. 152
List of Figures

Figure 1 (a) Heat Map of Cyclist Crashes (red Pts.) and Pedestrian Crashes (yellow Pts.) (b) Cyclist Trips at City of Vancouver TAZs

Figure 2 Hierarchy of the Analysis Variables

Figure 3 Pedestrian (a) and Bike (b) Networks Characterizations at the different TAZs

Figure 4 Bike Network Coverage for the City of Vancouver’s TAZs

Figure 5 Linearity (a) Straight Links (b) Non-Straight Links (Curved or Irregular)

Figure 6 Straight Sidewalk Links (Right Image) and Non-Straight Bike Network Links (Left Image)

Figure 7 Average Slopes of the Bike Network Links

Figure 8 City of Vancouver Road Classes

Figure 9 (a) Traffic Signals Density and (b) Bus Stops Density within City of Vancouver TAZs

Figure 10 Residential, commercial, and recreational areas in the City of Vancouver

Figure 11 Multivariate distribution of the Pedestrian and Cyclist Crashes

Figure 12 Vancouver’s Active Transportation Hot Zones a) Mahalanobis Distance Method b) EB PSI Method

Figure 13 Results of the Consistency with FB Mahalanobis Distance Method

Figure 14 Safe Zones in the City of Vancouver

Figure 15 SZs’ Recurrence for each Trigger Variable

Figure 16 Boulevards for Residential and Commercial Areas (Image Courtesy of NACTO)
Figure 17 High Household and Employment Densities at Downtown Vancouver (Map data: Google, Google Maps) .................................................................................................................................................................................................................. 136

Figure 18 High Density of Arterial-Collector Roads at Downtown HZs .............................................. 138

Figure 19 Traffic Calming Measures at Local Streets and Residential Areas in Some SZs (Map data: Google, Google Maps) .............................................................................................................................................................................................................. 139

Figure 20 Bus-Bike Interactions when entering/exiting bus stops (Map data: Google, Google Maps) ............................................................................................................................................................................................................................................................................................................................................................................................................... 143

Figure 21 Recommendations from SZs for Cyclist safety on Transit Network (Map data: Google, Google Maps) ............................................................................................................................................................................................................................................................................................................................................................................................................... 144

Figure 22 Absence of Bike Facilities on Some Main Roads in the HZs Such as Howe St., Thurlow St., W. Broadway, and W.12th (Map data: Google, Google Maps) .............................. 145

Figure 23 Absence of Pedestrian and Bike Facilities on Some HZs’ Roads (Map data: Google, Google Maps) ............................................................................................................................................................................................................................................................................................................................................................................................................... 146

Figure 24 Shared Space (Map data: Google, Google Maps) ................................................................. 146

Figure 25 Active Commuters’ Safety Issues at Intersections (Map data: Google, Google Maps) ............................................................................................................................................................................................................................................................................................................................................................................................................... 148

Figure 26 Examples of Well Treated Minor-Major Intersections (Map data: Google, Google Maps) ............................................................................................................................................................................................................................................................................................................................................................................................................... 149

Figure 27 Continuous Paths for Pedestrians and Cyclists (Map data: Google, Google Maps).. 150

Figure 28 Midblock Crossings for Cyclist and Pedestrian paths (Map data: Google, Google Maps) ............................................................................................................................................................................................................................................................................................................................................................................................................... 150

Figure 29 Jaywalking due to the Absence of Midblock Facilities and Long Distance between Intersections (Map data: Google, Google Maps) ............................................................................................................................................................................................................................................................................................................................................................................................................... 151
List of Symbols

\( \theta_i \)  
Expected Crash Frequency

\( \mu_i \)  
Predicted Crash Frequency

\( \psi_s \)  
Spatial Variation Proportion

\( \sigma^2 \)  
Variance

\( \beta_i \)  
Parameter

\( \lambda_i \)  
Poisson Parameter for Crashes Distribution

\( k \)  
Dispersion Parameter

\( u_i \)  
Random Error Term

\( S_i \)  
Spatial Error Term

\( Y \)  
Collision Frequency

\( \chi^2 \)  
Chi-Square
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Active Transportation</td>
</tr>
<tr>
<td>CM</td>
<td>Crash Models</td>
</tr>
<tr>
<td>EB</td>
<td>Empirical Bayes</td>
</tr>
<tr>
<td>FB</td>
<td>Full Bayesian</td>
</tr>
<tr>
<td>ICBC</td>
<td>Insurance Company of British Columbia</td>
</tr>
<tr>
<td>GLM</td>
<td>Generalized Linear Modeling</td>
</tr>
<tr>
<td>PSI</td>
<td>Potential of Safety Improvement</td>
</tr>
<tr>
<td>PLN</td>
<td>Poisson-Lognormal</td>
</tr>
<tr>
<td>RPPLN</td>
<td>Random Parameters Poisson-Lognormal</td>
</tr>
<tr>
<td>SPLN</td>
<td>Spatial Poisson-Lognormal</td>
</tr>
<tr>
<td>TAZ</td>
<td>Traffic Analysis Zone</td>
</tr>
<tr>
<td>HZ</td>
<td>Hot Zone</td>
</tr>
<tr>
<td>SZ</td>
<td>Safe Zone</td>
</tr>
<tr>
<td>SD</td>
<td>Scaled Deviance</td>
</tr>
<tr>
<td>DIC</td>
<td>Deviance Information Criteria</td>
</tr>
</tbody>
</table>
Acknowledgements

It is beyond words to thank ALLAH, The Almighty, for everything.

I can’t thank my Mom, Dad, and Sister enough for their continuous support and sacrifice throughout this long journey.

I offer my enduring gratitude to Dr. Tarek Sayed for his invaluable contributions to this work and for inspiration.

I owe special thanks to Dr. Mohamed H. Zaki, Dr. Alex Bigazzi, and Dr. Maged Senbel for their help and advice.

Lastly, I’d like to thank all my BITSAFS lab colleagues, from whom I’ve learned a lot.
بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Al-Quraan – Surat AlFatiha – Verse 1
Chapter 1: Introduction

This chapter provides a general introduction to the thesis, and it includes three main parts: motivation, problem statement and thesis structure. The motivation part presents background information that is essential to appreciate the significance of the research problems. The problem statement part discusses the research objectives and issues. Lastly, the chapter concludes by describing the thesis structure which presents a roadmap for the thesis components.

1.1 Motivation

Road crashes are a leading cause of death globally, especially among those aged 15–29 years old (WHO, 2015). Traffic injuries and fatalities are known for the heavy burden that they place on national economies and households. Due to the growing recognition of the enormous toll exacted by road traffic crashes, the United Nations announced this decade (2011-2020) as a Decade of Action for Road Safety. In September 2015, representatives of states attending the United Nations General Assembly identified measures, such as increasing safety standards in cars, reducing drunk-driving, exploring sustainable transport, etc., that aim at halving the global number of deaths and injuries from road traffic crashes by 2020.

From that standpoint, many cities worldwide are recognizing the vital role that active transportation can play in creating safer, as well as healthier and more sustainable and livable communities. This motivates city authorities to apply various policies that would encourage this growing trend of transportation. Active transportation, with walking and cycling being its main components, is any human-propelled mode of transportation. Pedestrians are the largest single
traveller group since almost everyone walks as a part of his/her journey. Also, more people are cycling in the past decade for commuting, work, and leisure activities (Cycling UK, 2012; USA Today, 2014).

Despite some associated risks of exposure to traffic and air pollution (de Nazelle et al., 2011), the promotion of active transportation presents a promising strategy. Active transportation not only addresses the problems of energy consumption, environmental pollution, and climate change, but also provides substantial health benefits (de Hartog et al., 2010). Physical inactivity is a primary contributor to the constant elevation in rates of obesity, heart disease, diabetes, stroke, and other chronic health conditions (CDC, 2011). It is estimated that there are 3.2 million deaths worldwide per year attributable to physical inactivity (Mathers et al., 2009). Active transportation can overcome car dependence as well as increase physical activity levels (Lindsay et al., 2011). Moreover, shifting to such sustainable modes would reduce traffic congestion and vehicle crashes (Litman, 2010). The total economic benefits of active transportation in Canada can reach up to $7 billion/year at a mode share as low as 15% (Campbell and Margaret, 2004).

Regardless of the aforementioned benefits of active transportation, cyclists and pedestrians are vulnerable road users. They are usually subjected to an elevated level of perceived and actual injury risk and discomfort, which may discourage commuters from using active modes and see them as less safe than driving (Buehler et al., 2016; Lawson et al., 2012). Almost quarter of all deaths on the world’s roads is among those with the least protection, i.e., pedestrians (22%) and cyclists (4%) (Toroyan, 2013). For various types of community settings, safety concerns deter one in five Canadians from walking or cycling (Canadian Medical Association, 2013). Although
vulnerable road users usually account for a low number of reported crashes, they still account for most of the fatalities. In British Columbia, active commuters\(^1\) represent 5.5% of the injuries and 25% of the fatalities due to motor vehicle crashes (Transport Canada, 2007). The total cost to BC society of pedestrian and cycling injuries and fatalities amounts to around $1,175 million per year (The BC Cycling Coalition, 2015). Locally, in the city of Vancouver, active trips accounted for approximately 20% of all trips (City of Vancouver, 2012) and around 3% of the reported crashes between years 2007 and 2012. However, they represented approximately 50% of the fatalities due to crashes with motor vehicles over this period (ICBC, 2013). It is, therefore, essential for city officials to first develop an efficient and safe active transportation networks before encouraging more road users to shift to active commuting.

In view of this, the City of Vancouver adopted the “Transportation 2040” plan in 2012 (City of Vancouver, 2012), which provides a long-term strategic vision for transportation within the City of Vancouver. The Transportation 2040 plan includes targets to considerably increase the proportion of trips made by sustainable transport, i.e., walking, cycling, and transit, while working in parallel towards zero transportation-related fatalities and emissions. A key component of this plan is creating an active transportation network and environment that is safe and convenient for commuters of all ages and abilities.

Accordingly, the need is growing for advocating proactive planning strategies that are capable of assessing the safety and operation of active transportation networks. Although numerous studies investigated the factors that influence active transportation safety, most of those studies were

\(^1\) In this research active commuters mainly refer to cyclists and pedestrians
undertaken at a micro level. These micro-level studies looked at active commuting within specific roadway sections or intersections allowing practitioners to propose safety countermeasures and supporting policies. Although these policies can be beneficial in increasing the safety of the investigated sites, the micro-level approach has some limitations. First, it is considered a reactive approach, in which practitioners usually have to wait for crashes to occur at particular sites before proposing and implementing any interventions. Moreover, the micro-level approach focuses on specific sites with common characteristics, which makes it less suitable for the long-term strategic planning of entire zones or transportation networks.

This emphasizes the need for models that can act as viable decision-support tools for proactive safety assessment and policy planning of active transportation. Macro-level crash models can play such role efficiently. Crash models (CMs) are statistical models that can act as effective tools for proactive safety planning. In macro-level CMs, crashes are modeled as a function of wide area (e.g., neighborhood, traffic analysis zone, census tract, etc.) characteristics instead of limited road sections or intersections. An important output of the macro-level CMs is the ability to identify and rank, as well as propose appropriate countermeasures for, crash-prone areas by investigating the macro-level features associated with crashes. This can effectively contribute to early stage neighborhood planning.

In brief, the modeling of active commuters’ crashes on the macro-level with implications for transportation safety planning has received less attention compared to vehicle crashes. Therefore, developing accurate macro-level CMs for active transportation safety planning is the primary objective of this research. These models can provide guidelines for practitioners. They can
identify and evaluate policies to effectively integrate safety considerations with long-range metropolitan active transportation planning and decision-making. Identifying the crash hot and safe zones as well as the trigger variables for each, within the City of Vancouver, can be a prominent application of the developed macro-level CMs. Field studies along with literature consultation should help afterwards in suggesting appropriate safety measures and policy recommendations for the hot zones.

1.2 Problem Statement

The study of active commuters’ safety draws its importance from the physical vulnerability of this type of road user and the key role that walking and cycling play in a sustainable transportation system. Although some studies have discussed the safety of pedestrians and cyclists on the macro-level; these studies suffer from some limitations, and several research gaps still need to be covered. The research problems tackled in this thesis represent a number of novel developments in active transportation safety analysis and present solutions for various research issues as follow:

- The investigation of the associations between the cyclists’ safety and cycling correlates has been recently gaining an increased importance in the literature. However, macro-level safety models that include comprehensive zone characteristics along with better traffic exposure measures (e.g., bike kilometers travelled) were not investigated. Almost all of the previous studies used weak exposure proxies to represent cyclist ridership in the safety models, along with inadequate or unrepresentative zone characteristics. Such deficiency can lead to biased models of cyclist safety. Moreover, despite the fact that
various cyclist safety correlates were discussed in the literature, none of these studies investigated the impact of quantitative network indicators, extracted using graph-theory, on macro-level cyclist safety. This lays the ground for the first research objective:

“Develop macro-level models including actual traffic exposure measures (e.g., bike kilometers travelled) to evaluate the impact of bike network configuration (i.e., connectivity, continuity and topography), as well as various zonal characteristics (i.e., socio-demographics, land use, road network, and built environment), on cyclist safety”

• For walking, which is the other key element of active transportation, although it constitutes the highest proportion of active transportation trips, limited research was conducted to address the safety of pedestrians on the macro level. The safety correlates investigated in the previous studies were not comprehensive and were lacking many crucial zone characteristics as well as strong proxies for traffic exposure measure. In addition, the impact of quantitative indicators of the pedestrian network on pedestrian safety was not investigated before. Therefore the following research objective is designed to address this shortcoming:

“Develop macro-level models including strong traffic exposure proxies (e.g., walk trips) to evaluate the impact of pedestrian network configuration (i.e., connectivity, continuity, and topography), as well as various zonal characteristics (i.e., socio-demographics, land use, road network, and built environment), on pedestrian safety”

• It is essential to demonstrate that the modeling and estimation approaches that are used to analyze pedestrian and cyclist crashes are the most robust and are able to account for
critical statistical issues. This practical demand introduces the third research objective as follows:

“Develop crash models for pedestrians and cyclists using the state-of-the-practice empirical Bayesian approach as well as three state-of-the-art full Bayesian approaches, and assess the results in order to settle on the most robust modeling approach for macro-level analysis of cyclist and pedestrian safety”

- Two important issues that are usually overlooked when developing active transportation safety models are spatial correlation and mode (i.e., dependency between pedestrian and cyclist crashes) correlation. Limited studies have investigated models that account for both correlations concurrently. These few multi-response studies overlooked the dissimilarity between the two modes by using the same covariates for both modes. This formulates the next research objective as follows:

“Develop multivariate (multi-response) spatial models that allow the covariates to vary from one mode to another with the aim of accounting for the interactions between pedestrian and cyclist safety through flexible and robust macro-level active transportation crash models”

- Although some studies were conducted to investigate the identification of pedestrian and cyclist crash-prone zones, none of the previous studies investigated the combined active transportation crash hot zones (i.e., have higher crash frequency than normal) through accounting for the correlation between the underlying modes. Moreover, none of the previous studies attempted to investigate the active transportation safe zones (i.e., having
lower crash frequency than normal). Therefore, identifying the crash hot and safe zones for active transportation is a prominent application developed in this research. This is described in the fifth research objective as follow:

“Identify active transportation hot and safe crash zones within the City of Vancouver using the novel Mahalanobis distance method as well as the conventional PSI method, then undergo consistency tests to decide on the identification methodology with the best performance”

- Lastly, a comprehensive case study is required to demonstrate this research capacity to provide appropriate safety countermeasures for the identified crash hot zones. This lays the ground for the last research objective:

“Analyze the trigger variables within the City of Vancouver hot and safe crash zones, and then conduct field studies and literature consultation to propose policy recommendations for improving active transportation safety”

In summary, the principal objective of this research is to develop proactive assessment tools that can be used by transportation engineers and city planners to design safe and efficient active transportation networks. Such tools need to be comprehensive by including a wide range of variables that can be associated with active transportation safety. Accordingly, this research develops macro-level crash models that are capable of assessing the impact of traffic exposure, bike and pedestrian networks’ configuration, built environment, surrounding land use, socio-demographics, and road network on the safety of active commuters. The city of Vancouver traffic analysis zones are used as units of analysis in this research. Many of the covariates’
connections with active transportation safety are newly investigated (e.g., network configuration variables). The active transportation hot and safe crash zones within the city as well as the related trigger variables are then identified using a novel and robust approach. Lastly, policy recommendations are proposed according to the conducted trigger variables’ analysis, field study, and literature consultation.

1.3 Thesis Structure

This chapter covers the motivation behind this research, the problem statements and research objectives, and the research outline. Chapter 2 documents a comprehensive review of the literature. The review discusses crash modeling, safety correlates, and hot zones’ identification and remedy in the realm of transportation engineering. A narrow focus of the review is on relevant work in active transportation safety. Chapter 3 presents the data collection process and the extraction of the analysis variables employed in the crash models. Chapter 4 discusses the crash modeling methodologies followed in this research. Chapter 5 discusses the results of the crash models developed to evaluate the impact of bike and pedestrian networks’ configuration on cyclist and pedestrian safety respectively. Chapter 6 discusses the results of the crash models developed to evaluate the impact of various zone characteristics on cyclist and pedestrian safety. Chapter 7 comprises the univariate joint crash models as well as the multivariate crash models for pedestrian and cyclist safety, besides a comprehensive statistical discussion. Chapter 8 demonstrates the methodology of hot and safe crash zones’ identification as well as the ranking and locations of the hot and safe zones. Chapter 9 discusses the identification and analysis of hot and safe zones’ trigger variables as well as policy recommendations. Lastly, chapter 10 outlines the summary, contributions, limitations, and proposed future work.
Chapter 2: Literature Review

This chapter provides a comprehensive literature review of various issues associated with safety modeling. It also discusses and summarizes the results of previous studies that investigated active transportation safety. Lastly, this chapter reviews the different techniques for identifying crash hot zones in the literature.

2.1 Crash Models

Crash models (CMs) are mathematical models that relate the crash frequency at an entity (road location or traffic zone) to various attributes of this entity. They are developed using specific statistical techniques and are considered valuable tools with diverse applications such as: evaluating the effectiveness of safety improvement measures, detecting and ranking of crash-prone locations, and estimating the safety level of facilities (Sawalha and Sayed, 2001). This section comprises four subsections, i.e., statistical inference approaches of CMs, modeling techniques of CMs, unobserved heterogeneity issues, and macro-level analysis of CMs, which are discussed as follows.

2.1.1 Statistical Inference Approaches

There are two approaches to statistical inference nowadays; the frequentist approach and the Bayesian approach. The frequentist approach, or what is called point estimate inference, relies solely on the data observed and the classical likelihood; On the other hand, the Bayesian approach incorporates both likelihood and prior information (Washington et al., 2010). The Empirical Bayes (EB) and the Full Bayes (FB) approaches are two Bayesian approaches that are
now widely used in traffic safety modeling. The empirical Bayes (EB) and full Bayes (FB) methods assume that a probability distribution can be found before any data become available (prior distribution). Once information becomes available, the prior distribution is converted into a posterior distribution using Bayes theorem. The Bayesian approach provides substantial philosophical and practical advantages as argued by several researchers (e.g., Mitra and Washington, 2007). It is considered more suitable for safety modeling compared to the classical likelihood-based inference methods and has been more popular in the recent traffic safety research.

In particular, the EB approach is considered the state-of-practice inference approach in the field of traffic safety (Hauer et al., 2002). It is considered as a bridge between the frequentist approach and the fully Bayesian approach. The two most likely reasons for the popularity of the EB approach are (El Basyouny, 2011): i) the use of the maximum likelihood estimation method which is preprogrammed in many statistical softwares and ii) the simplicity of conducting the posterior analysis since closed-form functions are available for most common distributions. However, this advantage is also a limitation since the EB method cannot be used when the likelihood function is difficult to characterize (e.g. Poisson-Llognormal models). The EB approach is used to refine the estimate of the expected crash frequency of a location by combining the observed number of crashes with the predicted number of crashes from a safety performance function (SPF) of the same type of road facility (Hauer, 1997). Alternatively, the FB approach integrates the estimation of the SPF with the posterior distribution estimate of the expected crash frequency through a single step (Li et al., 2008); (El-Basyouny and Sayed, 2009); (El-Basyouny and Sayed, 2013). The EB method typically deals with the entire study period as a
single data point (either total or calculated as per year) and does not account for the spatio-temporal and multivariate model structures. The FB approach is more appealing since it treats each time period as an individual data point, allowing inference at more than one level (for hierarchical models) and multivariate as well as spatial/temporal analysis, which is one focus of this research. The full Bayes estimators were found to perform better than the EB estimators especially when working with small datasets characterized by an overall low mean accident frequency (Miranda-Moreno and Fu, 2007). More discussion about the differences between the full Bayesian and the empirical Bayesian approaches can be found in the hot zones’ identification section.

Bayesian analysis using Full Bayes (FB) hierarchical statistical models has become more popular due to its flexibility and its ability to use prior information, which results in improved parameter estimates. Recent years have witnessed enormous progress in numerical methods which, combined with the availability of free and easy to use software, have permitted the implementation of the full Bayes approach. Different from the classical models, FB models do not depend on the assumption of asymptotic normality. Sampling-based methods of FB estimation focus on estimating the entire density of parameters as compared to the traditional classical estimation methods which are intended for finding a single point estimate using maximum likelihood approach (Congdon, 2007). Although sometimes point estimates may be more convenient for practical applications, since they clearly suggest a single point, the FB approach has a significant advantage over the maximum likelihood estimation. The Bayesian estimation determines posterior density for each parameter under consideration. This density estimation is the outcome of a process where a long run or a series of long runs of samples are
taken from the posterior density based on the prior information about the parameter and data. Therefore, FB analysis can provide more accurate measures of uncertainty on the posterior distributions of the parameters’ estimates (El Basyouny and Sayed, 2009), which the frequentist approach can’t as it does not consider uncertainty in the correlation structures. That is why the frequentist approach sometimes yields an overestimation of the precision of the parameter estimates associated with the covariates. Consequently, the FB approach provides a considerable interpretive advantage since posterior estimates reflect the probabilities that the analyst is primarily interested in, the probability of the null hypothesis being true called a Bayesian Credible Interval (BCI) (Washington et al., 2005). On the other hand, classical confidence intervals of the parameter estimates provide the probability of observing data given that a parameter takes on a specific value. FB analysis was also found more suitable for spatial effects models due to its ability to implement complex correlation structures (Aguero-Valverde and Jovanis, 2008).

2.1.2 Modeling Techniques

Traditional modeling methods in the past used to consider the number of observed crashes at a site to be an unbiased estimate of the traffic safety at that site. Those conventional methods used to compare the mean crash frequency, due to various observations, to appropriate standard values in order to indicate what is considered the true level of traffic safety. However, the mean crash frequency is never known with complete accuracy, but it can only be approximately estimated using safety models. That has led many researchers to abandon the conventional analysis techniques because of the accompanied statistical uncertainty and because of their failure to acknowledge the effects of site-specific attributes (such as traffic and geometric characteristics)
on the crash occurrence (Sawalha and Sayed, 2006). Later, probabilistic methods for crash modeling emerged and were distinguished from ‘conventional’ methods in that any parameter (e.g., crash frequency at a location) is regarded as a random variable with a probability distribution (El Basyouny, 2011). Probabilistic methods recognize that crashes are rare random events and account for the stochastic effects in crash data. Within the probabilistic approach, there are several regression techniques to model road collisions.

Generalized linear modeling (GLM), which assumes a non-normal distribution error structure, is widely used for the development of CMs since conventional linear regression models lack the distributional property to adequately describe crashes. This inadequacy is due to the random, discrete, nonnegative, and typically sporadic nature that characterizes the occurrence of crashes (Miaou and Lum, 1993; Sawalha and Sayed, 2006 and 2001). A Negative Binomial (Poisson-Gamma) error distribution assumption has become the standard for micro- and macro-level GLM CMs; and plenty of models have been produced in this way (Hauer and Lovell, 1988); (Sawalha and Sayed, 2001; Wei and Lovegrove, 2013).

More recently, several researchers have proposed the use of the Poisson-Lognormal (PLN) model as an alternative to the Poisson-Gamma model for modeling crash data (Miaou et al., 2003; Lord and Miranda-Moreno, 2008; Aquero-Valverde and Jovanis, 2008). In particular, FB Poisson-Lognormal models were recommended over FB Poisson-Gamma models when assuming vague priors and whenever crash data characterized by low sample mean values were used for developing crash prediction models (Dominique Lord 2008).
In addition, a new area of research in traffic safety is the non-parametric data-driven approach. Few studies had attempted road safety analysis using such approaches such as Thakali et al. (2016) and Pan et al. (2017) studies.

2.1.3 Unobserved Heterogeneity

Missing covariates (referred to as unobserved heterogeneity), that could plausibly determine the likelihood of crash occurrence or severity level, can present serious specification problems for traditional statistical analyses. It can result in biased parameter estimates, erroneous inferences, and inaccurate crash predictions (Mannering et al., 2016). The models usually used in traffic safety are considered as global models, as the variables from these models are forced to have the same effect on all units or zones (Amoh-Gyimah, 2016). Such models may not be efficient enough for tracking the unobserved heterogeneity. Therefore using different techniques to enhance crash modeling, by incorporating, for example, random effects, or random parameters, or multi-responses, etc., were recommended in the literature. The regression models described in the section 2.1.2 can be extended to include any of the modeling techniques enhancements that are discussed as follow.

2.1.3.1 Random Effects Models

The random effects are mainly due to spatial or temporal correlations. In general, the literature shows that the CMs incorporating spatial/temporal effects performed better than those that did not since they can address the spatial/temporal heterogeneity. Temporal correlation exists due to the fact that crash data observations are likely correlated over time as many of the unobserved effects, associated with a specific site, will remain similar with time. Correspondingly, a spatial
correlation might occur since neighboring sites usually have similar environmental and geographic characteristics, and thereby form clusters with similar crash occurrences (El Basyouny, 2011).

Spatial correlation has been long-established as an important issue that should not be overlooked in the crash modeling process since crash data often include observations that are in close spatial proximity (Quddus, 2008). Incorporating spatial correlation has two main advantages: i) spatially correlated sites pool strength from neighboring sites, thereby improving model parameters’ estimation (Aguero-Valverde and Jovanis, 2008; Lord and Mannering, 2010); and ii) spatial dependence can be a surrogate for unknown and relevant covariates, thereby reflecting unmeasured factors (Chiou et al., 2014). Various studies have discussed CMs that incorporated spatial correlations with either road element-specific (intersections or road segment) or regional crashes. Time correlation was also incorporated in some safety studies and would have similar advantages to the spatial correlation, i.e., improving parameter estimates and accounting for the influence of the factors that change by time (e.g., from year to year). Wang and Abdel-Aty (2006) are among the first to address the spatial correlation among intersections at corridor-level as well as the temporal correlation among the data. They applied the generalized estimating equation (GEE) approach for modeling crash data for 476 signalized intersections from 41 corridors in the state of Florida. In addition, they investigated the longitudinal data for 208 signalized intersections over three years. Wang et al. (2013) conducted a spatio-temporal analysis to explore the relationship between traffic congestion and road accidents. They developed a series of classical count outcome models (random-effects negative binomial models) and spatio-temporal models using a full Bayesian hierarchical approach. The spatio-temporal
models better fit the data, and the level of traffic congestion was found positively associated with serious injuries. Also, Aguero-Valverde and Jovanis (2010) explored the spatial correlation in multilevel crash frequency models for different types of urban and rural road segments, and found out that spatial effects CMs performed better than the CMs without spatial effects. Brijs et al. (2008) studied the effect of weather conditions on daily crash counts using a discrete time-series model. They found out that several assumptions related to the effect of weather conditions on crash counts were significant in the data and that if serial temporal correlation is not accounted for in the model, this may produce biased results.

