TRANSIT OPERATIONS, CITY TRANSPORTATION PLANS AND OVERALL TRANSPORTATION NETWORK SAFETY

by

Md. Ahsanul Karim

B. Sc. Engg. (Civil), Bangladesh University of Engineering & Technology, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

The Faculty of Graduate Studies

(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

April 2011

Abstract

Modern transportation planning considers issues such as road congestion, pollution and mobility proactively. However, road safety is usually evaluated in a reactive manner only after the implementation of transportation plans, and when safety problems arise. Although this reactive road safety approach has been very effective, it is associated with significant costs on communities. Therefore, several researchers advocate a more proactive approach to road safety analysis. Several studies developed macro-level Collision Prediction Models (CPMs) that could assess the road safety in a proactive manner, and provide a safety planning decision support tool to community planners and engineers. However, these models have limitations as they do not target the safety evaluation of different goals of a typical city transportation plan. Therefore, the motivation for this research arises from the necessity of developing tools that could predict the safety effect of a typical city transportation plan such as changes in the transportation and transit network configurations, and ultimately could evaluate the safety estimates among alternatives of different transportation plans and policies.

The main goal of this thesis is to develop a set of macro-level collision prediction models to investigate the relationships between various transportation and sociodemographic characteristics, and the overall roadway safety. The developed models consider the Poisson variations and the heterogeneity (extra-variation) on the occurrence of collisions and the spatial effects on the distribution of collisions caused by the similarity in environment and geography of the neighbouring sites. It has been shown that the goodness of fit improved with the incorporation of spatial effect.

In this study, a set of zonal-level transit reliant and application-based collision prediction models were developed. Data from Metro Vancouver, British Columbia were used to develop models using a generalized linear modelling approach with a negative binomial error structure. Different transit-related variables were found to be statistically significant, namely bus stop density, percentage of transit-km traveled with regard to total vehicle-km traveled, percentage of commuters walking, percentage of commuters biking, and percentage of commuters using

transit. The CPMs related total, severe, and property damage only collisions to the implemental aspects related to the goals of long-term transportation plans.

Table of Contents

List of Tables

List of Figures

Acknowledgements

All praises go to the Almighty, the most Merciful, and the most Beneficent.

I offer my enduring gratitude to my supervisor, Dr. Tarek Sayed for the guidance, encouragement, patience, technical as well as financial support he extended to me during the course of this project. I would also like specially thank Dr. Mohamed Wahba, my co-supervisor, for enlarging my vision of engineering and providing coherent answers to my endless problems.

I also acknowledge the contribution of my colleague, Shweekar Ibrahim, who helped me learn the software WinBUGS, which I needed while performing my research.

My sincere thanks to Insurance Corporation of British Columbia (ICBC), Translink and The Census Canada for their data, without what this research would not have been possible.

Finally, but foremost, I would like to thank my parents, sisters, and brother for the unconditional love, support and encouragement they provided me in all my endeavours despite living in thousands of kilometres away.

...To my beloved parents...

1 Introduction

The main topic of this thesis is to investigate the relationships between various transportation and sociodemographic characteristics, and the overall roadway safety. This topic is motivated by the problem of a lack of reliable empirical tools for planners and city transportation authorities to evaluate the safety estimates among alternatives of different transportation plans and policies. The problem is addressed by developing several zonal level transit reliant and application-based collision prediction models using generalized linear modelling multivariate regression techniques.

1.1 Background

Today, transportation is the key infrastructure of a country. The economic status of a country depends on how well it is served by the roads, railways, airports, ports, pipelines, and shipping networks. Each of these modes has its own appeal and its own drawbacks. However, because of the ability to permit high speed to reach even the remotest part of a country, roadway is the most extensively used mode of transportation throughout the world. Similarly, road transportation is the major mode of transport in Canada (Transport Canada, 2010), and it is obvious that it will continue to be the prime mode of transport for decades to come. Though the development of a transportation system is an essential part of the human civilization, this growth has also caused safety problems. Therefore, considerable efforts are being devoted to overcome these safety problems, and transportation engineers are conducting passionate research to analyze and model collisions.

1.2 Statements of the problem

Road safety has increasingly been regarded as the primary transportation concern all over the world. Worldwide, around 1.3 million people lose their lives annually in road collisions. In the US, 10 million collisions occurred in 2007. In Canada, about 3000 fatalities and 200,000 injuries resulted from road collisions in 2007 (de Leur & Sayed, 2003; Transport Canada, 2010).

Some previous studies (Gaspers, 2004; Peden et al., 2004) have considered the magnitude of death toll due to road collisions to be a problem of epidemic nature. According to the '*World Report on Road Traffic Injury Prevention*' road collision is the ninth leading contributor to the global burden of disease and injury. This report also indicates that the economic cost of road collisions and injuries is estimated to be 1% to 2% of the gross national product (GNP) worldwide (Peden et al., 2004). If the existing trends continue, the *WHO Global Burden of Disease: 2004 Update* study predicts the following changes from 2004 to 2030 (WHO, 2008):

- Road traffic collisions will rise from the ninth leading cause of death globally in 2004 to the fifth in 2030.
- Road traffic collision deaths are projected to increase from 1.3 million in 2004 to 2.4 million in 2030 (representing more than 5% of all deaths).
- Road traffic injuries will become the second leading cause of death for low- income and middle - income countries.
- Road traffic deaths will increase on average by over 80% in low-income and middle - income countries and decline by almost 30% in high-income countries. This reduction in collisions in high-income countries will likely be caused by extensive research on this area.

While modern transportation planning addresses proactively such issues as road congestion, pollution and mobility, road safety has usually been dealt within a reactive manner. In other words, alternative transportation network scenarios and transportation policies are typically evaluated in the planning process based on their expected effects on congestion, pollution, and other impacts, but not on safety. The most favourable network alternative and supporting policies are then proposed and implemented, and road safety is assessed only after the transportation facility has been used and safety problems have arisen.

The traditional road safety improvement programs (RSIPs) generally focus on the identification, diagnosis, and remedy of existing collision-prone locations or "black spots". This is a reactive approach by nature and it has been very successfully applied for some time (Lovegrove, 2006). It uses statistical techniques such as empirical Bayes (EB), full Bayes, and micro or macro level collision prediction models (CPMs) to identify hazardous locations or black spots, characterized by abnormally high collision frequency or severity. Analyzing the associated collision problems, collision modification factors (CMFs) and benefit/cost (B/C) ratios are then used to identify the most cost effective safety countermeasure. However, the application of this reactive approach requires a considerable collision history, and that requirement is regretfully associated with significant costs on the communities as retrofitting countermeasures at the identified black spots in existing transportation network is usually costly. Hence, the transportation authorities as well as the road safety researchers have come into a consensus that a proactive approach need to be taken to address and mitigate this problem at the very early stage (de Leur & Sayed, 2003; Lovegrove & Sayed, 2006b). A proactive approach addresses road safety explicitly and focuses on predicting and improving the safety of the planned facilities at the very initial stage. The objective of the proactive road safety approach is to evaluate safety throughout the planning process and thereby minimize the road safety risk, and also prevent black spots from occurring. However, a major barrier to this proactive approach has been the lack of necessary empirical tools to evaluate road safety from a macro-level or planning perspective. In a study, (De Leur & Sayed, 2002) have highlighted this limitation and proposed a framework for proactive road safety planning, arguing that such an approach to road safety not only complements the traditional or reactive method but also prevents unsafe situations from arising in the first place.

There have been a few studies which aimed to develop models to predict the number of collisions for individual elements of the urban transportation network such as intersections and arterial road sections. However, few of these studies have developed models at a more aggregate zonal level that could be used as part of the transportation planning process. It is of interest to note some recent research by Lord (Lord, 2000) that aimed to develop a tool that would allow the estimation of traffic collisions on computerized or digital transportation networks during the planning process. To accomplish this aim, collision prediction models specifically created in that research were applied to two sample digital networks created with the help of (EMME/2). The results showed that it is possible to predict collisions on computerized transportation networks, but the accuracy is directly related to the precision of the traffic flow predictions from transportation planning software programs.

Studies that have tried to apply micro-level collision prediction models to evaluate safety in planning, have been largely obstructed by some inherent limitations of these micro-level CPMs (Ho, Nepomuceno, & Zein, 1998; Lord & Persaud, 2004). One of these limitations is due to the fact that usually micro-level CPMs predict the level of safety at a single location, such as an intersection or a road segment, and the traffic volume levels of these locations can be estimated relatively easily. Besides, Lovegrove (Lovegrove, 2006) argued that planning-level forecasts at any one location are usually inaccurate due to their regional screenline level of calibration, their longer-term timeframe, and their much broader focus (i.e. major road networks only). In addition to errors in forecasts, the degree of complexity and amount of programming required to incorporate micro-level CPMs into planning level analyses is convoluted, and does not endear itself to being used by practitioners. Therefore, there is a research gap between what is needed and what is available in terms of reliable empirical tools to facilitate the proactive engineering approach to road safety improvement programs.

Therefore, many road agencies are starting to adopt proactive road safety strategies at the macro-level aimed at reducing/eliminating collisions before problems emerge (de Leur & Sayed, 2003). This requires the development of tools that can estimate the impact of various transportation planning strategies. Recently, some studies by Lovegrove and Sayed (Lovegrove & Sayed, 2006a; Lovegrove & Sayed, 2006b; Lovegrove, 2006; Lovegrove & Sayed, 2007) developed and used community-based macro-level collision prediction models (CPMs) in proactive road safety improvement programs for the Greater Vancouver Regional District (GVRD). The research by Lovegrove et al. (Lovegrove, 2006; Lovegrove & Sayed, 2007) was successful in developing macro-level CPMs and describing the model use guidelines that provided a safety planning decision support tool to community planners and engineers. While their research provided adequate evidence of using macro-level CPMs as a reliable empirical tool, there are some other issues which have not been dealt with. The most important one is the lack of availability of collision prediction models that are based on the safety evaluation of different targets/goals of a city transportation plan.

1.3 City transportation plans

Every city has a transportation plan that aims at making the urban environment a great place to live as well as a safe, prosperous and environmentally friendly economic centre. New urban planning concepts are committed to meeting the safety and mobility needs of residents while reducing the reliance on cars and making massive investments in the development of alternatives to car, such as public transit, carpooling, car sharing and active forms of transportation such as walking and bicycling. The new concepts redefine the function and mode-assignment of streets by emphasizing on sustainable modes of transportation and accommodating the needs for safety and mobility.

The Metro Vancouver (formerly known as the Greater Vancouver Regional District- GVRD), which is home to half of British Columbia's population and workforce, is Canada's gateway to Asia-Pacific economies and includes the nation's third largest urban centre. The GVRD is currently in the process of updating its regional growth strategy. TransLink, Metro Vancouver’s regional transportation authority, is dedicated to creating and sustaining a transportation system that meets the needs of residents, business, and goods movers, in a manner that protects the environment and supports the economic and social objectives of the region (TransLink, 2008). To achieve this goal, TransLink intends to significantly reduce residents’ dependence on cars through massive investment in various forms of public transit, and active transportation, including the bus rapid transit service, sky trains, bikes and walking, and by encouraging such more appropriate uses for cars as carpooling and car sharing. TransLink has increased transit ridership by 38 per cent between 1998 and 2007 while population has increased by 13 percent (City of Vancouver, 2005). In addition to being just a driving force in economic development, public transit also represents a key strategy in sustainable development, because it provides an excellent method of reducing greenhouse gases (Société de Transport de Montréal, 2008).

The city authorities now want to see a change in people’s travel behaviour and want a transit-oriented city. But the safety consequences of this attempt is yet mostly undiscovered. So there is a need to develop models that could predict the zonal safety effect of improvement or change in the transportation and transit network configurations, thus providing one additional input to transportation planners as well as city authorities to comprehensively assess the future network scenarios. These models can also help transit agencies understand the factors that contribute to collision occurrence and allow them to take appropriate steps to provide effective transit services with minimal safety risk. Finally these models can give a more clear understanding on how the change in people’s travel behaviour, more reliance on transit as well as a transit-oriented society approach affect the overall zonal roadway safety.

1.4 Spatial effects on collision data

Most of the studies that dealt with macro level CPMs have considered the Poisson variations and the heterogeneity (extra-variation) on the distribution of collisions. Some micro level CPMs have also considered the spatial correlations or spatial effects, as neighbouring sites typically show similarity in environment and geography which results in the formation of clusters having similar collisions occurrences (Aguero-Valverde & Jovanis, 2008; El-Basyouny & Sayed, 2009). But the spatial effects at the zonal level have rarely been considered (Jegede, 1988; Keeler, 1994). These studies looked at the variations over an entire region, and examined the nature of collisions in particular zones as part of a larger spatial pattern, although the primary units were as large as cities rather than a single metropolitan area. Most of the studies mentioned before have supported the incorporation of spatial effects in CPMs as this incorporation not only improves the goodness of fit, but also it can explain enough variation that some of the model covariates would deem non-significant.

1.5 Objectives of the research

This study aims to bridge the gap related to the inability of micro-level CPMs to do proactive, planning-level safety evaluations as well as to investigate the safety impact of various goals/targets of a long term city transportation plan by developing application-based macro-level CPMs. These macro-level CPMs will allow the city transportation authorities and planners to evaluate the safety of typical transportation plans in a proactive manner, and will result in significant reductions in the level of road collision frequencies. Moreover, this proactive tool may also be able to complement and enhance the effectiveness and efficiency of traditional reactive black spot program.

The specific objectives of this research are as follows:

- To develop transit oriented CPMs to facilitate the evaluation of transit safety at the planning levels as well as to quantify the impact of various transit network variables on the overall roadway safety. The major North American cities now aim at changing people's travel behaviour which will result in more reliance on transit. This policy targets to change the average transit ridership, average transit trip length, transit network, etc. so that a transit oriented society may be developed. The transit reliant CPMs to be developed will evaluate the safety consequences of the changes in these transit related elements.
- To develop application-based macro-level (i.e. zone based) CPMs to evaluate the safety consequences of various goals/targets of a city transportation plan. These CPMs capture the interaction between the collision frequency of all traffic (transit and non-transit) of different type and severity and the implemental aspects related to the goals of long-term

transportation plans, such as travel demand management policy impacts and transit investment implications.

- To evaluate the spatial effects on the occurrence of collisions and to check whether the inclusion of spatial variables can improve the goodness of fit and inference capability of the developed CPMs.
- Finally, to make recommendations on practical guidelines for the use of the developed macro-level CPMs to promote consistency in proactive evaluation of the city transportation plan. Consistent practises of these guidelines will result in reliable and comparable estimates of level of safety of the overall transportation network of a zone.

In a nutshell, this research is largely motivated by the lack of exclusive and reliable tools for the city transportation planners to evaluate the safety consequences of different goals of the city transportation plan. The issue will be dealt with the development of various transportation goal-oriented application-based macro-level collision prediction models (CPM) using extensive data extraction and Generalized Linear Regression Modelling (GLM) regression techniques. The ultimate social and economic benefits of successful model development include building new empirical tools for city authorities to evaluate the safety of the city transportation plan, helping planners and engineers do proactive road safety planning, and enhancing the effectiveness in RSIPs significantly.

1.6 Structure of the thesis

The research work performed in this study is divided into different topics and presented in five chapters.

A brief introduction to the background and statement of the problem is presented in the Chapter One. The chapter also contains the objectives and importance of the study along with a brief description of the structure of the thesis.

Chapter Two presents the literature review of the related topics. It also describes and highlights the state of research on, and specific issues related tothe development and use of macro-level CPMs. It describes the magnitude of the road safety problem, reactive and proactive engineering approaches, current state of associated empirical tools, collision prediction model development, and previously performed works on collision prediction model. It also discusses the TransLink's Transport 2040 plan and the spatial effects on the occurrences of collisions.

In Chapter Three, the methodology of this study is described. This covers the database and variables selection, data extraction and processing, necessary assumptions, and the analytical methods of conceptual model development.

Chapter Four contains the developed macro-level transit reliant and application-based CPMs with necessary statistical parameters for goodness of fit. It also discusses the associations of these CPMs, interprets their outcomes, analyzes the spatial effects on those models, and compares the results with previous studies.

The summary findings of the research are presented in Chapter Five. This chapter also contains the research contributions, limitations, as well as the suggestions for further study.

An appendix section is attached at the end of this thesis.

2 Literature Review

2.1 Introduction

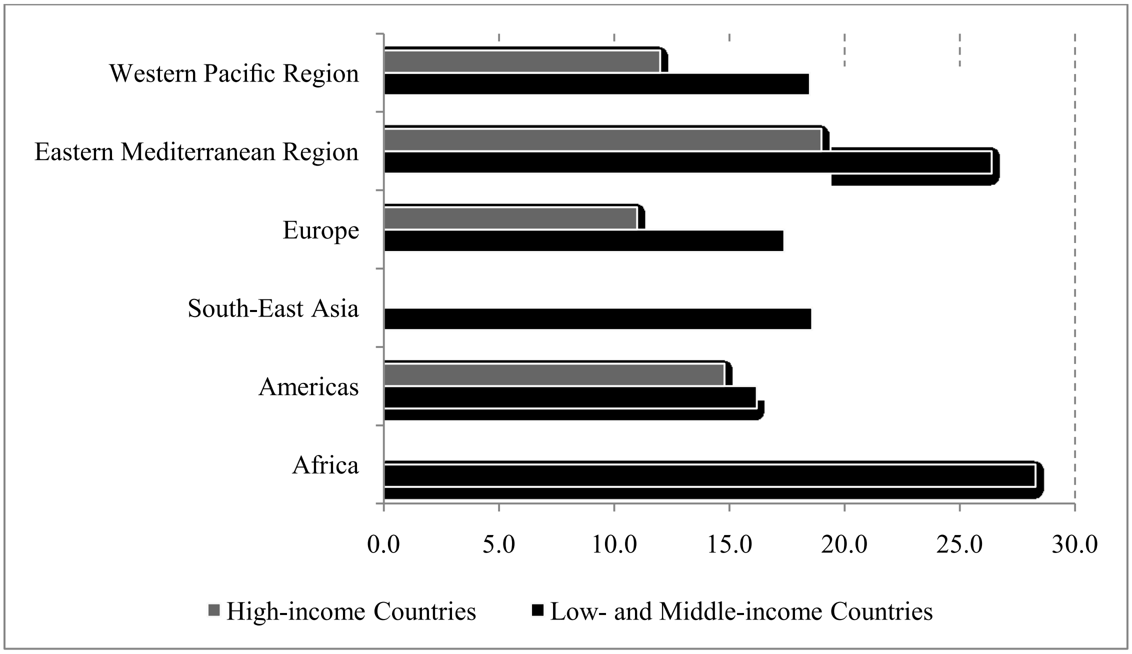
This chapter consists of a comprehensive literature review on the context and background of the study performed in this thesis. The chapter is composed of seven main sections. In section 2.2, the road collision situations of the world, North America and British Columbia are briefly summarized. Sections 2.3 and 2.4 discuss different reactive and proactive approaches of road safety estimates. Section 2.5 deals with the methodology, background, development and uses of macro-level zonal collision prediction models, and the relevant previous studies. The spatial effects on the occurrence of collisions and approaches to develop spatial models are presented in Section 2.6. It also contains some previous research on spatial correlation. Section 2.7 briefly presents the TransLink's Transport 2040 plan with its goals and strategies. Finally, section 2.8 reviews the definitions and discusses different terms and tools related to statistical analysis of collision data.

2.2 Road collision situation

2.2.1 Global perspective

The statistics of road collisions worldwide are important to understand the international nature of road safety problem. Worldwide, almost 16,000 people die every day from different types of injuries. Injuries account for 12% of the global disease, making it the main reason of death among people aged one to forty year, and at the same time the third cause of overall mortality. 25% of all these injuries are caused by road traffic collisions (Peden et al., 2004). Road collisions are liable for the death of around 1.3 million people all over the world. According to the *WHO World Report on Road Traffic Injury Prevention*, death rates in road collisions are higher in low-income and middle-income countries than in high-income countries. Altogether, low-income and middle-income countries accounted for 90% of all road traffic deaths. Figure 1 shows the road traffic injury mortality rates per one hundred thousand populations in WHO regions in the year 2002 (Peden et al., 2004).

Figure 1 Road traffic injury mortality rates per one hundred thousand populations in 2002



Source: (Peden et al., 2004)

Global, regional and country trends show that road traffic death rates have decreased in high-income countries since the 1960s and 1970s, although countries' rates vary greatly even within the same region. For example, in North America, from 1975 to 1998, road traffic fatality rate per one hundred thousand populations declined by 27% in the United States but by 63% in Canada. Meanwhile, rates in low-income and middle-income countries have increased substantially. In Asia, from 1975 to l998, road traffic fatality rates increased by 44% in Malaysia but by 243% in China. Two major studies predict that the trend towards increase in low-income and middle-income countries will continue, unless deliberate action changes it (Peden et al., 2004).

The economic cost of road collisions and injuries is estimated to be about 1% of GNP in developing countries, 1.5% in countries in economic transition and 2% in highly-motorized countries totalling the annual burden of economic costs at around US$ 518 billion. If an individual country is considered, the economic burden accounts for a loss of percentage of GNP ranging from 0.3% in Vietnam to almost 5% in both of Malawi and in Kwa-Zulu-Natal, South Africa. Table 1 shows the approximated road collision costs by region (Peden et al., 2004). In this table, the highly-motorized countries consist of Australia, Japan, New Zealand, North America, and the western European countries.

Table 1 Approximated road collision costs by region

Region	GNP, 1997 (US$ Billion)	Estimated Annual Cost	
		% of GNP	Costs (US$ Billion)
Africa	370	1	3.7
Asia	2454	1	24.5
Latin America and Caribbean	1890	1	18.9
Middle East	495	1.5	7.4
Central and Eastern Europe	659	1.5	9.9
Subtotal	5868		64.4
Highly-motorized Countries	22665		453.3
Total			517.7

Source: (Peden et al., 2004)

2.2.2 North American perspective

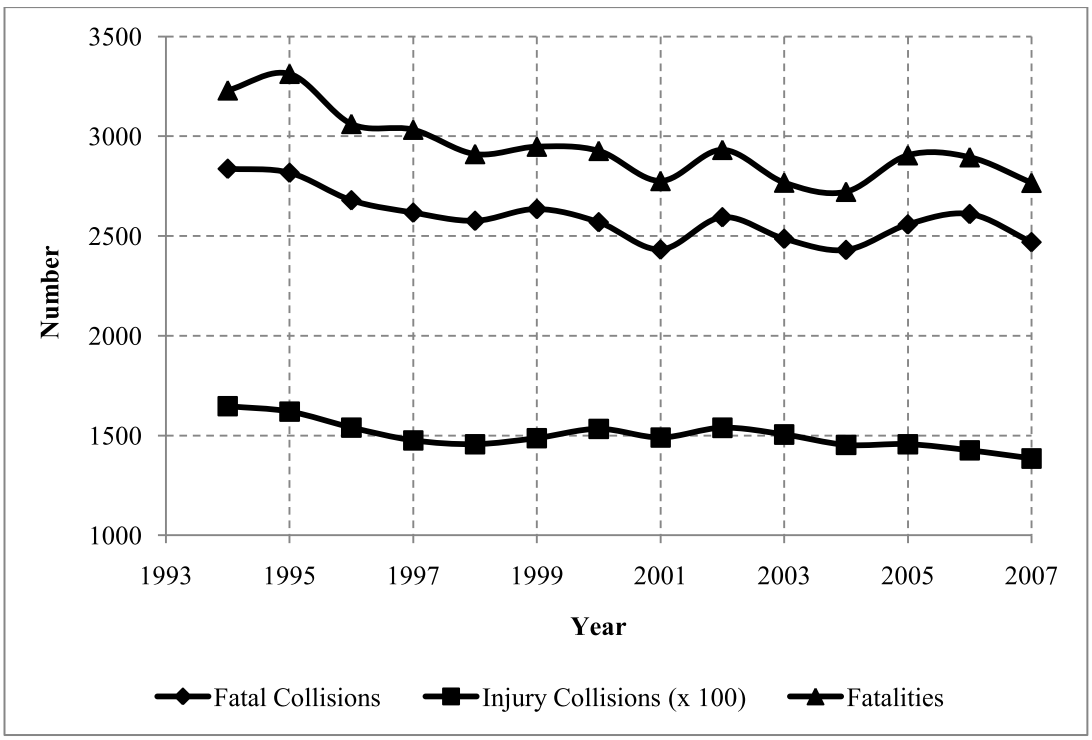
The condition of road traffic collisions in North America is relatively better than other regions except the European region. In general, since the 1960s, considerable reductions in the numbers and rates of fatalities in North America have been observed. This reduction in road traffic fatalities is considered to be the outcome of wide range of road safety research and studies, implementation of different safety measures, and enforcement of traffic laws. However, this reduction might not necessarily reflect an improvement in road safety for each individual. Table 2 presents the estimated mortality in North America caused by road collision injuries in the year 2002 for different sex and income groups (Peden et al., 2004).

Table 2 Estimated mortality caused by road collisions in 2002 (by sex and income level)

Indicator	Income Level	Male	Female
Mortality due to Road Collisions (Absolute Numbers)	All	100378	33405
	High	32610	15255
	Low/Middle	67768	18150
Mortality due to Road Collisions (Rate per 100,000 Populations)	All	23.9	7.7
	High	20.5	9.3
	Low/Middle	25.9	6.8
Proportions of All Deaths due to Injuries (%)	All	24.0	27.6
	High	28.5	29.6
	Low/Middle	22.2	26.1
Proportions of All Deaths (%)	All	3.2	1.2
	High	2.5	1.2
	Low/Middle	3.7	1.2

In the US, 10 million collisions occurred in 2007, and road collision is the major cause of death for people aged between six months to forty-five years. It is also the reason of more pre-retirement years of life lost there than even the years lost by the combined effects of heart diseases and cancer. Ultimately, road collision becomes the prime cause of loss of productive life (de Leur & Sayed, 2003). The risk of automobile injury per passenger mile traveled in the US is 0.91 deaths per billion passenger miles, and in case of rails roads, airlines and buses the corresponding values are 0.06, 0.03 and 0.02 respectively (Waller, 2000). It was also estimated that the human capital costs of road traffic collisions in 2000 in the US were US$ 230 billion (Peden et al., 2004).

Figure 2 Collisions and casualties in Canada during 1994-2007



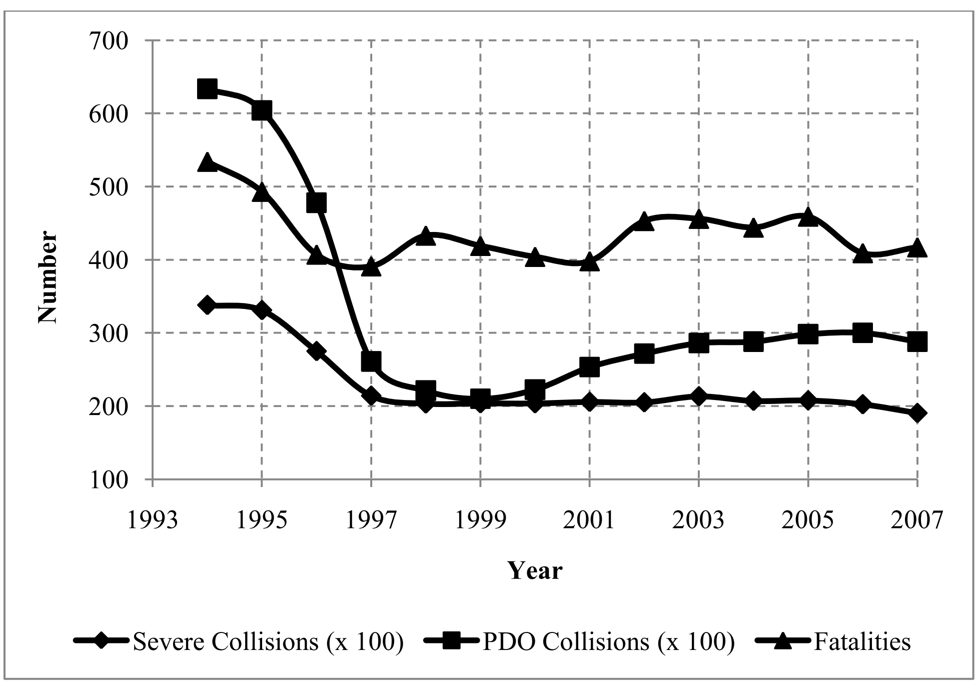
Source: (Transport Canada, 2010)

In Canada, about 3000 fatalities and 200,000 injuries resulted from road collisions in 2007 (Transport Canada, 2010). During the last fifty years, more people lost their lives in road collisions than the number of Canadians killed in two world wars. On an average, eight Canadians die in road collisions daily, and many more become seriously injured. Fortunately, the data during the period of 1994 to 2007 shows that the total numbers of fatal and injury collisions as well as the number of fatalities per year are decreasing, although the decreasing rate is relatively small. Figure 2 shows the collisions and casualties in Canada during 1994-2007. The economic burden of road collisions here is also enormous. Each year approximately $ 25 billion are lost due to road collisions (Transport Canada, 2010).

2.2.3 Local perspective

The province of British Columbia has adopted a policy to report most auto collision claims as part of the provincial mandate, and these claims are reported by the Insurance Corporation of British Columbia (ICBC). In a study, De Leur and Sayed showed that these claims can be used as a reliable proxy for reported collision data (De Leur & Sayed, 2000). According to the ICBC, in 2007, 47,870 traffic collisions were reported. Among them 17,914 were injury collisions and 372 were fatal collisions. The total reported number of fatally injured victims was 417. The number of people reported injured in 2007 was 25,064. In the year 2006 the value was 27,556. On average a fatality occurred in every 21.0 hours, and 68.7 people were reported injured each day in the province in 2007. About 90.9% of all collisions were attended by police (ICBC, 2008).

Figure 3 Collisions and casualties in BC during 1994-2007



Source: (ICBC, 2008)

Figure 3 shows the total numbers of severe and property damage only (PDO) collisions as well as fatalities in BC during the period of 1994-2007. This figure shows that from 1995 to 1996, total reported collisions dropped by 19% (from 93,490 to 75,288). The collisions fell by a further 37% (to 47,495) from 1996 to 1997. These considerable reductions maybe partially attributed to some massive development in transportation network during those years. Besides, some police jurisdictions ceased to submit collision report forms altogether during these years (ICBC, 2008).

2.3 Reactive road safety improvement program

Reactive road safety improvement programs (RSIPs) consist of identifying and treating the locations those are hazardous with respect to collisions. These potential hazardous locations are called collision prone locations. Traditional RSIPs generally focus on the identification, diagnosis, and remedy of existing collision-prone locations or "black spots". These collision prone locations or black spots are characterized by abnormally high collision frequency or severity compared to other locations. This approach is reactive by nature and it has been very successfully applied (Lovegrove, 2006). It uses statistical techniques such as empirical Bayes (EB), full Bayes, and micro or macro level collision prediction models (CPMs) to identify hazardous locations or black spots. Analyzing the associated collision problems, collision modification factors (CMFs) and benefit/cost (B/C) ratios are then used to identify the most cost effective safety countermeasure. However, the application of this reactive approach requires a considerable collision history, and unfortunately this requirement is associated with significant costs on the communities as retrofitting countermeasures at the identified black spots in existing transportation network is usually costly. Hence, the transportation authorities as well as the road safety researchers have come into a consensus that a proactive approach need to be taken to address and mitigate this problem at the very early stage of the transportation planning process (de Leur & Sayed, 2003; Lovegrove & Sayed, 2006b). A proactive approach addresses road safety explicitly and focuses on predicting and improving the safety of the planned facilities at the very initial stage. The objective of the proactive road safety approach is to address safety issues at the planning stage and thereby minimize the road safety risk and also prevent black spots from occurring.

2.4 Proactive road safety improvement program

Recently, a need has emerged to consider and evaluate the safety of regional or zonal transportation planning with a view to improving the overall safety of the transportation network. This strategy of considering road safety in the planning level has introduced the concept of safety conscious planning. Safety conscious planning aims at incorporating safety routinely throughout each step of the transportation planning process in an effective and comprehensive manner. This approach minimizes the road safety risk and also prevents collision prone or hazardous locations from occurring. However, efficiency and effectiveness of this approach mostly depend on the reliability of the empirical tools those are used to calculate the safety estimates at the planning stages. Although sufficient reliable empirical tools are available for the traditional reactive road safety engineering approach, only a few reliable proactive tools are found (de Leur & Sayed, 2003; Lovegrove, 2006). The following sub-sections discusses the available proactive empirical tools, including road safety audits, Sustainable Road Safety programs, road safety risk indices, and collision prediction models.

2.4.1 Road safety audits

In a proactive nature, road safety audits (RSAs) can be applied in road planning and design stages. Previous researchers have recommended RSAs for all regional and community-wide planning (de Leur & Sayed, 2003; Hadayeghi, Shalaby, & Persaud, 2003). RSAs consist of an audit or explicit safety evaluation of transportation projects of any size, and the benefits of this approach largely depend on the experience and professional judgment of the members of the team performing this audit. Sometimes, micro-level CPMs are used to enhance the effectiveness of this approach.