More specifically, some studies have addressed the spatial and temporal error correlation effects when modeling active modes’ crashes. Siddiqui et al. (2012) used a Bayesian spatial framework to model pedestrian and bicycle crashes in TAZs and investigate the effects of spatial correlation. Macro-level models for pedestrian and cyclist crashes were estimated as a function of variables related to roadway characteristics, and various demographic and socio-economic factors. They found that the Bayesian models with spatial correlation performed better than the models that did not account for spatial correlation. DiMaggio (2015) quantified the spatiotemporal risk of pedestrian and cyclist injury in New York City on the census tract level over a 10-year period. He identified the areas of increased risk and evaluated the role of socioeconomic and traffic-related variables in injury risk. Wang et al. (2016) investigated the association between pedestrian crash frequency and various predictor variables including roadway, socio-economic, and land-use features using Bayesian conditional autoregressive (CAR) models with seven different spatial weight features. Their results indicated that the spatial weight feature based on geometric centroid distance outperformed all the other spatial weight features.
2.1.3.2 Random Parameters Models

A contemporary approach for modeling the mean function advocates the use of random parameters (Anastasopoulos and Mannering, 2009). In the traditional models only one regression equation is fit to the dataset; but, on the other hand, a random parameters model builds up various regression equations for individual sites. This modeling technique has been considered by several studies for its added flexibility and intuitive application (e.g., Li et al., 2008; Milton et al., 2008). Random parameters can be viewed as an extension of the random intercept models since in addition to varying the model intercept, each estimated parameter in the random parameters’ model is allowed to vary across the dataset observations. This model focuses on explaining a part of the extra-variation through enhancing the mean function by accounting for the unobserved heterogeneity from one site/group of sites to another (El Basyouny, 2011).

Previous studies (e.g., Venkataraman et al., 2011) suggested that the random parameters models outperform the fixed parameters models due to their ability to efficiently account for the unobserved heterogeneity across individual observations or observation clusters. Milton et al. (2008) and Gkritza and Mannering (2008) were among the first to adopt the random parameters modeling approach in traffic safety analysis. Since then, there has been a growing body of traffic safety research focused on that approach, such as Anastasopoulos et al. (2012) study that used random-parameters tobit model to study the factors affecting highway accident rates for urban interstates. Their results show that the random-parameters model outperformed its fixed-parameters counterpart and had the potential to provide a fuller understanding of the factors determining accident rates on specific roadway segments.
Recently, few studies have addressed the issue of random parameters when modeling active modes’ crashes. Ukkusuri et al. (2011) developed a random parameter negative binomial model for predicting pedestrian crash frequencies at the census tract level. They reported the influences of a comprehensive set of variables describing the socio-demographic and built-environment characteristics on pedestrian crashes. Several parameters in the model were found to be random, which indicates their heterogeneous influence on the numbers of pedestrian crashes. Behnood and Mannering (2017) developed a random parameter multinomial logit model of bicyclist-injury severity, with heterogeneity in parameter means and variances, to explore the effects of a wide range of variables on bicyclist injury-severity outcomes. Model estimation results showed that many factors potentially affect the likelihood of severe injuries in bicycle/motor-vehicle crashes. Those factors included bicyclist and driver race and gender, alcohol-impaired bicyclists or drivers, older bicyclists, riding or driving on the wrong side of the road, drivers’ unsafe speeding, and bicyclist not wearing a helmet.

2.1.3.3 Multivariate (Multi-response) Models

Crashes have been typically analyzed by modeling each crash category separately, without taking into account the correlations that probably exist among different levels such as severity (e.g., fatal, major injury, minor injury or property damage only), or crash type (e.g., angle, head-on, rear-end, sideswipe or pedestrian-involved). These unaddressed correlations may be due to omitted variables, which can impact crash occurrence at all classification levels, or due to ignored shared information in the unobserved error terms (El Basyouny, 2011). Therefore, the univariate handling of correlated counts as independent can lead to imprecise road safety analysis. Several studies have applied multivariate models for estimating the traffic crash count
under various classifications, such as crash severity level (Park and Lord, 2007), or crash type (Mothafer et al., 2016).

Recently, multivariate crash models incorporating spatial effects have been attracting much interest in crash modeling. There has been an increasing effort to utilize the correlation between adjacent sites along with the correlation among crash types in order to mitigate problems resulting from the unobserved heterogeneity (Aguero-Valverde et al., 2016). Aguero-Valverde (2013) used a full Bayesian hierarchical approach to estimate crash frequency models on canton level. He compared a multivariate conditional autoregressive model to its univariate counterpart, and the results showed that the multivariate spatial model performed better than the univariate one. Narayananmoorthy et al. (2013) applied a modeling framework to predict injury counts at a Census tract level, based on crash data from Manhattan, New York. They proposed a spatial multivariate count model to jointly analyze the injury severity for pedestrian and cyclist crashes. Their study results emphasized the need to use multivariate modeling for the analysis of injury counts by road-user type and injury severity level while accommodating for spatial dependence.

Barua et al. (2014) and Islam et al. (2016) investigated the inclusion of spatial correlation in multivariate crash severity models. The models from the two studies were developed using severe (injury and fatal) and no-injury crashes data. The results advocated the use of multivariate Poisson-Lognormal (PLN) models including spatial effects over univariate PLN spatial models. Also, an interesting study by Aguero-Valverde et al. (2016) proposed a multivariate spatial model to concurrently model the frequency of different crash types (head-on, rear-end, and sideswipe). Their study results showed that the model that considered both multivariate and
spatial correlation had the best fit among the different models investigated. Furthermore, they found that the multivariate correlation played a stronger role than the spatial correlation when modeling the crash frequencies of different crash types.

To a smaller extent, few studies attempted incorporating random parameters in multivariate models. El-Basyouny and Sayed (2011) developed a multivariate Poisson-lognormal intervention model that was used for the analysis of crash counts by severity levels. The model was extended to incorporate random parameters to account for the correlation between sites within comparison-treatment pairs. The analysis revealed that incorporating such design features as matched comparison groups in the specification of safety performance functions can significantly improve the fit while reducing the estimates of the extra-Poisson variation. As well, such extended models can be used to account for heterogeneity due to unobserved road geometrics, traffic characteristics, environmental factors and driver behavior.

Anastasopoulos (2016) estimated random parameters multivariate tobit and zero-inflated count data models of accident injury-severity rates and frequencies, respectively. The proposed modeling approach accounted for unobserved factors that may vary systematically across segments with and without observed or reported accident injury-severities, thus addressing unobserved, zero-accident state and non-zero-accident state heterogeneity. Moreover, the multivariate setting allowed accounting for contemporaneous cross-equation error correlation for modeling accident injury-severity rates and frequencies as systems of seemingly unrelated equations. The tobit and zero-inflated count data modeling approaches addressed the excessive amount of zeros inherent in the two sets of dependent variables (accident injury-severity rates
and frequencies, respectively), which are – in nature – continuous and discrete count data, respectively, that are left-censored with a clustering at zero. Bhat et al. (2017) proposed a spatial random coefficients flexible multivariate count model to examine the number of pedestrian injuries by injury severity at the spatial level of census tract. They acknowledged the different risk factors for different types of pedestrian injuries and accounted for unobserved heterogeneity in the risk factor effects. They also recognized the multivariate nature of the injury counts by injury severity level within each census tract.

Few studies have addressed the correlation among crash modes when modeling active transportation safety. Nashad et al., (2016) developed a bivariate crash model by adopting a copula-based negative binomial model for pedestrian and cyclist crash frequency. Their proposed approach accommodated potential heterogeneity (across zones) in the dependency structure. They used three years of pedestrian and cyclist crash count data at the Statewide Traffic Analysis Zone (STAZ) level and investigated its association with proxy traffic exposure measures, socio-economic characteristics, road network characteristics, and land use attributes. The authors then compared the performance of the copula crash model with the independent crash models, which confirmed the importance of incorporating the dependence between pedestrian and cyclist crashes in the macro-level analysis. Also, Heydari et al., (2016) introduced a flexible Bayesian multivariate modeling technique using a Dirichlet process mixture, which they used to account for the correlation among pedestrian and cyclist crashes through a heterogeneous correlation structure (Heydari et al., 2017). The model investigated the effects of various factors such as built environment characteristics on pedestrian and cyclist injury counts at signalized intersections in Montreal. Their proposed latent class multivariate model helped in addressing
unobserved heterogeneity through its latent class component. They showed that such a flexible model better captured the underlying complex structure of the correlated data, resulting in a more accurate multivariate crash model.

Two recent studies have attempted to address both spatial and mode correlations concurrently when modeling active transportation crashes. Lee et al. (2015) estimated multivariate and univariate crash models based on traffic analysis zones (TAZs) and compared their performance. The models were developed for motorized and non-motorized crash types (i.e., motorists, pedestrians and cyclists) using proxy traffic exposure variables along with several socio-economic and road facility variables. They found that the multivariate spatial model outperformed the univariate spatial ones and that the spatial error component played an important role in significantly improving the model performance. Also, Huang et al. (2017) proposed a multivariate spatial model to simultaneously analyze the occurrence of motor vehicle, bike, and pedestrian crashes at urban intersections. They found that the multivariate spatial model outperformed the univariate spatial model and the multivariate aspatial model. Their results confirmed the highly correlated heterogeneous residuals in modeling crash risk among motor vehicles, bicycles, and pedestrians. Regarding spatial correlation, they found that the estimates of the variance for spatial correlations of all three crash modes in the multivariate and univariate models were statistically significant. However, the correlations for spatial residuals between different crash modes at adjacent sites were not statistically significant.
2.1.4 Macro-level Analysis

Macro-level CMs address crashes on a wide area level (e.g., traffic analysis zones, census level, and neighborhood). The macro-level CMs are based on cross-sectional data analysis, and it is assumed that the non-modeled variables will not have much impact on the model inference. However, it is recommended to include areas that have similar characteristics and to collect as much data as possible about the studied area in order to reduce the random factors affecting the crash risk. The way in which data is extracted and aggregated influences whether underlying causal mechanisms are concealed or revealed in the resulting models (Davis, 2004). Many studies have shown that using improperly aggregated macro-level data, correlated by geographic area, may conceal associations between crashes and predictor variables, leading to biased results (Kmet et al., 2003; Davis, 2004). Aggregation bias is a phenomenon that can produce a statistically significant association suggesting a causal relationship that is in direct contradiction to the actual underlying association (Lovegrove, 2005). In order to minimize the aggregation bias, it is recommended to use as large data set as possible, stratify classes of possible models, and then identify a member of each class that best fits the data (Davis, 2004); (Lovegrove, 2005).

Recently, few studies have attempted to combine macro-level and micro-level approaches in a nested model form in order to reduce the effects of both ecological and atomistic fallacies on crash modeling (e.g. Heydari et al., 2017; Lee et al., 2018).

Macro-level CMs require rigorous statistical analysis for identifying the appropriate variables and their associations with crashes. It is extremely important for the purpose of long-range planning that the independent variables used in the macro-level CMs are generally available, meaningful for planning purposes, and have statistical sound and reliable relationship with
crashes (Lovegrove, 2005). In other words, variables are needed to be relevant to simultaneous transportation planning and account for factors that would possibly affect safety planning research to identify the relevant measures.

The above-mentioned issues need to be considered while developing the macro-level CMs for this research since the absence of reliable data sources and practical techniques that can provide relevant future values of explanatory variables, shall bring about macro-level CMs of little practical value for road safety and planning purposes.

2.2 Crash Hot Zones’ Identification

Crash hot zones are those zones with higher crash frequency than normal. Hot zones’ identification is essential to improve the safety of road networks. The majority of traditional identification/ranking methods have relied on historical traffic crash records to obtain an estimation of safety for diverse traffic entities. These methods included the crash frequency method, the crash rate method, the crash severity method, the safety index method, and the rate quality control method (Huang et al., 2009).

Those simple (somehow naïve) statistical methods have serious limitations despite their simplicity, particularly when dealing with short-term period data (e.g., 2 to 3 years or less). The methods are subject to the regression to the mean problem along with its inability of examining the crash dispersion (Elvik, 2007). Ranking the sites by the extent to which the crash risk exceeds what is normal for sites with similar characteristics can be more useful. Nevertheless,
naive methods are incapable of deducing the “normal” expected crash rate since it does not account for various heterogeneities associated with the diverse site features.

To overcome the drawbacks of naive ranking methods, several studies have proposed model-based ranking approaches for hotspot identification. Elvik (2007) presented an extensive review of the state-of-the-art approaches to hot spot identification back then and concluded that the empirical Bayes approach (EB) was the most reliable method. Cheng and Washington showed that the EB approach was the most consistent and reliable method for identifying sites with promise using simulation experiments (Cheng and Washington, 2005) and innovative and robust evaluation criteria (Cheng and Washington, 2008). Moreover, Lan and Persaud (2011) showed that hotspot ranking techniques based on Bayesian methodologies are superior to those that simply rely on the observed crash count.

A good indicator of the expected safety benefits following a treatment, under the empirical Bayesian framework, is the potential for safety improvement (PSI) or accident reduction potential (Sayed and Rodriguez, 1999). The PSI indicator is able to identify which locations are likely to generate the highest return on investments. Using this indicator, hotspots are ranked according to their PSI values, which are estimated as the difference between the observed crash frequency $\theta_i$ and the predicted normal crash frequency for similar sites ($\mu_i$). The use of PSI to rank sites for required safety improvements is commonly adopted in the literature. This measure is considered among those that produce the most reliable performances for network screening according to the American Highway Safety Manual, and it is currently known as “excess expected average crash frequency with EB adjustment” (AASHTO, 2010). The PSI method
continues to be the state-of-the-practice for identification and ranking of crash-prone locations and has been applied in many recent studies (e.g., Yang et al., 2016; Cheng et al., 2017).

While the EB approach is demonstrably better suited to reliably estimate the safety of traffic sites than the simple statistical methods, it may be subject to several limitations under certain conditions. Such limitations include the requirement for a large sample size of data to develop safety performance functions (SPFs), the lack of flexibility in defining underlying distributions for observed crashes, the fact that it can produce only point estimations of expected crashes, and the absence of consideration of the penalty for an over-parameterized model. For example, in a standard before-and-after evaluation using the EB approach, an external (or previously obtained) safety performance function (SPF) is used for each candidate site to provide the prior estimate of safety. However, in general hot spot identification studies, it may sometimes require all sites in a road network to be screened and the SPF to be fitted purely from the local data. In such a condition, the EB approach may be criticized for implicitly using the data twice. That is, the data are first used to estimate the parameters in SPF and, once these values are determined, the observed crash count of each site is utilized to make inference about the posterior estimation (Carlin and Louis, 2000). Also, the EB approach may be inadequate to explicitly account for the “uncertainty” of associations of covariates and safety (Miaou and Lord, 2003). In the EB approach, the uncertainty, which is used to indicate the confidence level of the estimated standard safety, is merely represented by a vague over-dispersion term and it is identical for all sites in model prediction. Thus, once the SPF is calibrated, the point estimates of all covariate effects will be assumed to be true without any uncertainty. However, the covariate effects in the SPF are typically estimated from crash data and thus are subject to uncertainty. Hence, the
posterior estimate of safety may be more accurate if all uncertainties associated with each covariate effect on safety could be fully taken into account.

Because of these limitations of the EB method, the fully Bayes (FB) approach has been proposed to identify and rank hotspots. The essential characteristic of the FB analysis is its explicit use of probability for quantifying uncertainty in the inferences based on statistical data analysis. Thus, by accommodating all the uncertainties in the model, the typical expected crash rate could be represented by a fitted distribution and the final safety estimate for a specific site is obtained by averaging the possible values relative to the distribution (Miaou and Lord, 2003). It is clear that the EB approach is a special case of FB that arises when an FB analysis is simplified, for instance, by assuming that the safety effect of covariates such as lane width is known without any uncertainty. A further advantage of the FB method is its flexibility in explicitly representing the hierarchical models. Hierarchical structures exist extensively in crash data because of the data collection and clustering process. Several recent studies have indicated that crash frequency may be better modeled by explicitly accounting for the spatial and temporal heterogeneities using hierarchical modeling technique (e.g., Huang et al, 2008, Song et al., 2006). Hence, there is a need to explore the use of FB hierarchical models in site ranking and hotspot identification.

Schluter et al. (1997) were the first to apply the FB method to ranking hazardous sites. They proposed a Bayesian hierarchical Poisson-Gamma model to rank high-risk sites among 35 intersections by using criteria such as the posterior probability of selecting the worst site and the posterior mean (PM) of crashes. Bossche et al. (2003) used a Bayesian binomial hierarchical model to rank hazardous intersections for bicycles in a small university town in Belgium based
only on crash data. They used the PM of crashes to rank the sites and concluded that there is no such thing as the correct ranking because of the stochastic character of bicycle crashes. Miranda-Moreno and Fu (2007) explored the differences between EB and FB through a simulation study in which they used both methods to calculate the PM to rank sites. They found that the FB estimators performed better when working with small sample data sets characterized by a low mean accident frequency. For larger datasets (e.g., more than 300 sites), the two approaches performed similarly. Moreno and Fu (2007) mentioned that “This is not surprising because, for small samples, the estimated SPFs for the EB method may not be reliable, resulting in errors in the estimation of expected crashes; however, the FB method can carry that uncertainty to the final estimation”. Huang et al. (2009) evaluated the FB method for hotspot ranking using crash records from 1997 to 2006 of 582 four-legged signalized intersections in Singapore. Expected crashes from EB and FB obtained from the past 3 years of crash data were used to identify hotspots. The treated the average of the observed crash counts in the entire 10-year period as a true mean for evaluation of the two methods, and sensitivity and specificity were used as evaluation criteria. That could be problematic because ten years of crash counts may still be subject to regression to the mean (RTM); however, they concluded that the FB approach outperforms the EB in correctly identifying hot spots.

A prominent application of the full Bayesian hotspot identification approach is the multivariate identification method. Crashes of different severity or types might be highly correlated. This correlation would be ignored when separate rankings by crash severity/type are made for each outcome level. Multivariate extensions of the simple Poisson distribution, such as the multivariate Poisson log-normal (MVPLN) model have been proposed in the literature to account
for this correlation. Consequently, some hotspot identification techniques that were developed for univariate analyses have been adapted for the multivariate setting. For instance, Aguero-Valverde and Jovanis (2009) proposed to rank hotspots by excess crash cost using the expected crash frequency as estimated by the MVPLN models. A comparison of this method with the excess crash cost estimated by the independent models indicated significant differences between the multivariate and univariate approaches. El-Basyouny and Sayed (2009) extended the probability-of-excess method, which can identify locations with a very high probability of exceeding the long-term mean of all sites, to the multivariate framework. Moreover, El-Basyouny and Sayed (2013) proposed a multivariate ranking of hotspots based on statistical depth functions, which are tools for non-parametric multivariate analysis and provide center-outordering of multivariate data.

Recently, there has been some research that aims at detecting hot zones for active transportation modes on the macro-level. However, these efforts focused on either analyzing the total active transportation crashes without considering possible correlation between crash outcomes (Tasic and Porter, 2016), or analyzing the active modes’ crashes taking in consideration the correlation between them, but the hot zones for each active mode were identified independently (Lee et al., 2015); (Nashad et al., 2016).

2.3 Active Transportation Safety

Jacobsen (2003) and Robinson (2005) studies are among the earliest studies that attempted to investigate active commuters’ safety on a wide area level. Jacobsen (2003) examined the relationship between the number of vulnerable road users (pedestrians and cyclists) and their
crashes with motor vehicles based on five data sets from different locations worldwide. Results showed that, at the population level, the number of motorists colliding with vulnerable road users would increase at approximately 0.4 power of the number of people cycling or walking. Robinson (2005) studied three datasets from Australia and concluded similar results to Jacobsen's study. The following subsections focus on the factors that are associated with pedestrian and cyclist safety.

2.3.1 Pedestrian Safety Correlates

Several studies attempted to investigate pedestrian safety on a macro-level. This was applied on various levels of area aggregation such as census tract (Cottrill and Thakuriah, 2010; Ukkusuri et al., 2011; Abdel-Aty et al., 2013), traffic analysis zone (TAZ) (Abdel-Aty et al., 2013; Wang et al., 2013; Lee et al., 2015b; Wang et al., 2016), and block group (Noland et al., 2013; Abdel-Aty et al., 2013). In those previous studies, macro-level crash models attempted to relate pedestrian crashes to a variety of explanatory zonal features. A summary of significant associations from the previous macro level (and some micro-level) studies investigating pedestrian safety is as follows.

Traffic volume has always been a significant predictor of pedestrian crashes (Dumbaugh and Li, 2010; Lascala et al., 2000; Lee and Abdel-Aty, 2005; Loukaitou-Sideris et al., 2007; Wier et al., 2009). Vehicle miles traveled (Abdel-Aty et al., 2013; Lee et al., 2015b) as well as average annual daily traffic (Loukaitou-Sideris et al., 2007; Wier et al., 2009) were found positively associated with pedestrian crashes. On the other hand, pedestrian exposure has been also critical for pedestrian crash modeling. Pedestrian exposure reflects the opportunity for a risky
pedestrian-vehicle interaction to occur. However, it is difficult to be measured directly since this would involve tracking the movements of all people at all times (Wang et al., 2016; Greene-Roesel et al., 2007). Greene-Roesel et al. (2007) summarized five common surrogate metrics used to describe pedestrian exposure; including population, number of pedestrians, number of trips, distance traveled, and time spent traveling. Wang et al. (2016) used population as a surrogate for pedestrian exposure, and it had a positive effect on pedestrian crashes. Amoh-Gyimaha et al. (2016) also found that the population and percentage of commuters walking to work had a positive association with the number of pedestrian crashes.

Socio-economic and demographic factors are also among pedestrian crash predictors. An area’s socioeconomic deprivation level was found associated with pedestrian crashes (Cottrill and Thakuriah, 2010; Graham and Glaister, 2003; Loukaitou-Sideris et al., 2007; Siddiqui et al., 2012). This is usually measured by proxy factors such as the percentage of households without vehicles, the level of household income, and the unemployment rate. Median household income was found to be negatively associated with pedestrian crashes (Siddiqui et al., 2012; Lee et al., 2013), while the percentage of residents living below the poverty line (Wier et al., 2009; Lee et al., 2015a) was associated with increase in pedestrian crashes. Households without vehicles were also found to have a positive association with pedestrian crashes (Noland et al., 2013; Lee et al., 2015b; Siddiqui et al., 2012). Similarly, areas with a higher proportion of uneducated residents showed a positive association with pedestrian crashes (Ukkusuri et al., 2011), while a higher proportion of high school graduates showed a negative association (LaScala et al., 2000). Demographic features, such as population (Kim et al., 2006; Ukkusuri et al., 2011; Lee et al., 2015b), population density (Loukaitou-Sideris et al., 2007; Siddiqui et al., 2012), employment
and population (Wier et al., 2009; Siddiqui et al., 2012), and employment density (Loukaitou-Sideris et al., 2007) were all found positively associated with pedestrian crashes. Also, population characteristics had significant associations with pedestrian crashes (Demetriades et al., 2004; Fontaine and Gourlet, 1997; Johnson et al., 2004). The number of vulnerable road users, including children and older people, was found to be significantly correlated with pedestrian crashes. A higher number of pedestrian crashes was associated with a higher density of children (Abdel-Aty et al., 2013) and a lower percentage of the resident population aged 65 and older (Wier et al., 2009).

Moreover, land use features could influence pedestrian activity and potentially affect crashes. It was observed that pedestrian crashes were likely to increase with more residential land areas (Hadayeghi et al., 2007; Loukaitou-Sideris et al., 2007; Siddiqui et al., 2012; Wier et al., 2009). Other land use activities that generate traffic such as commercial, industrial, retail, and parks\(^2\) were also found to have a positive association with pedestrian crashes (Kim et al., 2006; Ukkusuri et al., 2011; Pulugurtha et al., 2013). Wang and Kockelman (2013) introduced land use entropy and found that balanced land development had a mildly positive impact on reducing severe crashes, and could serve as a countermeasure for curbing pedestrian fatalities. Amoh-Gyimaha et al. (2016) also found that mixed land use had a positive association with pedestrian crashes. Wang et al. (2016) concluded that pedestrian crashes were higher in TAZs with medium land use intensity than in TAZs with low and high land use intensity.

\(^2\) Parks were found to have contradicting associations with pedestrian crashes among the different referenced studies.
Significant correlations were also found between pedestrian crashes and road facilities (Loukaitou-Sideris et al., 2007; Wier et al., 2009; Abdel-Aty et al., 2013; Lee et al., 2015b). Traffic engineers found that the intersections with higher number of pedestrian crossings had led to higher probabilities of vehicle-pedestrian crashes (Siddiqui et al., 2012; Abdel-Aty et al., 2013). Lengths of different road types, as well as speed limits, had significant effects on pedestrian crashes (Quddus, 2008; Kim et al., 2010; Abdel-Aty et al., 2013; Lee et al., 2015a, b). Most recently, Wang et al. (2016) investigated the association between pedestrian crash frequency and various roadway variables and indicated significant factors including length of major arterials, length of minor arterials, road density, average intersection spacing, and percentage of 3-legged intersections.

Lastly, several previous studies investigated the impact of various network indicators on motorized and non-motorized traffic safety. Moeinaddini et al. (2014) estimated the relationship between urban street indicators and passenger fatalities. Their model showed that more blocks per area were correlated with more passenger transport fatalities. In addition, their model showed that more nodes per selected areas are associated with more passenger transport fatalities. Consequently, they concluded that more simple large blocks with fewer nodes would be correlated with fewer fatalities. Zhang et al. (2015) investigated the associations between road network structure and non-motorist accidents. The results of their study suggested that: “1) there would be fewer non-motorist-involved crashes, if the network was more centered on major roads; 2) a network with a higher average number of intersections on the shortest path connecting each pair of roads tended to experience fewer crashes involving pedestrians and bicyclists; 3) and the more clustered road networks into several sub-core networks, the lower the
non-motorist crash count” (Zhang et al., 2015). Dai and Jaworski (2016) evaluated pedestrian crashes using network-based spatial techniques, and they conducted built environment audit within hotspots. They found that road gradient change was safer to pedestrians while road curvature showed a non-significant association. Cai et al. (2016) analyzed macro-level pedestrian and bike crashes incorporating spatial spillover effects. They found that pedestrian crashes were positively correlated with signalized intersection density, log of sidewalks length, and the proportion of local roads among other variables. Guo et al. (2017) investigated the effect of road network patterns on pedestrian safety. They developed a global integration index to quantify the road network structure. Their results indicated that higher global integration was associated with more pedestrian-vehicle crashes; and that the irregular pattern network was proved to be safest in terms of pedestrian crash occurrences, whereas the grid pattern was the least safe. They also found that the conditional auto-regressive (CAR) model with neighborhood structure, based on road network connectivity, had better goodness-of-fit, implying the importance of accurately accounting for spatial correlation when modeling spatially aggregated crash data.

2.3.2 Cyclist Safety Correlates

Recently, there have been increasing efforts to study cyclist crashes on aggregated levels. Different units of analysis were adopted in previous studies including census tracts (Narayanamoorthy et al., 2013), grid-based structures (Gladhill and Monsere, 2012), and TAZs (Siddiqui et al., 2012; Wei and Lovegrove, 2013). TAZs have been the most common unit of analysis in macro-level studies. It is a spatial aggregation of census blocks, and its size is usually a function of socio-economic data.
Several studies have discussed the relationship between cyclist crashes and various explanatory variables on the macro-level. Lovegrove (2007) developed a community-based bike–auto crashes model using negative binomial regression for the Greater Vancouver Regional District. The model showed that bike crashes were associated with bike mode share in rural areas. Kim et al. (2007) explored various crash, roadway, land, and environmental factors contributing to the injury severity of cyclists, who were involved in cyclist-motorist crashes in North Carolina. Their results showed that several factors increased the probability of a fatal injury in a crash, including head-on crashes, speeding, inclement weather, no streetlights, morning peak, truck involvement, intoxicated drivers/cyclists, and age. Kim et al. (2010) also discussed results from macro level CMs, which suggested that demographic variables, accessibility variables (e.g. number of dead ends, number of intersections, etc.), bus route length, and number of intersections were positively associated with bike-auto crashes. Wei and Lovegrove (2013) developed community-based negative binomial models for bike–auto crashes. The models showed an increase in the cyclist crashes with the increase of the total lane kilometers, bike lane kilometers, bus stops, traffic signals, intersection density, and arterial–local intersections percentage.