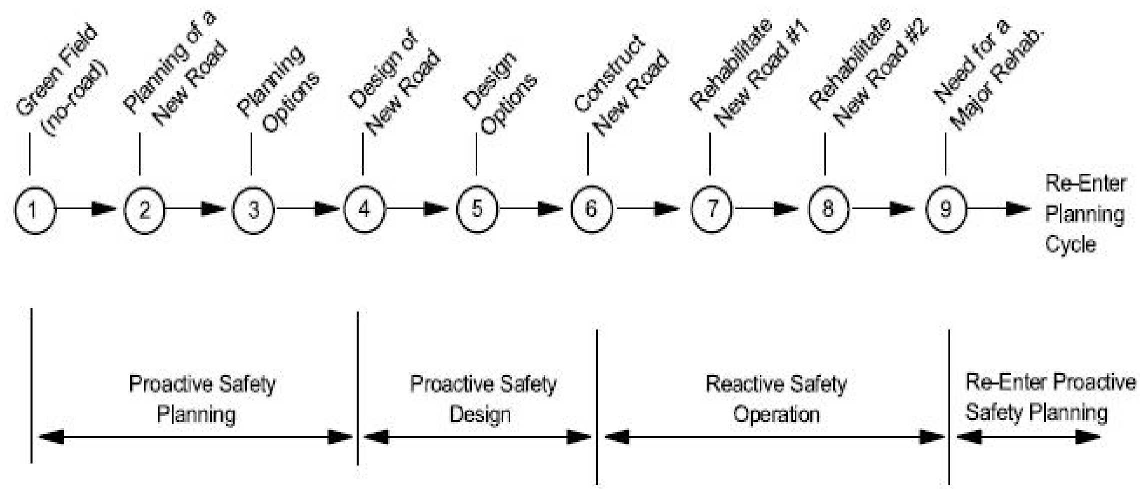
2.4.2 Dutch Sustainable Road Safety program

The Dutch Sustainable Road Safety (SRS) program was aimed at reducing collisions in a proactive manner. This program considered that the root causes of the present road safety problems were the results of absence of an adequate community planning (Lovegrove, 2006). This program was introduced in 1986 and it was significantly based on transportation demand management (TDM) strategies (Khondaker, 2008). Lovegrove has argued that this program provides some important pieces of information and clues for the development of macro-level collision prediction models, although no such CPMs were developed in this program (Lovegrove, 2006).

2.4.3 Proactive road safety planning framework

De Leur and Sayed (de Leur & Sayed, 2003) have shown the opportunity to provide safety inputs into each stage of the planning process. It starts from the beginning of the process (when no road exists) through to the design stage. The first opportunity to provide an explicit safety influence within the planning process is by influencing the planning decisions used to develop solution options. They proposed several guiding principles, to enable decisions to be influenced by safety requirements. A methodology was also proposed there to augment safety evaluation both quantitatively and qualitatively. Figure 4 below shows the transportation planning process with safety input at each stage.

Figure 4 Transportation planning process with safety input



Source: (de Leur & Sayed, 2003)

To quantify the safety impact like other researchers (Haddon, 1980; Hauer, Ng, & Lovell, 1988), De Leur and Sayed (de Leur & Sayed, 2003) considered the three fundamental elements to describe road safety risk - exposure (exposure to hazards), probability (likelihood of encountering hazard), and consequence (severity of hazard if encountered), and assumed the following relationship:

$$\text{Risk} = \text{function of } \{\text{exposure, probability, and consequence}\}$$

Here exposure represents a measure to quantify the “exposure” of road users to potential hazards, probability represents a measure to quantify the chance of a vehicle being involved in a collision, and consequence represents a measure to quantify the severity level resulting from potential collisions. The framework to incorporate safety in planning level by De Leur and Sayed is found to be practical, manageable and brief but sufficient to provide a general guideline for proactive transportation planning. However, the subjective nature of this framework requires additional modification in this area to provide a planning level collision prediction model (Khondaker, 2008).

2.4.4. Micro level planning CPMs

Persaud et al. (Persaud & Dzbik, 1993) introduced a more objective measure of long range safety evaluation using collision prediction models (CPMs). A few researches have been found which have developed micro-level collision prediction models to predict the number of collisions with respect to individual elements of the urban transportation network such as intersections and arterial road sections (Kiattikomol, Chatterjee, Hummer, & Younger, 2008; Lord, 2000). However, there is no denying the fact that studies those have tried to apply micro-level collision prediction models to evaluate safety in planning, have been largely obstructed by some inherent limitations of these micro-level CPMs. One of these limitations is due to the fact that usually micro-level CPMs predict the level of safety at a single location, such as intersection or road segment, and the traffic volume levels of these locations can be estimated relatively easily. Besides, Lovegrove and Sayed (Lovegrove & Sayed, 2006a) argued that planning-level forecasts at any one location are usually inaccurate due to their regional screen-line level of calibration, their longer-term timeframe, and their much broader focus (i.e. major road networks only). In addition to errors in forecasts, the degree of complexity and amount of programming required to incorporate micro-level CPMs into planning level analyses is convoluted, and does not endear itself to being used by practitioners. Therefore, there is a research gap between what is needed and what is available in terms of reliable empirical tools to facilitate the proactive engineering approach to road safety improvement programs.

2.4.5 Macro level planning CPMs

Due to the inherent weakness of micro-level planning collision prediction models, many road agencies are starting to adopt proactive road safety strategies at the macro-level aimed at reducing/eliminating collisions before problems emerge (de Leur & Sayed, 2003). This requires the development of tools that can estimate the impact of various transportation planning strategies. Recently, some studies (Lovegrove & Sayed, 2006a; Lovegrove & Sayed, 2006b; Lovegrove, 2006; Lovegrove & Sayed, 2007) have developed and used community-based macro-level collision prediction models (CPMs) in proactive road safety improvement programs for the Metro Vancouver area (previously known as Greater Vancouver Regional District-GVRD). The research by Lovegrove et al. (Lovegrove, 2006; Lovegrove & Sayed, 2007) was successful in developing macro-level CPMs and describing the model use guidelines that provided a safety planning decision support tool to community planners and engineers. While their research provided enough evidence of using macro-level CPMs as a reliable empirical tool, there are some other issues which have not been dealt with. The most important one is the lack of availability of collision prediction models those are solely based on the safety evaluation of different targets/goals of a city transportation plan. Hence, the target of this research is to develop some macro-level (zone-level) transit reliant and application based collision prediction models to explicitly evaluate the safety estimates of different goals of a city transportation plan.

2.5 Zone level collision prediction models techniques

The following sub-sections will discuss the theory of the development of zone level collision prediction models, and the quality of database based on the previous studies. These sub-sections will also include discussions on how these models can be used for zone specific prediction, and how to identify and rank the collision prone zones.

2.5.1 Development of zone level CPM

Conventional linear modeling approach is usually not used in the development of collision prediction models as traffic collisions are discrete, non-zero, non-negative and over-dispersed (Buyco & Saccomanno, 1988; Hauer et al., 1988; Jovanis & Chang, 1986; Miaou & Lum, 1993). However, generalized linear modeling (GLM) approach has the advantages over conventional linear modeling approach with respect to overcoming the limitations associated with the later approach (Hauer et al., 1988; Sawalha & Sayed, 2006; Sayed & Rodriguez, 1999). Different computer packages are available to model collision data assuming a wide range of probability distributions including Poisson, negative binomial, gamma, etc. The non-linear model form can also be linearised with the help of several built-in link functions in these software packages.

The GLM approach can consider either negative binomial or Poisson error structure. Poisson error structure is relatively simpler to calculate as in Poisson distribution the mean of the data set is equal to its variance. But most of the times, researchers advocated for using negative binomial error structure as collision data are usually over-dispersed (Kulmala, 1996; Sawalha & Sayed, 2006). The reasons behind this over-dispersion might be more than one (Miaou & Lum, 1993). Traffic collisions are complex phenomena and it is really hard to sort out all of the variables those can explain occurrences of a collision. There may be some qualitative variables which cannot be quantified and hence cannot be used in collision prediction models. Uncertainties in traffic exposure data and traffic variables might be another reason of this over-dispersion. Moreover, over-dispersion might also be caused by the non-homogeneous roadway environment, from where collision data come.

The quality of data is very important for the development of CPMs. Data quality may be affected by various reasons (Lovegrove, 2006). In smaller communities, a statistically significant sample size might not be available. Sometimes, collisions data formats are not appropriate for downloading or analysis. Besides, collisions, especially the non-severe collisions, may be under reported. Lastly, there may be error while reporting and compiling collision data. However, fortunately several jurisdictions in North America have adopted policy to report most auto collision claims as part of the provincial mandate, and these claims can be used as a reliable proxy for reported collision data (De Leur & Sayed, 2000).

2.5.2 Zone specific prediction using empirical Bayes refinement

The empirical Bayes (EB) approach is used to refine the estimates from the collision prediction models for any specific location. This approach uses two types of clues: the predicted number of collisions found from the developed CPM, and the observed number of collisions from real data (Hauer et al., 1988; Sawalha & Sayed, 1999; Sawalha & Sayed, 1999).

According to Sawalha and Sayed (Sawalha & Sayed, 1999), the EB estimates of the expected number of collisions can be calculated as:

$$EB_{safety\ estimate} = \alpha \text{ x } E(\Lambda) + (1-\alpha) \text{ x } Observed\ data \tag{1}$$

Where,

$$\alpha(weight\ parameter) = \frac{1}{1 + \frac{Var\left[E(\Lambda)\right]}{E(\Lambda)}} \tag{2}$$

E (Λ) = predicted number of collisions for that location (found from CPM)

Observed data = observed numbed of collisions for that location

The variance of safety estimate,

$$Var(\Lambda) = \frac{\left[E(\Lambda)\right]^2}{\kappa} \tag{3}$$

Therefore, the EB safety estimate can also be expressed as

$$EB_{safety\ estimate} = \left[\frac{E(\Lambda)}{\kappa + E(\Lambda)}\right](\kappa + Observed\ data) \tag{4}$$

The variance of the EB $_{\text{safety estimate}}$ can be expressed by the following equation:

$$Var\left(EB_{safety\ estimate}\right) = \left[\frac{E(\Lambda)}{\kappa + E(\Lambda)}\right]^2 (\kappa + Observed\ data) \tag{5}$$

2.5.3 Identification of collision prone zones using zone level CPMs

Collision prone zones have significantly higher number of collisions with respect to a specific norm. Sawalha and Sayed advocated for using the EB refinement method to identify the collision prone locations because of the inherent randomness in occurrences of collisions (Sawalha & Sayed, 1999). Based on this study, the technique to identify the collision prone zones is described below.

The predicted number of collisions and its variance for a specific location or zone are first estimated using the outcome of a suitable collision prediction model. The prior distribution of the imaginary reference population of that location is thus found. This prior distribution is assumed to follow a gamma distribution. The parameters of this prior distribution are α and β. These two parameters can be expressed as:

$$\beta = \frac{E(\Lambda)}{Var(\Lambda)} \text{ and } \alpha = \beta \text{ x } E(\Lambda) = \kappa \tag{6}$$

Then the observed collision data of this location and the dispersion parameter of the CPM (κ) are used to calculate the $EB_{\text{safety estimate}}$ and the variance of this $EB_{\text{safety estimate}}$ using equations (4) and (5). The posterior distribution of collisions of that location is thus found. This posterior distribution is also assumed to follow another gamma distribution with parameters α_1 and β_1. The parameters can be expressed by:

$$\beta_1 = \frac{EB_{safety\ estimate}}{Var\left(EB_{safety\ estimate}\right)} = \frac{\kappa}{E(\Lambda)} + 1 \text{ and } \alpha_1 = \beta_1 \text{ x } EB_{safety\ estimate} = \kappa + Observed\ data \tag{7}$$

The probability distribution function of this posterior distribution will be:

$$f_{EB_{safety\ estimate}}(\lambda) = \frac{\left[\frac{\kappa}{E(\Lambda)} + 1\right]^{(\kappa + Observed\ data)} \lambda^{(\kappa + Observed\ data - 1)} e^{-\left[\frac{\kappa}{E(\Lambda)} + 1\right]\lambda}}{\Gamma(\kappa + Observed\ data)} \tag{8}$$

The final step consists of comparing the values of the EB $_{\text{safety estimate}}$ of the location to the regional average or norm of zones with identical traits. Usually the 50^{th} percentile (P_{50}) of the prior distribution's probability density function, or the CPM prediction, [E (Λ)] is used as the specified norm. P_{50} can be calculated as:

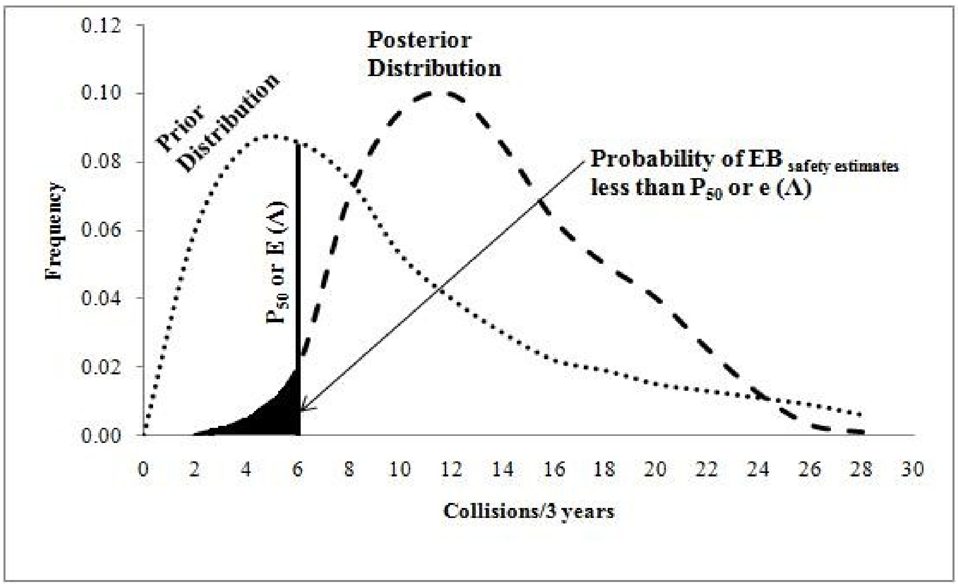
$$\int_0^{P_{50}} \frac{\left[\frac{\kappa}{E(\Lambda)}\right]^{\kappa} \lambda^{\kappa - 1} e^{-\left[\frac{\kappa}{E(\Lambda)}\right]\lambda}}{\Gamma\kappa} d\lambda = 0.5 \tag{9}$$

The zone is considered collision prone if there is a significant probability, δ (value of δ is usually 0.95 or higher), that the EB $_{\text{safety estimate}}$ exceeds the P_{50} or E (Λ) value. This comparison is usually calculated through use of the integral of the posterior probability density function in the range from 0 (zero) to P_{50}, or 0 to E (Λ). So, the zone is collision prone, if:

$$\left[1-\int_{0}^{E(\Lambda)\ or\ P_{50}}\frac{\left[\frac{\kappa}{E(\Lambda)}+1\right]^{(\kappa+Observed\ data)}\lambda^{(\kappa+Observed\ data-1)}e^{-\left[\frac{\kappa}{E(\Lambda)}+1\right]\lambda}}{\Gamma(\kappa+Observed\ data)}d\lambda\right]\geq\delta \tag{10}$$

Figure 5 shows a diagram to demonstrate the process of identification of collision prone zone graphically.

Figure 5 Identification of collision prone zones using EB safety estimate



Source: (Sawalha & Sayed, 1999)

2.5.4 Ranking of collision prone zones

After identifying the collision zones, they should be ranked to find out the most vulnerable collision prone location. This ranking order is also needed when adequate funding is not available to treat all the collision prone locations. Sawalha and Sayed (Sawalha & Sayed, 1999) have described two methods to calculate the rank. Usually, the difference between the EB $_{\text{safety estimate}}$ and the predicted collision frequency from CPM for the collision prone location, E (Λ), is calculated to achieve high cost effectiveness. This difference is called the potential collision reduction (PCR), and can be expressed by:

$$PCR = EB_{safety\ estimate} - E(\Lambda) \tag{11}$$

To assess the risk of collision involvement, the collision risk ratio (CRR) can also be calculated. A higher CRR indicates a higher risk of collision involvement. CRR can be calculated by:

$$CRR = \frac{EB_{safety\ estimate}}{E(\Lambda)} \tag{12}$$

Sawalha and Sayed (Sawalha & Sayed, 1999) have advocated for combining these two ranking criteria by sorting out collision prone locations based on the sum of these two ranks in ascending order. According to the requirement of the transportation authority, equal or different weight may be used for these two ranking methods.

2.5.5 Previous work on CPMs

During the last couple of years, several studies have been undertaken in North America to recognize the relevance of safety in the transportation planning process (De Guevara, Washington, & Oh, 2004; Hadayeghi et al., 2003; Hadayeghi, Shalaby, & Persaud, 2007; Lovegrove & Sayed, 2006a; Washington, Schalkwyk, Meyer, Dumbaugh, & Zoll, 2006). These research studies were mostly concentrated on providing methodology and safety planning decision-support tools, for assessing safety of alternative network planning initiatives proactively.

Recently, some studies incorporated variables representing transit characteristics to address transit as well as the whole transportation network safety issues. For example, Jovanis et al. (Jovanis, Schofer, Prevedouros, & Tsunokawa, 1991) analyzed 1800 collisions in Chicago to identify factors contributing to collisions involving transit buses. The analysis indicated that the gender of drivers did not contribute to collision occurrence and age had a negative impact on collision involvement. This study also found that experience with the transit agency was strongly associated with collision occurrence as drivers with 3 to 6 years of experience were significantly overrepresented in collisions. Moudon et al. (Moudon, Hess, & Matlick, 2004) analyzed pedestrian safety and developed a logistic regression model for pedestrian collision locations, relating the odds of a pedestrian collision to socioeconomic, demographic, and transit demand factors. This study found that bus stop usage, 24-hour traffic volume and the amount of building area in retail uses within one-quarter mile of the center of a 'Pedestrian Collision Locations' were positively related to pedestrian collisions. Cheung et al (Cheung, 2007) attempted to investigate the safety effects of the presence of transit on the arterial transportation network by developing arterial-level collision prediction models for collisions involving all motor vehicles (transit and non-transit). These models incorporated characteristics applicable to urban transit planning in Toronto, Ontario. The developed models suggested that average annual daily traffic, transit frequency, segment length, presence of on-street parking, and percentage of near-sided bus stops were associated with increased frequency of collisions, whereas percentage of far-sided stops and average stop spacing were linked with reduced collision frequency.

Shahla et al. (Shahla, Shalaby, Persaud, & Hadayeghi, 2009) presented the relationship between public transit service configurations and the overall safety performance of signalized intersections in Toronto, Ontario. This was a micro model with results that could be compared with macro models. The models showed that annual average daily traffic, public transit and pedestrian traffic volumes, turn movement treatments, and transit features (such as public transit stop location, mode technology, and availability of transit signal priority technology) were significantly associated with public transit–related collisions at signalized intersections. Intersections with public transit service also tended to experience more collisions than otherwise similar intersections.

The previous studies investigated the arterial road safety or overall intersection safety due to the presence of transit network. However, little work was found that showed the effects of various transit network elements on the safety situation of the overall zonal transportation network. Besides, some vital policy related transit elements such as average (normalized) transit trip length or average (normalized) transit ridership were not considered in these studies. Therefore, there is a need to have a clear understanding on how this transit-oriented society approach changes the zonal travel behaviour and affects the overall zonal transportation safety.

A literature search found no study on collision prediction model (CPM) that exclusively developed to investigate the safety impact of various goals/targets of a long term city transportation plan, although some previous studies developed zonal level collision prediction models those can be used in road safety planning applications. According to Saccomanno and Fu (Saccomanno, Fu, & Roy, 2007), in 1997 Nassar and Saccomanno created a micro model that established the long-term potential for collisions at a given location by using variables such as annual average daily traffic, length of road section, and traffic signals on road sections. This model, like others, was appropriate at the project or corridor level. Only a few studies developed models at a more aggregate zonal level which were appropriate for transportation planning. Hadayeghi et al. (Hadayeghi et al., 2003) have stated that Levine, in 1995, conducted one of these studies, consisting of a model that related collisions in zones in Hawaii to demographic characteristics. De Leur and Sayed (de Leur & Sayed, 2003) proposed a framework for proactive road safety planning, arguing the importance of applying road safety evaluation in planning process to avoid future unsafe situations. Lord (Lord, 2000) sought a tool that could estimate traffic collisions on digital transportation networks during the planning process. Hadayeghi et al. (Hadayeghi et al., 2003) developed a series of macro-level collision prediction models those would estimate the number of collisions in planning zones in the city of Toronto, Ontario, as a function of zonal characteristics.

Using a refined methodology with generalized linear regression, Lovegrove and Sayed (Lovegrove & Sayed, 2006a; Lovegrove & Sayed, 2006b; Lovegrove, 2006; Lovegrove & Sayed, 2007) developed 47 CPMs, each significantly associated with one or more of 22 community-based explanatory variables and clearly showed how to use those in planning levels. Lovegrove et al. (Lovegrove, Lim, & Sayed, 2010) also used these macro-level models to evaluate the road safety of the 2005-2007 Three Year Regional Transportation Plan for the Greater Vancouver Regional District (GVRD) and found that the 3-Year Plan would be safer than the base scenario.

2.6 Spatial effects on collision prediction models

Recent studies have found a significant effect of spatial correlation on the occurrences of all types of collisions (Aguero-Valverde & Jovanis, 2008; El-Basyouny & Sayed, 2009; Wang & Abdel-Aty, 2006). The following paragraphs discuss the background of this spatial correlation, the methodology to develop spatial models, and some previous work on collision prediction models where spatial effects have been incorporated.

2.6.1 Background of spatial correlations

The development of macro-level collision prediction models should consider four types of effects on the distribution of collisions (Aguero-Valverde & Jovanis, 2008; El-Basyouny & Sayed, 2009). First, Poisson variation is considered due to the random, discrete, and non-negative nature of collision data. Second, heterogeneity or extra variation or over-dispersion is considered because of the within-site effects which reflect the individual characteristics of each location. These two effects are generally considered in negative binomial models. But there are two more effects. The temporal correlation effect is partially removed by aggregating the collision data over multiple years (e.g. in this study collision data will be aggregated over a three year period from 1994 to 1996). But the spatial correlation effect, is hardly considered while developing the macro-level collision prediction models. Spatial effect is observed due to the fact that neighbouring sites typically have similar environmental and geographic characteristics, and ultimately results in a cluster that has similar collision occurrence.

2.6.2 Development of spatial models

Different approaches have been found from the previous studies that addressed the spatial correlations at zonal level (Aguero-Valverde & Jovanis, 2008; El-Basyouny & Sayed, 2009; Yuan, Yuan, & Huang, 2010). Based on the study of Aguero-Valverde et al. (Aguero-Valverde & Jovanis, 2008), El-Basyouny and Sayed (El-Basyouny & Sayed, 2009) have shown that first-order neighbouring structure can be used to sort out the neighbouring zones. This approach includes all zones that are directly connected with the zone being dealt with. The first order zones are directly connected with the zone under consideration, the second order zones are directly related with the first order neighbours, and the third order zones are directly connected with the second order neighbours. This order of neighbours is fully based on the topology of zones and does not account for the effect of distance between zones. Zones within a group are spatially correlated, and zones from different groups are assumed to be statistically independent.

Development of spatial models is somewhat similar to that of negative binomial models. El-Basyouni and Sayed (El-Basyouny & Sayed, 2009) have shown three different types of spatial models- Gaussian conditional autoregressive (CAR) models, MM models, and extended MM models. Gaussian CAR models are widely used for modeling spatial effects.

2.6.3 Previous work on spatial models

Limited numbers of research have been conducted incorporating the spatial effects on collision prediction models. The scope of those studies varies from states, counties, traffic analysis zones, census wards, etc. Huang et al. (Huang, Abdel-Aty, & Darwiche, 2010) described that most of the previous studies dealt with the road collisions to different explanatory variables. These variables mainly consists of disaggregated mileages of different road types, road density, intersection density, posted speed, vehicle miles traveled (VMT), volume-to-capacity ratios, area, population, households, age groups, land use, employment, income, etc.

In 1995, Levine et al. (Levine, Kim, & Nitz, 1995a) included spatial component in their analysis for the data from Honolulu, Hawaii, for motor vehicle collisions. They geocoded the collision data to the nearest intersection or ramp, and then calculated various spatial statistics. In another study, the same authors (Levine, Kim, & Nitz, 1995b) developed a spatial model at the census block level using weighted average of neighbouring locations. They examined the spatial correlation between trip generations and motor vehicle collisions for the City and County of Honolulu. However, the study considered a normal distribution of error structure of collision data.

Aguero-Valverde and Jovanis (Aguero-Valverde & Jovanis, 2006) developed county-level full Bayes hierarchical models with spatial and temporal effects using injury and fatal collision data for Pennsylvania for the period 1996 – 2000. They compared those FB models to traditional negative binomial (NB) models. They observed no evidence of spatial correlation in fatal collisions, but found significant spatial effects on the occurrences of injury collisions. The authors also provided a mechanism to reduce the effect of spatial correlation. They concluded that FB spatial models might be more useful where units are smaller and closer, and therefore, the probability of spatial correlation is higher.

Wang and Abdel-Aty (Wang & Abdel-Aty, 2006) investigated the temporal and spatial correlations among the collision data of 476 signalized intersections in Florida. They used generalized estimating equations with negative binomial link function. They adopted independent, exchangeable and auto-regression correlation structures, and found that the auto-regression structure has the maximum correlation value of 0.63 for any two successive intersections along a corridor. They also observed that the correlation between any two intersection decreases with the increase in spacing between them.

In another study in 2008, Aguero-Valverde and Jovanis (Aguero-Valverde & Jovanis, 2008) used a full Bayes hierarchical approach with conditional autoregressive method to assess the importance of including spatial effects on collision prediction models. They used data from 865 rural two lane segments of Centre County in Central Pennsylvania and part of District 2.0, and found that significantly high proportions of the total variability could be explained by the spatial correlations. They also found that some significant covariates of negative binomial models became non-significant after incorporating the spatial correlations.

El-Basyouny and Sayed (El-Basyouny & Sayed, 2009) investigated the spatial effects on the collision prediction models using the data of 281 road segments in Vancouver, British Columbia. The authors developed Poisson-log normal and negative binomial model, and compared those two models with three spatial models developed using three spatial modeling techniques- the Gaussian conditional auto-regressive, the multiple membership and the extended multiple membership approach. They found that the incorporation of spatial effects not only improved the goodness of fit, but also it could explain enough variation that one of the model covariates, presence of undivided cross-section, would deem non-significant. They also observed comparatively smaller coefficient value of AADT under spatial models. They argued that smaller value of the exponent of AADT indicated that not considering the spatial variables yielded a bias associated with model misspecifications.

Huang et al. (Huang et al., 2010) developed county-level Bayesian spatial model to examine the spatial effects on occurrence of collisions in Florida using a conditional autoregressive prior. They used the data from 67 counties and observed significant spatial correlations in occurrences of collisions across adjacent counties. They also found that the spatial distribution of collisions was associated with land use features, population density, travel activities, and transportation infrastructure such as highway density.

In summary, the few studies discussed above observed significant spatial effects in collision prediction models. But further investigation is needed as very few studies have dealt with zonal-level spatial models.

2.7 City transportations plan: TransLink's *Transport 2040* plan

Metro Vancouver is Canada's gateway to Asia-Pacific economies and includes the nation's third largest urban centre. TransLink is Metro Vancouver's regional transportation authority. TransLink has adopted a thirty-year long transportation plan for Metro Vancouver with a view to creating and sustaining a transportation system that meets the needs of residents, business, and goods movers, in a manner that protects the environment and supports the economic and social objectives of the region (TransLink, 2008). The plan is named as *'Transport 2040'* which contains six transportation goals along with four strategies to achieve those goals. To achieve these goals, TransLink intends to significantly reduce residents' dependence on cars through massive investment in various forms of public transit and active forms of transportation, including biking and walking. TransLink also encourages the inhabitants of Metro Vancouver to use autos more appropriately by carpooling and car sharing. TransLink has increased transit ridership by 38 per cent between 1998 and 2007 while population has increased by 13 percent (City of Vancouver, 2005). The city authority now wants to see more change in people's travel behaviour and wants a fully transit-oriented city. Among the six goals, the first five goals are related to the transportation network. Those six goals are mentioned below (TransLink, 2008):

- Goal 1: Greenhouse gas emissions from transportation are aggressively reduced, in support of federal, provincial and regional targets
- Goal 2: Most trips are by transit, walking and cycling
- Goal 3: The majority of jobs and housing in the region are located along the Frequent Transit Network
- Goal 4: Traveling in the region is safe, secure, and accessible for everyone
- Goal 5: Economic growth and efficient goods movement are facilitated through effective management of the transportation network
- Goal 6: Funding for TransLink is stable, sufficient, appropriate and influences transportation choices

'Transport 2040' also sets out four strategies to meet the goals mentioned above in a way that protects what the residents of Metro Vancouver value the most (TransLink, 2008). The strategies are presented below:

- Strategy 1: Make early investments that encourage development of communities designed for transit, cycling, and walking
- Strategy 2: Optimize the use of the region's transportation assets and keep them in good repair
- Strategy 3: Build and operate a safe, secure, and accessible transportation system

- Strategy 4: Diversify revenue sources and pursue new and innovative ways to fund transportation

2.8 Definitions of statistical terms and tools

The occurrence of collision can be considered a random event with respect to time or distance. Collisions can be counted and they are always non negative. The occurrence of collisions usually follows either Poisson or negative binomial distribution. A collision is, in theory, the result of a Bernoulli trial. Each time a vehicle enters an intersection, a highway segment, or any other type of entity (a trial) on a given transportation network, it will either collide or not collide. For purposes of consistency a collision is termed a "success" while failure to collision is a "failure". For the Bernoulli trial, a random variable, defined as X, can be generated with the following probability model: if the outcome *w* is a particular event outcome (e.g. a collision), then X (*w)* = 1 whereas if the outcome is a failure then X (w*)* = 0. To see whether the collision data from a particular site follow random behavior or deterministic in nature, statistical analysis is the only available tool.

It is necessary to be familiar with the statistical terms related to collision studies and computation of collision prediction model. Therefore in this section, the important statistical terms are described in brief.

Mean: A descriptive statistic, measuring central tendency which is calculated by dividing the sum of a set of scores by the number of scores

Standard Deviation: A measure of dispersion within a set of data, calculated from the square root of the variance, to give a value in the same range as raw scores. The standard deviation is the spread of scores around the mean of the sample.

Normal Distribution: Symmetrical division of measuring values around an average value is the normal distribution. The number of measuring values is at a maximum around the average value and decreases away from the average value. The division can be presented in a so-called normal distribution curve.

Poisson Distribution: A frequency distribution giving the probability that *n* points (or events) will be (or occur) in an interval (or time) *x*, provided that these points are individually independent and that the number occurring in a subinterval does not influence the number occurring in any other non overlapping subinterval.

Poisson distribution assumes that mean is equal to variance.

Negative Binomial Distribution: It is the distribution arising from an experiment consisting of a sequence of independent trials, subject to several constraints.

Negative Binomial Distribution takes care of over dispersed data and assumes that mean is less than variance.

Correlation: A procedure which measures the strength of the relationship between two (or more) sets of measures which are thought to be related.

Regression: The general process of predicting one variable from another by statistical means, using previous data.

Generalized Linear Model (GLM): It is a flexible generalization of ordinary least squares regression. It relates the random distribution of the measured variable of the experiment (the *distribution function*) to the systematic (non-random) portion of the experiment (the *linear predictor*) through a function called the link function.

Link Function: The link function provides the relationship between the linear predictor and the mean of the distribution function.

Confidence Level: The probability used to describe the degree of certainty, e.g., a 95% confidence level.

t-Test: A parametric statistical test of the difference between the means of two samples.

Analysis of Variance (ANOVA): It is a statistical procedure which allows the comparison of the means and standard deviations of three or more groups in order to examine whether a significant relation exists between variables. It differs from the t-test as it can test for differences among many groups, not just two groups.

Degrees of Freedom: The term degree of freedom is used to denote the number of independent comparisons which can be made between the members of a sample.