In terms of socio-economic variables, Siddiqui et al. (2012) showed that population, employment, and median household income were positively associated with cyclist crash frequency. Regarding land use, the increase in commercial land use was found positively associated with the frequency of cyclist crash frequency and injury (Narayanamoorthy et al., 2013); (Vandenbulcke, 2014). Also, Amoh-Gyimah et al. (2016) found that the increase in
residential area percentage, industrial area percentage, and land use balance mix was positively associated with cyclist crashes.

As for travel demand variables, bike traffic (Strauss et al., 2013; Miranda-Moreno et al., 2011), as well as vehicle traffic (Hamann and Peek-Asa, 2013), were found positively associated with cyclist crash frequency. Regarding traffic control variables, low-speed streets were found negatively associated with cyclist crashes, while high-speed streets yielded an increased number of cyclist crashes (Siddiqui et al., 2012). Also, traffic signal density was found positively associated with cyclist crashes (Wei and Lovegrove, 2013) (Chen, 2015).

For road network features, Chen et al. (2012) showed that the installation of bike lanes did not lead to additional crashes, but a possible increase in the number of cyclists instead. On the other hand, vehicle lanes and intersection density were found positively associated with the frequency of cyclist crashes (Siddiqui et al., 2012); (Wei and Lovegrove, 2013). For different types of bike facilities, the off-road bike lanes were found safer than the on-road ones (Teschke et al., 2012) (Hamann and Peek-Asa, 2013) (Reynolds et al., 2009). Regarding street elements, bus stop density was positively associated with cyclist-motorist crash frequency (Wei and Lovegrove, 2013) (Strauss et al., 2013). Moreover, Chen (2015) showed that the parking sign density was positively associated with cyclist crash frequency. Street lighting was also included in previous studies for modeling cyclist injury severity (Kim et al., 2007) (Klop and Khattak, 1999), where darkness was found to significantly increase injury severity.

More recently, Chen (2015) indicated that the zonal bike crash frequencies are spatially correlated. His study included some cycling related variables that had not been investigated in
prior studies such as bike trips vs. total trips, zonal mean of the driving speed limits, zonal mean slopes, and densities of street trees and parking signs. He found that the zonal mean of driving speed limits, the total number of trips, length of on-arterial bike lanes, the entropy of mixing land use, and the density of traffic signals were positively associated with cyclist crashes. Kaplan and Prato (2015) modeled the frequency and the severity of crashes involving cyclists and motorists in the Copenhagen region by estimating a link-based multivariate Poisson-lognormal model. The model underlined the relevance of infrastructure, land use, and spatial effects to explain the variation in the number of cyclist-motorist crashes. They showed that the number of crashes was nonlinearly related to the average bike and vehicle daily traffic, which confirmed the safety in numbers hypothesis (Elvik, 2009; Jacobsen, 2003). Kaplan and Prato (2015) also found that bike paths were associated with fewer crashes, which was in line with studies that showed that cycling facilities increased actual and perceived safety (de Rome et al., 2014); (Kaplan et al., 2014). That disagreed with past literature which assumed that cycling in mixed traffic was safer than cycling on bike infrastructure (Pucher et al., 1999); (Rodgers, 1997). Lastly, Prato et al. (2016) showed that the positive associations with traffic exposure were nonlinear and complied with the “safety in numbers” hypothesis (Jacobsen, 2003).

2.4 Summary

This chapter presented an overview of the various crash modeling and hot zones identification techniques that have been commonly applied in the literature. Moreover, this chapter provided a narrow focus on active transportation safety correlates. There are many areas related to the subjects discussed in this chapter that either deserve further attention or have not been dealt with
adequately in the traffic safety literature. The upcoming chapters of the thesis provide a more in-depth analysis of these areas.
Chapter 3: Data Collection

This chapter presents the data sources as well as the process of extracting the various analysis variables.

3.1 Data Sources

Zone-level CMs are developed in this study based on 134 TAZs in the City of Vancouver. These zones represent discrete geographical areas that are defined by similar land use or specific zoning/geographic features and are used to generate trips to the network based on the expected level of activity. Although it is generally desirable to use major road network components as boundaries, often adjacent homogenous land uses dictate that the zone be expanded to incorporate multiple blocks.

Explanatory variables that are related to zone characteristics and network configuration are included in the CMs. Walk trips, bike kilometer travelled, and vehicle kilometers travelled are incorporated in the models as traffic exposure variables. The data needed for the analysis variables is compiled using the ArcGIS software which is a geographic information system for working with maps and geographic information. The software is mainly used for processing and visual representation of the data after extracting it (i.e. the data) from five main sources:

1. Insurance Corporation of British Columbia, a public automobile insurance company, provided the crash data for a 5 years period (2009-2013). Only pedestrian-motorist and cyclist-motorist crashes are included in the analysis, as shown in Figure 1a. A 5 years period is selected to collect an adequate sample size. The sample included 3 severity levels, i.e. fatality, injury, and property
damage only. However, the total number of pedestrian crashes as well as the cyclist crashes is included in the analysis in order not to disperse the sample size.

2. Translink, the Metro Vancouver transportation authority, provided the geo-coded files for the city of Vancouver road network, pedestrian network, bike network, land use, and TAZ boundaries. Moreover, Translink provided the output of an Emme2 transportation planning model for the travel demand in Metro Vancouver in the year 2011. Translink used the 2011 household travel survey and land uses as inputs to calibrate the model, and the 2011 cordon counts to validate the model assignments.

3. Acuere Analytics provided the Vancouver Cycling Data Model (VCDM 2011), as shown in Figure 1b. The VCDM used the bike count occurring between years 2005 and 2011 to estimate the annual average daily bike traffic (AADB) over the city of Vancouver bike network in 2011 (El Esawey et al., 2015). The available data covered more than 810,000 hourly volumes over seven years. The model was efficient in estimating the AADB on most of the bike network links (more than 70% of the network).

4. The open data catalogue of the City of Vancouver (http://vancouver.ca/your-government/open-data-catalogue.aspx) provided the city built environment data (i.e. transit stops, traffic signals, actuated signals, trees, and light poles) as well as the contour map of the city.

5. Census Canada provided the socio-economic data (i.e. employment, population, and household data) of the City of Vancouver according to the 2011 census.
Figure 1 (a) Heat Map of Cyclist Crashes (red Pts.) and Pedestrian Crashes (yellow Pts.) (b) Cyclist Trips at City of Vancouver TAZs
3.2 Analysis Variables

Before discussing the methodology of developing the macro-level models, the variables that can be incorporated in the crash models are first identified in this section. Figure 2 shows the hierarchy of the variables that are investigated during the analysis. The analysis variables are divided into four main categories; crashes, traffic exposure, network configuration, and zonal characteristics. Each category incorporates a large set of variables as shown in Figure 2. These are further discussed in the following subsections. The aggregation process of the data for the various variables, at the zone level, is conducted using the ArcGIS software as discussed below.
Figure 2 Hierarchy of the Analysis Variables
3.2.1 Crashes

Cyclist-motorist and pedestrian-motorist crashes are the only dependent variables. In this study, crashes are aggregated at the different TAZs according to their geo-spatial locations. Zero pedestrian crashes are found in six TAZs and zero cyclist crashes are found in two TAZs.

A methodological question arises regarding how to assign the crashes that occur along the zonal boundaries. Different approaches were attempted in the literature. Lovegrove and Sayed (2007) aggregated boundary data in an automatic geospatially precise way to develop a macro-level crash distribution model for the Great Vancouver Regional District. Wei (2010) made use of five boundary data aggregation approaches (the one-to-one ratio method, the half-to-half ratio method, the geospatial method, the vehicle kilometer travelled (VKT) ratio method, and the total lane kilometers (TLKM) ratio method) for the geocoded data on the TAZ boundaries. Wang et al. (2012) aggregated the boundary crashes into TAZs manually, and proposed a new crash model by dividing crashes into the on-system road and off-system road crashes. In this research, a new approach is used for aggregating the crashes on boundary, where the bike kilometers travelled (BKT) ratio is selected as a method of aggregation for the cyclist-motorist crashes, and walk trips (W) ratio is used as a method of aggregation for the pedestrian-motorist crashes. In this method, boundary cyclist crashes are distributed between the adjacent TAZs according to the ratio of the BKT (or W in case of pedestrian crashes) in those zones. This method is similar to the VKT ratio method but the BKT or W ratios are used instead of the VKT due to their better association with cyclist or pedestrian crashes as shown in the literature review and as shall be concluded from the analysis afterwards.
3.2.2 Traffic Exposure

VKT represents the exposure due to the motorist trips, while BKT represents the exposure due to the bike trips. Since it is very difficult to represent all kilometer trips travelled by pedestrians in a specific area to an acceptable degree of accuracy; the modeled walking trips at each TAZ are used as a measure of exposure for pedestrian trips. BKT is obtained using the Vancouver cycling data model, which provided the modeled cyclist trips on the city of Vancouver segments (El Esawey et al., 2015). The trip count at each segment is multiplied by the corresponding segment length to obtain BKT, which is then aggregated for each TAZ. Walking trips and VKTs for the different TAZs are obtained directly using the Emme2 model at a zone level. Although a significant amount of time and resources have been involved in the Emme2 model development, some applications may not be allowed due to the model level of accuracy and assumptions. Translink Regional Transportation Model User’s guide states that the model best suits regional and macro-level strategic planning, but it requires careful interpretation if to be used on the micro-level.

3.2.3 Network Configuration

The graph theory concept originated from the “Seven Bridges of Konigsberg” problem, which was solved by Euler in the 18th century. Garrison and Marble (1962) applied graph theory to transportation networks and were able to develop indices for connectivity. Also, Kansky (1963) presented indices that characterized network connectivity and complexity. More recently, Gattuso and Mirello (2005) were able to evaluate the topology and geography of metro networks in some European cities and New York City-based on graph indicators. Derrible and Kennedy (2010) presented a new methodology to draw metro networks as graphs, and accordingly they
were able to create directness and structural connectivity indicators. Quintero et al. (2014) introduced a novel approach to redraw transit networks as graphs, and hence they were able to include new connectivity indicators. Those indicators were used in developing macro-level CMs to assess the safety of Metro Vancouver transit network (Quintero et al., 2013). They found that crashes were significantly associated with transit network properties such as connectivity, overlapping degree and the local index of Transit Availability. As well, the models showed a significant relationship between crashes and some transit physical and operational attributes such as the number of routes, frequency of routes, bus density, length of bus and 3+ priority lanes. The graph theory measures had also been applied to the field of transportation planning in several other studies (Xie and Levinson 2007; Derrible and Kennedy 2009; Rodrigue et al. 2009).

A few studies used graph theory measures to explain individuals’ non-motorized travel behavior. Dill (2004) evaluated various measures of connectivity for the purpose of increasing walking and bicycling. Four measures were applied to the census tracts in Portland metropolitan region. Though positively correlated with active ridership, the measures did not consistently assign the same level of connectivity for a tract. Dill and Voros (2007) found significant differences between connected node ratios and people who biked during a certain period of time. Berrigan et al. (2010) explained individuals’ non-motorized behavior by measuring the link-node ratio and other graph indices for a local street grid within short buffers around survey respondents’ home addresses. Network quality and connectivity were evaluated at a micro level in former studies by investigating the individual discontinuities in the on-street bike facilities (Krizek and Roland 2005; Birk and Geller 2006; Barnes and Krizek 2005). Tal and Handy (2012) showed that
“accounting for actual pedestrian connectivity, particularly the connections to schools and other public facilities, can lead to both better planning and more accurate research with respect to the conditions that promote walking”. Lastly, Lundberg and Weber (2014) analyzed the connectivity and network perceptions for non-motorized transport and university populations. They examined local bike and pedestrian networks in the vicinity of the University of Alabama campus to assess the utility of these networks for travel to the university by students and employees. They found that increases in connectivity can be expected to lead to an increase in non-motorized travel.

In order to extract graph indicators for the bike and pedestrian networks’ configuration; we need first to characterize the networks into their basic links and nodes. Links represent the homogenous network segments (e.g., bike lane in the case of bike network or pedestrian in the case of pedestrian network); while the nodes represent the connections (intersections) between two different separate links.

Since a zonal level of aggregation is required, a technique is developed for splitting the entire network between the zones. The links and nodes are distributed among the different TAZs according to their geospatial location. If a link is found to pass through two zones, then it is divided between those two adjacent zones using a weight, which is relative to the link’s length within each zone. Figure 3 shows the city of Vancouver bike and pedestrian networks’ characterization into their basic links and nodes. After the graph characterization of both the bike and pedestrian networks; several network indicators can be extracted using graph theory. Three main categories of network configuration are used in this research (i.e. connectivity, directness, and topography).
Figure 3 Characterization of the (a) Pedestrian Network and (b) Bike Network at the different TAZs

First, for the network connectivity, several indicators can be quantified as follows:

- Network Density
- Intersection Density
- Degree of Connectivity
- Coverage
- Degree of Complexity
Network density (NetD) is the ratio between the length of the links and the corresponding TAZ area. This measure was used in previous studies for measuring the connectivity of pedestrian and cyclist networks (Mately et al., 2001; Jennifer Dell, 2004). Intersection density is calculated as the ratio between the number of intersections within the zonal network and the area of the corresponding TAZ. This indicator aggregates all the intersections without distinguishing between their different types. Another commonly used indicator is the degree of connectivity (Conn), which represents the ratio between the actual number of links in a TAZ and the maximum possible number of links in the TAZ ($l_{\text{max}}$). According to graph theory, the maximum possible number of links within a planar graph is calculated using equation 1, where $n$ is the number of nodes within a graph.

$$Conn = \frac{\text{Number of Links in TAZ}}{\text{Maximum Number of links in TAZ}}, \quad l_{\text{max}} = 3(n-2) \quad (1)$$

The value of Conn is bounded between 0 and 1. A completely connected network will have a Conn equal to 1, while a completely disconnected network will have a Conn equal to 0. The Conn indicator has been used in previous studies for evaluating transit networks (Derrible and Kennedy, 2011; Quintero et al., 2013). However, two deficiencies are noticed upon applying the indicator to this research. First, the equation form that is used to calculate $l_{\text{max}}$ would vary from a zone to another according to the graph shape (i.e. if the graph is not planar). Moreover, equation 1 is not valid when the number of nodes is less than 3. Therefore, an indicator that may better represent the active transportation network connectivity is suggested. The indicator is called coverage (Cov) and is shown in equation 2.
Cov simply assumes that \( l_{\text{max}} \) is the total number of street links in the TAZ. This assumption is usually more practical than the one in equation 2, because the maximum possible number of links not often exceeds the number of the street links in any zone. However, this indicator is reliable only if the complete layer of the road network is available, and if not, Conn can be used. Figure 4 shows the bike network along with heat maps for bike network coverage at the different TAZs.

The last indicator in this category is complexity (Comp), which represents the degree of network complexity. Comp is defined as the average number of links per node (Kansky, 1963), as shown in equation 3:

\[
\text{Comp} = \frac{\text{Number of Links in TAZ}}{\text{Number of Nodes in TAZ}}
\]

Networks with a complex structure will have high Comp value, while those with simple structures will have low Comp value.
As for the second category of graph indicators, i.e. Directness, three aspects are studied in this research:

- Continuity
- Orientation
- Linearity

Sheltma (2012) defined linearity within a network as the ratio between a crow line and its effective straight line. For calculating orientation and continuity, Sheltma (2012) proposed manual methodologies by counting every turn along the key bike routes in case of orientation, and by counting every crossing along the key bike routes in case of continuity. Nevertheless, such methodologies are inconvenient for macro-level studies as they would take considerable time and effort.
Therefore, new ways to calculate directness need to be applied in the context of the macro-level research. Three graph indicators are used: average edge length, average length per vertex, and linearity. Both average edge length and average length per vertex represent the continuity of the network, while the linearity indicator can represent both the linearity and the orientation of the network. Average edge length is calculated as the ratio between the total length of a zonal network and the number of links in the corresponding zone, while average length per node is calculated as the ratio between the total length of a zonal network and the number of nodes within the zone (Kansky, 1963).

Linearity within a network was defined by Sheltma (2012) as the ratio between the effective straight line and crow line. In this research, linearity is calculated using equation 4 as the ratio between the modified network length and the original network length in the TAZ, where the modified network is a hypothetical network in which all links are straight (maintaining the original nodes). A lower value of Lin represents more non-linearity in the existing network. Figure 5 shows the difference between the straight and non-straight links, where non-straight links represent any curved or irregular link.

\[
Lin = \frac{\text{Total Length of the Modified network in TAZ}}{\text{Total Length of the Network in TAZ}}
\] (4)
To modify the non-straight links to straight ones, the network can be exported from ArcGIS to AutoCAD software to manage the separation of the non-straight links from the straight links, and then modifying them into straight ones. Afterwards, the network is imported back to ArcGIS to aggregate the length of the modified network links and measure the linearity at each TAZ using equation 4.
Lastly, for the topography category, the average weighted slope of the network and the length of the network at each TAZ can be included as indicators in this category. The total length of the zonal network represents the size of the network infrastructure within a TAZ. The average weighted slope of the zonal network gives an indication of the average absolute steepness of the network within each zone.

The total length of the network is calculated by aggregating all the network links, regardless of their types, within each TAZ. The average weighted slope of the network in each TAZ is calculated according to the following steps. First, the absolute grades along each link are averaged to compute the average slope of each link, as shown in Figure 7, using the contour map of the city of Vancouver. Afterwards, the slope at each link is given a weight relative to its length. Finally, the average weighted slope of the links (in %) is calculated for each TAZ as shown in equation 5; where \( l \) represents the link length and \( s \) represent the link’s slope.

\[
\text{Average Weighted Slope in TAZ} = \frac{\sum_{i=1}^{n} l_i * s_i + l_2 * s_2 + \ldots + l_n * s_n}{\sum_{i=1}^{n} l_1 + l_2 + \ldots + l_n} \tag{5}
\]
3.2.4 Zonal Characteristics

A big set of non-graph zone characteristics variables are also investigated in this research to identify their impact on cyclist and pedestrian safety. As for the socio-demographic variables (i.e. population, employment, and household), they are already provided by the Emme2 model in an aggregated form at the different TAZs. For the rest of the variables, the aggregation and splitting of the various elements among the different TAZs is undertaken using the ArcGIS software.

The total lengths of the freeway, arterial, collector, and local roads can be included in the models as road class variables. The road segments of similar class are represented as links, and their
lengths are then split and aggregated at each TAZ. The proportion of each road class of the total road network length is then calculated for each TAZ. Figure 8 shows the city of Vancouver road network classes.

The length of the on-street bike links is aggregated for each TAZ, then the on-street and off-street bike link proportions are calculated. OnSt_Prop represents the proportion of the bike facilities, which are shared with the road, out of the whole bike network; while OffSt_Prop represents the proportion of the separated bike facilities.

![City of Vancouver Road Classes](image)

**Figure 8 City of Vancouver Road Classes**

Built environment elements include the aggregated number of traffic signals, actuated signals, bus stops, and light poles at each TAZ. These can then be divided by the corresponding TAZ
area to get the traffic signal, bus stop, and light pole densities. Figure 9 shows the distribution of the traffic signal and bus stop densities along the city of Vancouver TAZs.

Figure 9 (a) Traffic Signals Density and (b) Bus Stops Density within City of Vancouver TAZs
For the land use category, the areas of commercial, residential, and recreational zonings are aggregated for each TAZ to obtain the total area of each land use type. Figure 10 shows the residential, commercial, and recreational areas within the city of Vancouver. The land use areas can then be divided by the corresponding TAZ area to get the land use density of each type for each TAZ.

![Figure 10 Residential, commercial, and recreational areas in the City of Vancouver](image)

Table 1 provides the definitions and descriptive statistics of the analysis variables as follows.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crashes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CColl</td>
<td>Cyclist-Motorist Crashes over 5 Years</td>
<td>12.71</td>
<td>13.48</td>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td>PColl</td>
<td>Pedestrian-Motorist Crashes over 5 Years</td>
<td>15.45</td>
<td>11.45</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td><strong>Exposure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VKT</td>
<td>Vehicle Kilometer Travelled</td>
<td>4290.43</td>
<td>3315.10</td>
<td>189.46</td>
<td>22288.79</td>
</tr>
<tr>
<td>BKT</td>
<td>Bike Kilometer Travelled</td>
<td>1047.78</td>
<td>2102.07</td>
<td>0</td>
<td>21462.77</td>
</tr>
<tr>
<td>EXP&lt;sub&gt;c&lt;/sub&gt;</td>
<td>(VKT*BKT) for Each Zone</td>
<td>6.47E06</td>
<td>25.9E06</td>
<td>0</td>
<td>292.7E06</td>
</tr>
<tr>
<td>EXP&lt;sub&gt;w&lt;/sub&gt;</td>
<td>(VKT*W) for Each Zone</td>
<td>16.10E06</td>
<td>14.73E06</td>
<td>173375</td>
<td>70.96E06</td>
</tr>
<tr>
<td>W</td>
<td>Walk Trips</td>
<td>3971.64</td>
<td>2677.49</td>
<td>247.11</td>
<td>13906.56</td>
</tr>
<tr>
<td><strong>Socio-Economic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EmpD</td>
<td>Employment Density (Employment/Zone Area)</td>
<td>12236.26</td>
<td>26399.07</td>
<td>84.54</td>
<td>170910</td>
</tr>
<tr>
<td>HhsD</td>
<td>Household Density (Households/Zone Area)</td>
<td>416.52</td>
<td>436.35</td>
<td>0</td>
<td>2141.88</td>
</tr>
<tr>
<td>PopD</td>
<td>Population Density (Population/Zone Area)</td>
<td>8391.82</td>
<td>6995.85</td>
<td>0</td>
<td>33658.9</td>
</tr>
<tr>
<td><strong>Land Use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResD</td>
<td>Residential Density (Residential Areas/Zone Area)</td>
<td>0.34</td>
<td>0.20</td>
<td>0</td>
<td>0.67</td>
</tr>
<tr>
<td>CommD</td>
<td>Commercial Density (Commercial Areas/Zone Area)</td>
<td>0.08</td>
<td>0.11</td>
<td>0</td>
<td>0.58</td>
</tr>
<tr>
<td>RecD</td>
<td>Recreational Density (Recreational Areas/Zone Area)</td>
<td>0.10</td>
<td>0.13</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>Built Environment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SigD</td>
<td>Signal Density (Number of Signals/Zone Area)</td>
<td>14.26</td>
<td>18.43</td>
<td>0</td>
<td>110.55</td>
</tr>
<tr>
<td>StopD</td>
<td>Transit Stops Density (Number of Stops/Zone Area)</td>
<td>24.28</td>
<td>23.62</td>
<td>0</td>
<td>162.24</td>
</tr>
<tr>
<td>Ped_Act</td>
<td>Proportion of Pedestrian Actuated Traffic Signal (Number of Pedestrian Actuated Signals/Total Traffic Signals)</td>
<td>0.39</td>
<td>0.32</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PoleD</td>
<td>Light Poles Density (Number of Poles/Zone Area)</td>
<td>665.81</td>
<td>373.71</td>
<td>24.32</td>
<td>2188</td>
</tr>
</tbody>
</table>

*Table Continues in the following page*
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Road Facility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ArtColl_Prop</td>
<td>Arterial-Collector Roads Proportion (Arterial + Collector Roads Length/Road network Length)</td>
<td>0.35</td>
<td>0.21</td>
<td>0.12</td>
<td>1</td>
</tr>
<tr>
<td>Loc_Prop</td>
<td>Local Roads Proportion (Local Roads Length/Road Network Length)</td>
<td>0.64</td>
<td>0.21</td>
<td>0</td>
<td>0.87</td>
</tr>
<tr>
<td>OffSt_Prop</td>
<td>Proportion of Off-Street Bike Links (Total Length of Off-Street Bike Links/Road Network Length)</td>
<td>0.00014</td>
<td>7.51x10^{-5}</td>
<td>0</td>
<td>0.0009</td>
</tr>
<tr>
<td><strong>Bike Network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CConn</td>
<td>Degree of Bike Network Connectivity</td>
<td>0.38</td>
<td>0.11</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>InterD</td>
<td>Intersections Density</td>
<td>74.28</td>
<td>33.80</td>
<td>6.08</td>
<td>235.95</td>
</tr>
<tr>
<td>CCov</td>
<td>Degree of Bike Network Coverage</td>
<td>0.34</td>
<td>0.19</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CNetD</td>
<td>Bike Network Density</td>
<td>5.38</td>
<td>3.78</td>
<td>0</td>
<td>21.91</td>
</tr>
<tr>
<td>CComp</td>
<td>Complexity of the Bike Network</td>
<td>0.99</td>
<td>0.14</td>
<td>0</td>
<td>1.66</td>
</tr>
<tr>
<td>CLin</td>
<td>Linearity</td>
<td>0.98</td>
<td>0.03</td>
<td>0.84</td>
<td>1</td>
</tr>
<tr>
<td>CAvgEdLen</td>
<td>Bike Network Average Edge Length</td>
<td>0.13</td>
<td>0.05</td>
<td>0</td>
<td>0.57</td>
</tr>
<tr>
<td>CAvgVerLen</td>
<td>Average Node Length</td>
<td>0.13</td>
<td>0.054</td>
<td>0</td>
<td>0.62</td>
</tr>
<tr>
<td>CSlope</td>
<td>Average Weighted Slope for Bike Network</td>
<td>2.52</td>
<td>0.90</td>
<td>0.63</td>
<td>6.65</td>
</tr>
<tr>
<td>CLen</td>
<td>Total Length of Bike Network Links</td>
<td>3.37</td>
<td>2.52</td>
<td>0</td>
<td>17.40</td>
</tr>
<tr>
<td><strong>Pedestrian Network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>InterD</td>
<td>Intersections Density</td>
<td>74.28</td>
<td>33.79</td>
<td>6.07</td>
<td>235.95</td>
</tr>
<tr>
<td>PConn</td>
<td>Degree of Pedestrian Network Connectivity</td>
<td>0.47</td>
<td>0.058</td>
<td>0.32</td>
<td>0.70</td>
</tr>
<tr>
<td>PCov</td>
<td>Degree of Pedestrian Network Coverage</td>
<td>0.95</td>
<td>0.21</td>
<td>0.68</td>
<td>2.41</td>
</tr>
<tr>
<td>PNetD</td>
<td>Pedestrian Network Density</td>
<td>0.015</td>
<td>0.0034</td>
<td>0.0037</td>
<td>0.025</td>
</tr>
<tr>
<td>PComp</td>
<td>Complexity of the Pedestrian Network</td>
<td>1.36</td>
<td>0.16</td>
<td>0.94</td>
<td>1.72</td>
</tr>
<tr>
<td>PLin</td>
<td>Linearity</td>
<td>0.61</td>
<td>0.16</td>
<td>0.23</td>
<td>0.99</td>
</tr>
<tr>
<td>PAvgEdLen</td>
<td>Pedestrian Network Average Edge Length</td>
<td>109.18</td>
<td>23.45</td>
<td>57.90</td>
<td>242.43</td>
</tr>
<tr>
<td>PAvgVerLen</td>
<td>Average Length Per Vertex</td>
<td>148.14</td>
<td>31.63</td>
<td>65.14</td>
<td>282.84</td>
</tr>
<tr>
<td>PSlope</td>
<td>Average Weighted Slope for Pedestrian Network</td>
<td>3.01</td>
<td>1.77</td>
<td>0.53</td>
<td>14.76</td>
</tr>
<tr>
<td>PLen</td>
<td>Total Length of Pedestrian Network Links</td>
<td>12</td>
<td>8.78</td>
<td>0.95</td>
<td>54.30</td>
</tr>
</tbody>
</table>
Chapter 4: Crash Models

This chapter discusses the crash modeling and statistical inference techniques that are followed in this research. The chapter is comprised of two main sections, i.e., EB approach and FB approach, which are elaborated as follow.