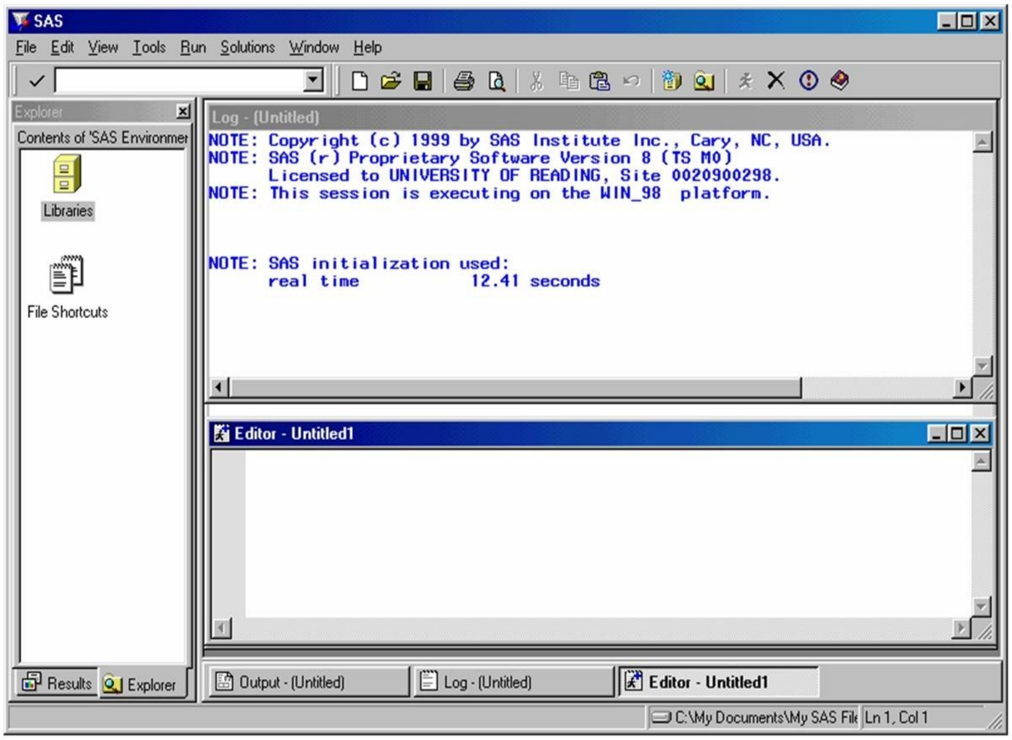
Coefficient of Correlation (R^2): A statistic value showing the degree to which two variables are related. It can vary from -1 (perfect negative correlation) through 0 (no correlation) to +1 (perfect positive correlation).

R_α^2**:** It is one kind of correlation coefficient and is expressed by $R_\alpha^2 = 1-(\alpha/\alpha_{max})$, where α = dispersion parameter estimated in the negative binomial model and α_{max} = dispersion parameter estimated in the same negative binomial model with only an intercept term and a dispersion parameter.

Chi-Squared Test: The appropriate test to investigate independent frequency counts taken from a sample is the Chi-Squared test. Pearson's chi-square test is the best-known of several chi-square tests.

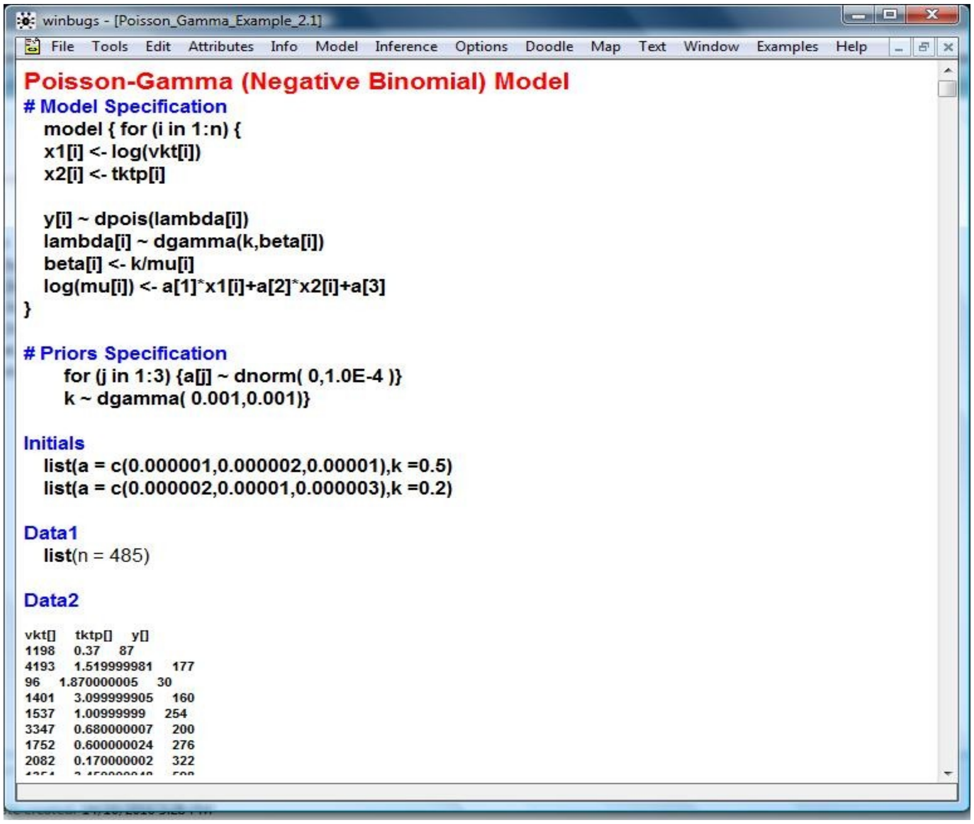
Statistical Software- SAS: SAS (Statistical Analysis System) is based on the use of an enhanced text editor for program and data input. It is a vast package, with extensive facilities for data management and statistical analysis. SAS comprises integrated modules that allow the graphical display and the analysis of data in areas including those of statistics, econometrics and operational research. It is suitable for applications development, data management and reporting. Commands are entered and amended using the SAS editor and are stored in memory. The resulting job can then be saved and/or submitted for processing. SAS includes SAS/ANALYST which uses a spreadsheet data entry window and aids to basic data analysis. The default image is as shown in Figure 6.

Figure 6 Default SAS window



Statistical Software- WinBUGS: It is the Windows interface of Open-BUGS. WinBUGS is used to sample the posterior distribution as well as to estimate the parameters of any regression model using Bayesian analysis. It can do practical Markov chain Monte Carlo (MCMC) computations for a wide variety of Bayesian models. Figure 7 shows the default image of WinBUGS.

Figure 7 Default WinBUGS window



2.9 Summary

The literature review presented in this chapter provided some basic background information for the topic of traffic safety and issues related with collision databases, model developments and identification of collision prone locations. The literature review also discussed that the conventional linear regression approach is not appropriate for the analysis of collision data. The negative binomial distribution is considered to be more appropriate and was used in the majority of the previous studies since it allows the variance to be larger than the mean. The methodology, background, development and uses of zone-level spatial collision prediction models, and the relevant previous studies are also discussed. Besides, the TransLink's *Transport 2040* plan with its goals and strategies is also mentioned here.

3 Methodology

3.1 Introduction

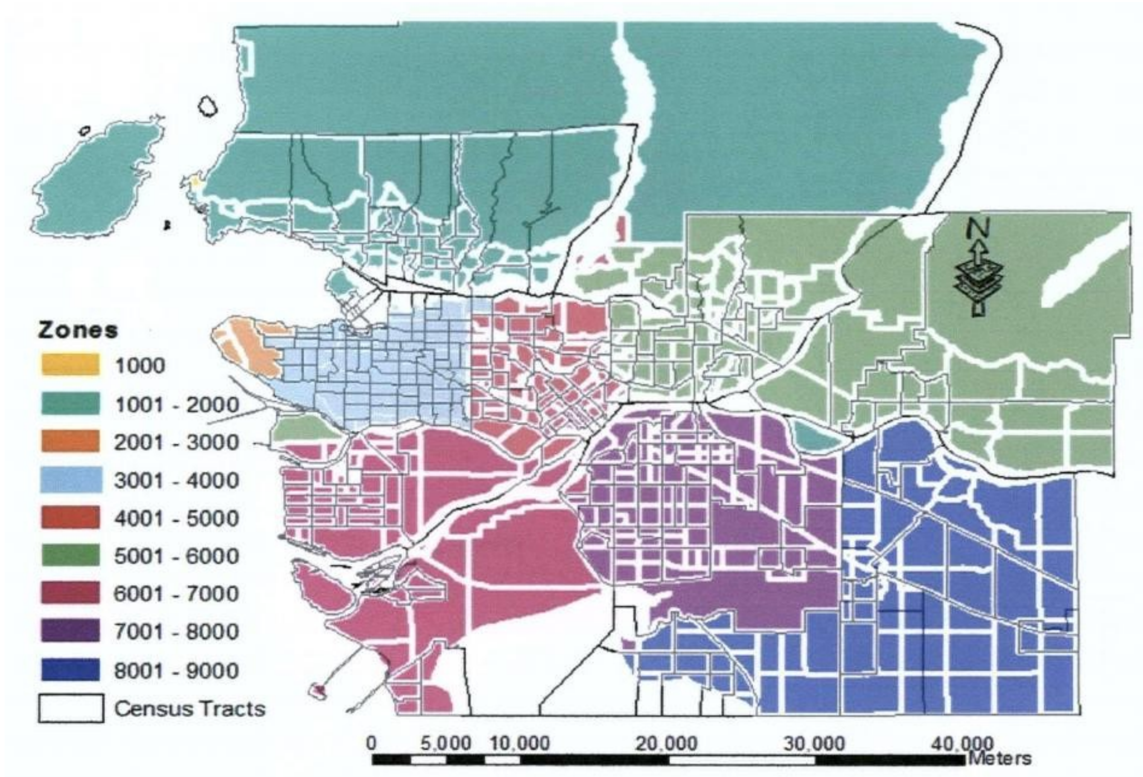
This chapter contains the data extraction approach and the techniques of developing collision prediction models. It also discusses the types of models, the background of application based models and their relationships with the goals of city transportation plan, model form and error structure, variable selection, approaches of the development of collision prediction models with and without spatial effects, goodness of fit statistics and refinement of the developed models by outlier analysis.

3.2 Geographic scope and data extractions

The data extracted for the development of collision prediction models relate to the geographic area shown in Figure 8 and correspond to the year 1996, unless otherwise stated. The area comprises the Metro Vancouver (formerly known as the Greater Vancouver Regional District-GVRD) in the Province of British Columbia. The land area of the Metro Vancouver is roughly 3,000 square kilometres, and is comprised of 21 member municipalities. The population of this area was 2 million in the year 1996 (Census Canada., 1996). Lovegrove has observed that most of this population has traditionally lived and worked in the western, urban communities clustered around the Central Business District (CBD) (Lovegrove, 2006).

Aggregation of the data into representative zones was done to enable zone-level CPM development in accordance with the objectives of this thesis. The aggregation unit was based on the TAZs used in the GDRD's *Emme/2* transportation planning model. These GVRD's Emme/2 TAZ sizes worked satisfactorily in ensuring adequate data points in each zone. Besides the zone boundaries have been chosen in a way that these overlap as closely as possible with the boundaries of census tracts and municipalities (Lovegrove, 2006).

Figure 8 Metro Vancouver Emme/2 model TAZ and census tract boundaries in 1996



Source: (Lovegrove, 2006)

Similar to Lovegrove and Sayed (Lovegrove & Sayed, 2006a), the data used in this study were extracted from three main sources:

- The first source, TransLink, provided geocoded files of land use, road network, and zone – census tract boundaries for 485 zones (TransLink, 2003). It also provided spreadsheets converted from Emme/2 travel demand data files based on a 1996 morning peak hour scenario.
- The second source, Census Canada provided sociodemographic, employment, and mode split data for each zone for the census year 1996, chosen as the base year because the 2001 census occurred during a transit strike in Vancouver (rendering its data unusable for the purposes of this research) (Census Canada., 1996).The third source, the Insurance Corporation of British Columbia (ICBC) provided geocoded files for over 250 000 collision claims in the GVRD for the years 1996, 1997, and 1998 (ICBC, 2003). Three years of the collision data were used to reduce the randomness in the data and to quantify

collision types for which data are typically sparse (e.g., bicycle or pedestrian collisions). The auto-mobile insurance companies in most jurisdictions in North America are privately owned, and obtaining claims data would be difficult. In British Columbia, however, ICBC, a public automobile insurance company, handles most insurance claims and centrally warehouses the data. The availability of actual geocoded claim data (in addition to police report files) from ICBC is considered a great advantage in overcoming many of the traditional unreported, unattended, and (or) incomplete collision data problems associated with local municipal collision databases (Lovegrove & Sayed, 2006a).

A total of 58 explanatory variables were available that could be included in models development. Previously, Lovegrove and Sayed (Lovegrove & Sayed, 2006a) used some of these variables to develop macro-level CPMs. A correlation matrix was created to quantitatively express the one-to-one relationships among all variables- dependent and explanatory. Among the correlated variables, scatter diagrams showed that the collision data and relevant explanatory variables had non-linear relationships. The correlation matrix was also examined to check for the existence and the nature of the co-linearity relationship among explanatory variables.

Out of the 58 available variables, 30 variables were chosen for the development of CPMs based on the following criteria: (Table 3 shows the variable definition and descriptive statistics)

- relevance to planning and transit
- availability at the planning stage
- relevance to collisions as demonstrated in earlier research
- easiness of forecasting

Table 3 Variable definition and descriptive statistics (485 zones)

Variable	Description	Unit	Avg.	Min.	Max.	St Dev.
ALKP	Percentage of arterial lane km per total lane km	%	31.73	4	85	15.14
AUTOP	Percentage of commuters driving or car pooling	%	73.99	0	95.9	20.43
AVHKT	Average auto trip length	auto-km/person	19.07	0	2253.64	127.85
BIKP	Percentage of commuters biking	%	1.51	0	11.4	1.76
BSD	No of bus stops	no.	15.98	0	51	9.92
BSKD	No of bus stops per unit area	no./Ha	0.16	0	2.6	0.21
BSKT	Ratio of Number of bus stops to transit-km travelled	no./bus-km	1.07	0	22	1.76
BTKT	Average transit trip length	bus-km/person	0.72	0	76.67	4.59
BUSP	Percentage of commuters using transit	%	14.25	0	40.5	8.15
CLKP	Percentage of collector lane km per total lane km	%	29.84	15	35	3.2
EMP	Employed workers in total labour force [(Employed population/ population aged between 15-65)*100]	%	90.89	67	98	4.5
HBNCP	Ratio of number of hard curves (>45^0) to number of conflicts	%	12.79	0	43	9.24
I3WP	Percentage of 3-way intersections to all types of intersections	%	52.19	0	100	23.64
IALP	Percentage of arterial-local intersections to all types of intersections	%	17.46	0	67	13.63
INCD	Average income	$100,000	0.77	0	6.76	0.47
INTD	Number of intersections per unit area	no./Ha	0.46	0.02	2.12	0.28
INTKD	Number of intersections per lane km	no.	1.03	0.1	3.41	0.32
LLKP	Percentage of local lane km per total lane km	%	38.34	0	64	12.69
NHD	Number of households per unit area	units/Ha	12.58	0	121.8	15.2
PARP	Participation in labour force [(population aged between 15-65/total population)*100]	%	61.17	19	80	8.87

Variable	Description	Unit	Avg.	Min.	Max.	St Dev.
PEDP	Percentage of commuters walking	%	7.16	0	50.3	10.18
POPD	Population density	no./Ha	34.86	0	280.1	35.57
RTKT	Average transit ridership (occupancy)	person/bus-km	12.17	0	96.07	14.01
RVHKT	Average auto ridership (occupancy)	person/auto-km	0.59	0	46.16	2.16
SBNCP	Ratio of number of soft curves (<45^0) to number of conflicts	%	11.42	0	71	8.96
SIGD	Number of signalized intersections per unit area	no./Ha	0.03	0	0.63	0.07
TCD	Number of commuters per unit area	no./Ha	15.44	0	145	19.31
TKTP	Ratio of transit-km traveled to vehicle-km travelled [(TKT/VKT)*100]	%	0.96	0	8.1	1.01
VKT	Vehicle (transit and auto) km travelled (extracted from Emme/2 output)	km	3566.13	0	17579	2585.25
WKGD	Workers per resident (worker/population)	no./pop [fraction]	0.31	0	5.49	0.73
T3	Number of total collisions (transit and non-transit) over 3 years	no.	516.96	0	2885	508.68
S3	Number of severe collisions (fatal and injury for transit and non-transit) over 3 years	no.	113.44	0	848	128.74
PDO3	Number of property damage only collisions (transit and non-transit) over 3 years	no.	403.52	0	2243	392.94

It should be noted that many variables have a minimum of zero, but that would not affect the model parameter estimation as out of total 485 zones, only a few zones have variables with a zero value. The descriptive analysis of the collision and explanatory variables reveals some interesting facts and trends about the GVRD transportation system elements, urban travellers' socio-demographics and the road network collisions, as follows:

- Transit-km traveled (TKT) is highly and positively correlated (0.53) with auto-km traveled (VHKT) which is logical as TKT tends to grow where VHKT or vehicle-km traveled (VKT) grows and vice versa.

- Workers per resident (WKGD) and ratio of transit-km traveled to vehicle-km travelled (TKTP) are positively correlated (0.5), meaning that if the average number of workers per residents increases, the demand and usage for transit also increases.
- Correlation between percentage of commuters biking (BIKP) and percentage of commuters using transit (BUSP), and correlation between BIKP and transit kilometre traveled (TKT) are moderate and positively associated. These outcomes suggest that Bike-and-Ride initiative should be complemented with more transit coverage. For the shift towards a more sustainable transportation system, TransLink and the GVRD are trying to encourage people to use hybrid set of transport modes, with more emphasis on non-motorized modes.
- Correlation between percentage of commuters using transit (BUSP) and percentage of commuters using autos (AUTOP) is negatively associated. Correlation between percentage of commuters biking (BIKP) and AUTOP is also negative. However, these two correlation values are quite small. So, providing more transit and bike friendly network may make a moderate modal shift from auto to transit and auto to bike, but it is difficult to make such conclusions based on the data used in this study.
- Correlation between percentage of commuters walking (PEDP) and ratio of transit-km traveled to vehicle-km travelled (TKTP) is 0.54. So, providing walk friendly environment will also encourage the percentage of transit usage. The result satisfies the intuition because of the fact that almost every transit trip needs some amount of walking to and from the bus stops as bus trip usually is not a 'door-to-door' service.
- Number of bus stops per total lane kilometre (BSKP) is positively correlated with ratio of transit-km traveled to vehicle-km travelled (TKTP). It satisfies the general requirement of more bus stops if the frequency and amount transit network are high among the whole network. Correlation between BSKP and average number of working people per residents (WKGD) is also positive and the value is moderate which implies that the more WKGD, the higher demand of more bus stops.
- Increase in family size (FS) decreases percentage of commuters using transit (BUSP) as well as other sustainable mode, such as percentage of commuters walking (PEDP). The correlation between FS and PEDP is -0.66. So, obviously FS is positively and

significantly correlated with percentage of commuters using autos (AUTOP), and the correlation value is 0.52.

3.3 Modeling approach and methodology of CPMs

This section deals with the types of application based collision prediction models with corresponding explanatory variables, the background of the classifications of variables, CPMs and their relationships with the goals of city transportation plan, model form and linearization of the model form, need for a non-normal error structure, process for variable selection with order, method to develop CPMs, goodness of fit statistics and refinement of the developed models by outlier analysis.

3.3.1 Types of CPMs

2 different classes of collision prediction models were developed. The first class related the overall roadway safety to the transit elements and developed 'Transit Reliant CPM'. The corresponding explanatory exposure variable in this model type is VKT and the explanatory exponential variables are RTKT or RVHKT, BTKT or AVHKT, BSKT or BSD or BSKD. The major North American cities now aim at changing in people's travel behaviour which will result in more reliance on transit. This policy targets to change the average transit ridership, average transit trip length, transit network, etc. so that a transit oriented society may be developed. The transit reliant CPM could evaluate the safety consequences of the changes in these transit related elements.

The second class of model developed 'Application Based CPMs' that incorporated the interaction between the collision frequency of all traffic (transit and non-transit) of different type and severity with other variables those are related to achieving different goals/targets of a city transportation plan. A total of 7 types of application based models were developed. The models were named based on the target to be achieved and the corresponding variables for each type were distributed based on the relevance with achieving that target. The model types and corresponding variables are shown in Table 4.

Table 4 Different types of application based CPMs and corresponding variables

Types	Model Name	Corresponding Variables
1	Change in Exposure	VKT
2	GHG Emission Reduction/Sustainable Environment	VKT; TKTP, BUSP, PEDP, BIKP
3	Change in Modal Share	VKT; AUTOP, BUSP, PEDP, BIKP
4	Change in Network Alignment	VKT; HBNCP, SBNCP, INTKD, I3WP
5	Economic Growth	VKT; WKGD, INCD, PARP or EMP
6	Change in Accessibility/Mobility	VKT; ALKP or CLKP or LLKP, SIGD, I3WP, IALP
7	Urban Sprawl	VKT; POPD or NHD or TCD, INTD, SIGD, IALP, WKGD

3.3.2 Application based CPM background

The '*Transport 2040*' plan of TransLink sets out strategies to meet the transportation challenges GVRD may face in the next thirty years. This plan identifies six key goals, and proposes strategies to reach those goals having more reliance on people-oriented and sustainable modes, including public transit, biking and walking (TransLink, 2008).

Exposure plays the vital role in the occurrence of all types of collisions. Many previous studies have found a strong and positive relationship between the number of all types of collisions and exposure expressed in either the total number of vehicles (AADT) or total vehicle kilometre travelled (VKT) (Cheung, Shalaby, Persaud, & Hadayeghi, 2008; Lovegrove & Sayed, 2006a; Lovegrove, 2006; Lovegrove & Sayed, 2007). Every city has a target to reduce the exposure by reducing the amount of autos and making the auto users use transit. As a result, same number of people will move but the amount of AADT or VKT will be less which will make the transportation network safer (Société de Transport de Montréal, 2008).

The *Transport 2040* plan aims to aggressively reduce the GHG emission (TransLink, 2008). It is the first of the six goals. It can be achieved by reducing the proportion of autos in total vehicle stream. As a result, the percentage of transit will be increased and at the same time the modal share of the active form of transportation- walking and biking will be increased.

The second goal of the *Transport 2040* plan is to have a drastic modal shift from auto to transit, walking and biking (TransLink, 2008). As almost every transit trip needs some amount of walking or biking, this modal shift is expected to be distinctively found in a less modal share in auto mode and hence a higher share in transit, walking and biking mode.

Changing the existing network alignment is a common phenomenon in a live city and it depends on the projected or current usage of the zonal road network. Usually any transportation plan aims at aligning the road network in line with the corridor having the maximum density of jobs and housing. This type of corridor is a prerequisite of a successful transit route. According to the third goal of the *Transport 2040*, the transit network will be aligned to the corridor having maximum job and housing density (TransLink, 2008).

The prime target of each transportation plan is to maintain the movement of people and goods in the most efficient way so that the economic growth can be enhanced. TransLink's fifth goal is to have an economic growth in GVRD through efficient management of the transportation network (TransLink, 2008).

Accessibility and mobility of a roadway depends on the functional type of that road. A part of TransLink's fourth goal is to make the transportation network more accessible (TransLink, 2008). Arterial road has higher mobility than collector or local road as arterial enjoys higher speed. But it has the least accessibility among these three. To increase accessibility in the arterial road, more intersections will be needed. Again, intersections situated on the arterial are usually signalized. Local-arterial intersections and percentage of 3-way intersections play an important role in roadway safety, as found by Lovegrove et al (Lovegrove & Sayed, 2006a).

The target of urban containment is to increase the number of inhabitants in a zone keeping the zonal boundary unchanged. Cities are adopting this initiative to have a greater density of people especially around the arterial or transit routes. TransLink third goal is partially related to urban containment (TransLink, 2008). Because of this urban containment, the amount of intersections will be increased as more local roads will be developed. As a result the percentage of local-arterial intersections will also be increased. Besides, some unsignalized intersections will be signalized due to more demand of vehicle.

3.3.3 CPM form

Total, severe and property damage only collision prediction models for all traffic (transit and non transit) were developed in this studies. Usually two main options for estimating the model parameters are used:

- The conventional linear regression approach, which assumes a normal distribution error structure; and
- The GLM approach, which assumes a non-normal error structure (usually Poisson or negative binomial) (Sayed & Rodriguez, 1999).

Several studies have shown that collision data are random, discrete, non negative and over dispersed (Buyco & Saccomanno, 1988; Hauer et al., 1988; Jovanis & Chang, 1986; Miaou & Lum, 1993). So the conventional linear regression approach is not appropriate to model traffic collisions (Sawalha & Sayed, 2006). As the GLM has the advantage of overcoming the limitations associated with the use of conventional linear regression (Hauer et al., 1988; Sawalha & Sayed, 2006; Sayed & Rodriguez, 1999), it is used in this study. Sawalha and Sayed (Sawalha & Sayed, 2006) have suggested that the mathematical form for any CPM should satisfy two conditions. First, the model should yield logical result, indicating that the model must not predict any negative number of collisions and it must predict zero collision frequency when the exposure value is zero (in this study the exposure variable is vehicle kilometre travelled or VKT). The next condition is that there must exist a known link function to linearize the model form for the purpose of coefficient estimation. These two conditions are met if a model form is chosen in which there exists the product of powers of the exposure variable multiplied by an exponential incorporating the remaining explanatory variables. Therefore, the recommended model form for the expected collision frequency at zone can be mathematically expressed as:

$$E(Y) = a_0 * VKT^{a_1} * \exp\left[\sum_i b_i X_i\right] \tag{13}$$

where E (Y) is the predicted collision frequency (over 3 years); a_0 is intercept, a_1, and b_i are model parameters; VKT is the external exposure variable and X_i is the explanatory variable (Hadayeghi et al., 2003; Sawalha & Sayed, 2006). By definition, the number of collision becomes zero only if the exposure variable, VKT, becomes zero.

In the modeling process, a log-linear transformation is made (refer to equation 14). This is the reason the model is called generalized linear model even though the final model form is multiplicative.

$$\ln\left[E(Y)\right] = \ln\left\{a_0 * VKT^{a_1} * \exp\left[\sum_i b_i X_i\right]\right\}$$

$$\text{or, } \ln\left[E(Y)\right] = \ln(a_0) + a_1 \ln(VKT) + \left[\sum_i b_i X_i\right] \tag{14}$$

3.3.4 Error structure

It has been mentioned in the earlier section that the GLM approach can consider the error structure of collision occurrences either Poisson or negative binomial. According to Sawalha and Sayed (Sawalha & Sayed, 2006), if Y is considered the random variable expressing the collision frequency at a given zone during a specific time period (in this study the time period covered 3 years from 1994 to 1996), and let y be a certain realization Y, the mean of Y, shown by Λ, is itself considered a random variable (Kulmala, 1996). For Λ = λ, Y is considered to follow Poisson distribution with parameter λ (i.e. mean = variance = λ). Therefore, the density function is

$$P(Y = y \mid \Lambda = \lambda) = \frac{\left(\lambda^y e^{-\lambda}\right)}{y!} \tag{15}$$

The corresponding mean and variance are

$$E(Y \mid \Lambda = \lambda) = \lambda \tag{16}$$

$$Var(Y \mid \Lambda = \lambda) = \lambda \tag{17}$$

As each zone has its own regional characteristics with a unique mean collision frequency, Λ, Hauer et al (Hauer et al., 1988) have shown that for an imaginary group of locations with similar characteristics, Λ follows a gamma distribution with parameters κ and κ/μ. Here, κ is the shape parameter. The corresponding density function is

$$f_\Lambda(\lambda) = \frac{\left(\kappa/\mu\right)^{\kappa} \lambda^{\kappa-1} e^{-\left(\kappa/\mu\right)\lambda}}{\Gamma\kappa} \tag{18}$$

with a mean and variance of

$$E(\Lambda) = \mu \tag{19}$$

$$Var(\Lambda) = \frac{\mu^2}{\kappa} \tag{20}$$

Sawalha and Sayed (Sawalha & Sayed, 2006) have shown that the distribution of Y around E (Λ) = μ ultimately follows negative binomial distribution with the following density function, mean and variance respectively:

$$P(Y = y) = \frac{\Gamma(\kappa + y)}{\Gamma(\kappa) y!} \left(\frac{\kappa}{\kappa + \mu} \right)^{\kappa} \left(\frac{\mu}{\kappa + \mu} \right)^{y} \tag{21}$$

$$E(Y) = \mu \tag{22}$$

$$Var(Y) = \mu + \frac{\mu^2}{\kappa} \tag{23}$$

The last two equations (equations 22 and 23) show that the expected value of collision frequency is usually less than its variance, indicating the fact that collision data are generally over-dispersed. The only exception is when $\kappa \to \infty$, only in which case the distribution of Λ is concentrated at a point and the negative binomial distribution becomes identical to the Poisson distribution (Kulmala, 1996). In this study, the maximum likelihood estimate of this shape parameter, κ, will be determined by an iterative process in statistical analysis software - SAS.

3.3.5 Model development

The estimation of model parameters was based on the methodology showed in Sawalha et al. (Sawalha & Sayed, 2006). First, model parameters were estimated assuming that the error structure followed Poisson distribution. Then the dispersion parameter (σ_d) was calculated using the following formula (McCullagh & Nelder, 1989)

$$\sigma_d = \frac{Pearson\ \chi^2}{n - p} \tag{24}$$

where, n is the number of data points (observations), p is the number of model parameters, and Pearson χ^2 is expressed by

$$Pearson\ \chi^2 = \sum_{i=1}^{n} \frac{\left[y_i - E\left(Y_i\right)\right]^2}{Var\left(Y_i\right)} \tag{25}$$

where, y_i is the observed number of collisions in zone i; $E(Y_i)$ is the predicted number of collisions for the zone i as obtained from the CPM; and Var (Y_i) is the variance of the observed collisions in zone i. Pearson χ^2 statistic follows the χ^2 distribution with n-p-1 degrees of freedom. If the value of σ_d was found to be significantly greater than 1.0, the collision data were over dispersed, and a further analysis with a negative binomial error structure was done. In this study, the parameters were estimated by maximum-likelihood (Hauer et al., 1988) process in the statistical analysis software SAS.

3.3.6 Variable selection

For any type of model, first all the corresponding variables of that particular type were used to get the t-statistics and the variables were then arranged in an ascending order of t-values. Then forward stepwise procedure was applied by which the variables were added to a model one by one, as described in (Sawalha & Sayed, 2006). In this procedure, each model was developed by one exposure variable, and then subsequent exponential variable was added at a time based on the change in model fit due to the added variable. The exposure variable ensured that all models met the "zero exposure = zero collision risk" prediction principle. The decision to retain a variable in the model was based on several criteria. First, the t-statistic of the parameter estimate (equal to the parameter estimate divided by its standard error or by having a square root of Wald statistic) was significant at the 95% confidence level (>1.96). Second, the addition of the variable to the model caused a significant drop in SD at the 95% confidence level (>3.84) (Sawalha & Sayed, 2006). Lastly, the variable showed little correlation with any of the other independent variables of that model type. For example, ALKP is highly correlated with CLKP (-0.99) and LLKP (-0.79). Hence, only one of these three variables could be included in any one CPM.

3.3.7 Goodness-of-fit measures

Three statistical measures were applied to check the goodness of fit of the developed models. According to McCullagh and Nelder, the Pearson χ^2 and Scaled Deviance were used to assess the model goodness-of-fit (McCullagh & Nelder, 1989). The formula for the Pearson χ^2 statistic has been defined before (model development section). The scaled deviance (SD) measured twice the difference between the maximized logarithm-likelihoods of the studied model and those of the full or saturated model. The full model has as many parameters as the number of observations, so the model fits the data perfectly. Therefore, the full model, which possesses the maximum logarithm-likelihood achievable under the given data, provides a baseline for assessing the goodness-of-fit of an intermediate model with p parameters. The SD for negative binomial data distribution is as follows (McCullagh & Nelder, 1989):

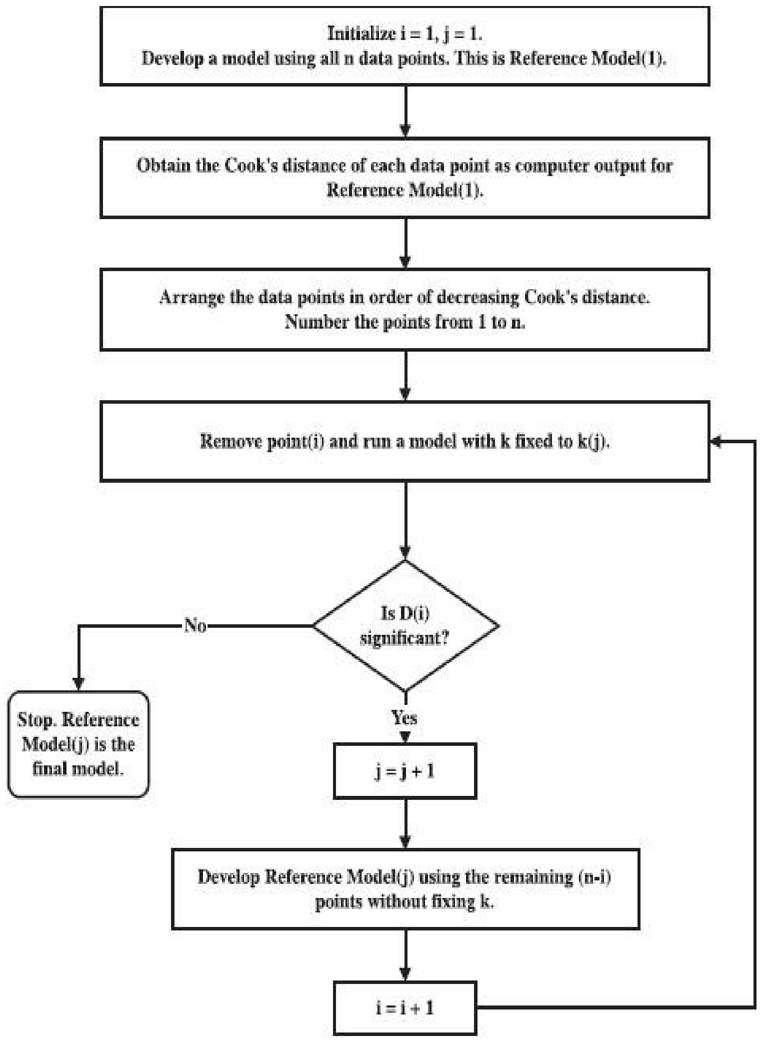
$$SD = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{E(Y_i)}\right) - (y_i + \kappa)\ln\left(\frac{y_i + \kappa}{E(Y_i) + \kappa}\right)\right] \tag{26}$$

where, κ is the shape parameter. Both SD and the Pearson χ2 have χ2 distributions for normal theory linear models, but they are asymptotically χ2 distributed with n-p-1 degrees of freedom (df) for other distributions of the exponential family. For a statistically significant model, both the Pearson χ^2 and SD value must be less than Critical $\chi^2_{(\text{significance, df})}$. At the same time, the t-statistics of all parameters should be significant at the 95% level of confidence (Lovegrove & Sayed, 2006a).