4.1 EB Approach

4.1.1 Crash Models for Pedestrians and Cyclists

The generalized linear modeling (GLM) form used for EB CMs should generally satisfy two conditions (Sawalah and Sayed, 2006). First, it should not yield negative results in terms of not predicting negative crashes and also should predict zero crashes for zero traffic exposure (when there are no vehicles or bikes on the road). As well, there needs to be a link function to transform the model into a linear form. Based on empirical studies (Miaou and Lum, 1993); (Sawalha and Sayed, 2001), a negative binomial GLM model form that is able to handle the overdispersion in the count data is commonly used. It includes an exposure measure (e.g. vehicle kilometers traveled) raised to some power and multiplied by an exponential function including the other non-exposure explanatory variables. The GLM model can be expressed mathematically as shown in equation 6. The number of pedestrian/cyclist crashes in zone i is assumed to follow a Poisson distribution with parameter $\mu_i$ which itself is considered a random variable with a gamma-distributed error term.

$$\mu_i = a_0 V^{a_1} V^{a_2} exp (\Sigma b_j x_j)$$

(6)
Where \( \mu_i \) is the predicted crash frequency, \( V \) is the measure of the traffic exposure (e.g. VKT, BKT, \( W \)), \( x_j \) are any other explanatory variables (e.g. network indicators, land use, etc.), and \( a_0 \), \( a_1 \), \( a_2 \), and \( b_j \) are parameters of the model.

The recommended procedure to add the explanatory variables into the CM is a forward stepwise procedure (Sawalha and Sayed, 2001). Variables are added one by one, and their significance is tested. Variables representing exposure must be included first. Once all the variables are evaluated, two statistical measures are used to assess the goodness of fit of the GLM models, including Pearson chi square (\( \chi^2 \)) and scaled deviance (SD) statistics. For a well-fitted model and a relatively large number of observations, the expected value of Pearson \( \chi^2 \) and SD will be approximately equal to the number of degrees of freedom (df) (Sawalha and Sayed, 2001). The SAS software is used to develop the CMs, undergo the significance tests for the explanatory variables, and assess the goodness of fit of the developed models.

### 4.1.2 EB Estimation

The estimate of the expected number of collisions at a TAZ is refined by combining the observed number of crashes at the TAZ with the predicted number of crashes obtained from CM to yield a more accurate, TAZ-specific safety estimate. The EB safety estimate of the expected number of crashes at any TAZ can be calculated by using equation 7:

\[
\theta_i = \alpha \times \mu_i + (1-\alpha) \times \text{count} \tag{7}
\]

Where
\( \theta_i \) : EB expected safety estimate.

\( \mu_i \) : predicted crash frequency from the developed CM.

\[ \alpha = \frac{k}{k + \mu_i} \]

\( k \) : over dispersion parameter of the developed CM.

### 4.2 FB Approach

#### 4.2.1 Univariate (Uni-response) Models for Pedestrian and Cyclist Crashes

Area-based Poisson-lognormal models are used in this research within the FB framework. These models are able to handle the overdispersion in the count data. They are also capable of accounting for the unobserved heterogeneity by allowing random effects and random parameters in the model structure as shown in the following subsections. The development of the models in this research followed some procedures similar to that described by El-Basyouny (2010).

#### 4.2.1.1 Poisson Lognormal Model

\( Y_i \) is assumed to be the number of pedestrian/cyclist crashes in zones \( i \), and \( Y \) is assumed to follow a Poisson distribution with parameter \( \lambda \) which itself is considered a random variable. The overdispersion caused by the unobserved or unmeasured heterogeneity is accounted for as shown in equation 8, where \( u_i \) is related to site-specific attributes and heterogeneity, which follows a lognormal distribution.

\[
\ln \lambda_i = a_0 + a_1 \ln(EXP_{1i}) + a_2 \ln(EXP_{2i}) + \sum_n \beta_n X_{ni} + u_i
\]

(8)
Where $a_0$ is the intercept value, $a_1$, $a_2$, and $\beta_n$ are model parameters, $EXP_{1i}$ and $EXP_{2i}$ are traffic exposure variables (VKT with BKT in case of cyclist crash model, and VKT with W in case of pedestrian crash models), $X_{ni}$ represents explanatory variables (e.g. network indicators, land use, etc.), $u_i$ accounts for the spatially unstructured random error among zones and follows a lognormal distribution as implied by equation 9.

$$u_i \sim \text{normal}(0, \sigma_u^2) \quad (9)$$

### 4.2.1.2 Spatial Poisson Lognormal Model (SPLN)

A random effects model can be formulated by introducing a spatial effect term as shown in equation 10, where $s_i$ is a spatially structured conditional autoregressive term for zone $i$.

$$\ln \lambda_i = a_0 + a_1 \ln(EXP_{1i}) + a_2 \ln(EXP_{2i}) + \sum \beta_n X_{ni} + u_i + s_i \quad (10)$$

Spatial autocorrelation is a technical term for the fact that spatial data from near sites are more likely to be similar than data from a distant site (O’Sullivan and Unwin, 2014). The first-order neighbor approach is used in which all zones that are directly connected with the zone being dealt with are included (Karim et al., 2013). Often, for fitting complex models in spatial statistics context, Bayesian methods are advocated because they are computationally more convenient than likelihood-based methods (Torabi, 2012). This study used first-order neighbors to define the neighboring structure. The purpose of performing the spatial analysis is to model the spatial factors and their correlations across zones. The spatial effects of the random error in this research can be accounted for by Gaussian CAR techniques and $S_i$ is calculated according to equation 11.
Where $\sigma^2_{\delta}$ is the spatial variation, and $n_i$, $C(i)$, and $S_{-i}$ represent the number of neighbors of zone $i$, the set of neighbors of zone $i$, and the set of all spatial effects except $S_i$, respectively. Equation 11 is based on an adjacency-based proximity measure, where the conditional variance is inversely proportional to the number of neighboring zones, and the conditional mean is the mean of the adjacent spatial effects.

The spatial effects are assessed by computing the spatial variation proportion out of the total variation according to equation 12.

$$\Psi_s = \frac{\sigma^2_{S}}{\sigma^2_{S} + \sigma^2_u}$$  \hspace{1cm} (12)$$

Where $\sigma^2_{S}$ is the marginal variance of $s$, and it can be directly estimated from the posterior distribution of $s$. $\sigma^2_u$ is the variance due to the spatially unstructured random error. It should be noted that although including spatial effects usually increases the models’ goodness of fit, it can sometimes affect the estimation of the parameters by making some variables non-significant even if they were significant in the CMs without spatial effects (Karim et al., 2013).

**4.2.1.3 Random Parameters Poisson Log-normal (RPPLN) Model**

The PLN model consist of a set of fixed parameters. However, the impact of the explanatory variables on crash counts may vary across zones or group of zones. It is possible that some
variables may have higher/lower impact on certain zone/cluster of zones than the other. To address this issue, the zonal variations can be accounted for by allowing the regression coefficients in the aforementioned PLN model to vary randomly from one zone cluster to another.

Typically, the $i$ TAZs belong to $G$ mutually exclusive groups. Assuming that the $i^{th}$ TAZ belongs to a group $g(i) \in \{1, 2, \ldots, G\}$. This leads to a RPPLN model by groups, which are (i.e., the groups) defined by the city of Vancouver 23 neighborhoods.

\[
\ln(\lambda_i) = \beta_{g(i),0} + \beta_{g(i),1}\ln(EXP_{1i}) + \beta_{g(i),2}\ln(EXP_{2i}) + \beta_{g(i),3}X_{i,3} + \ldots + \beta_{g(i),n}X_{i,n} + u_i \tag{13}
\]

Where $\beta_{g(i),n} \sim N(\beta_n, \sigma_n^2)$, $n = 0, 1, \ldots, n$. It should be noted that a random parameter $\beta_{g(i),n}$ is used when the posterior estimate deviation $\sigma_j^2$ is significantly larger than 0; otherwise, the parameter $\beta_n$ is fixed across groups. In this study, a total of $i = 134$ TAZs in the city of Vancouver belonging to $g = 23$ neighborhoods.

4.2.1.4 Priors Specification

The specification of prior distribution of the parameters is required before the FB estimates. Prior distribution reflects the prior knowledge about the considered parameters. The prior may be informative or vague according to the availability of prior information. Generally, a diffused normal distribution with a zero mean and a large variance is the most commonly used prior to estimate the regression parameters (El-Basyouny and Sayed 2009). For $\sigma_u^2$ and $\sigma_j^2$, the commonly used prior is a gamma distribution with parameters ($\epsilon, \epsilon$), where the value of $\epsilon$ is a
small number, e.g. 0.001 (Karim, Wahba and Sayed 2013). For $\sigma_s^2$, the prior distribution of $\sigma_s^2$ is assumed to be a gamma distribution with parameters $(1+\Sigma l_i/2, 1+n/2)$, where $l_i$ is the term contributed by each zone and is calculated as in equation 14.

$$l_i = n_is_i(s_i - \bar{s}_i)$$  \hspace{1cm} (14)

### 4.2.2 Multivariate (Multi-Response) Models for Pedestrian and Cyclist Crashes

In order to undergo the multivariate modeling of pedestrian and cyclist crashes, there are two approaches; the standard approach and the mixed approach (Wright, 1998).

The standard approach is to include the same set of covariates for the two modeled crash types. This can be achieved by either including all the available explanatory variables that would impact the pedestrian or cyclist crashes in one model; or by including only the explanatory variables that are expected to commonly impact both the pedestrian and cyclist crashes. However, both ways suffer from shortcomings. If all the possible explanatory variables are included, it is expected that some of these variables will be irrelevant for one type of crashes, which would affect the precision of the model estimates. On the other hand, if only the common variables that are expected to commonly affect both modes’ crashes are included, important associations may be missed due to omitting essential variables, which would bias the model estimates. The standard approach can work well when modeling multivariate crash severity levels, which are usually explained by the same set of covariates. However, this approach seems to be inadequate for modeling different crash types/modes that should be explained using mixed covariates (i.e. different exposure measures and explanatory variables).
Accordingly, another approach for multivariate modeling can be proposed, which allows including a different set of covariates for each modeled crash type in order to overcome the aforementioned shortcomings. This approach is widely used for fitting mixed effects, and is also a valuable tool for multivariate analysis. Capabilities of the mixed approach, which are lacking in the standard multivariate procedures, include, but not limited to, the ability to use observations that have incomplete responses; and the ability to handle non-standard (e.g., multiple designs) multivariate models (Wright, 1998).

\( Y_i \) is assumed to be the number of pedestrian/cyclist crashes in zones \( i \), and \( Y \) is assumed to follow a Poisson distribution with parameter \( \lambda \) which itself is considered a random variable.

\[ Y_i^k \sim \text{Poisson}(\lambda_i^k) \]  

\[ \ln(\lambda_i^k) = a_0^k + a_1^k \ln(\text{EXP}_{1i}^k) + a_2^k \ln(\text{EXP}_{2i}^k) + \sum_n \beta_n^k x_{ni}^k + u_i^k + s_i^k \]  

For \( k \) as the mode type, \( a_0^k \) is the intercept value, \( a_1^k \), \( a_2^k \), and \( \beta_n^k \) are model parameters, \( \text{EXP}_{1i}^k \) and \( \text{EXP}_{2i}^k \) are the traffic exposure variables (VKT with BKT in the case of cyclist crash model and VKT with W in the case of pedestrian crash model), and \( x_{ni}^k \) represents explanatory variables. The superscript “\( k \)” is added to the exposure and explanatory variables to indicate that the independent variables can vary from one mode to another (i.e. can be missing in the model of one mode but present in the other). Through this model structure, we have the ability to change the covariates included in the crash model for each crash type. The overdispersion caused by the unobserved or unmeasured heterogeneity is accounted for using \( u_i^k \) which is related to zone-
specific attributes. It represents spatially unstructured random error and follows a multivariate normal distribution. $s_i^k$ is the spatial component that represents structured correlated random effects suggesting that zones that are close to each other are correlated. The spatial effects can be accounted for by multivariate Gaussian CAR techniques.

The procedure of defining the multivariate normal and multivariate CAR error structures is similar to that done by (Islam et al., 2016) and as illustrated in equations 17 and 18. For the multivariate k-dimensional (only 2 modes in this model) normal error, the diagonal elements of the variance–covariance matrix $\Sigma$ represent the variances, and the off-diagonal elements represent the covariances. For model estimation, the following prior is used: $\Sigma^{-1} \sim \text{Wishart} (I, k)$, where $I$ is the $k \times k$ identity matrix.

$$u_i^k \sim \text{Normal}(0, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \quad (17)$$

For the multivariate CAR model, the vector of spatially correlated i neighborhood is

$$(S_{1i}, S_{2i}) | (S_{-1i}, S_{-2i}) \sim \text{Normal} (\overline{s_{ki}}, \omega), \quad \omega = \begin{pmatrix} \sigma_{s11} & \sigma_{s12} \\ \sigma_{s21} & \sigma_{s22} \end{pmatrix} \quad (18)$$

Where, $(S_{1i}, S_{2i})$ denotes the neighborhoods of the $k \times n$ matrix $s_i^k$, excluding the $i^{th}$ neighborhood; $n_i$ is the number of neighborhoods adjacent to a neighborhood, $\omega$ is the variance–covariance matrix for spatial correlation. The diagonal elements of the covariance matrix $\omega$ represent the spatial variance. The off-diagonal elements represent the spatial covariance of
different crash modes. For model estimation, the prior is assumed as $\omega^{-1} \sim \text{Wishart} (I, k)$, where $I$ is the $k \times k$ identity matrix.

### 4.2.3 FB Estimation

WinBUGS tool, which is statistical software for Bayesian analysis using Markov chain Monte Carlo (MCMC) methods, is used to sample the posterior distribution as well as to estimate the parameters. MCMC methods can sample from the joint posterior distribution repeatedly. This technique generates sequences (chains) of random points, the distributions of which converge to the target posterior distributions. A burn-in sample is used for the purpose of monitoring convergence and then excluded. Parameter estimation, performance evaluation, and inference are obtained by the following iterations.

Two chains are used to run each model in WinBUGS, and 20,000 MCMC iterations are discarded as burn-in sample. Afterwards, 20,000 iterations are performed for each chain. The summary statistics of each chain are then estimated from WinBUGS and the convergences of the developed models are thoroughly checked to ensure that the posterior distribution has been established for parameter sampling. Convergence can be checked in several ways. First, two or more parallel chains with diverse starting values are tracked so that full coverage of the sample space is ensured. Brooks–Gelman–Rubin statistic is also used to check the convergence of multiple chains, where convergence occurs if the value of the Brooks–Gelman–Rubin statistic is less than 1.2 (El-Basyouny and Sayed, 2009). Moreover, convergence can be checked by visually inspecting the MCMC trace plots of the model parameters. Finally, the ratios of the Monte Carlo errors relative to the respective standard deviations of the estimates can be
calculated as a measure of convergence. As a rule of thumb, convergence occurs when these ratio values are less than 0.05. In addition to convergence, the significance of the parameter estimates is tested at the 95% level using the credible intervals.

4.2.4 Models Comparison

The FB models can be compared using the Deviance Information Criteria (DIC), which is a measure of model complexity and fit. Generally, the model with smaller DIC outperforms the models with larger DIC and offers the best short-term predictions. Spiegelhalter et al. (2002) proposed the DIC as a measure of model complexity and fit. Let $D$ denote the un-standardized deviance of the postulated model then:

$$DIC = \bar{D} + P_D \quad ; \quad P_D = \bar{D} - \bar{D} \quad (19)$$

Where $\bar{D}$ is the posterior mean of $D$, $\bar{D}$ is the point estimate obtained by substituting the posterior means of the model’s parameters in $D$ and $P_D$ is a measure of model complexity estimating the effective number of parameters.

As a goodness of fit measure, DIC is a Bayesian generalization of Akaike’s information criteria (AIC) that penalizes larger parameter models. DIC $\approx$ AIC in cases with weak prior information (Breheny, 2012). El-Basyouny (2010) showed that DICs are additive under independent models and priors. According to Spiegelhalter et al. (2002), the models with DIC/AIC difference within the minimum value lower than two deserve to be considered as equally well, while models with
values ranging within 2-7 show considerably less support for the model with higher DIC. Values higher than 7 completely omit the model with higher DIC.
Chapter 5: Evaluating the Impact of Network Configuration on Cyclist and Pedestrian Safety

This chapter discusses the results of the crash models developed to evaluate the impact of the graph-based bike and pedestrian network indicators, discussed in section 3.2.3, on cyclist and pedestrian crashes. Although few macro-level studies looked into the association between networks’ patterns and the safety of vulnerable road users (e.g. Zhang et al., 2015; Wei and Lovegrove, 2012), almost none have investigated the impact of a comprehensive set of indicators, representing bike and pedestrian network structures, on active transportation safety.

GLM CMs are first developed to investigate the significant variables within the network indicators’ categories (i.e., connectivity, directness, and topography). Afterwards, a model is developed to merge the variables from all the categories into one CM in order to yield the best predictability. To address potential multi-collinearity when more than one attribute is included in the combined CMs it is ensured that the independent variables in those CMs are not strongly, or even moderately, correlated. The procedure for selecting the variables for the combined models is a forward stepwise procedure. The order in which the variables are added is based on their P-value, from lowest to highest. Whether to add or remove a variable in the model is decided based on the parameter’s statistical significance. FB models are then developed to validate the results of the GLM models and to account for the spatial effects. The aforementioned process is conducted for pedestrian crashes and then replicated for cyclist crashes.
5.1 Pedestrian Safety Models

Using the methodologies discussed in chapter 4, EB and FB macro-level CMs are developed incorporating traffic exposure and pedestrian network indicators. Table 2 shows the developed negative binominal GLM CMs, while Table 3 shows the FB models incorporating spatial effects. The goodness of fit statistics and the explanatory variables significance are provided in both tables. The models show good fit; with almost all the explanatory variables being statistically significant at the 5% level.

5.1.1 Traffic Exposure Indicators

Different pedestrian exposure measures, such as population, employment, and walk trips, were tested to represent pedestrian exposure along with the vehicle exposure measure (i.e. VKT). Walk trips measure was found to produce the best fit models for pedestrian crashes and the highest exponent for pedestrian exposure in the CMs, that’s why it is used in this study to represent the pedestrian exposure. Walk trips and vehicle kilometer travelled variables are used as the main exposure variables in all the developed models. Positive non-linear associations are found between pedestrian crashes and both vehicle kilometers travelled and walk trips. These results are plausible and in agreement with previous studies by Amoh-Gyimah et al. (2016), Chen et al. (2016), and Cai et al. (2016). The exponents of the exposure variables are less than one, which supports the safety in numbers hypothesis introduced by Jacobsen (2003). Specifically, more pedestrian trips would lead to less pedestrian crashes per trip. Therefore, programs such as walk to work (or walk to school) day as well as other share the road training campaigns can be efficient in promoting pedestrian trips and reducing pedestrian crash rates.
Walk trips exponent has higher value than VKT exponent, which means that the pedestrian exposure contributes more to the pedestrian-motorist crashes than the motorized traffic exposure.

### 5.1.2 Network Graph Indicators

The CMs’ results for the different graph indicators’ categories are discussed as follow.

#### 5.1.2.1 Network Connectivity

The CMs show that there is a positive association between the pedestrian-vehicle crashes and most of the pedestrian network connectivity measures (i.e., InterD, Conn, Comp, and NetD). The positive association between intersection density and pedestrian crashes reported in this study is similar to that reported in Siddiqui et al. (2012). This can be attributed to the higher pedestrian-vehicle interactions at the intersections, which would lead to an increased crash risk.

The degree of connectivity is related to the number of pedestrian network links and network configuration, and is found to be positively associated with pedestrian crashes. This is likely due to the fact that more links between the nodes would lead to higher exposure to pedestrian-vehicle conflicts and consequently higher crash potential. Similar results were found for transit networks, where the degree of connectivity was found positively associated with crash frequency (Quintero et al., 2013). As for network density, a higher value indicates more link density and, presumably, better zone coverage and network connectivity (Dill, 2004). Therefore, it is plausible that the relationship between network density and crashes is similar to the one between the degree of connectivity and crashes. The positive associations between pedestrian crashes and network connectivity, network density, and intersection density are similar to the results of a previous
study by Chao et al. (2009). They found that higher crash risk for cyclist and pedestrian is associated with higher intersection density, sidewalk density, and sidewalk connectivity. The results are contrary to Mirenda-Moreno et al. (2011) study, where they found that sidewalk density is negatively associated with pedestrian crashes. A model is also developed to evaluate the impact of pedestrian network complexity on the frequency of pedestrian crashes. The model resulted in a positive association between complexity and pedestrian crashes. Lastly, the coverage indicator quantifies the amount of street links that are covered by pedestrian network. The negative association between pedestrian network coverage and pedestrian crashes can be attributed to the extensive presence of separated pedestrian paths in some zones or to the fact that more sidewalks covering the streets would intuitively give better protection to pedestrians. This result agrees with a recent study by Yu (2016), in which he found that the sidewalk coverage of the street network was negatively associated with pedestrian crashes.

### 5.1.2.2 Network Directness

The average edge length and average length per vertex variables are found to be negatively associated with pedestrian crashes. These results imply that longer links without hindrances or discontinuities is presumably more convenient and safer to the pedestrians. This association agrees with Quintero et al. (2013) safety study on Metro Vancouver transit network.

The pedestrian network linearity is also found to be negatively associated with pedestrian crashes. This result shows that irregular sidewalks are less safe for pedestrians. This can be attributed to pedestrians potentially bypassing the irregular parts of the sidewalk network by
walking on the streets without any protection from the motorized traffic, which would lead to higher crash risk.

5.1.2.3 Network Topography

The zonal length of the pedestrian network is found to be negatively associated with the pedestrian-vehicle crashes. This agrees with a recent study by Yu (2015) that concluded that more pedestrian infrastructure would improve pedestrian safety. The results are on contrary to the results reported by Cai et al. (2016), who used sidewalk length as a traffic exposure for pedestrian crashes. Also, a negative association is found between the weighted slope of the zonal pedestrian network and pedestrian crashes. This may be explained as vehicles usually reduce their speeds at climbing slopes and exert more caution to vulnerable road users at descending slopes, which would lower crash risk. This result is consistent with a study by Chen and Zhou (2016), in which they found that the proportion of steep areas within zones was negatively associated with pedestrian crashes.
Table 2 Negative Binomial GLM Analysis Estimates for Pedestrian Network Graph Indicators

<table>
<thead>
<tr>
<th>Model Coll=</th>
<th>$K$</th>
<th>df</th>
<th>SD</th>
<th>$X^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network Connectivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0005$W^{0.75}VKT^{0.44} \exp(0.004\text{IntersectDen})$</td>
<td>5.58</td>
<td>130</td>
<td>138.02</td>
<td>136.24</td>
</tr>
<tr>
<td>0.0006$W^{0.74}VKT^{0.41} \exp(29.10\text{PNetDen}^*)$</td>
<td>5.34</td>
<td>130</td>
<td>137.99</td>
<td>137.10</td>
</tr>
<tr>
<td>0.0042$EXP^{0.53} \exp(-0.58\text{PCov})$</td>
<td>4.54</td>
<td>131</td>
<td>139.77</td>
<td>145.88</td>
</tr>
<tr>
<td>0.011$W^{0.71} \exp(0.83\text{PComp})$</td>
<td>3.70</td>
<td>131</td>
<td>139.43</td>
<td>129.31</td>
</tr>
<tr>
<td>0.31$VKT^{0.32} \exp(2.55\text{PConn})$</td>
<td>2.38</td>
<td>131</td>
<td>143.63</td>
<td>153.48</td>
</tr>
<tr>
<td><strong>Combined Model</strong> $0.001W^{0.71}VKT^{0.46} \exp(0.004\text{IntersectDen}-0.54\text{PCov})$</td>
<td>5.81</td>
<td>129</td>
<td>137.07</td>
<td>137.42</td>
</tr>
<tr>
<td><strong>Network Directness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0014$W^{0.76}VKT^{0.47} \exp (-0.0091\text{PAvgEdLen})$</td>
<td>5.88</td>
<td>130</td>
<td>136.68</td>
<td>138.98</td>
</tr>
<tr>
<td>0.0006$W^{0.80}VKT^{0.52} \exp (-0.0065\text{PAvgVerLen})$</td>
<td>5.85</td>
<td>130</td>
<td>137.18</td>
<td>139</td>
</tr>
<tr>
<td>0.046$W^{0.75} \exp (-0.86\text{PLin})$</td>
<td>3.69</td>
<td>131</td>
<td>140.30</td>
<td>130.96</td>
</tr>
<tr>
<td><strong>Combined Model</strong> $0.003W^{0.76}VKT^{0.43} \exp (-0.010\text{PAvgEdLen}-0.52\text{PLin}^*)$</td>
<td>6.06</td>
<td>129</td>
<td>136.36</td>
<td>138.68</td>
</tr>
<tr>
<td><strong>Network Topography</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0002$W^{0.72}VKT^{0.65} \exp (-0.037\text{PLen})$</td>
<td>6.52</td>
<td>130</td>
<td>135.85</td>
<td>138</td>
</tr>
<tr>
<td>0.001$W^{0.78}VKT^{0.49} \exp (-0.097\text{PSlope})$</td>
<td>5.93</td>
<td>130</td>
<td>138.72</td>
<td>136.18</td>
</tr>
<tr>
<td><strong>Combined Model</strong> $0.0003W^{0.73}VKT^{0.65} \exp (-0.035\text{PLen}-0.087\text{PSlope})$</td>
<td>7.46</td>
<td>129</td>
<td>135.55</td>
<td>138.69</td>
</tr>
<tr>
<td><strong>Mutual Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0002$W^{0.79}VKT^{0.66} \exp (-0.031\text{PLen}+0.0029\text{IntersectDen}-0.089\text{WSlope})$</td>
<td>7.69</td>
<td>128</td>
<td>133.60</td>
<td>134.27</td>
</tr>
</tbody>
</table>

* Significant at the 10% level
All other parameters are significant at the 5% level or higher

5.1.2.4 Combined FB CMs with Spatial Effects

FB CMs incorporating spatial effects are developed to confirm the GLM models’ results. The effects of the variables in the FB models are consistent with the results of the GLM CMs. All the
variables are found significant except one variable (Lin). This is likely due to the inclusion of the spatial effects, which was demonstrated previously to have a negative impact on some variables significance even though it improves the CM fit (Karim et al., 2013). The spatial effects are found substantial in all the FB models ($\psi >> 0.50$), as shown in Table 3. This highlights the importance of considering spatial effects in the macro-level CMs that investigate the impact of pedestrian network graph indicators on pedestrian safety, which agrees with a few former pedestrian safety studies (Wang et al., 2016; Amoh-Gyimah et al., 2016; Cai et al., 2016).

Table 3 FB Analysis Estimates for Pedestrian Network Graph indicators

<table>
<thead>
<tr>
<th>Variable</th>
<th>Connectivity</th>
<th>Continuity</th>
<th>Topography</th>
<th>Mutual Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Intercept</td>
<td>-6.65</td>
<td>0.96</td>
<td>-6.81</td>
<td>1.06</td>
</tr>
<tr>
<td>EXP</td>
<td>0.58</td>
<td>0.055</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>W</td>
<td>-</td>
<td>-</td>
<td>0.84</td>
<td>0.11</td>
</tr>
<tr>
<td>VKT</td>
<td>-</td>
<td>-</td>
<td>0.39</td>
<td>0.090</td>
</tr>
<tr>
<td>IntersectDen</td>
<td>0.0032*</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCov</td>
<td>-0.59*</td>
<td>0.33</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PAvgEdLen</td>
<td>-</td>
<td>-</td>
<td>-0.005*</td>
<td>0.003</td>
</tr>
<tr>
<td>PLin</td>
<td>-</td>
<td>-</td>
<td>-0.15**</td>
<td>0.40</td>
</tr>
<tr>
<td>PLen</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PSlope</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.986</td>
<td>0.002</td>
<td>0.986</td>
<td>0.002</td>
</tr>
<tr>
<td>DIC</td>
<td>790.90</td>
<td>789.10</td>
<td>784.72</td>
<td>784.62</td>
</tr>
</tbody>
</table>

- Not Applicable  *Significant at 10% level  ** Non significant at the 10% level  
All other parameters are significant at the 5% level or higher

### 5.2 Cyclist Safety Models

Using the methodologies discussed in chapter 4, EB and FB macro-level CMs are developed incorporating traffic exposure and bike network indicators. Table 4 shows the developed...
negative binomial GLM CMs, while Table 5 shows the FB models incorporating spatial effects. The goodness of fit statistics and the explanatory variables significance are provided in both tables. The models show good fit; with almost all the explanatory variables being statistically significant at the 5% level.