3.3.8 Outlier analysis

Outliers are data points having unusual or extreme values in comparison to the rest of the data. Outliers may be caused by irregularities or errors that occur during data collection or recording, or when the data are genuinely different from the rest (Sayed & Rodriguez, 1999). In this study, the outlier analysis was performed via the Cook's distance to identify extremely influential outliers. The larger the Cook's distance value, the stronger its influence on the model. The outlier analysis was based on the method described in Sawalha and Sayed (Sawalha & Sayed, 2006). First, the data were sorted in descending order according to the Cook's distance values. Then, in a stepwise procedure, the points with the largest cook's distance values were removed, assessing the change in the scaled deviance (SD). Points causing a significant drop in SD were extremely influential outliers. As the scaled deviances of two negative binomial models with different κ value cannot be compared directly, this process of outlier analysis required that the value of κ for the negative binomial model with n data points should be kept the same for the negative binomial model with n-1 data points. Then the difference in SD was compared with $\chi^2_{(\text{significance, df})}$ to assess whether the removed data point was an extremely influential outlier. Based on the study of Sawalha and Sayed ((Sawalha & Sayed, 2006)), the following figure (Figure 9) shows the procedure to detect and remove outliers.

Figure 9 Detection and removal of outliers from NB models



Source: (Sawalha & Sayed, 2006)

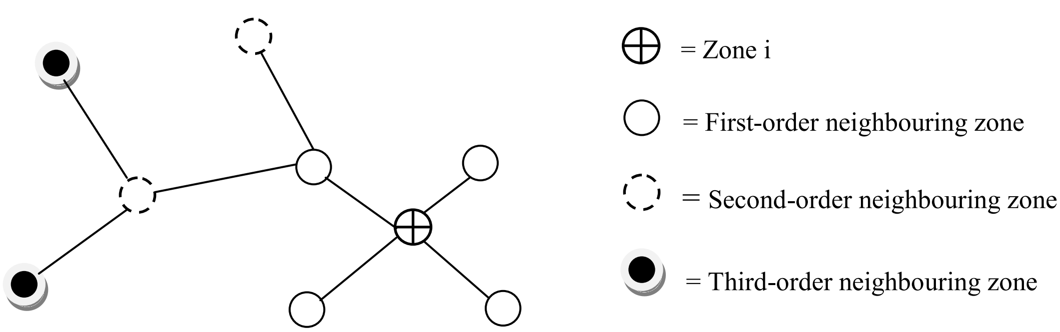
3.4 Spatial Poisson-gamma CPMs

Spatial Poisson-Gamma CPMs investigated the spatial effects on the developed application based collision prediction models. The following paragraphs of this section describe the methods of data preparation for spatial analysis, form and error structure of spatial models, Gaussian conditional autoregressive models development, parameter estimation techniques using full Bayes approach, selection of prior distribution and initial values, goodness of fit statistics, and techniques of comparing negative binomial and spatial models.

3.4.1 Data preparation

In order to explore the spatial effects on collisions, data of 485 zones of Metro Vancouver (formerly known as Greater Vancouver Regional District- GVRD) were used in this study. A map was obtained using the GIS software from which the adjacent zones of each zone were identified. Different approaches were found from the previous studies that addressed the spatial correlations at zonal level (Aguero-Valverde & Jovanis, 2008). Based on the study of Aguero-Valverde et al. (Aguero-Valverde & Jovanis, 2008), El-Basyouny and Sayed (El-Basyouny & Sayed, 2009) used first-order neighbours to define the neighbouring structure. The same approach was used here where all zones those were directly connected with the zone being dealt with were included. The schematic structure of neighbouring zones is shown in Figure 10. In this figure, the first order zones were directly connected with the zone under consideration, the second order zones were directly related with the first order neighbours, and the third order zones were directly connected with the second order neighbours. This order of neighbours was fully based on the topology of zones and does not account for the effect of distance between zones. The zone with its first order neighbours formed a group and in this study this method resulted in 146 groups. The number of zones per group was found to vary from 1 to 7. Zones within a group were spatially correlated, and zones from different groups were assumed to be statistically independent.

Figure 10 Neighbour structure definition



3.4.2 Spatial models

The development of spatial models in this study followed the procedure shown in El-Basyouni and Sayed (El-Basyouny & Sayed, 2009). Let Y_i is the number of collisions at zone i and it has shown before that Y is assumed to follow Poisson distribution with parameter λ. The mean of Y, shown by Λ, is itself considered a random variable. To account for the overdisperson caused by unobserved or unmeasured heterogeneity, it is assumed that

$$\lambda_i = u_i \exp(\mu_i) \tag{27}$$

where, u_i is related to site-specific attributes, and exp (μ_i) represents multiplicative random effects. It has been shown before that exp (μ_i) follows a gamma distribution for negative binomial model. Spatial Poisson models can be found by introducing a spatial effect in equation (27). So,

$$\lambda_i = u_i \exp(\mu_i)\exp(s_i) \tag{28}$$

where, s_i is the spatial component suggesting that sites that are close to each other should have correlations among them.

To get a fair comparison of CPMs with and without spatial effects, similar model form was used. So, the model form was (Aguero-Valverde & Jovanis, 2008):

$$\ln \lambda_i = \ln\left(a_{0_i}\right) + a_{1_i} \ln\left(VKT_i\right) + \sum_i b_i X_i + \mu_i + s_i \tag{29}$$

where, a_{0i} is intercept, a_{1i}, and b_i are model parameters; VKT_i is the external exposure variable, X_i is the explanatory variable, μ_i is heterogeneity among zones, and s_i is spatially correlated random effect for zone i. El-Basyouni and Sayed (El-Basyouny & Sayed, 2009) showed three different types of spatial models- Gaussian conditional autoregressive (CAR) models, MM models, and extended MM models. The first type of model was developed in this study.

3.4.3 Gaussian conditional autoregressive (CAR) models

Gaussian CAR models are widely used for modeling spatial effects. In this study, the inclusion of spatial effects following Gaussian CAR techniques was based on the methodology found in El-Basyouni and Sayed (El-Basyouny & Sayed, 2009). Let n_i, C (i), and s_{-i} represent the number of neighbours of zone i, the set of neighbours of zone i, and the set of all spatial effects except s_i, respectively. So, s_i in equation (28) will be:

$$s_i \mid s_{-i} \sim \left(\overline{s_i}\ ,\ \sigma_s^2 \big/ n_i\right),\ \overline{s_i} = \sum_{j \in C(i)} s_j \big/ n_i \tag{30}$$

where, σ^2_s is the spatial variation. Equation (30) was based on adjacency-based proximity measure; $w_{ij} = 1$ if zones i and j were first order neighbouring zones, and $w_{ij} = 0$ otherwise. The conditional variance is inversely proportional to the number of neighbouring zones, and the conditional mean is the mean of the adjacent spatial effects (El-Basyouny & Sayed, 2009).

3.4.4 Full Bayes estimation technique

Aguero-Valverde et al. (Aguero-Valverde & Jovanis, 2008) have argued that Bayesian analysis is more suitable for spatial and spatiotemporal models as this analysis approach can easily implement complex correlation structures. Full Bayes (FB) hierarchical models are quite flexible and FB can also offer several other advantages while dealing with collision frequencies. It can explicitly use the prior information which ultimately results in improved parameter estimates. It not only deals with the uncertainty associated with parameter estimates but also can provide exact measures of uncertainty on the posterior distributions of these parameters. This is an extra advantage of FB approach over the Empirical Bayes (EB) as the later approach usually does not consider the uncertainty in correlation structures. This limitation of EB approach sometimes yields to overestimation of the precision of parameters associated with the covariates. The authors have also indicated that Bayes methods provide confidence (credible) intervals which satisfy the intuition and coincide with common sense of interpretations (Aguero-Valverde & Jovanis, 2008).

In this study, Markov chain Monte Carlo (MCMC) methodology was applied using WinBUGS 2.2.0, the Windows interface of Open-BUGS. WinBUGS was used to sample the posterior distribution as well as to estimate the parameters. MCMC methods could sample from the joint posterior distribution repeatedly. This technique generated sequences (chains) of random points, the distribution of which converged to the target posterior distributions. A subsample was used for the purpose of monitoring convergence. Then the subsample was excluded as a burn-in sample. Parameter estimation, performance evaluation, and inference were obtained by the rest of the iterations (El-Basyouny & Sayed, 2009).

3.4.5 Selection of prior distribution

Obtaining FB estimates requires the specifications of prior distributions for the parameters. These prior distributions reflect the prior knowledge about the parameters under consideration. Based on the availability of prior information, prior may be informative or uninformative (vague). Informative prior is formulated with the help of the prior information. Bedrick et al. (Bedrick, Christensen, & Johnson, 1996) discussed the issue of informative priors in generalized linear models (GLM) and got tractable posteriors. They considered two priors- conditional mean priors and data augmentation priors. Both of those priors had the same form as the likelihood. El-Basyouny et al. (El-Basyouny & Sayed, 2009) indicated that Schluter et al. (Schluter, Deely, & Nicholson, 1997) also discussed various aspects of priors in collision data analysis. On the other hand, uninformative or vague priors are usually used when the prior information is not available.

El-Basyouny et al. (El-Basyouny & Sayed, 2009) discussed that diffused normal distribution is the most commonly used prior to estimate the regression parameters. This distribution has a zero mean and large variance. For the dispersion parameter, κ, under the negative binomial model, the commonly used prior is gamma distribution with parameters (ϵ, ϵ) or parameters (1, ϵ), where the value of ϵ is a small number. In this study the parameters of gamma distribution were considered (ϵ, ϵ) and 0.001 was used as the value of ϵ.

Among three types of spatial models, Gaussian conditional autoregressive models were developed in this study. El-Basyouny et al. (El-Basyouny & Sayed, 2009) have shown that equations (30) and (31) are equivalent:

$$f\left(s_1, s_2, \ldots\ldots, s_n \mid \sigma_s^2\right) \propto \sigma_s^{-n} \exp\left\{-\sum n_i s_i \left(s_i - \overline{s_i}\right) \big/ 2\sigma_s^2\right\} \tag{31}$$

Equation (31) provides the correct likelihood function for σ_s^2. The prior distribution of σ_s^2 was assumed to be a gamma distribution with parameters ($1 + \Sigma l_i/2$, $1 + n/2$), where l_i was the term contributed by each zone. l_i was calculated by

$$l_i = n_i s_i \left(s_i - \overline{s_i}\right) \tag{32}$$

Based on El-Basyouny et al. (El-Basyouny & Sayed, 2009), the marginal variance of spatial effects, σ^2_{sm}, was approximated by the following formula:

$$\sigma_{sm}^2 \approx \frac{2\sigma_s^2}{\overline{n}} \tag{33}$$

where,

$$\bar{n} = \frac{\sum n_i}{n} \tag{34}$$

3.4.6 Parameter estimation

It has been discussed in the previous section that the parameters of the application based collision prediction models with spatial effects were estimated with WinBUGS 2.2.0. Two chains were used to run each model. Based on Aguero-Valverde et al. (Aguero-Valverde & Jovanis, 2008), 1,000 to 5,000 MCMC iterations were discarded as burn-in samples. Then 20,000 iterations for each chain were performed. The summary statistics of each chain were then estimated from WinBUGS. The convergences of the developed models were thoroughly checked. The next section describes the way to check the convergence.

3.4.7 Goodness of fit measure

It is of great importance to monitor the convergence because it ensures that the posterior distribution has been "obtained". Ultimately it indicates that the parameter sampling should begin. Usually two or more parallel chains with diverse starting values are tracked so that full coverage of the sample space is ensured as well as the convergence occurs. El-Basyouny and Sayed used Brooks-Gelman-Rubin (BGR) statistic to check the convergence of multiple chains (El-Basyouny & Sayed, 2009). The similar approach was used in this study. If the value of BGR statistic was found to be less than 1.2, convergence occurred. Besides, MCMC trace plots for the model parameters was directly be found in WinBUGS and by visual inspection of these plots, convergence was checked. This study also used the ratios of the Monte Carlo errors relative to the respective standard deviations of the estimates as a measure of checking the convergence. It is a thumb rule that convergence occur when these ratio values are less than 0.05. Moreover, the significances of the parameter estimates were tested at 95% confidence level.

3.4.8 Comparison of CPMs with and without spatial effect

The purpose of performing spatial analysis was to model the spatial factors and their correlations across zones. The effects of spatial correlation were calculated by computing the proportion of total variation that was caused by spatial variation, and the corresponding formula was (El-Basyouny & Sayed, 2009):

$$\Psi_s = \frac{\operatorname{var}(s)}{\operatorname{var}(s) + \kappa} \tag{35}$$

where, var (s) is the marginal variance of s, and it can be directly estimated from the posterior distribution of s. Κ is the dispersion parameter of the developed negative binomial model. The proportion of total variability caused by spatial variation for Gaussian CAR model was checked. Significant spatial correlation exists where the value of Ψ_s is found to be greater than 0.5 (Aguero-Valverde & Jovanis, 2008; El-Basyouny & Sayed, 2009). In this case, the neighbours of zones with high predicted collision frequencies will likely to have high predicted collision frequencies and vice versa.

CPMs with and without spatial effects were also compared based on their values of dispersion parameters, κ. It is a rule of thumb that the model having greater κ value is the better among the two (CPMs with and without spatial effects).

Previous literature has found that inclusion of spatial correlation sometimes affects the estimation of parameters by making some variables 'not significant' although those variables have been significant in the negative binomial CPMs. Moreover, sometimes the regression coefficient of any highly significant variable, such as AADT, is reduced or changed due to the incorporation of spatial effects (Aguero-Valverde & Jovanis, 2008; El-Basyouny & Sayed, 2009). To check these two phenomena, every pair of parameter estimates of two corresponding models was also checked.

3.5 Summary

Based on the methodology described in the previous sections, data was extracted and prepared. Then applications based collision prediction models with and without spatial effects were developed. Generalized linear modeling approach was applied considering a negative binomial error structure. The goodness of fit of the developed models was checked at the 95% level of confidence. Outlier analysis was also performed to remove the outliers and make the models more robust. Then spatial models were developed and compared with the corresponding negative binomial models. Having developed the models with the methodology described in this chapter, the outputs are shown in Chapter Four.

4 Results

4.1 Introduction

This chapter contains the developed transit reliant and other application based collision prediction models based on the methodology presented in Chapter Three. The models considering the spatial effects are also discussed in it. Different groups of models are shown here for total, severe and property damage only collisions with their corresponding goodness of fit statistics. The values and signs of the exponents of each variable are also discussed and compared with the previous studies. Besides, analysis of possible casual mechanisms is also presented with explanations.

4.2 Transit reliant CPM

The group of transit reliant collision prediction models is presented in Table 5. Three models for total, severe and PDO collisions were developed.