5.2.1 Traffic Exposure Indicators

Two traffic exposure measures are investigated in this category, i.e. BKT, and VKT. Crash frequency is found non-linearly positively associated with all the investigated types of traffic exposure. These results are intuitive and consistent with several previous studies (Prato et al., 2016) (Haman and Peek-Asa, 2013) (Miranda-Moreno et al., 2011). The exponents of the exposure variables are less than one, which support the “safety in numbers” hypothesis (Jacobsen, 2003), and in line with the results form Prato et al. (2016) study. The results also agree with the studies that concluded the presence of positive associations between congestion and lower crash rates (e.g. Reynolds et al., 2009). The BKT has a higher exponent value than VKT, which shows the higher impact of BKT, than VKT, on cyclist-motorist crashes.

In order to incorporate the bike network indicators into the CMs, three exposure measures are investigated (i.e. BKT only, both BKT and VKT, and EXP) to develop three models for each network indicator. The CM that yields the highest statistical significance for the network indicator is the one reported in the results.

5.2.2 Network Graph Indicators

The CMs’ results for the different categories of graph indicators’ are discussed as follows.
5.2.2.1 Connectivity

The CMs show that there is a positive association between the cyclist-vehicle crashes and the various bike network connectivity measures (i.e., InterD, Conn, Cov, and NetD). The positive association between intersection density and bike crashes reported in this study is analogous to that reported in Siddiqui et al. (2012), Strauss et al. (2013), and Wei and Lovegrove (2013) studies. This can be attributed to the higher cyclist-vehicle interactions at the intersections, which would lead to an increased crash risk.

To the best of the authors’ knowledge, the safety effects of bike network’s degree of connectivity, coverage, density, and complexity has not been investigated in previous studies. The degree of connectivity and the degree of coverage are related to the number of bike network links, network configuration, and bike network coverage; and are found to be positively associated with bike crashes. This is likely due to the fact that more links between the nodes would lead to higher exposure to cyclist-vehicle conflicts, and consequently higher crash potential. Similar results were found for transit networks, where the degree of connectivity was found positively associated with crashes frequency (Quintero et al., 2013). As for network density, a higher value indicates better zone coverage and network connectivity (Dill, 2004). Therefore, it is intuitive that the relationship between network density and crashes is similar to the one between degree of connectivity and crashes. Lastly, a model is developed to evaluate the impact of bike network complexity on the frequency of cyclist crashes. Although the model resulted in a positive association between complexity and cyclist crashes, the complexity variable is found statistically non-significant, and therefore not reported in the results.
5.2.2.2 Directness

The average edge length is found to have a significant negative association with cyclist crashes, while the average length per vertex is found to be statistically non-significant. These results imply that a higher average edge length, which indicates longer links without hindrances or discontinuities, is presumably more convenient and safer to the cyclists. The association between average edge length as well as average length per vertex and crashes agrees with a safety study done by Quintero et al. (2013) on Metro Vancouver transit network. On the other hand, linearity is found positively associated with cyclist crashes. This positive association may be attributed to the tendency of motorists and cyclists to speed more or pay less attention on straight links, which would increase crash risk. However, such result needs further investigation and validation.

5.2.2.3 Topography

The zonal length of the bike network is found negatively associated with the cyclist-vehicle crashes. This agrees with recent studies that concluded that more bike infrastructure would increase cyclist safety (de Rome et al., 2014), (Kaplan et al., 2014), (Prato et al., 2015). Also, a negative association is found between the weighted slope of the zonal bike network and cyclist crashes. This may be explained as cyclists usually reduce their speeds at climbing slopes and be more attentive at descending slopes, which would lower crash risk. This result is consistent with a study by Chen (2015), who used a zonal mean slope variable to represent the average absolute slopes of the TAZs. Although the variable used in Chen’s study was non-significant at 5% level, it showed negative association between zonal mean slope and cyclist crashes. However a more recent study by Thomas et al. (2017) found similar results to this thesis results.
Three CMs (A, B, and C) are built to combine the significant cycling network attributes. The relationships between the various network attributes and cyclist crashes in the combined models are found similar to those in the models relating the attributes individually to cyclist crashes.

### 5.2.2.4 Combined FB CMs with Spatial Effects

Combined FB CMs incorporating spatial effects are developed and spatial effects are found highly significant (i.e. $\psi >> 0.50$) in all the developed models indicating the importance of spatial correlation. This highlights the necessity of accounting for spatial autocorrelation in the macro-level CMs that investigate the impact of bike network graph indicators on cyclist safety. Model A is found to have the best fit (lowest DIC) among the three reported combined models, which agrees with the GLM results. The inclusion of spatial effects caused two network indicators (i.e. network density and average weighted slope) to become non-significant. All the other network indicators are found significant at either 5% or 10% levels.
Table 4 Negative Binomial GLM Analysis Estimates for Cyclist Network Graph Indicators

<table>
<thead>
<tr>
<th>Model Coll=</th>
<th>$K$</th>
<th>df</th>
<th>SD</th>
<th>$X^2$</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.184BKT$^{0.643}$</td>
<td>2.52</td>
<td>131</td>
<td>142.57</td>
<td>125.16</td>
<td>$a_0$, BKT &lt;0.001</td>
</tr>
<tr>
<td>0.239VKT$^{0.484}$</td>
<td>1.31</td>
<td>132</td>
<td>149.82</td>
<td>132.40</td>
<td>$a_0$=0.081, VKT&lt;0.001</td>
</tr>
<tr>
<td>0.040BKT$^{0.59}$VKT$^{0.22}$</td>
<td>2.43</td>
<td>130</td>
<td>142.60</td>
<td>118.71</td>
<td>$a_0$, BKT &lt;0.001, VKT=0.0084</td>
</tr>
<tr>
<td>0.016EXP$^{0.45}$</td>
<td>2.27</td>
<td>131</td>
<td>144.42</td>
<td>117.33</td>
<td>$a_0$, EXP &lt;0.001</td>
</tr>
<tr>
<td><strong>Cycling Network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Connectivity Indicators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.016 BKT$^{0.58}$VKT$^{0.28}$ exp(0.006InterD)</td>
<td>2.63</td>
<td>129</td>
<td>141.88</td>
<td>111.10</td>
<td>$a_0$, BKT, VKT&lt;0.001, InterD=0.0038</td>
</tr>
<tr>
<td>0.012BKT$^{0.61}$VKT$^{0.27}$exp(1.80Conn)</td>
<td>2.56</td>
<td>129</td>
<td>141.83</td>
<td>117.48</td>
<td>$a_0$, BKT &lt;0.001, VKT=0.001, Conn=0.018</td>
</tr>
<tr>
<td>0.010 EXP$^{0.46}$exp(0.65CCov)</td>
<td>2.28</td>
<td>130</td>
<td>143.17</td>
<td>112.76</td>
<td>$a_0$, EXP &lt;0.001, Cov=0.061</td>
</tr>
<tr>
<td>0.0098 EXP$^{0.46}$exp(0.046CNetD)</td>
<td>2.38</td>
<td>130</td>
<td>143.01</td>
<td>111.80</td>
<td>$a_0$, EXP &lt;0.001, NetD=0.009</td>
</tr>
<tr>
<td><strong>Directness Indicators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.28BKT$^{0.63}$exp(-4.76CAvgEdLen)</td>
<td>2.38</td>
<td>130</td>
<td>142.31</td>
<td>125.13</td>
<td>$a_0$=0.005, BKT &lt;0.001, AvgEdLen=0.024</td>
</tr>
<tr>
<td>0.23*10$^{-4}$ BKT$^{0.65}$VKT$^{0.18}$exp(7.81CLin)</td>
<td>2.70</td>
<td>129</td>
<td>142.82</td>
<td>122.79</td>
<td>$a_0$, BKT &lt;0.001, VKT=0.023, Lin=0.001</td>
</tr>
<tr>
<td><strong>Topography Indicators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.020 BKT$^{0.61}$VKT$^{0.31}$ exp(-0.063CLen)</td>
<td>2.50</td>
<td>129</td>
<td>143.07</td>
<td>121.32</td>
<td>$a_0$, BKT, VKT&lt;0.001, L=0.049</td>
</tr>
<tr>
<td>0.053BKT$^{0.58}$VKT$^{0.24}$ exp(-0.17CSlope)</td>
<td>2.46</td>
<td>129</td>
<td>140.73</td>
<td>115.68</td>
<td>$a_0$, BKT &lt;0.001, VKT=0.003, WSlope=0.029</td>
</tr>
<tr>
<td><strong>Mutual Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A) 0.19*10$^{-4}$ BKT$^{0.58}$ VKT$^{0.25}$ exp(0.005InterD+7.51CLin-0.14CSlope)</td>
<td>3.03</td>
<td>127</td>
<td>141.39</td>
<td>121.23</td>
<td>$a_0$, BKT &lt;0.001, VKT=0.0017, Lin=0.001, InterD=0.014, WSlope=0.067</td>
</tr>
<tr>
<td>B) 0.032BKT$^{0.60}$ VKT$^{0.26}$exp(1.64CConn-3.58 CAvgEdLen-0.17CSlope)</td>
<td>2.70</td>
<td>127</td>
<td>140.37</td>
<td>117.27</td>
<td>$a_0$, BKT &lt;0.001, VKT=0.002, Conn=0.03, AvgEdLen=0.09, WSlope=0.02</td>
</tr>
<tr>
<td>C) 0.032EXP$^{0.52}$ exp(0.043CNetD-0.08CLen)</td>
<td>2.56</td>
<td>129</td>
<td>143.38</td>
<td>118.14</td>
<td>$a_0$, EXP &lt;0.001, NetD=0.01, L=0.008</td>
</tr>
</tbody>
</table>
### Table 5 FB Analysis Estimates for Cyclist Network Graph Indicators

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>SD</th>
<th>2.5%</th>
<th>97.5%</th>
<th>2.5%</th>
<th>97.5%</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-11.96*</td>
<td>2.09</td>
<td>0.98</td>
<td>0.98</td>
<td>-4.21*</td>
<td>0.78</td>
<td>-5.75</td>
<td>-2.66</td>
</tr>
<tr>
<td>BKT</td>
<td>0.46*</td>
<td>0.054</td>
<td>0.35</td>
<td>0.56</td>
<td>0.46*</td>
<td>0.05</td>
<td>0.36</td>
<td>0.57</td>
</tr>
<tr>
<td>VKT</td>
<td>0.39*</td>
<td>0.086</td>
<td>0.22</td>
<td>0.56</td>
<td>0.39*</td>
<td>0.08</td>
<td>0.22</td>
<td>0.55</td>
</tr>
<tr>
<td>EXP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>InterD</td>
<td>0.0035*</td>
<td>0.001</td>
<td>0</td>
<td>0.007</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CConn</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.06**</td>
<td>0.64</td>
<td>-0.19</td>
<td>2.31</td>
</tr>
<tr>
<td>CNetD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.016</td>
<td>0.018</td>
</tr>
<tr>
<td>CAvgEdLen</td>
<td>-</td>
<td>-</td>
<td>5.07*</td>
<td>0.016</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CLin</td>
<td>7.96*</td>
<td>1.95</td>
<td>4.14</td>
<td>11.79</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CLen</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CSlope</td>
<td>-0.034</td>
<td>0.065</td>
<td>-0.16</td>
<td>0.09</td>
<td>-0.06</td>
<td>0.06</td>
<td>-0.19</td>
<td>0.07</td>
</tr>
<tr>
<td>DIC</td>
<td>741.66</td>
<td></td>
<td></td>
<td></td>
<td>745.37</td>
<td></td>
<td></td>
<td>745.83</td>
</tr>
<tr>
<td>ψ</td>
<td>0.992</td>
<td>0.001</td>
<td>0.988</td>
<td>0.995</td>
<td>0.992</td>
<td>0.001</td>
<td>0.995</td>
<td>0.992</td>
</tr>
</tbody>
</table>

*Significantly different from zero at 5%
**Significantly different from zero at 10%
Chapter 6: Evaluating the Impact of Zonal Characteristics on Cyclist and Pedestrian Safety

This chapter discusses the results of the crash models that are developed to evaluate the impact of the non-graph zonal characteristics, discussed in section 3.2.4, on cyclist and pedestrian crashes. The CMs incorporate many variables that are related to socio-economics, land use, built environment, and road facility in order to investigate the impact of these characteristics on active transportation safety. Traffic exposure variables are incorporated in the CMs, i.e. bike kilometers travelled (BKT), or walk trips (W), along with vehicle kilometers travelled (VKT). The incorporation of an extensive set of active transportation correlates along with traffic exposure measures highlights the importance of this chapter and its contribution to the literature of active transportation safety.

GLM CMs are first developed to look into the significant correlates within each of the investigated categories (i.e., socio-economic, land use, built environment, and road facility). Afterwards, a model is developed to merge the variables from all the categories into one CM in order to yield the best predictability. To address the potential multi-collinearity, when more than one attribute is included in the combined CMs, it is ensured that the independent variables in those CMs are not strongly, or even moderately, correlated. The procedure for selecting the variables for the combined models is a forward stepwise procedure. The order in which the variables are added is based on their P-value, from lowest to highest. Whether to add a variable to the model or remove it is decided based on the parameter’s statistical significance. FB models
are then developed to validate the GLM models and to account for the spatial effects. This is conducted first for the pedestrian CMs and then replicated for the cyclists CMs as follows.

### 6.1 Pedestrian Safety Models

Using the methodologies discussed in chapter 4, EB and FB macro-level CMs are developed for pedestrian safety incorporating traffic exposure and non-graph zonal characteristics. Table 6 shows the developed negative binomial GLM CMs, while Table 7 shows the FB models incorporating spatial effects. The goodness of fit statistics and the explanatory variables significance are provided in both tables. The models show good fit; with almost all the explanatory variables being statistically significant at the 5% level. W and VKT are the main traffic exposure variables in all the models, as demonstrated in the previous chapter, and the results are discussed as follow.

#### 6.1.1 Traffic Exposure Indicators

Walk trips and vehicle kilometer travelled variables are used as the main exposure variables in all the developed models. Positive non-linear associations are found between pedestrian crashes and both vehicle kilometers travelled and walk trips. These results are plausible and in agreement with previous studies by Amoh-Gyimah et al. (2016), Chen et al. (2016), and Cai et al. (2016). The exponents of the exposure variables are less than one, which supports the safety in numbers hypothesis introduced by Jacobsen (2003). More pedestrian trips would lead to less pedestrian crashes per trip. Walk trips exponent has higher value than VKT exponent, which means that pedestrian exposure contributes more to the pedestrian-motorist crashes than the vehicle traffic exposure. Therefore, programs such as walk to work (or walk to school) day as well as other
share-the-road training campaigns can be efficient in promoting more pedestrian trips and reducing pedestrian crash rates.

6.1.2 Non-Graph Zonal Characteristics

The CMs’ results for the different categories of non-graph zonal characteristics are discussed as follows.

6.1.2.1 Socio-Economic Model

The socio-economic CM is primarily based on the explanatory variables extracted from census data. The model reveals positive associations between pedestrian crashes and both the employment and household densities. The results are reasonable since the aforementioned variables can be considered as surrogate measures for traffic exposure, thereby explaining their positive associations with pedestrian crashes. The result for the employment density agrees with previous studies by Siddiqui et al. (2012) and Cai et al. (2016).

6.1.2.2 Land Use Models

The models in this category incorporate explanatory variables that refer to land zonings within the TAZs. The results in Tables 6 and 7 show that the increase in residential and recreational area densities is associated with a decline in the frequency of pedestrian crashes. The association of the recreational area density with less pedestrian crashes is intuitive because these areas usually provide off-street and continuous paths for active transportation commuters reducing the conflict risk between the vulnerable commuters and vehicles. This result is consistent with a study by Ukkusuri et al. (2011), who found a negative association between parks total area and pedestrian
crashes. The negative association between the residential area density and pedestrian crashes can be explained by the ongoing traffic calming measures applied by city of Vancouver to promote active transportation and limit motorized traffic at the residential neighborhoods (http://vancouver.ca/streets-transportation/traffic-calming-and-safety.aspx). On the other hand, the increase in the commercial area density is found associated with the increase in pedestrian crashes. This can be attributed to the side street activities that raise the potential risk of a pedestrian going into conflict with motorized traffic. The association between commercial areas and pedestrian safety is in line with two previous studies (Kim et al., 2006; Ukkusuri et al., 2011).

6.1.2.3 Built Environment Models

Built environment variables refer to the elements that are physically present on the pedestrian and road networks. The models show that pedestrian crashes are positively associated with transit stop, traffic signal, and light pole densities. More traffic signals imply the presence of higher number of wide intersections that usually incorporate complex vehicle and pedestrian maneuvers elevating the probability of crash occurrence, which agrees with previous studies by Siddiqui et al. (2012) and Cai et al. (2016). Also, the presence of bus stops indicates the occurrence of interactions between buses, vehicles, and pedestrians, which is also expected to increase pedestrian crash risk; this agrees with the results from (Lee et al., 2015a). An unexpected finding is the positive association between light pole density and the pedestrian crashes. This can be attributed to the higher pedestrian volume (higher exposure) on the streets that have better lighting (more light poles), especially at night time. However, this result requires further validation. On the other hand, the proportion of pedestrian actuated traffic signals is
found negatively associated with pedestrian crashes. This is logical since such facilities are meant to provide safer pedestrian crossing experience.

### 6.1.2.4 Road Facility Models

For this category, higher proportion of arterial plus collector roads is found to be positively associated with pedestrian crash frequency. This can be attributed to the higher speed and heavier traffic on these types of roads, which would increase the risk of severe interactions and conflicts between pedestrians and vehicles. On the other hand, a decline in the pedestrian-motorist crash frequency is found associated with higher proportion of local roads. A likely reason for such negative association is the relatively low speeds on local roads, which would result in the increase in drivers’ attentiveness, and, therefore, would reduce conflict potential. It can also be attributed to the City of Vancouver’s traffic calming measures. The former results agree with previous studies conducted by Wang and Kockelman, (2013) and Siddiqui et al. (2012).

### 6.1.2.5 Combined FB CMs with Spatial Effects

FB models are then developed for the combined CMs. The effects of the variables in the FB models are consistent with the results from the GLM CMs. All the variables are found significant at 5% level except four variables; RecD, ResD, PedAct, and PolesD. This is probably due to the inclusion of spatial effects, which was shown in a previous study to potentially negatively impact the significance of some variables (Karim et al., 2013). That’s why the models without spatial effects (GLM CMs) were preliminarily estimated to be able, after developing the FB models, to capture the impact of spatial effects on the parameters’ significance. The spatial effects are found
substantial in all the FB models ($\psi >> 0.50$) as shown in Table 7. This highlights the importance of considering spatial effects in the macro-level CMs that investigate the impact of non-graph zonal characteristics on pedestrian safety, which agrees with a few former studies investigating pedestrians safety (Wang et al., 2016; Amoh-Gyimah et al., 2016; Cai et al., 2016).

Table 6 Negative Binomial GLM Analysis Estimates for Pedestrian Non-Graph Zonal Characteristics

<table>
<thead>
<tr>
<th>Model Coll=</th>
<th>$K$</th>
<th>AIC</th>
<th>df</th>
<th>SD</th>
<th>$X^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socio-Economic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0008$W^{0.61}VKT^{0.54}$exp($2.8\times10^{-5}HhsD^*+0.56\times10^{-5}EmpD$)</td>
<td>5.88</td>
<td>881.7</td>
<td>129</td>
<td>138.40</td>
<td>143.79</td>
</tr>
<tr>
<td><strong>Land Use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0007$W^{0.69}VKT^{0.47}$exp($1.40CommD$)</td>
<td>5.88</td>
<td>876.9</td>
<td>130</td>
<td>139.13</td>
<td>140.29</td>
</tr>
<tr>
<td>0.0014$W^{0.30}VKT^{0.42}$exp($-1.07RecD-0.74ResD$)</td>
<td>6.13</td>
<td>873.9</td>
<td>129</td>
<td>138.59</td>
<td>142.86</td>
</tr>
<tr>
<td><strong>Built Environment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0006$W^{0.63}VKT^{0.54}$exp($0.0077SigD+5.9PoleD$)</td>
<td>6.75</td>
<td>866.6</td>
<td>129</td>
<td>137.76</td>
<td>142.93</td>
</tr>
<tr>
<td>0.0006$W^{0.70}VKT^{0.50}$exp($0.0066TransitD-0.30PedAct$)</td>
<td>6.17</td>
<td>872.4</td>
<td>129</td>
<td>137.65</td>
<td>139.29</td>
</tr>
<tr>
<td><strong>Road Facility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0005$W^{0.72}VKT^{0.47}$exp($0.81Art_Coll$)</td>
<td>6.09</td>
<td>872.8</td>
<td>130</td>
<td>138.76</td>
<td>144.98</td>
</tr>
<tr>
<td>0.0012$W^{0.72}VKT^{0.47}$exp($-0.79Loc_Prop$)</td>
<td>6.02</td>
<td>873.5</td>
<td>130</td>
<td>138.66</td>
<td>144.59</td>
</tr>
<tr>
<td><strong>Mutual Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0008$W^{0.59}VKT^{0.51}$exp($6.86PoleD-0.79RecD+0.006SigD^*$)</td>
<td>6.94</td>
<td>864.3</td>
<td>128</td>
<td>137.35</td>
<td>144.47</td>
</tr>
</tbody>
</table>

* Significant at the 10% level
All other parameters are significant at the 5% level or higher
Table 7 FB Analysis Estimates for Pedestrian Non-Graph Zonal Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Land Use</th>
<th>Built Environment</th>
<th>Road Facility</th>
<th>Mutual Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>Estimate</td>
<td>SD</td>
<td>Estimate</td>
</tr>
<tr>
<td>Intercept</td>
<td>-7.30</td>
<td>-6.98</td>
<td>0.92</td>
<td>-7.07</td>
</tr>
<tr>
<td>VKT</td>
<td>0.43</td>
<td>0.34</td>
<td>0.09</td>
<td>0.35</td>
</tr>
<tr>
<td>W</td>
<td>0.73</td>
<td>0.86</td>
<td>0.10</td>
<td>0.78</td>
</tr>
<tr>
<td>EmpD</td>
<td>0.72x10^-5</td>
<td>0.35x10^-5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HhsD</td>
<td>4.6x10^-5</td>
<td>2x10^-5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RecD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ResD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CommD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SigD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TransitD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PoleD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PedAct</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Art_Coll</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Loc_Prop</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ψ</td>
<td>0.96</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>DIC</td>
<td>787.54</td>
<td>788.96</td>
<td>787.84</td>
<td>787.93</td>
</tr>
</tbody>
</table>

- Not Applicable
* Non significant at the 5% level
All other parameters are significant at the 5% level or higher
6.2 Cyclist Safety Models

Using the methodology discussed in chapter 4, macro-level CMs for cyclist safety are developed incorporating traffic exposure and non-graph zonal characteristics. Table 8 shows the developed negative binomial GLM CMs, while Table 9 shows the FB models incorporating spatial effects. The goodness of fit statistics and the explanatory variables significance are provided in both tables. The models show good fit; with almost all the explanatory variables being statistically significant at the 5% level. BKT and VKT are the main traffic exposure variables in all the models, and the results are discussed as follow.

6.2.1 Traffic Exposure Indicators

Two traffic exposure measures are investigated in this category, i.e. BKT, and VKT. Crash frequency is found non-linearly positively associated with all the investigated types of traffic exposure. These results are intuitive and consistent with several previous studies (Prato et al., 2016; Haman and Peek-Asa, 2013; Miranda-Moreno et al., 2011). The exponents of the exposure variables are less than one, which support the “safety in numbers” hypothesis (Jacobsen, 2003), and in line with the results from Prato et al. study (2016). The results also agree with the studies that concluded the presence of positive associations between congestion and lower crash rates (e.g. Reynolds et al., 2009). The BKT has a higher exponent value than VKT, which shows the higher impact of BKT, than VKT, on cyclist-motorist crashes.
6.2.2 Non-Graph Zonal Characteristics

The CMs’ results for the different categories of non-graph zonal characteristics are shown as follow.

6.2.2.1 Socio-Economic Models

Socio-economic CMs are primarily based on the explanatory variables extracted from census data. The models reveal positive associations between cyclist crashes and the population, employment, and household densities. The results are reasonable since the aforementioned variables can be considered surrogate measures for traffic exposure, thereby explaining their positive associations with cyclist crashes. The results are in agreement with previous studies by Siddiqui et al. (2012) and Prato et al. (2016).

6.2.2.2 Land Use Models

The models in this category incorporate explanatory variables that refer to land zonings within the TAZs. The results in Table 8 show that the increase in residential and recreational area densities is associated with a decline in the number of cyclist crashes. The result for the recreational areas is logical because these areas usually provide off-street and continuous paths for active transportation commuters reducing the conflict risk between these vulnerable commuters and vehicles. The negative association between residential area density and cyclist crashes can be justified by the ongoing traffic calming measures applied by city of Vancouver to promote active transportation and limit motorized traffic at the residential neighborhoods. This is done using speed humps, diverters, traffic circles, etc. to reduce the speed of the traffic and
control its navigation within residential areas. However, this finding needs further investigation since it is found non-significant in the FB spatial models, as shall be shown later.

On the other hand, the increase in commercial area density is found associated with an increase in cyclist crashes. This can be attributed to the side street activities that raise the potential risk of a cyclist going into conflict with motorized traffic. The association between commercial areas and cyclist safety agrees with two previous studies (Narayamoorthy et al., 2013) (Vandenbulcke et al., 2014).

6.2.2.3 Built Environment Models

Built environment variables refer to the elements that are physically present on the bike and road networks. The models show that cyclist crashes are positively associated with both transit stop and traffic signal densities, which agrees with previous studies by Strauss et al. (2013) and Wei and Lovegrove (2013). More traffic signals imply the presence of more wide intersections that usually include complex vehicle and bike maneuvers elevating the probability of crash occurrence. On the other hand, the presence of bus stops indicates the occurrence of interactions between buses, vehicles, and bikes, which is also speculated to increase cyclist crash risk.

An unexpected finding is the positive association between light pole density and cyclist crashes. This can be attributed to the higher bike volume (higher exposure) or higher speeds on the streets with better lighting (more light poles), especially at night time. Chen (2015) found a similar result in his macro-level model for cyclist safety, although it was non-significant at the 5% level.
6.2.2.4 Road Facility Models

For this category, higher proportion of arterial roads, as well as higher proportion of arterial plus collector roads, are found positively associated with cyclist crash frequency. This can be explained by the higher speeds and the heavier traffic on these types of roads, which would increase the risk of conflict occurrence between cyclists and vehicles. On the other hand, a decline in cyclist-motorist crash frequency is found associated with higher proportion of local roads. A likely reason for such negative association is the relatively low speeds on local roads and traffic calming measures, which would result in an increase in the drivers’ attentiveness, and, therefore, reduce conflicts’ potential. The former results agree with previous studies conducted by Chen (2015) and Siddiqui et al. (2012). The proportion of the off-street bike links is found negatively associated with cyclist crashes. This result is reasonable and agrees with several previous studies suggesting that separating bike traffic from motorized traffic would likely improve cyclist safety (Reynolds et al., 2009).

CMs are developed to combine the significant variables that are within each of the aforementioned categories. Afterwards, a model is developed to merge the variables from all the categories into one joint CM in order to yield better predictability. The procedure for selecting the variables for both the combined and joint models is a forward stepwise procedure. The order in which the variables are added is based on their P-value, from lowest to highest. Whether to add a variable to the model or remove it is decided based on the parameter’s statistical significance. The combined models showed good fit with almost all the explanatory variables being statistically significant. The mutual model combined various attributes from all the
different model categories, i.e. BKT, VKT, SigD, HhsD, OffSt-Prop, RecD, and is found to have the best fit among all the combined CMs, with all its variables significant as shown in Table 8.

6.2.2.5 Combined Models FB CMs with Spatial Effects

FB models are then developed for both the combined and the mutual CMs. The parameter estimates of the FB models are not exactly the same as the GLM models. The spatial effects are found substantial in all the FB CMs ($\psi >> 0.50$), as shown in Table 9. This highlights the importance of considering spatial effects in the macro-level CMs that investigate the impact of non-graph zonal characteristics on cyclist safety, which agrees with a few former studies (Siddiqui et al., 2012; Cheng, 2015; Prato et al., 2016). The variables’ associations in the FB models are consistent with the results from the GLM CMs; and all the variables are found significant, except ResD, which is affected by the incorporation of spatial effects causing it to be non-significant.
### Table 8 Negative Binomial GLM Analysis Estimates for Cyclist Non-Graph Zonal Characteristics

<table>
<thead>
<tr>
<th>Model Crashes =</th>
<th>K</th>
<th>df</th>
<th>X²</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposure Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.184BKT^{0.043}</td>
<td>2.27</td>
<td>131</td>
<td>125.16</td>
<td>142.57</td>
</tr>
<tr>
<td>0.239VKT^{0.484}</td>
<td>1.31</td>
<td>132</td>
<td>132.40</td>
<td>149.82</td>
</tr>
<tr>
<td><strong>Combined Model A</strong></td>
<td>0.040BKT^{0.59}VKT^{0.22}</td>
<td>2.43</td>
<td>130</td>
<td>118.71</td>
</tr>
<tr>
<td><strong>Socio-Economic Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0170BKT^{0.58}VKT^{0.32}exp(6.76x10^{-6}EmpD)</td>
<td>2.55</td>
<td>129</td>
<td>116.87</td>
<td>142.47</td>
</tr>
<tr>
<td>0.013 BKT^{0.33}exp(4.64x10^{-2}HhsD)</td>
<td>2.60</td>
<td>129</td>
<td>125.22</td>
<td>142.47</td>
</tr>
<tr>
<td>0.015 BKT^{0.59}VKT^{0.31}exp(2.44x10^{-5} PopD )</td>
<td>2.56</td>
<td>129</td>
<td>125.53</td>
<td>142.24</td>
</tr>
<tr>
<td><strong>Combined Model B</strong></td>
<td>0.004BKT^{0.58}VKT^{0.48}exp(8.09x10^{-6}EmpD+5.52x10^{-5}HhsD )</td>
<td>2.85</td>
<td>128</td>
<td>120.05</td>
</tr>
<tr>
<td><strong>Land Use Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.035BKT^{0.57}VKT^{0.28}exp(-0.55ResD*)</td>
<td>2.46</td>
<td>129</td>
<td>111.39</td>
<td>142.49</td>
</tr>
<tr>
<td>0.055BKT^{0.61}VKT^{0.17}exp(-2.35RecD)</td>
<td>2.91</td>
<td>129</td>
<td>131</td>
<td>144.20</td>
</tr>
<tr>
<td>0.016BKT^{0.57}VKT^{0.33}exp(1.83 CommD)</td>
<td>2.64</td>
<td>129</td>
<td>111.57</td>
<td>142.25</td>
</tr>
<tr>
<td><strong>Combined Model C</strong></td>
<td>0.051BKT^{0.41}VKT^{0.23}exp(-0.56ResD*-2.38RecD)</td>
<td>3.01</td>
<td>128</td>
<td>125.79</td>
</tr>
<tr>
<td><strong>Built Environment Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.010BKT^{0.39}VKT^{0.38}exp(0.015SigD)</td>
<td>3.03</td>
<td>129</td>
<td>120.79</td>
<td>142.44</td>
</tr>
<tr>
<td>0.012BKT^{0.58}VKT^{0.43}exp(0.010StopD)</td>
<td>2.77</td>
<td>129</td>
<td>115.04</td>
<td>141.93</td>
</tr>
<tr>
<td>0.012BKT^{0.31}VKT^{0.38}exp(0.00068PoleD)</td>
<td>2.72</td>
<td>129</td>
<td>117.13</td>
<td>141.93</td>
</tr>
<tr>
<td><strong>Combined Model D</strong></td>
<td>0.010BKT^{0.54}VKT^{0.28}exp(0.015SigD)</td>
<td>3.03</td>
<td>129</td>
<td>120.79</td>
</tr>
<tr>
<td><strong>Road Facility Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.016BKT^{0.57}VKT^{0.38}exp(0.80ArtColl_Prop)</td>
<td>2.63</td>
<td>129</td>
<td>117.93</td>
<td>142.72</td>
</tr>
<tr>
<td>0.037 BKT^{0.57}VKT^{0.33}exp(-0.74Loc_Prop)</td>
<td>2.56</td>
<td>129</td>
<td>117.31</td>
<td>142.61</td>
</tr>
<tr>
<td>0.06 BKT^{0.63}VKT^{0.13}exp(-978OffSt_Prop*)</td>
<td>2.48</td>
<td>129</td>
<td>127.95</td>
<td>143.19</td>
</tr>
<tr>
<td><strong>Combined Model E</strong></td>
<td>0.023 BKT^{0.64}VKT^{0.21}exp(1.17ArtColl_Prop-1884OffSt_Prop)</td>
<td>2.94</td>
<td>128</td>
<td>133.09</td>
</tr>
<tr>
<td><strong>Mutual Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01 BKT^{0.63}VKT^{0.33}exp(3.79x10^{-5}HhsD-1.15RecD+0.01SigD-1315.5OffSt_Prop)</td>
<td>3.75</td>
<td>126</td>
<td>140.86</td>
<td>143.93</td>
</tr>
</tbody>
</table>

*Significant at the 10% level.
All other parameters are significant at the 5% level or higher

K: Over-dispersion parameter  \( df \): degrees of freedom \( X^2 \): Pearson chi square \( SD \): Scaled Deviance
### Table 9 FB Analysis Estimates for Cyclist Non-Graph Zonal Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Combined Model A</th>
<th>Combined Model B</th>
<th>Combined Model C</th>
<th>Combined Model D</th>
<th>Combined Model E</th>
<th>Mutual Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.61</td>
<td>0.66</td>
<td>-4.51</td>
<td>0.77</td>
<td>-3.54</td>
<td>0.73</td>
</tr>
<tr>
<td>BKT</td>
<td>0.44</td>
<td>0.053</td>
<td>0.48</td>
<td>0.058</td>
<td>0.51</td>
<td>0.05</td>
</tr>
<tr>
<td>VKT</td>
<td>0.36</td>
<td>0.083</td>
<td>0.41</td>
<td>0.090</td>
<td>0.33</td>
<td>0.086</td>
</tr>
<tr>
<td>EmpD</td>
<td>-</td>
<td>-</td>
<td>8.14x10^-6</td>
<td>3.52x10^-6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HhsD</td>
<td>-</td>
<td>-</td>
<td>5.41x10^-5</td>
<td>2.03x10^-5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RecD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.99</td>
<td>0.49</td>
</tr>
<tr>
<td>ResD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.040*</td>
<td>0.38</td>
</tr>
<tr>
<td>SigD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ArtColl_Prop</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>OffSt_Prop</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0.993</td>
<td>0.001</td>
<td>0.915</td>
<td>0.10</td>
<td>0.991</td>
<td>0.001</td>
</tr>
<tr>
<td>DIC</td>
<td>746.84</td>
<td>746.07</td>
<td>742.49</td>
<td>739.75</td>
<td>744.01</td>
<td>737.22</td>
</tr>
</tbody>
</table>

- Not Applicable
* Non significant at the 10% level
All other parameters are significant at the 5% level or higher.
Chapter 7: Univariate Joint Crash Models, Multivariate Mixed Crash Models, and Statistical Discussion

This chapter comprises the univariate joint crash models and the multivariate crash models for pedestrian and cyclist safety in addition to a comprehensive statistical discussion.

7.1 Univariate Joint Crash Models

The review of the earlier cross-comparison studies (as discussed thoroughly in the upcoming sections) shows some contradictions regarding which modeling approach would yield the best performance for macro-level safety models. This requires rigorous research to support the conclusions of the previous studies. This section contributes to the literature by incorporating a comprehensive set of original covariates to advocate the best modeling technique for cyclist and pedestrian macro-level CMs. In this section, all the explanatory variables (traffic exposure, network configuration, and zonal characteristics) are attempted to be incorporated in a joint cyclist crash model and a joint pedestrian crash model. The best fit models are presented. GLM models are developed first, as demonstrated in section 4.1, and the procedure for selecting the variables is a forward stepwise procedure. The order in which the variables are added is based on their P-value, from lowest to highest. Whether to add a variable to the model or remove it is decided based on the parameter’s statistical significance. Three different FB crash modeling approaches (PLN, SPLN, and RPPLN) are then investigated, as demonstrated in section 4.2, for pedestrian and cyclist crashes to select the best FB modeling approach and to examine the superiority of the SPLN modeling technique.
Table 10 shows the GLM parameter estimates and 95% confidence intervals for the developed best-fit joint CMs.

### Table 10 Negative Binomial GLM Estimates for the Pedestrian and Cyclist Crash Joint Models

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cyclist Crash Model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ($\beta_0$)</td>
<td>-3.24</td>
<td>0.66</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>BKT ($\beta_1$)</td>
<td>0.57</td>
<td>0.05</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>VKT ($\beta_2$)</td>
<td>0.24</td>
<td>0.08</td>
<td>0.0017</td>
</tr>
<tr>
<td>SigD ($\beta_3$)</td>
<td>0.01</td>
<td>0.00</td>
<td>0.0007</td>
</tr>
<tr>
<td>Hhs ($\beta_4$)</td>
<td>0.20</td>
<td>0</td>
<td>0.0015</td>
</tr>
<tr>
<td>OffSt_Prop ($\beta_5$)</td>
<td>-1422.67</td>
<td>446.80</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>CEdge ($\beta_6$)</td>
<td>-2.65</td>
<td>1.16</td>
<td>0.0221</td>
</tr>
<tr>
<td>CSlope ($\beta_7$)</td>
<td>-0.16</td>
<td>0.06</td>
<td>0.0094</td>
</tr>
<tr>
<td>CommD ($\beta_8$)</td>
<td>1.39</td>
<td>0.73</td>
<td>0.0588</td>
</tr>
<tr>
<td>Over-Dispersion (K)</td>
<td>4.87</td>
<td>0.0392</td>
<td>-</td>
</tr>
<tr>
<td>AIC</td>
<td>816.93</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Pedestrian Crash Model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept ($\beta_0$)</td>
<td>-7.6696</td>
<td>0.7454</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>W ($\beta_1$)</td>
<td>0.6508</td>
<td>0.0679</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>VKT ($\beta_2$)</td>
<td>0.6592</td>
<td>0.0709</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>PLen ($\beta_3$)</td>
<td>-0.027</td>
<td>0.008</td>
<td>0.0004</td>
</tr>
<tr>
<td>SigDen ($\beta_4$)</td>
<td>0.0078</td>
<td>0.0026</td>
<td>0.0028</td>
</tr>
<tr>
<td>PSlope ($\beta_5$)</td>
<td>-0.0874</td>
<td>0.0247</td>
<td>0.0004</td>
</tr>
<tr>
<td>Over-Dispersion (K)</td>
<td>8.25</td>
<td>0.024</td>
<td>-</td>
</tr>
<tr>
<td>AIC</td>
<td>847.39</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Tables 11 and 12 show the parameter estimates and the 95% credible intervals for the three developed FB joint CMs as well as the models’ performance statistics. Most of the parameter estimates are found significant as the 95% credible intervals were bounded away from zero. The
FB CMs outperform the GLM CMs as the DICs of all the FB CMs are found lower than the AICs of the corresponding GLM CMs.

For the FB cyclist crash models, the SPLN model performs best (with the lowest DIC) indicating the importance of accounting for spatial correlation in the data when capturing the variability in cyclist crash frequency. Following the SPLN model, the RPPLN model has better performance than the PLN model. The PLN model shows the least performance among the three developed FB CMs, which is expected since the PLN model does not adequately account for unobserved heterogeneity.

For the pedestrian FB crash models, the SPLN model also performs the best, with the lowest DIC, showing that accounting for spatial correlation is essential when capturing the variability in pedestrian crash frequency. Following the SPLN model, the RPPLN model performance is not “considerably” better than the PLN model since the DIC values’ difference is less than two.

In general, few studies in the literature cross-compared the aforementioned modeling approaches on the macro-level. Xu and Huang (2015) utilized crash data for 738 TAZs in the county of Hillsborough, Florida to develop spatial effects and random parameters models. The analysis demonstrated that the conditional autoregressive (CAR) spatial-effects model outperformed the random-parameter negative binomial (RPNB) model for severe crashes. Moreover, they found out that the semi-parametric geographically weighted Poisson regression model, which was capable of accounting for the spatial correlation in crash data, performed best with the lowest mean absolute deviance and Akaike information criterion measures. Truong et al (2016) also
employed spatial-temporal effects model as well as random parameters model based on crash data for 63 provinces of Vietnam. The results showed that spatiotemporal model with conditional autoregressive prior model outperformed the random intercepts negative binomial (RINB) model and the RPNB model. They did not find a significant difference between the performance of the RINB models and the RPNB models. Amoh-Gyimah et al. (2016) presented different estimation methods to model pedestrian and cyclist crashes. They compared the results of a RPNB models with the results of a non-spatial negative binomial (NB) model and a Poisson-Gamma CAR model. They found out that the RPNB model performed best with the lowest mean absolute deviation, mean squared predicted error and Akaike information criterion measures.

The results of the research in hand are consistent with Xu and Huang (2015) and Truong et al. (2016) studies, which showed that the spatial-effects CMs outperformed the random-parameters and random-intercepts CMs on the macro-level. The results, however, are different than the conclusions of Amoh-Gyimah et al. (2016) study that showed that the random-parameters CM performed better than the CAR CM on the macro-level. Following the model comparison results of the research in hand, it is advocated to firstly address the spatial correlation when macro-level cyclist or pedestrian crash data are collected, and then account for the residual unobserved heterogeneity through other complementing alternatives.

Variations in the parameters’ estimates and significance can be noticed across the models investigated in this research, which confirm the different mechanisms by which the three models capture the unobserved heterogeneity. Nevertheless, the variables’ associations are consistent across all the investigated models. As for the SPLN model, the spatial effects are found highly
significant (i.e. $\psi >> 0.5$). This highlights the importance of accounting for spatial correlation in the macro-level CMs investigating cyclist and pedestrian safety. Note that the inclusion of spatial correlation caused one network indicator (i.e., average weighted slope for cyclist CM) to become non-significant although it was significant for the other models. This agrees with previous studies that reported that the consideration of spatial correlation can affect the estimation of parameters by making some variables non-significant (Karim et al., 2013). As for the RPPLN models, it can be noted that the posterior estimate deviation is significantly larger than 0 for the significant variables showing that varying the parameters by neighborhood groups is efficient in tracking some of the unobserved heterogeneity.
### Table 11 Parameter Estimates and 95% Credible Intervals for FB Cyclist Crash Joint Models

<table>
<thead>
<tr>
<th>Variables</th>
<th>PLN</th>
<th>RPPLN</th>
<th>SPLN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>95% credit intervals</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>upper</td>
<td></td>
</tr>
<tr>
<td>Intercept ($\beta_0$)</td>
<td>-2.655</td>
<td>0.816</td>
<td>-3.838 -0.653</td>
</tr>
<tr>
<td>BKT ($\beta_1$)</td>
<td>0.577</td>
<td>0.061</td>
<td>0.464 0.655</td>
</tr>
<tr>
<td>VKT ($\beta_2$)</td>
<td>0.168</td>
<td>0.080</td>
<td>0.138 0.706</td>
</tr>
<tr>
<td>SigD ($\beta_3$)</td>
<td>0.015</td>
<td>0.004</td>
<td>0.010 0.304</td>
</tr>
<tr>
<td>Hhs ($\beta_4$)</td>
<td>0.227</td>
<td>0.044</td>
<td>0.142 0.314</td>
</tr>
<tr>
<td>OffSt_Prop ($\beta_5$)</td>
<td>-1357.0</td>
<td>417.2</td>
<td>-2160.0 -524.2</td>
</tr>
<tr>
<td>CEdge ($\beta_6$)</td>
<td>-3.387</td>
<td>1.190</td>
<td>-5.901 -1.070</td>
</tr>
<tr>
<td>CSlope ($\beta_7$)</td>
<td>-0.166</td>
<td>0.070</td>
<td>-0.322 -0.031</td>
</tr>
<tr>
<td>CommD ($\beta_8$)</td>
<td>1.001</td>
<td>0.777</td>
<td>-0.576 2.490</td>
</tr>
<tr>
<td>$\sigma_0^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_3^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_4^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_5^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_6^2$</td>
<td>0.24</td>
<td>0.05</td>
<td>0.16 0.35</td>
</tr>
<tr>
<td>$\sigma_7^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DIC</td>
<td>744.94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 12 Parameter Estimates and 95% Credible Intervals for FB Pedestrian Crash Joint Models

<table>
<thead>
<tr>
<th>Variables</th>
<th>PLN</th>
<th>RPPLN</th>
<th>SPLN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>95% credit intervals</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>upper</td>
<td></td>
</tr>
<tr>
<td>Intercept ($\beta_0$)</td>
<td>-7.774</td>
<td>0.77</td>
<td>-9.38 -6.27</td>
</tr>
<tr>
<td>VKT ($\beta_1$)</td>
<td>0.668</td>
<td>0.075</td>
<td>0.51 0.81</td>
</tr>
<tr>
<td>W ($\beta_2$)</td>
<td>0.661</td>
<td>0.068</td>
<td>0.52 0.79</td>
</tr>
<tr>
<td>SigD ($\beta_3$)</td>
<td>0.008</td>
<td>0.002</td>
<td>0.003 0.013</td>
</tr>
<tr>
<td>PLen ($\beta_4$)</td>
<td>-0.027</td>
<td>0.008</td>
<td>-0.042 -0.01</td>
</tr>
<tr>
<td>PSlope ($\beta_5$)</td>
<td>-0.090</td>
<td>0.026</td>
<td>-0.142 -0.04</td>
</tr>
<tr>
<td>$\sigma_{0u}^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_{1u}^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_{2u}^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_{3u}^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_{4u}^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_{5u}^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_u^2$</td>
<td>0.13</td>
<td>0.027</td>
<td>0.086 0.193</td>
</tr>
<tr>
<td>$\sigma_e^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\psi$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DIC</td>
<td>787.39</td>
<td>789.31</td>
<td>783.89</td>
</tr>
</tbody>
</table>
7.2 Comparison of the Cyclist and Pedestrian Safety Correlates’ Effects

It is obvious from section 7.1, as well as chapters 5 and 6, that the pedestrian and cyclist safety correlates have similar effects on crash frequency of the two modes except for two variables (i.e., linearity and coverage). This consistency in the safety correlates’ effects, as well as the statistical significance of the parameters, gives further confidence in the analysis results and makes it intuitive to account for the possible interactions/correlations between pedestrian and cyclist crashes when addressing active transportation safety. This demonstrates the need for multivariate crash models to rigorously investigate active transportation safety, which takes us to the next section.

7.3 Multivariate Mixed Crash Models

Table 13 shows the parameter estimates and the 95% credible intervals for the multivariate mixed crash model that was developed using the methodology discussed in section 4.2.2. The developed multivariate mixed CM accounts for mode and spatial correlations and allows the covariates to vary from one mode to another providing the model with higher flexibility. The mode correlation due to spatial effects is highly significant indicating the importance of accounting for both the spatial and the mode correlations when investigating active transportation safety. A more comprehensive discussion is presented in the following section.
Table 13 FB Multivariate Joint Crash Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>95% Credible Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedestrian Crash Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept 1</td>
<td>-7.29</td>
<td>0.82</td>
<td>0.500 - 1.883</td>
</tr>
<tr>
<td>W</td>
<td>0.68</td>
<td>0.09</td>
<td>0.378 - 0.753</td>
</tr>
<tr>
<td>VKT</td>
<td>0.57</td>
<td>0.09</td>
<td>0.0026 - 0.0135</td>
</tr>
<tr>
<td>SigD</td>
<td>0.008</td>
<td>0.002</td>
<td>-0.0438 - 0.01174</td>
</tr>
<tr>
<td>PLen</td>
<td>-0.027</td>
<td>0.007</td>
<td>-0.0121 - 0.0111</td>
</tr>
</tbody>
</table>

Cyclist Crashes Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>95% Credible Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 2</td>
<td>-3.09</td>
<td>0.80</td>
<td>0.354 - 0.585</td>
</tr>
<tr>
<td>BKT</td>
<td>0.46</td>
<td>0.066</td>
<td>0.0632 - 0.454</td>
</tr>
<tr>
<td>VKT</td>
<td>0.26</td>
<td>0.10</td>
<td>0.0069 - 0.024</td>
</tr>
<tr>
<td>SigD</td>
<td>0.014</td>
<td>0.004</td>
<td>0.066 - 0.26</td>
</tr>
<tr>
<td>Hhs</td>
<td>0.17</td>
<td>0.048</td>
<td>-0.184 - 0.087</td>
</tr>
<tr>
<td>CSlope</td>
<td>-0.056</td>
<td>0.069</td>
<td>-5.218 - 0.123</td>
</tr>
<tr>
<td>CEdge</td>
<td>-2.65</td>
<td>1.29</td>
<td>-223.6 - 168.4</td>
</tr>
<tr>
<td>OffSt_Prop</td>
<td>-32.17</td>
<td>98.67</td>
<td>-8.217 - 2.515</td>
</tr>
<tr>
<td>CommD</td>
<td>1.004</td>
<td>0.84</td>
<td>-0.96 - 0.88</td>
</tr>
<tr>
<td>Corr_u</td>
<td>0.66</td>
<td>0.35</td>
<td>0.20 - 0.99</td>
</tr>
<tr>
<td>Corr_s</td>
<td>0.66</td>
<td>0.35</td>
<td>0.20 - 0.99</td>
</tr>
</tbody>
</table>

7.4 Statistical Discussion on the Univariate and Multivariate Crash Models

This section provides insights on various statistical issues that are related to the developed univariate and multivariate crash models. Two categories of crash models are developed; traffic exposure models that only contain the basic exposure variables (usually used for applications such as hot zones identification) and joint models that include all the available explanatory variables (for the purpose of prediction, capturing associations, and assessing marginal effects and safety changes). The traffic exposure model is initially developed to explicitly assess the contribution of the traffic exposure variables to the crash model and to monitor the change in the
performance of different crash model structures upon adding more explanatory variables. Tables 14 and 15 show the estimation results for both the exposure and joint crash models using different error and correlation structures, i.e. univariate and multivariate heterogeneity-only models as well as univariate and multivariate heterogeneity plus spatial effects models. Table 16 shows the variance-covariance matrices for the different models. The conclusions are discussed as follow:

7.4.1 Spatial Correlation Effects
For the univariate exposure and joint crash models, the spatial effects are significant (i.e. \( \psi > 0.50 \)). Incorporating the spatial effects improves the univariate models’ performance, resulting in significantly lower DIC. The same applies to the multivariate exposure and joint models, in which the models incorporating the spatial effects are found to have better performance than the models without the spatial effects. This agrees with several previous studies that showed that accounting for spatial effects would improve the performance of the multivariate and univariate crash models (e.g., Wang and Kockelman, 2013; Siddiqui et al., 2012).

7.4.2 Mode Correlation Effects
Accounting for the correlation among active modes’ crashes improves the performance (lower DIC) of the crash models. This applies to both the exposure and joint crash models. The heterogeneity-only multivariate model and the multivariate model with both heterogeneous and spatial effects outperform the corresponding univariate models. This result affirms the conclusions of the Huang et al. (2017) and Lee et al. (2015) studies that showed that accounting
for correlation among different crash modes significantly improved the crash model performance.

The correlation among modes for spatial effects (Corr_s) is found significant and higher than the correlation among modes for heterogeneous effects (Corr_u). Corr_u is found significant for the models incorporating heterogeneity effects only while non-significant for models incorporating both spatial and heterogeneous effects, which is plausible due to the stronger impact of the spatial effects compared with the heterogeneous effects (i.e., $\psi > 0.50$) as observed in the univariate models incorporating heterogeneous and spatial effects. This result (Corr_s being more significant than Corr_u) agrees with the macro level study of Lee et al. (2015) and is in contrast to Huang et al. (2017) micro-level study. It suggests that the main correlations among crash modes exist between adjacent zones rather than individual zones.

**7.4.3 Spatial Correlation Effects Compared to Mode Correlation Effects**

The results show that for traffic exposure crash models, mode correlation has a higher impact on model performance than spatial correlation (i.e., DIC of the third model is lower than the DIC of the second model). This agrees with a study by Jonathan et al. (2016) in which they found that the multivariate correlation played a stronger role than the spatial correlation when modeling the crash frequencies, in terms of different crash types, using traffic exposure variables.

However, in the case of the joint crash model, spatial correlation has a higher impact on the models’ performance than the mode correlation (i.e., DIC of the second model is lower than the DIC of the third model). This result agrees with Wang and Kockelman (2013) who found that for
a crash model incorporating traffic exposure and explanatory variables, a spatial univariate model could perform better than a multivariate “aspatial” model.

A plausible explanation is that, on the one hand, the impact of mode correlation on the model performance is high when there are few independent variables included in the crash model, since it would considerably contribute towards the reduction of the omitted variable bias, but this contribution becomes weaker after adding more relevant explanatory variables. On the other hand, the impact of spatial correlation is almost stable and high in both the exposure and joint crash models and not considerably affected by adding/omitting the explanatory variables.

### 7.4.4 Omitted Variables Effects

The DIC of the univariate exposure crash models is considerably higher than that of the univariate joint crash models and the parameter estimates are different especially for the pedestrian crashes’ model. This shows that missing relevant explanatory variables can degrade the model’s performance and highlights the conclusion of Mannering et al. (2016) that omitting important explanatory variables would bias the model results. However, that negative impact on model performance is much lower in the case of the multivariate crash models, as there is no such large difference between the DIC of the exposure and joint crash models. This implies that accounting for the dependency between responses can contribute to reducing the effects of omitted variable bias.

In general, the joint crash models’ performance is found better than that of the exposure crash models, suggesting the importance of the added covariates in tracking the unobserved
heterogeneity. This is also confirmed by the reduction in the value of correlation terms (for both the multivariate unstructured heterogeneity-only models and the multivariate spatial effects models) when shifting from exposure to joint crash models. This supports El-Basyouny and Sayed (2009) statement that AADT-only models may suffer from omitted variables bias since the unobserved heterogeneity from other factors, known to influence crash frequency, ends up in the correlation structure.

Lastly, it is worth pointing out that it is impossible to include all relevant variables in any model; therefore, omitted variable bias is unavoidable. I did our best to incorporate all the available and relevant macro-level data for the City of Vancouver, which is more comprehensive and unique than any previous study. Nevertheless, there are other covariates that believably can improve the models’ performance and reduce the omitted variable bias. Those factors can be, not limited to, weather condition, speed, crash location (midblock, at intersection, etc.), time (day or night), etc. These factors are recommended to be included in future research. Also, some unobserved heterogeneity approaches such as random-parameters and latent class approaches (Mannering and Bhat, 2014) can mitigate some of the adverse impacts of omitting significant explanatory variables.
Table 14 FB Traffic Exposure Crash Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate Models With Heterogeneity</th>
<th>Univariate Models With Heterogeneity and Spatial Effects</th>
<th>Multivariate Model With Heterogeneity</th>
<th>Multivariate Model With Heterogeneity and Spatial Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Cyclist Crash Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept 1</td>
<td>-2.91</td>
<td>0.58</td>
<td>-3.61</td>
<td>0.66</td>
</tr>
<tr>
<td>BKT</td>
<td>0.52</td>
<td>0.07</td>
<td>0.44</td>
<td>0.053</td>
</tr>
<tr>
<td>VKT</td>
<td>0.22</td>
<td>0.08</td>
<td>0.36</td>
<td>0.083</td>
</tr>
<tr>
<td><strong>Pedestrian Crash Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept 2</td>
<td>-7.00</td>
<td>1.15</td>
<td>-7.36</td>
<td>0.88</td>
</tr>
<tr>
<td>W</td>
<td>0.78</td>
<td>0.09</td>
<td>0.85</td>
<td>0.10</td>
</tr>
<tr>
<td>VKT</td>
<td>0.38</td>
<td>0.06</td>
<td>0.36</td>
<td>0.08</td>
</tr>
<tr>
<td>$\psi_s$</td>
<td>-</td>
<td>-</td>
<td>0.98</td>
<td>0.001</td>
</tr>
<tr>
<td>Corr_u</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Corr_s</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total DIC</strong></td>
<td>1563.96</td>
<td></td>
<td>1535.93</td>
<td></td>
</tr>
</tbody>
</table>

-Not Applicable

* Non-significant at the 10% level
** Significant at the 10% level
All other parameters are significant at the 5% level or higher

$\psi$: Spatial Variation Proportion

Corr_u: Mode correlation due to heterogeneity effects.
Corr_s: Mode correlation term due to spatial effects.
## Table 15 FB Joint Crash Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate Models With Heterogeneity</th>
<th>Univariate Models With Heterogeneity and Spatial Effects</th>
<th>Multivariate Model With Heterogeneity</th>
<th>Multivariate Model With Heterogeneity and Spatial Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Pedestrian Crash Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept 1</td>
<td>-7.77</td>
<td>0.77</td>
<td>-8.01</td>
<td>1.03</td>
</tr>
<tr>
<td>W</td>
<td>0.66</td>
<td>0.06</td>
<td>0.83</td>
<td>0.11</td>
</tr>
<tr>
<td>VKT</td>
<td>0.66</td>
<td>0.07</td>
<td>0.51</td>
<td>0.10</td>
</tr>
<tr>
<td>SigD</td>
<td>0.008</td>
<td>0.002</td>
<td>0.008</td>
<td>0.003</td>
</tr>
<tr>
<td>PLen</td>
<td>-0.027</td>
<td>0.008</td>
<td>-0.028</td>
<td>0.010</td>
</tr>
<tr>
<td>PSlope</td>
<td>-0.09</td>
<td>0.026</td>
<td>-0.068</td>
<td>0.034</td>
</tr>
<tr>
<td>Cyclist Crashes Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept 2</td>
<td>-3.51</td>
<td>0.63</td>
<td>-3.07</td>
<td>0.76</td>
</tr>
<tr>
<td>BKT</td>
<td>0.55</td>
<td>0.05</td>
<td>0.45</td>
<td>0.058</td>
</tr>
<tr>
<td>VKT</td>
<td>0.28</td>
<td>0.07</td>
<td>0.27</td>
<td>0.09</td>
</tr>
<tr>
<td>SigD</td>
<td>0.016</td>
<td>0.004</td>
<td>0.015</td>
<td>0.004</td>
</tr>
<tr>
<td>Hhs</td>
<td>0.22</td>
<td>0.05</td>
<td>0.177</td>
<td>0.046</td>
</tr>
<tr>
<td>CSlope</td>
<td>-0.15</td>
<td>0.06</td>
<td>-0.067**</td>
<td>0.070</td>
</tr>
<tr>
<td>CEdge</td>
<td>-3.11</td>
<td>1.20</td>
<td>-2.89</td>
<td>1.34</td>
</tr>
<tr>
<td>OffSt_Prop</td>
<td>-1149</td>
<td>428.10</td>
<td>-3.37*</td>
<td>31.7</td>
</tr>
<tr>
<td>CommD</td>
<td>1.02*</td>
<td>0.76</td>
<td>0.883*</td>
<td>0.76</td>
</tr>
<tr>
<td>v</td>
<td>-</td>
<td>-</td>
<td>0.96</td>
<td>0.05</td>
</tr>
<tr>
<td>Corr_u</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Corr_s</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total DIC</td>
<td>1532.97</td>
<td></td>
<td>1522.05</td>
<td></td>
</tr>
</tbody>
</table>

-Not Applicable  * Non-significant at the 10% level  **Significant at the 10% level  All other parameters are significant at the 5% level or higher
Table 16 Variance-Covariance Matrices

<table>
<thead>
<tr>
<th>Crash Mode</th>
<th>Pedestrian Crash</th>
<th>Cyclist Crash</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposure Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multivariate Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Crash</td>
<td>0.25 (0.04)</td>
<td>0.198 (0.05)</td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td>0.198 (0.05)</td>
<td>0.50 (0.08)</td>
</tr>
<tr>
<td><strong>Univariate Spatial Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U error term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pedestrian Crash</td>
<td>0.003 (0.0005)</td>
<td></td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td></td>
<td>0.003 (0.0004)</td>
</tr>
<tr>
<td>S error term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pedestrian Crash</td>
<td>0.27 (0.03)</td>
<td></td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td></td>
<td>0.40 (0.05)</td>
</tr>
<tr>
<td><strong>Multivariate Spatial Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U error term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Crash</td>
<td>0.014 (0.03)</td>
<td>-0.009 (0.018)</td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td>-0.009 (0.018)</td>
<td>0.016 (0.02)</td>
</tr>
<tr>
<td>S error term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Crash</td>
<td>0.63 (0.21)</td>
<td>0.74 (0.19)</td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td>0.74 (0.19)</td>
<td>1.63 (0.31)</td>
</tr>
<tr>
<td><strong>Joint Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multivariate Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Crash</td>
<td>0.15 (0.02)</td>
<td>0.05 (0.03)</td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td>0.05 (0.03)</td>
<td>0.27 (0.05)</td>
</tr>
<tr>
<td><strong>Univariate Spatial Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U error term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pedestrian Crash</td>
<td>0.003 (0.0005)</td>
<td></td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td></td>
<td>0.02 (0.02)</td>
</tr>
<tr>
<td>S error term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pedestrian Crash</td>
<td>0.25 (0.04)</td>
<td></td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td></td>
<td>0.30 (0.04)</td>
</tr>
<tr>
<td><strong>Multivariate Spatial Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U error term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Crash</td>
<td>0.03 (0.03)</td>
<td>-0.012 (0.02)</td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td>-0.012 (0.02)</td>
<td>0.11 (0.09)</td>
</tr>
<tr>
<td>S error term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Crash</td>
<td>0.41 (0.19)</td>
<td>0.31 (0.16)</td>
</tr>
<tr>
<td>Cyclist Crash</td>
<td>0.31 (0.16)</td>
<td>0.57 (0.45)</td>
</tr>
</tbody>
</table>
Chapter 8: Hot and Safe zones Identification

This chapter shows the rankings and locations of the identified hot and safe zones in the city of Vancouver after demonstrating the identification methodologies.

There are several ways to detect and rank the crash-prone zones. The HZID methodology depends mainly on the technique used for building the macro-level CM. This section discusses how to identify hot and safe zones for active transportation. The standard method for identifying crash hotspots is to rank them in a descending order of the performance metric considered for the safety analysis, and the threshold for the identified hotspots is typically determined according to the amount of funding available for safety improvements. Usually, the development threshold is limited to the top 1–10% of sites (Cheng and Washington. 2008; Lan and Persaud, 2011). In this research, the approaches used for identifying the hot and safe zones are the practice-ready EB PSI method and the state-of-the-art FB Mahalanobis distance method as discussed in the upcoming sections. The development threshold is determined to be 10%. The crash models used in the identification approaches are the basic traffic exposure models, without any explanatory variables incorporated. The explanatory variables are kept out of the model in order to be systematically studied afterwards as plausible factors for the zones’ safety/unsafety; which is discussed in details in the next chapter.

8.1 The EB PSI Method

In the EB PSI method, the ranking criterion typically applies the PSI measure, which is based on the difference between the expected and the predicted crash frequency at each site. The PSI is
calculated as shown in equation 20. The zones are then ranked to ensure that those in most need of treatment are given priority, and the zones with higher PSI are considered more crash-prone.

\[ PSI_i = \theta_i - \mu_i \]  \hspace{1cm} (20)

Where \( \theta_i \) is the EB expected safety estimate, and \( \mu_i \) is the predicted crash frequency from the developed CM.

The PSIs can be calculated for cyclist and pedestrian crashes individually. In order to get a hot zone ranking for active transportation, the PSIs from both modes need to be combined. The cyclist and pedestrian crashes in Vancouver, British Columbia are found to have approximately similar costs (Button, 2013; Urban Systems et al., 2012). This agrees with a study by Miller et al. (2004) that comprehensively compared the cyclist and pedestrian crash costs. Therefore, for a univariate method such as EB PSI, it is plausible to add the PSIs due to cyclist and pedestrian crashes to come up with an active transportation hot zone ranking, as shown in equation 21. The resulting measure \("PSI_{Active \ Transportation}"\) can be then used to rank the active transportation hot zones in a descending order, and the top 10% hot zones are investigated afterwards.

\[ PSI_{Active \ Transportation} = PSI_{Pedestrian} + PSI_{cyclist} \]  \hspace{1cm} (21)

### 8.2 The Mahalanobis Distance Method

A multivariate crash model, which takes into account the covariance of different crash modes, e.g., cyclist and pedestrian crashes, leads the probability density function to elliptic boundaries.
According to Sacchi et al. (2015), “the probability density is high for ellipses near the multivariate center ($\mu_i$) and low for ellipses further away. For instance, the Euclidean distance from the multivariate center of the distribution in Figure 11 to the expected crash frequency of site 1 ($\theta_1$) is shorter than the distance to site 2 ($\theta_2$). If $\mu_i$ is assumed to be the long-term average crash frequency for sites similar to site 1 and 2, then, by applying the PSI concept, we would assume that site 2 is more crash-prone than site 1. However, for this distribution, the variance for cyclist crashes (y-axis) is less than the variance for the pedestrian crashes (x-axis), so $\theta_2$ would be “closer” to $\mu_i$ than $\theta_1$ in the sense that it is more likely to observe a site with crash frequency equal to $\theta_2$ than one equal to $\theta_1$ (the contour with higher probability that contains $\theta_2$ is nested within the contour that contains $\theta_1$ with lower probability density).”

To address the distance between $\theta_i$ and the distribution of crash frequency at similar sites, in a multivariate context, a measure other than PSI is required. The measure, known as the

![Figure 11 Multivariate distribution of the Pedestrian and Cyclist Crashes](image-url)
Mahalanobis distance, provides a way to account for the variance of each variable and the covariance between variables. Mahalanobis distance accounts for the variance difference in each direction and for the covariance between variables. It reduces to Euclidean distance for uncorrelated variables with unit variance. The Mahalanobis distance was first used to account for outliers in the development of safety performance functions (El-Basyouny and Sayed, 2010), and was then proposed by Sacchi et al. (2015) to investigate accident-prone intersections using severity and property-damage-only crash data modeled jointly. Sacchi et al. (2015) investigated the Mahalanobis distance method within-method consistency, and its consistency against other methods such as the EB PSI method and the posterior mean (PM) of collision frequency method. Their work demonstrated that the Mahalanobis distance outperformed the other methods and was capable of extending the univariate concept of PSI in a multivariate setting efficiently. The Mahalanobis distance, for multivariate normal models, is given by equation 22.

$$d^2 = (\lambda_i^* - \mu_i^*) \Sigma^{-1} (\lambda_i^* - \mu_i^*)$$  \hspace{1cm} (22)

Where $\lambda_i^* = (\ln(\lambda_i^1), \ln(\lambda_i^2), \cdots, \ln(\lambda_i^k))$ and $\mu_i^* = (\ln(\mu_i^1), \ln(\mu_i^2), \cdots, \ln(\mu_i^k))$

$\ln(\mu_i) = \ln(\lambda_i) - u_i \cdot s_i$ (the notations are the same as discussed in equation 16), $k$ is the number of modes, and $\Sigma$ is the covariance matrix. Notice that if $\Sigma$ is the identity matrix, then the Mahalanobis distance reduces to the standard Euclidean distance between $\theta_i^*$ and $\mu_i^*$.

TAZs are ranked in the descending order of their Mahalanobis distances and the top 10% hot zones are investigated. It is worth mentioning that only the zones with both cyclist and pedestrian crashes exceeding the corresponding coordinates of $\mu_i$ are considered active transportation hot
zones (i.e. zones in quadrant 1). In fact, as shown in Figure 11, the cyclist-pedestrian plane can be divided into four quadrants denoted I, II, III and IV. The zones in quadrant I represent the zones with both cyclist and pedestrian crashes exceeding both coordinates of $\mu_i$. The zones in quadrant II represent the zones with only cyclist crashes exceeding the cyclist coordinate of $\mu_i$ while the pedestrian crashes do not. The zones in quadrant III represent the zones in which both cyclist and pedestrian crashes do not exceed the coordinates of $\mu_i$ and considered the safest (safe) zones for active transportation. The zones in quadrant IV represent the zones with only pedestrian crashes exceeding the pedestrian coordinate of $\mu_i$ while the cyclist crashes do not. It is worth mentioning that the Mahalanobis distance allows the detection of the safe-prone zones as well as the crash-prone zones.

### 8.3 Hot Zones Identification Results for PSI and Mahalanobis Distance Methods

Figure 12 illustrates the ranking lists for the top 10 % hot zones due to both the PSI and Mahalanobis distance methods. The figure shows that although there are some common zones among the two lists, a considerable difference in the ranking results can still be noticed. This agrees with several previous studies (e.g. Aguero-Valverde and Jovanis, 2009) that showed that a multivariate approach can yield substantially different HZID and ranking results from a univariate approach.

Figure 12 can be used by road agencies, such as the City of Vancouver, to further investigate the factors contributing to active transportation crashes in the hot zones. According to the Mahalanobis distance method, the hot zones are clustered in three main neighborhoods, i.e., Downtown, Strathcona, and Mount Pleasant, along with zone 122. Seven hot zones (Six in the
case of EB PSI method), out of the thirteen identified, are within the downtown area (surrounded by a circle in the figure), which mean that the downtown area requires rigorous countermeasures. Overall, all the identified hot zones would require policy recommendations to improve active transportations safety there.
Figure 12 Vancouver’s Active Transportation Hot Zones a) Mahalanobis Distance Method b) EB PSI Method
8.4 Consistency Tests

Two tests are applied to evaluate the consistency of the HZID methods: a within-method consistency test and a between-methods consistency test. The within method-consistency test involves two diagnostic criteria; sensitivity and specificity, as established in previous studies (e.g. Elvik, 2007; Huang et al., 2009). A critical concern of this test is to identify the “truly” positive (hazardous) and negative (safe) locations, which requires that the truth to be known as a priori. To satisfy this requirement, and as done in some previous studies (e.g. Cheng and Washington, 2008; Washington and Cheng, 2005), the mean of the 5 years crash data is used as the true Poisson mean (TPM) of the TAZ. As such, TAZs with higher TPMs (highest 10%) are assumed to be the truly hazardous TAZs in this simulation. The sensitivity and specificity can be then calculated according to equations 23 and 24.

\[
Sensitivity = \frac{Number \ of \ Correct \ Positives \ Identified}{Total \ Number \ of \ True \ Positives} \quad (23)
\]

\[
Specificity = \frac{Number \ of \ Correct \ Negatives \ Identified}{Total \ Number \ of \ True \ Negatives} \quad (24)
\]

The between-methods consistency test \((Tc)\), proposed in the literature (Cheng and Washington, 2008), involves comparing and evaluating two ranked lists of the same dataset from different HZID methods using the evaluation criteria shown in equation 25. The test is designed to assess the homogeneity of the identified zones among the different HZID methods applied. The greater the number of hot zones that are similarly identified in both methods, the more consistent the HZID methods with each other.
\[ Tc = \{HZ_1, HZ_2, \ldots, HZ_{i\delta}\}_{n,j1} \cap \{HZ_1, HZ_2, \ldots, HZ_{i\delta}\}_{n,j2} \]  

where \( HZ_n \) is the \( n^{th} \) ranked site identified as a hot zone, \( i \) is the total number of sites in the dataset, \( \delta \) is the threshold of high-risk sites identified (e.g., \( \delta=0.1 \) corresponds to the top 10% of zones being identified as hot zones), and \( j1 \) and \( j2 \) represent the ranking method(s) being compared.

The within-methods consistency test shows that the Mahalanobis distance method outperforms the EB PSI method, which agrees with the conclusion of Sacchi et al. (2015) study. The sensitivity for the Mahalanobis distance method is 53.84% compared to 45.15% for the PSI method. The specificity for the Mahalanobis distance method is 95.04% compared to 94.21% for the PSI method. The ranked list obtained using the PSI method is then matched to the list from the Mahalanobis distance method in order to estimate the between-methods consistency at different \( \delta \) values (2.5%, 5%, 7.5%, and 10%). Figure 13 shows the results of the between-methods consistency, in which substantial inconsistency between FB Mahalanobis distance method and EB PSI method is noticed at the different percentages of hot zones identified. One approach to increase the consistency between the two methods is to standardize the EB PSIs. This can be conducted using equation 26.

\[
PSI_{Active\ Transportation} = \frac{PSI_{Pedestrian} - \text{Average (PSI}_{Pedestrian})}{\text{StdDev(PSI}_{Pedestrian})} + \frac{PSI_{Cyclist} - \text{Average (PSI}_{Cyclist})}{\text{StdDev(PSI}_{Cyclist})} 
\]  

Standardizing the EB PSI values led to more consistent identification results with the Mahalanobis distance method as shown in Figure 13. The consistency in the hot zones
identification results reached as high as 62%. Nevertheless, inconsistency still exists between both approaches especially for low percentages of hot zones identified. In fact, the typical threshold of 10% of sites used by highway agencies seems to provide different ranked lists depending on whether the univariate or the multivariate approach is used, and whether the empirical Bayes or full Bayes technique is used. This agrees with several previous studies, e.g., (Cheng et al., 2017; Sacchi, et al., 2015; Aguero-Valverde and Jovanis, 2009).

![Figure 13 Results of the Consistency with FB Mahalanobis Distance Method](image)

8.5 Active Transportation Safe Zones

The Mahalanobis distance method is then applied to identify the safe zones for active transportation in the City of Vancouver. Investigating those zones would help in identifying the elements that make them safe for active commuting. Zone one is the safest for active transportation, which was expected due to the presence of extensive bike and pedestrian facilities in this zone. The rest of the zones are ranked and illustrated in Figure 14.
Figure 14 Safe Zones in the City of Vancouver
Chapter 9: Policy Recommendations

This chapter encompasses two sections; the first section identifies the hot and safe zones’ trigger variables, while the second section discusses the suggested remedies and policy recommendations.

9.1 Hot and Safe Zones’ Trigger variables

The values of the safety correlates at each hot zone (HZ), as well as at each safe zone (SZ), are compared to their corresponding regional mean (listed in Table 1) to identify the trigger variables. A regional mean is the average of the specific variable value for all zones in the study area. The City of Vancouver regional statistics (mean and standard deviation) for the variables used in this study are shown in Table 1 (pages 59 and 60). This method was first applied by Lovegrove and Sayed (2007) to detect trigger variables for vehicle crashes. The variables’ values that are found to be significantly different than the regional statistics are identified as trigger variables. One sample student’s t-test with $n - 1$ degrees of freedom is used and if the t-statistic (shown in equation 27) is found significant at 95% confidence level or higher, then the variable is considered a trigger variable at a HZ or a SZ.

$$t = \frac{(\bar{x} - x)}{(s / \sqrt{n})}$$  \hspace{1cm} (27)

Where: $t$ is the t-statistic, $\bar{x}$ is the sample mean, $s$ is the sample standard deviation, $x$ is the specified value, and $n$ is the sample size.
Table 17 shows the trigger variables at the different HZs. The signs beside each variable indicate if the variable has a negative or positive impact on active commuters’ crashes respectively based on the analyses that were discussed in chapters 5 and 6. Many of the explanatory variables were triggering in most of the HZs, which affirms the importance of the explanatory variables extracted for this research. The trigger variables are categorized into three sets: land use, traffic demand, and traffic supply. It is noted that the downtown cluster incorporates the largest portion of trigger variables, indicating that considerable modifications and budget need to be directed towards this vital area. The trigger variables that are common within most of the downtown cluster zones are more signal density, transit stop density, arterial-collector roads, major intersections, household density, commercial area density, employment density, active network connectivity, as well as less local roads, residential area density, recreational area density, pedestrian network continuity, actuated signals, and active network infrastructure. As for the Mount Pleasant Cluster, the common trigger variables are more signal density, major intersections, household density, commercial area density, active network connectivity, as well as less sidewalk continuity, actuated signals, and residential area density. For the Strathcona cluster, the common trigger variables are less active network length, cyclist network continuity, and actuated signals. Lastly, for zone 122, the trigger variables are more major intersections, and less separated bike infrastructure.

On the other hand, Figure 15 shows the various trigger variables for the SZs. The figure shows the frequency of the safe zones that are affected by each trigger variable to highlight the common attributes among the SZs. The signs beside each variable indicate if the variable has a negative or positive impact on active transportation safety based on the analyses that were discussed in
chapters 5 and 6. Less transit density as well as less signal density are the most recurrent trigger variables as they are triggering in all the SZs. The commercial area density and employment density come afterwards, as they are found trigger variables in twelve zones. The rest of the trigger variables are ranked in a descending order according to Figure 15.

The recognition of the zonal safety issues is a crucial step to realize prospective suitable remedies. The trigger variables can be used together along with field studies and literature consultation to understand the overall safety issues within each hot zone. The SZ and HZ trigger variables are discussed in detail in the upcoming section along with a discussion of the plausible remedies and policy recommendations.
<table>
<thead>
<tr>
<th>Category</th>
<th>Trigger Variable</th>
<th>Downtown Cluster Zones</th>
<th>Mount Pleasant Cluster Zones</th>
<th>Strathcona Cluster Zones</th>
<th>Zone 122</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10 16 17 19 20 30 31 51 81 94 79 28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Land Use</strong></td>
<td>Residential Density (-)</td>
<td>- ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ - - ✓ ✓ - ✓ ✓ - ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Commercial Density (+)</td>
<td>- ✓ ✓ ✓ ✓ ✓ - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recreational Density (-)</td>
<td>- ✓ ✓ ✓ ✓ ✓ ✓ - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ - - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td><strong>Traffic Demand</strong></td>
<td>Household Density (+)</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Employment Density (+)</td>
<td>- ✓ ✓ ✓ ✓ ✓ ✓ - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td><strong>Traffic Supply</strong></td>
<td>Arterial-Collector</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roads (+)</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ - - - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Local Roads (-)</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Length (-)</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pedestrian</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connectivity (-)</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cyclist Continuity (-)</td>
<td>- - - - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paths Prop. (-)</td>
<td>- - - - ✓ ✓ - - ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Signal Density (+)</td>
<td>- ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transit Density (+)</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Major Intersection Prop. (+)</td>
<td>- ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connectivity (+)</td>
<td>- ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actuated</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Signals Prop. (-)</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td>- ✓ - ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
</tbody>
</table>
9.2 Discussion and Remedies

This subsection discusses the potential remedies that can be applied to the identified hot zones. Policy recommendations are suggested based on the analysis of the hot and safe zones’ trigger variables as well as field studies and literature consultation. The policy recommendations are
categorized into three main classes, i.e., land use management, traffic demand management, and traffic supply management as follow.

### 9.2.1 Land Use Management

Land use, including access to destinations and daily life activities’ support, has been recognized for being a significant factor in the traffic safety literature review. Table 17 shows that commercial area density is identified as a positive trigger variable in nine HZs, while nine HZs are characterized by having lower recreational or residential area density than the regional mean. These results agree with our aforementioned analysis in chapter 6 and the previous studies that demonstrated the negative impact of high commercial area density on active transportation safety, likely due to side street activities. Similarly, Figure 15 shows that twelve SZs have a lower commercial area density than average; while residential area density is identified as a trigger variable in seven SZs, and recreational density is identified as a trigger variable in six SZs. This shows the plausible importance of land use mix.

Smart growth (Compact Cities) policies can promote safer active transportation by integrating land use and transportation planning as follows (Alliance for biking and walking, 2016):

- Locate schools and well distributed small parks within walking/cycling distance of a large number of homes.

- Apply traffic calming measures at the residential/commercial areas such as traffic circles, median diverters, and curb bulges. The city of Vancouver applied many of these measures in the residential areas, which explain why the residential areas is a trigger

- Encourage employment related businesses and entertainment centers to be located within a 15-minute commute by public transportation, cycling, or walking.
- Apply traffic safety countermeasures at the commercial (and residential) areas such as boulevards designs, which separate large streets into parallel urban realms, buffering the commercial/residential street edge from the high-speed throughway using frontage roads and multi-way operations (NACTO, 2013) Also, textured or pervious pavements that are flushed with the curb can be utilized to reinforce the active transport priority operation of the street as well as delineating a narrow carriageway and a non-linear path for travel.

Figure 16 Boulevards for Residential and Commercial Areas (Image Courtesy of NACTO)

Table 18 shows a summary of the land use safety effects in the City of Vancouver’s SZs and HZs, as well as the plausible countermeasures. Those suggested land use management policies
need to be considered in the Downtown, Mount Pleasant, and Strathcona HZs’ clusters as revealed by Table 17.

Table 18 Summary of Land Use Safety Effects and Policy Recommendations

<table>
<thead>
<tr>
<th>Land Use Indicator</th>
<th>Summary of the Safety Effects in HZs and Szs</th>
<th>Policy Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial Area Density</td>
<td>• A positive trigger variable in 70% of the HZs&lt;br&gt;• 92% of the Szs have lower commercial area density than regional mean</td>
<td>• Apply Boulevard designs&lt;br&gt;• Apply textured or pervious pavements</td>
</tr>
<tr>
<td>Residential Area Density</td>
<td>• 70% of the HZs have lower residential area density than regional mean&lt;br&gt;• Residential area density is a trigger variable in 53% of the Szs</td>
<td>• Apply traffic calming measures such as traffic circles, median diverters, and curb bulges, etc.&lt;br&gt;• Foster neighborhood identity to enhance AT environment and sense of community&lt;br&gt;• Locate well distributed small parks within walking/cycling distance of a large number of homes&lt;br&gt;• Encourage community centers within 15-minute commute by public transportation, cycling, or walking</td>
</tr>
<tr>
<td>Recreational Area Density</td>
<td>• 70% of the HZs have lower recreational area density than regional mean&lt;br&gt;• Recreational density is a trigger variable in 46% of the Szs</td>
<td></td>
</tr>
</tbody>
</table>

9.2.2 Traffic Demand Management

Traffic demand usually increases due to the growing number of residents, employees, and tourists, as well as due to the new developments within the cities. Employment and households are regarded as the main elements of traffic demand in this research. High employment and household densities are identified as positive trigger variables for seven HZs. This agrees with the literature review, in which it has been recognized that the increase in employment and household densities is usually associated with an increase in active commuters’ crashes, especially if there is no proper bike/pedestrian infrastructure present. Figure 17 illustrates an example of the high household and employment densities at the Downtown HZ cluster.
Twelve SZs have lower employment density than the regional mean. Also, lower household density than the regional mean is observed in eleven SZs. Some recommendations regarding demand management for a safer active commuting experience can be as follow:

- Share the road campaigns along with trainings for employees and residents.
- Bike to work/school days, walk to work/school days, and other population health-sponsored programs.
- Implementation of better bike/pedestrian infrastructure.
- Supporting bike-share programs.
- Reduced use of single-occupant vehicle through carpool programs.
- Employer-sponsored transportation demand management programs for employees, such as telework, flex-time work schedules, van-pooling, or incentives for employees to use transit and active transportation such as employer-subsidized transit plans.
Table 19 shows a summary of the traffic demand effects in the SZs and HZs, as well as the plausible countermeasures. The mentioned traffic-demand management policies are expected to support the safety in numbers effect (Jacobsen, 2003) and mitigate the negative effects of the high employment and household densities as well as the plausible increase of motorized traffic exposure. Those demand management strategies are mostly required in the Downtown and Strathcona clusters as revealed by Table 17.

Table 19 Summary of Traffic Demand Safety Effects and Policy Recommendations

<table>
<thead>
<tr>
<th>Traffic Demand Indicator</th>
<th>Summary of the Safety Effects in HZs and SZs</th>
<th>Policy Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household Density</td>
<td>A positive trigger variable in 54% of the HZs</td>
<td>Educational programs and campaigns for the residents such as share the road campaigns, bike or walk to work/school days, and other population health-sponsored programs</td>
</tr>
<tr>
<td></td>
<td>85% of the SZs have lower household density than regional mean</td>
<td>Support carpool and bike-share programs and make it accessible</td>
</tr>
<tr>
<td>Employment Density</td>
<td>A positive trigger variable in 46% of the HZs</td>
<td>Employer-sponsored programs such as telework, flex-time work schedules, van-pooling, or incentives to use transit and active transportation</td>
</tr>
<tr>
<td></td>
<td>92% of the SZs have lower employment density than regional mean</td>
<td>Bike to work days, walk to work days, and other employment health-sponsored programs</td>
</tr>
</tbody>
</table>

9.2.3 Traffic Supply Management

This subsection discusses policy recommendations for traffic supply management. It addresses three main components, i.e., road network, transit network, and active transportation network.

9.2.3.1 Road Network

The road network’s configuration and conditions can either enhance or diminish the walkability and bikeability of a community as well as its traffic safety. Hierarchical street patterns (freeway-
arterial-collector-local) contribute to crashes since they divert traffic to high-speed roads that have large intersections and do not usually consider other modes of transportation (U.S department of housing and urban development, 2016).

In this research, the high proportion of arterial and collector roads is identified as a positive trigger variable in nine HZs. This agrees with the literature review that demonstrated the negative impact of major roads on active transportation safety in contrast to local roads. Figure 18 shows an example of the high arterial-collector roads’ density at the Downtown HZs cluster.

![Figure 18 High Density of Arterial-Collector Roads at Downtown HZs](image)

On the other hand, ten SZs are characterized by their low proportion of arterial and collector roads, while nine SZs are of higher local roads’ proportion than the regional mean. Local roads are identified as a trigger variable in the SZs plausibly due to the traffic calming measures.
implemented by the City of Vancouver, as exemplified in Figure 19. Overall, treatments towards sustainable street network can include retrofitting the major roads (e.g., arterials and collectors) by limiting traffic speeds, creating narrower cross sections, providing turn lanes, and providing well-designed pedestrian crossings and bike lanes (NACTO, 2015). Moreover, landscaping can be utilized for making the big roadways look “visually” narrower to calm traffic (Harkey and Zegeer, 2004). A higher-level solution can be retrofitting the street network itself to switch it from a grid structure, which was found in the traffic safety literature to be more crash-prone, to a more organic structure (Zhang et al., 2015).

Figure 19 Traffic Calming Measures at Local Streets and Residential Areas in Some SZs (Map data: Google, Google Maps)
Besides street network hierarchy, a vital element of the road network is the street intersections. The intersections’ design and operation should permit the pedestrians and cyclists to cross major streets at minimal conflicts with the motorized traffic. In particular, signalized intersections can provide unique opportunities as well as challenges for livable communities and complete streets. Traffic signals provide control of motorized and non-motorized traffic with extensive benefits. Where signalized intersections are closely spaced, signals can be used to control vehicle speeds by providing appropriate signal progression on a corridor (U.S department of housing and urban development, 2016), which can be utilized either for motorized or non-motorized traffic benefit. Traffic signals may impose challenges on non-motorized users when there are significant vehicle volumes as those may severely conflict with the existing pedestrian and cyclist movements.

Eight HZs suffer from higher signal density as well as higher proportion of major intersections than the regional means. This is plausibly due to the complex interactions between the different road users within the intersection especially if proper measures, such as left-turn lanes, leading pedestrian/cyclist intervals, bike boxes, and crosswalks, are absent at the intersections. Eleven HZs are characterized by the low proportion of actuated signals, which affirms that the lack of suitable measures can negatively affect cyclist and pedestrian safety.

Conversely, all the SZs are found to have lower traffic signal density than the regional mean. Also, ten SZs are characterized by having a high proportion of actuated traffic signals than the regional mean. In order to improve the safety of active commuters at the major and signalized intersections, some recommendations can be put forward as follow:

- Install cyclist/pedestrian actuated signals (Urban Systems and Region of Peel, 2016).
➢ Install cyclist/pedestrian leading intervals (Osama et al., 2015).

➢ Adjust signal timing to consider the safety of all users as well as their convenience so as not to hinder active commuters with overly long waits or insufficient crossing time.

➢ Make sure that signalized intersections’ design and operation are consistent with the MUTCD recommendations.

➢ Use new technologies for monitoring and managing active commuters’ movements, e.g., automated traffic control for pedestrian safety (Harkey and Zegeer, 2004).

➢ Use road delineation and painting to add capacity and address safety issues. Bike boxes, crosswalks, and left turn lanes are some elements to be considered.

➢ In some cases, roundabouts and non-conventional intersections can offer safer and more convenient intersection treatment than traffic signals, so such alternatives can be considered but with caution.

Table 20 shows a summary of the street network safety effects in the SZs and HZs, as well as the plausible countermeasures.
### Table 20 Summary of Road Network Safety Effects and Policy Recommendations

<table>
<thead>
<tr>
<th>Road Network Indicator</th>
<th>Summary of the Safety Effects in HZs and SZs</th>
<th>Policy Recommendations</th>
</tr>
</thead>
</table>
| **Street Network Hierarchy** | - Arterial-collector roads proportion is a trigger variable in 70% of the HZs, and lower than regional mean in 70% of the SZs  
- 70% of the SZs have higher local roads proportion than the regional mean | - Enhance the major roads’ safety by limiting traffic speeds, creating narrower cross sections, providing turn lanes, and providing well-designed pedestrian crossings and bike lanes  
- Use landscaping for making the big roadways look “visually” narrower to calm traffic  
- Retrofit the street network structure to more organic structures |
| **Street Intersections** | - 62% of the HZs have higher proportion of major intersection and traffic signal density than regional mean  
- 85% of HZs have lower proportion of actuated signals than regional mean  
- 100% of SZs have lower traffic signal density than regional mean, while 77% of SZs have higher proportion of actuated signals than regional mean | - Install cyclist/pedestrian actuated signals  
- Install cyclist/pedestrian leading intervals  
- Use new technologies for monitoring and managing active commuters’ maneuvers  
- Use road delineation and painting to add capacity and address safety issues  
- Check non-conventional intersections design |

### 9.2.3.2 Transit Network

Transit is an essential and integral element of any active transportation development plan. Active commuters’ safety around transit network elements (e.g., bus stops) needs to be rigorously investigated. Transit stop density is identified as a positive trigger variable in eight HZs, while found lower than the regional mean for all the SZs. These effects agree with several studies in the literature review as well as our aforementioned analysis results.

It is noticed that, in the City of Vancouver, buses and bikes usually share the same lane, which can cause stress to the cyclists during bus maneuvers at the stops, as illustrated in Figure 20. Moreover, pedestrians and cyclists sometimes jaywalk/jaycycle to catch a bus, especially if there
are no midblock crossings near the transit stop, which would increase the crash risk with conflicting traffic.

Figure 20 Bus-Bike Interactions when entering/exiting bus stops (Map data: Google, Google Maps)

Two recommendations for improving cyclist safety on the transit network are illustrated in Figure 21, in which bike lanes are separated from bus lanes, or painted at the locations of expected severe interaction with traffic such as parking and transit stops. Midblock crossings near transit stops can be also important for both pedestrian and cyclist safety. More comprehensive countermeasures for active transportation safety near transit are also proposed by transit street design guides (e.g., NACTO, 2016). Providing transit boarding islands for reducing interactions between transit and cyclist/pedestrians at transit stops is one of those recommendations. Also, transit corridors can be built to foster an active commuting scale for walking and cycling to complement public transit (NACTO, 2016).
Table 21 shows a summary of the transit network safety effects in the SZs and HZs, as well as the plausible countermeasures.

Table 21 Summary of Transit Network Safety Effects and Policy Recommendations

<table>
<thead>
<tr>
<th>Transit Network Indicator</th>
<th>Summary of the Safety Effects in HZs and SZs</th>
<th>Policy Recommendations</th>
</tr>
</thead>
</table>
| Transit Stops             | • 62% of the HZs have higher transit stop density as a positive trigger variable, while all the SZs have lower transit density than regional mean | • Separate/paint the bike lanes at transit stops locations  
• Install midblock crossings near transit stops  
• Consider building transit boarding islands and transit corridors |
9.2.3.3 Active Transportation Network

Active transportation network is the third element of traffic supply and is a key element due to the focus of this research.

The low level of bike and pedestrian facilities is identified as a trigger variable in nine HZs, mainly in downtown and Strathcona clusters. More specifically, the low level of separated bike facilities is a trigger variable in five HZs. Figures 22 and 23 illustrate examples of active transportation facilities’ absence on some HZs’ roads.

Figure 22 Absence of Bike Facilities on Some Main Roads in the HZs Such as Howe St., Thurlow St., W. Broadway, and W.12th (Map data: Google, Google Maps)
In contrast, it is observed that six SZs have more bike and pedestrian facilities than the regional mean. More specifically, five SZs are found to have a higher share of separated bike facilities than the regional mean. Shared spaces are also observed at some locations in the SZs as shown in Figure 24.
In addition to the specialized infrastructure for cyclists and pedestrians, the directness, connectivity, and accessibility of active transportation networks are major attributes affecting active transportation safety (Miranda-Moreno et al., 2011, Vandenbulcke et al., 2014). Eight HZs are found to have higher pedestrian and bike network connectivity than the regional mean. In addition, ten HZs are found to have lower pedestrian network continuity than the regional mean, while seven HZ are of lower bike network continuity than the regional mean.

AT network connectivity is considered an issue in many HZs plausibly due to the high network complexity as well as the high intersection density without having appropriate separation from the motorized traffic or other suitable treatments. It is to be noted that the minor intersections (e.g., with alleyways) in the HZs are not well treated for active commuters (e.g., no crosswalks, bike paths or proper signage). Also, some minor-major intersections do not contain any active commuter facilities (e.g., no actuated signals, crosswalks, bike paths, or proper signage). This can lead to jay-movements and severe interactions with motorized traffic at those intersections. Figure 25 illustrates some examples of these issues at the alleyways and minor-major intersections.
In contrast, eight SZs are found to have lower AT network connectivity than the regional mean along with a low proportion of intersection density and major intersections. It can be noticed that the existing minor-major intersections are well treated in many of the SZs due to measures such as bike paths, refuge medians, crosswalks, proper signage, and actuated signals, as shown in Figure 26. These can be considered as mitigative measures when high intersection density is inevitable.
As for AT network continuity, the pedestrian network continuity is found higher than the regional mean in six SZs, while the bike network continuity is found higher than the regional mean in four SZs. On the other hand, the pedestrian network continuity is found lower than the
regional mean in ten HZs, and bike network continuity is found less than the regional mean in seven HZs. Figure 27 shows examples of continuous bike and pedestrian links in the Szs.

![Figure 27 Continuous Paths for Pedestrians and Cyclists (Map data: Google, Google Maps)](image)

The safety of active transportation network can be fostered in the HZs by providing continuous active transportation network along with frequent midblock crossings and moderate intersection density as shown in Figure 28.

![Figure 28 Midblock Crossings for Cyclist and Pedestrian paths (Map data: Google, Google Maps)](image)

It is considered a necessity to couple the AT network continuity with midblock crossings, otherwise continuity would be rather a disadvantage than an advantage, as jay-movements may
occur more frequently due to the long distance between intersections, as exemplified in Figure 29.

Figure 29 Jaywalking due to the Absence of Midblock Facilities and Long Distance between Intersections
(Map data: Google, Google Maps)

Table 22 shows a summary of the active transportation networks’ safety effects in the SZs and HZs, as well as the plausible countermeasures.
### Table 22 Summary of Active Transportation Network Safety Effects and Policy Recommendations

<table>
<thead>
<tr>
<th>Active Transportation Network Indicator</th>
<th>Summary of the Safety Effects in HZs and SZs</th>
<th>Policy Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infrastructure</strong></td>
<td>Low level of bike and pedestrian facilities is a trigger variable in 70% of the HZs</td>
<td>Build more bike and pedestrian infrastructure with higher levels of separation from motorized traffic</td>
</tr>
<tr>
<td></td>
<td>High level of bike and pedestrian facilities is a trigger variable in 46% of the SZs</td>
<td>Consider shared space designs</td>
</tr>
<tr>
<td><strong>Connectivity</strong></td>
<td>62% of the HZs have higher connectivity than regional mean, while 62% of the SZs have lower connectivity than regional mean</td>
<td>Treat minor intersections by considering the installation of crosswalks, bike pathways, actuated signals, hybrid beacons, and proper signage</td>
</tr>
<tr>
<td><strong>Continuity</strong></td>
<td>Pedestrian network continuity is lower than regional mean in 77% of the HZs, and bike network continuity is lower than regional mean in 54% of the HZs</td>
<td>Redesign intersections to be simple right-angled</td>
</tr>
<tr>
<td></td>
<td>Pedestrian network continuity is higher than regional mean in 46% of the SZs, while the bike network continuity is found higher than regional mean in 31% of the SZs.</td>
<td>Install separated and continuous facilities for active transportation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduce intersection density along with installation of frequent midblock crossings</td>
</tr>
</tbody>
</table>

In general, as for the pedestrian network, it should be well-planned, in good condition, clear, and safe for walking. General provisions for sidewalks include pathway width, slope, and space for street furniture, utilities, and landscaping. Small blocks with complete sidewalk network that is accessible and made of high-quality materials are attributes that help to create a positive walking environment. Ideally, both sides of the street should have sidewalks in order to have a high level of sidewalk continuity (Urban systems and Region of Peel, 2016). As for the bike network, designated bike facilities should be provided on all major roads to attain the highest level of directness and connectivity in the network as these streets are typically where destinations are located. Providing direct routes that connect to key destinations will ensure that cycling travel
time is competitive with driving. When it is impractical to supply bike facilities on a major street, some guidelines need to be used to determine whether it is suitable to supply facilities on a parallel local street (U.S department of housing and urban development, 2016). Cities are encouraged to build up a bike network comprised of primary routes, supplemented with sub-routes that can offer connections between those primary routes (Urban systems and Region of Peel, 2016). The Cycling in Cities Program at the University of British Columbia revealed that in order to augment the cycling modal share in urban areas, the spacing for bike routes’ network with designated facilities should not exceed 500 meters (Winters et al., 2010). It is also essential that gaps within the bike network are identified and prioritized. A cyclist stumbling on an unforeseen gap is forced to either continue through potentially hazardous conditions or detour to a safer route which often requires local knowledge as well as time and effort.

Overall, it is crucial for a well-planned pedestrian and bike networks to be accessible and usable by a large section of people, including seniors, people with disabilities, and parents with children. The design of the active transportation environment should accommodate the distinctive needs of these groups and to provide better circulation along with the highest benefits for safety and ridership (Urban systems and Region of Peel, 2016).

Lastly, it should be noted that the recommendations in this subsection (Discussion and Remedies) are based on the conclusions of various advocacy and empirical studies as well as limited field studies. It is, however, necessary to conduct additional and more rigorous micro-level research to affirm the effectiveness of those recommendations in improving active transportation safety.
Chapter 10: Conclusions

This chapter presents the thesis conclusions and it encompasses three main components; summary, contributions, along with limitations and future work.

10.1 Summary

The primary goal of this research is to provide transportation engineers, planners, and policymakers with tools that can be used to improve cyclist and pedestrian safety. The study used extensive GIS data from the City of Vancouver 134 traffic analysis zones to develop empirical macro-level crash models (CMs) incorporating variables related to socio-economics, land use, built environment, and road facility. Some of those variables were investigated for the first time in macro-level safety studies such as actuated signals, separated bike lanes, etc. In addition, indicators for bike and pedestrian networks’ configuration were originally developed on the macro-level using graph theory.

The cyclist CMs were developed using two main traffic exposure variables, i.e., vehicle kilometer travelled and bike kilometer travelled, while the pedestrian CMs were developed using vehicle kilometer travelled and walk trips. Previous studies in the literature used less representative proxies for cyclist and pedestrian exposure (e.g. Wei and Lovegrove, 2013; Amoh-Gyimah et al., 2015). Furthermore, some former studies did not even use an exposure measure for motorized traffic (e.g. Siddiqui et al., 2012; Chan, 2015). This research overcomes such limitation that could lead to biased results, since traffic exposure is one of the main factors that affect crash occurrence.
This research attempted to develop reliable and rigorous macro-level cyclist as well as pedestrian CMs. The state-of-the-practice EB technique along with the state-of-the-art FB technique was employed for statistical inference and modeling in this research. Three types of FB models were tested, i.e., PLN, RPPLN, and SPLN. The statistical comparison indicated that the SPLN model outperformed the other models. A considerable proportion of the total variability was explained by the spatial correlation under the SPLN model, which indicates that ignoring spatial correlation may result in a biased inference when modeling AT safety.

GLM EB univariate CMs as well as SPLN FB univariate CMs were developed to study the associations of the active transportation safety correlates with cyclist and pedestrian crashes. The results showed that cyclist/pedestrian crashes were non-linearly and positively associated with the traffic exposure variables, i.e., BKT, VKT, and W respectively. The exponents of the exposure measures were less than one supporting the “safety in numbers” hypothesis. The exponents of W and BKT were higher than that of VKT showing the higher effect of the non-motorized exposure on active transportation safety. The results also showed that the increase in cyclist/pedestrian crashes was associated with the increase in socio-economic attributes such as employment and household densities, and built environment attributes such as transit stop and traffic signal densities. Regarding land use, a positive association was found between cyclist/pedestrian crash frequency and commercial area density, while both residential and recreational areas’ densities had negative associations with active commuters’ crashes. For road network facilities, higher cyclist/pedestrian crash frequency was found associated with more arterial and collector roads proportion, while a decline in those crashes was found associated with the increase in local roads proportion. Cyclist crashes were negatively associated with the
off-street bike links proportion, and pedestrian crashes were negatively associated with the pedestrian actuated traffic signals. Bike and sidewalk networks’ connectivity indicators (except pedestrian network coverage) were all found positively associated with cyclist/pedestrian crashes on the contrary of the continuity (except pedestrian network linearity), infrastructure, and topography indicators of the active transportation network, which were found negatively associated.

The cyclist and pedestrian safety correlates have almost similar effects for all the investigated variables, which led us to investigate the development of a full Bayesian multivariate model using the mixed approach. The mixed approach can account for mode and spatial correlations while flexibly incorporating different explanatory variables for each crash type. The developed multivariate models showed the significance of the correlation (due to spatial effects) between pedestrian and cyclist crashes indicating the importance of addressing such dependency when investigating active transportation safety.

A statistical discussion that investigated different statistical issues regarding the developed univariate and the multivariate crash models was then presented in this thesis. Basic traffic exposure CMs as well as joint CMs incorporating traffic exposure and explanatory variables were developed for cyclist and pedestrian crashes. For each of these two CM categories, unstructured random error term and spatially structured conditional autoregressive term were incorporated in the univariate and multivariate model forms. The multivariate exposure and joint models performed better (significantly lower DIC) than the corresponding univariate models. This was true for the models accounting for unstructured effects only as well as those accounting
for both structured and unstructured effects. Also, the univariate and multivariate models incorporating both spatially structured and unstructured effects outperformed the corresponding models not incorporating the spatially structured effects. For both the exposure and joint crash models, a significant correlation was observed between cyclist and pedestrian crashes. The mode correlation was found to have a higher impact on model performance than the spatial correlation in the case of exposure crash models, but have a lower impact in the case of joint crash models. Also, the performance of the joint crash models was found significantly better than the performance of the basic exposure crash models, and the mode correlation was lower in the joint models than in the exposure models. This showed that the variables added to the joint model helped to reduce the omitted variables bias and unobserved heterogeneity.

The research then presented a framework for identifying, diagnosing, and treating active transportation hot zones, where the city of Vancouver was used as a case study. Active transportation hot and safe zones were identified and ranked using the Mahalanobis distance method. The HZs were found spatially clustered in four clusters, with the downtown being the largest cluster including seven zones out of the thirteen hot zones identified. Consistency tests were undertaken to demonstrate the superiority of the state-of-the-art FB Mahalanobis distance HZID method over the state-of-the-practice EB PSI HZID method and show the disparity between the two methods in the identification and ranking results. Trigger variables were then statistically identified for the various hot and safe zones. City of Vancouver’ downtown area incorporated most of the trigger variables, which implies the rigorous efforts needed to treat the active transportation network there. The other HZs’ clusters incorporated trigger variables at various extents.
Lastly, three levels of policy recommendations, i.e., land use management, traffic demand management, and traffic supply management, are suggested through literature consultation, trigger variable analysis, and limited field studies. In brief, for a zone to be safely bikeable and walkable, it is suggested to design small blocks with a broad variety of land uses and sustainable street networks as well as apply diverse demand management programs. An added bonus would be servicing those blocks with a continuous, well-connected, and accessible active transportation network, along with a safely separated and easily accessible transit network.

10.2 Contributions

The thesis introduces several contributions to the traffic safety literature, which are related to four central themes. The four themes are AT safety correlates, AT crash modeling, AT hot zones identification, and AT policy recommendations.

Three main contributions related to AT safety correlates are as follow:

1. The developed crash models incorporated representative exposure indicators for the motorized and non-motorized traffic rather than using weak proxies as in the previous macro-level studies.

2. The crash models developed for active commuters’ safety incorporated critical explanatory variables that were not investigated in the former macro-level studies. The variables included AT network indicators such as connectivity, topography, and continuity features, as well as several built environment, socio-economic, land use, and road facility features.
3. The impact of the various safety correlates on pedestrian and cyclist crashes were evaluated to investigate their consistency among the different crash modes.

As for AT crash modeling, the following contributions are highlighted:

4. State-of-the-practice EB CMs as well as state-of-the-art FB CMs were developed to investigate the effects of the AT safety correlates and assess the consistency of the results among the two approaches.

5. Three FB modeling approaches were investigated, i.e., PLN, RPPLN, and SPLN, using a comprehensive set of covariates to select the FB crash modeling approach that would yield the best performance when evaluating active commuters’ safety.

6. A mixed modeling approach was developed to allow more flexibility for the multivariate CMs. This approach allowed the covariates to vary among crash modes (i.e., cyclist and pedestrian crashes) while accounting for the correlation between those modes and spatial effects.

7. A comprehensive investigation was conducted regarding the impact of mode and spatial correlations on the performance of active commuters’ macro-level CMs. The changes in mode and spatial correlations, when more explanatory variables are added, were also monitored.

Regarding hot zones identification, the following two contributions are achieved:

8. A methodology for hot and safe zones identification, based on the concept of Mahalanobis distance, was proposed for the macro-level safety analysis of active transportation for the first time.
9. The proposed FB Mahalanobis distance method was used to calibrate the performance of the state-of-the-practice EB PSI method for active transportation HZID.

Lastly, for the policy recommendations that were suggested for the city of Vancouver, the following contributions are highlighted:

10. Trigger variables were detected for the identified active transportation hot and safe zones within the city of Vancouver.

11. A diagnosis and remedy framework was established based on the trigger variables’ analysis, field studies, and literature consultation in order to upgrade active transportation safety within the City of Vancouver.

10.3 Limitations and Future Work

This thesis can contribute to a better understanding of active transportation safety analysis. However, this thesis is not without limitations and several areas of further research can be pursued as follows.

Regarding crash data, some limitations can be observed such as unreported crashes due to low severity and absence of records for cyclist/pedestrian crashes. The analyzed crash data included all the three levels of crash severity, i.e., fatality, injury, and property damage only, as the crash data was not categorized so as not to disperse the limited sample. Also, only one method of boundary crashes aggregation was applied in this research and no further validation was conducted, which should be addressed in future research.
As for crash modeling, despite the advantages of the macro-level crash models, they might overlook the effects of some essential micro-level safety covariates. Also, the developed crash models were not validated or applied to other active transportation environments to assess their transferability. The environment, traffic volume, and user behavior of active transportation in the City of Vancouver can be different from other cities in Europe or Asia for example. Therefore, it is worthy to investigate the transferability of the developed crash models to other pedestrian and cyclist environments to assess the diverse perspectives that may impact active transportation safety. Moreover, endogeneity can be an issue in the developed CMs. Endogeneity can be due to measurement errors, simultaneous causality, or omitted variables. For example, some important policy variables in this study, such as zonal length, can plausibly suffer from endogeneity due to simultaneous causality with crash frequency or correlation with omitted covariates. This can lead to biased parameter estimates and erroneous marginal effects. Endogeneity should be tested, and econometric approaches such as instrument variables methods can then be applied to account for the statistical effects relating to endogeneity. Two-level modeling techniques can also be attempted. Incorporating temporal correlation in the active modes CMs can contribute to reducing the unobserved heterogeneity. Besides, accounting for the crash severity levels using an extended multivariate structure is expected to improve the model performance.

Regarding explanatory variables, it should be noted that the crash data were collected for five years (2009-2013) while the associated explanatory variables and traffic exposure measures were measured at only one year; i.e., the median year, 2011. Also, it was hard to use walk kilometer travelled (WKT) as a pedestrian exposure measure in the CMs due to the absence of a comprehensive count of walk trips, and so relying on the EMME model output instead.
Collecting walk trips count data in order to deduce a WKT indicator as a traffic exposure measure for pedestrian crashes is expected to improve the CMs’ analysis results. It is also worthy to investigate more variables related to bike and pedestrian network indicators, such as those mentioned in Hong et al. (2016) study, as well as variables that represent road network patterns. Besides, incorporating more variables associated with active commuters’ activities and safety, such as weather and different types of bike/pedestrian facilities, needs to be considered. This is expected to improve the macro-level crash models performance and reduce the omitted variable bias.

For the hot zones identification, the consistency tests that were used to compare the HZID methods lacked before-after crash data that can be used to assess the long-term consistency of the HZID methods. It is necessary to apply more rigorous consistency tests along with more comprehensive crash data to further affirm the superiority of the FB Mahalanobis distance method over the EB PSI method.

The effects of some policy recommendations were not comprehensively analyzed or quantified in this research to determine their applicability and suitability to the hot zones. Also, costs may vary widely depending on the type and the scope of the facility/activity proposed. Therefore, additional site-specific testing of the recommendations, as well as a comprehensive benefit-cost analysis, is required. In addition, investigating more trigger variables is essential in order to propose the most efficient active transportation policies. Monitoring the variation in the crash rates, after applying the suggested remedies, can validate the results of this research.
Lastly, examining the associations between bikeability/walkability scores or indices and cyclist/pedestrian crashes, as well as investigating the associations of real traffic exposure measures with graph and non-graph indicators can be interesting topics for future research.
Bibliography


Noland, R.B., Klein, N.J. and Tulach, N.K., 2013. Do lower income areas have more pedestrian casualties?. Accident Analysis & Prevention, 59, pp.337-345.

Nelson, A., & Allen, D., 1997. If you build them, commuters will use them: Association between bike facilities and bike commuting. Transportation Research Record: Journal of the Transportation Research Board, (1578), 79–83.


Urban systems and Region of Peel, 2016. Pedestrian *and bicycle design guidance report*.


U.S department of housing and urban development, 2016. *Creating walkable and bikeable communities*.