Table 5 Transit reliant CPM

Model ID	Model Form	κ	df	SD	Pearson χ^2	$\chi^2_{(0.05,df)}$	t-statistics
1.1.0	T3 = 0.5885 VKT $^{0.7537}$ exp [2.6308 BSKD + 0.0055 RTKT – 0.0227 BTKT]	2.17	450	490	450	500	Constant=-1.3, VKT=14, BSKD=9, RTKT=3, BTKT=-3, κ=16
1.2.0	S3 = 0.0219 VKT $^{0.9873}$ exp [3.0285 BSKD – 0.0584 BTKT]	1.16	471	548	440	523	Constant=-8, VKT=18, BSKD=9, BTKT=-4, κ=16
1.3.0	PDO3 = 1.0297 VKT $^{0.6599}$ exp [2.2423 BSKD + 0.0066 RTKT – 0.0205 BTKT]	2.03	457	499	503	508	Constant=0.1, VKT=12, BSKD=8, RTKT=4, BTKT=-3, κ=16

All of the variables in all three models (1.1.0, 1.2.0 and 1.3.0) were found to be significant at the 5% significance level. As expected, the coefficients of vehicle-km traveled (VKT) in all of the three models were positive indicating that with the increase in traffic exposure, the number of total, severe and PDO collisions increase at a decreasing rate (as coefficient value is less than

1.0) which appear to confirm the intuitive expectation. Other studies also found similar result. Lovegrove and Sayed (Lovegrove & Sayed, 2006a) developed macro-level (i.e., neighbourhood or traffic zone level) collision prediction models using data from 577 neighbourhoods across the Metro Vancouver (formerly known as Greater Vancouver Regional District- GVRD) and observed the dominance and positive influence of VKT on the occurrence of total and severe collisions. Lovegrove et al (Lovegrove, Lim, & Sayed, 2010) also checked the application of these macro level collision prediction models to evaluate the road safety of a *TransLink's 2005-2007 3-year Plan* and found positive correlation of total and severe collisions with VKT. Hadayeghi et al (Hadayeghi et al., 2003) developed zonal level yearly models for total and severe collisions to evaluate the safety of urban transportation systems in the City of Toronto and found positive exponents for VKT. In another study, Hadayeghi et al (Hadayeghi et al., 2007) developed a series of zonal-level collision prediction models that are consistent with conventional models commonly used for urban transportation planning with data for the city of Toronto and observed a positive and highly significant relationship between total and severe collisions and VKT. Similarly, de Guevara et al (De Guevara et al., 2004) developed planning level collision prediction models for the Tucson metropolitan region of Arizona and found that injury type of collisions are positively associated with VKT. The exponent value of VKT for total collisions is smaller than that for the severe collisions, which is also similar to the findings of from Lovegrove and Sayed (Lovegrove & Sayed, 2006a) who dealt with the data from Metro Vancouver, but different from the findings from Hadayeghi et al (Hadayeghi et al., 2003; Hadayeghi et al., 2007) who used the data from Toronto. Although the exponent of VKT for total collisions is smaller than that for the severe collisions, the intercept term for total collisions is much higher than the intercept term for severe collisions. So, for any increase in VKT, the increase in total collision will be higher than the increase in severe collision, which follows the intuition as total collisions consist of both of the severe and property damage only collisions.

The positive values of bus stop density (BSKD) in all of the models presented in Table 5 suggest that more total, severe and PDO collisions could be expected in a traffic zone where the bus stop density is higher (considering all other explanatory variables remained fixed). This may be because of the fact that increased number of bus stops in a zone results in more frequent stop-and-go events by transit vehicles, thus increasing the frequency of merging conflicts between the transit vehicles and general traffic. No previous study has been found that directly related the bus stop density with zonal level total, severe or PDO collision frequencies of all traffic (transit and non-transit both). But in a study dealing with the data set from Toronto area, Cheung et al (Cheung et al., 2008) showed that bus stop density was positively associated with zonal-level transit involved collisions. They also observed that wider spacing between consecutive stops (corresponding to lower density of bus stops along the transit line) is less likely to cause arterial-level all traffic (transit and non-transit both) collisions than tighter stop spacing (corresponding to higher density of bus stops). Although this study does not exclusively deal with the transit collisions, based on the result from Toronto it may be said that the increase in bus stop density will likely to cause an increase in total, severe and PDO collisions of transit vehicle in Metro Vancouver. But further research is needed to prove this relationship.

The coefficient of the average transit trip length (BTKT) was negative. Besides, the negative coefficient for severe collisions was higher than the coefficients (negative) for total and PDO collisions. The reason behind this negative sign may be the more transit-km a passenger traveled, the more time he/she would remain inside the bus and ultimately he/she would be less exposed to collisions. Although no previous study has been found to directly support this explanation, it is obvious that more transit trip length will result in low boarding and unloading volumes in bus stops which will eventually reduce the pedestrian activity at the area surrounding the bus stops. There may be another indirect relationship of BTKT with collision frequencies. Larger transit trip length may be due to the reason of longer distance between two successive bus stops, i.e. due to less bus stop density (BSKD) and it has been explained before that BSKD is positively associated with total, severe and PDO collision frequencies.

The coefficient of average transit ridership or transit occupancy (RTKT) was found to be positive. This may be due to the reason that an increase in the number of person per transit-km traveled may increase the exposure to collisions for those persons (transit passengers) as almost every transit trip needs some amount of walking or biking. This outcome is supported by a previous study by Jovanis et al (Jovanis et al., 1991), although that study dealt with the collisions on any route. They found that average number of collisions that occurred on the route was positively associated with weekly average transit ridership and number of weekly bus trips. Besides, Cheung et al (Cheung et al., 2008) also observed that transit frequency was positively correlated with arterial-level all traffic (transit and non-transit both) collisions. It is somewhat logical to conclude that more transit frequency likely refers to more transit ridership. RTKT was found to be significant for total and PDO collisions. It was also significant for severe collisions, but after the outlier analysis it became non-significant. It should be observed that the coefficient of RTKT is quite small in comparison with the coefficient of VKT. The increase in RTKT should cause a substantial decrease in the percentage of people using autos, and this modal shift should also decrease the total number of autos in the zonal transportation network. It is expected that a reduction in the number of autos in a zone will also cause a decrease in VKT, and ultimately the safety effect of increased RTKT (more collisions) will be offset by the safety effect of reduced VKT (less collisions). So, the positive association of RTKT with collision frequencies does not necessarily mean that it will make the transportation network less safe.

All the independent variables in transit reliant collision prediction models are significant at the 5% level as their t-statistics are found to be more than 1.96. The variables selected for each type of models are of reasonable signs in terms of their expected tendency to increase or decrease collision frequency. These variables are also found to be significant determinants of collision frequency. The dispersion parameters in all of these three models are highly significant. Their large values confirm that the data are over dispersed relative to the Poisson distribution and that the negative binomial error structure is therefore justified. The ratios of deviance to degree of freedom are in all models also close to 1.0, which is another indication that the choice of the NB error structure appears to be appropriate for these models. One CPM (1.2.0) has scaled deviance value exceeding the χ^2 statistic, although the Pearson χ^2 value satisfies the condition. However, that SD value is within 5 percent of the target χ^2 statistic. Two models (1.1.0 and 1.3.0) have intercept values less than the t statistic (< 1.96), indicating that the constant term leading the CPM may not significantly differ from 1.0. But this does not pose significant difficulty as the significance of these intercepts is marginal. Similar results, in terms of slightly exceeding the χ^2 statistic by SD and non-significant intercept, were also noted by Hadayeghi et al. (Hadayeghi et al., 2003) and Lovegrove and Sayed (Lovegrove & Sayed, 2006a).

4.3 Application based CPMs

A series of zonal-level application based collision prediction models were developed. Table 6 presented the CPMs that dealt with the number of collision frequencies associated with the change in exposure.

Table 6 Application based CPMs: change in exposure

Model Name	Model ID	Model Form	κ	df	SD	Pearson χ^2	$\chi^2_{(0.05,df)}$	t-statistics
Change in Exposure	2.1.0	T3 = 3.8023 VKT $^{0.607}$	1.42	481	539	437	533	Constant=3.9, VKT=14, κ =17
	2.2.0	S3 = 0.0448 VKT $^{0.9555}$	1.02	469	547	410	520	Constant=-7, VKT=17, κ =16
	2.3.0	PDO3 = 4.0338 VKT$^{0.5676}$	1.47	479	535	521	531	Constant=4.2, VKT=14, κ =17

Zero exposure means zero collisions. So, collisions will increase with an increase in exposure. As expected, the coefficients of VKT in all of the models presented in Table 6 were positive indicating that with the increase in traffic exposure, the number of total, severe and PDO collisions increase which appear to confirm the intuitive expectations. Similar results were also found in case of transit reliant collisions. Besides, the exponent values of VKT for total collisions are also slightly smaller than those for the severe collisions. The case is different in case of the intercept values, i.e. the intercept in total collision models is much higher than the intercept in severe collision models. All of the intercepts, dependent variables and κ values in all three models were found to be significant at the 5% significance level. Although the scale deviances (SD) in all of these three models are slightly greater than the critical χ2 values, they are exceeding the limit by 1% in models 2.1.0 and 2.3.0, and by 5% in model 2.2.0. These findings are similar to the previous studies on Metro Vancouver data (Lovegrove & Sayed, 2006a).

Table 7 Application based CPMs: GHG emission reduction / sustainable environment

Model Name	Model ID	Model Form	κ	df	SD	Pearson χ^2	$\chi^2_{(0.05,df)}$	t-statistics
GHG Emission Reduction / Sustainable Environment	3.1.0	T3 = 0.3260 $VKT^{0.8581}$ exp [0.3959 TKTP]	1.72	476	526	487	528	Constant=-3, VKT=19, TKTP=9, κ=17
	3.1.1	T3 = 0.3577 $VKT^{0.8376}$ exp [0.0271 PEDP + 0.1608 TKTP + 0.0599 BIKP]	1.83	476	526	485	528	Constant=-2.8, VKT=19, PEDP=6, TKTP=4, BIKP=3, κ=17
	3.2.0	S3 = 0.0143 $VKT^{0.9934}$ exp [0.3046 TKTP]	1.05	473	551	512	525	Constant=-8.3, VKT=16, TKTP=6, κ=16
	3.3.0	PDO3 = 0.3597 $VKT^{0.8026}$ exp [0.1806 TKTP + 0.026 PEDP + 0.0729 BIKP]	1.90	476	524	495	528	Constant=-2.9, VKT=18, TKTP=4, PEDP=6, BIKP=3, κ=16

Table 7 showed the CPMs that dealt with the number of collision frequencies associated with the change in green house gas (GHG) emissions. It has been already described in Chapter Three that this study finds no data that directly expresses the change in GHG emission. So, some proxy variables have been used here. GHG emissions can be reduced by introducing the active modes of transportation- walking and biking as well as by encouraging the use of public transportation, i.e. by increasing the share of buses in the total stream of vehicles. In this case the increase in the percentage of transit in the total vehicle stream can be expressed by a proxy variable- TKTP, which is the percentage of transit kilometre travelled to the total vehicle kilometre travelled, along with another two variables that are closely related to transit operation- BIKP and PEDP. It has been already mentioned that almost every transit trip needs some amount of walking and biking.

The positive sign of TKTP satisfies intuitive expectations as the more TKT, the more probability of occurring conflicts between transit and autos or in between transit boarding and unloading passengers and autos. No previous study has been found that directly related TKTP with zonal level total, severe or PDO collision frequencies. But in a study dealing with the data set from Toronto area, Cheung et al (Cheung et al., 2008) showed that Transit-kilometre travelled was positively associated with zonal-level transit involved collisions. So the increase in collisions both for the transit and non-transit vehicles is most likely caused by the increase in transit collisions.

The positive sign for percentage of commuters biking (BIKP) in model may be due to the presence of some shared bikeways. This outcome is supported by a previous study done in Denmark. From a study carried out in Copenhagen to observe the effect of bike and other modes of traffic on the safety of road intersections and links, Jensen et al (Jensen, 2008) found that the construction of bike tracks resulted in 18-20 percent increase in bike traffic and 9-10 percent decrease in auto traffic. At the same time this construction of bike tracks was negatively associated with total number of collisions and injuries on the road links and positively associated with total and injury collisions at intersections, where as the combined effects of bike tracks on the road network in terms of total and injury collisions were found to be positive. In another study Brude and Larsson (Brude & Larsson, 1993) found a positive relationship between the number of bike users and total bike collisions at an intersection. BIKP was found to be significant in case of total and PDO collisions (3.1.1, 3.3.0), but it became insignificant for severe collisions after removing the outliers. So, the increase in total collisions due to the increase in the percentage of commuters using bike was likely to be caused by the increased number of property damage only collisions.

The coefficient of percentage of commuters walking (PEDP) was also positive. The reason behind this may be more pedestrian percentage refers to more pedestrian activity, more movement in the transportation network and in a mixed traffic condition (where auto, transit, bike, pedestrians- all are present) this leads to more exposure to collisions. But in a study dealing with the data set from 509 signalized intersections in the City of Montreal, Miranda-Moreno (Miranda-Moreno, Morency, & El-Geneidy, 2010) observed a statistically significant and positive relationship between pedestrian activity and collision frequency at different types of intersections. Shahla et al (Shahla et al., 2009) conducted a study to evaluate the overall safety performance of signalized intersections in Toronto and they observed that pedestrian volume is positively associated with total collisions at any intersections, and with total and transit involved collisions at transit-serviced intersections. Besides, in another study Brude and Larsson (Brude & Larsson, 1993) found a positive association of the number of pedestrians with total pedestrian collisions at an intersection. Similar to BIKP, PEDP was also found to be significant for total and PDO collisions (3.1.1, 3.3.0). It was also significant for severe collisions, but after the outlier analysis, it did not remain significant.

The positive sign of percentage of commuters using transit (BUSP) in model 3.1.2 also satisfied intuitive expectations as the more BUSP, the more probability of occurring conflicts between bus passengers and other vehicles during the time of boarding and off-boarding. Besides, an increase in the percentage of commuters using transit may also increase the exposure to collisions for those persons (transit passengers) as almost every transit trip needs some amount of walking or biking. This finding is supported by a previous study done by Shahla et al (Shahla et al., 2009), although they dealt with the transit and non-transit collisions at the intersections only. Using the data sets from different signalized intersections in Toronto area, the authors observed that public transit volume was significantly and positively correlated with total collisions at any intersections, and total and transit involved collisions at transit-serviced intersections.

The positive coefficients of TKTP, PEDP, BIKP and BUSP indicate their positive associations with total, severe and PDO collisions for transit and non-transit vehicles. But it should be observed that the coefficient values of these variables are comparatively much smaller than the exponent value of VKT. It is expected that the increase in any of these variables (TKTP, PEDP, BIKP and BUSP) will likely to decrease the percentage of people using autos. This modal shift from auto to transit, and auto to different active forms of transportation should decrease the total number of autos in the transportation network of a zone. Reduced number of autos in a transportation network should reduce the amount of VKT. As VKT has a bigger and positive coefficient, a decrease in VKT will largely decrease the number of collisions in the whole network. As a result, the number of total, severe and PDO collisions increased by the increased number of TKTP, PEDP, BIKP or BUSP will likely to be offset by the number of collisions decreased due to reduced VKT.

Table 7 shows that all of the intercepts and independent variables are significant at the 5% level of significance. The dispersion parameters are also highly significant. The variables show reasonable signs with respect to the association with collision frequencies. Only one CPM (3.2.0) has scaled deviance value slightly greater than the corresponding χ^2 statistic (exceeded by 5%), although its Pearson χ^2 value satisfies the condition. Previous studies also found such phenomenon of higher scaled deviance than the corresponding χ^2 statistic (Hadayeghi et al., 2003); (Lovegrove & Sayed, 2006a).

Table 8 shows the collision prediction models that express the number of total, severe and PDO collisions with respect to the percentage of people using different modes of a transportation system. These models can evaluate the change in number of collisions due to a change in modal share in the zonal transportation network.

Table 8 Application based CPMs: change in modal share

Model Name	Model ID	Model Form	κ	df	SD	Pearson χ^2	$\chi^2_{(0.05,df)}$	t-statistics
Change in Modal Share	4.1.0	T3 = 6.0279 $VKT^{0.7858}$ exp [- 0.0264 AUTOP]	2.04	468	509	435	519	Constant=5.3, VKT=19, AUTOP=-13, κ=16
	4.1.1	T3 = 1.5869 $VKT^{0.6049}$ exp [0.0536 BUSP]	1.93	473	516	469	525	Constant=1.39, VKT=15, BUSP=12, κ=16
	4.1.2	T3 = 0.5642 $VKT^{0.739}$ exp [0.0391 BUSP + 0.0251 PEDP]	1.99	477	523	479	529	Constant=-1.64, VKT=17, BUSP=8, PEDP=7, κ=16
	4.1.3	T3 = 2.0643 $VKT^{0.7983}$ exp [0.0146 PEDP - 0.0164 AUTOP + 0.0645 BIKP]	2.19	459	499	442	510	Constant= 1.5, VKT=18, PEDP=3, AUTOP=-5, BIKP=3, κ=16
	4.2.0	S3 = 0.2284 $VKT^{0.9687}$ exp [- 0.0233 AUTOP]	1.10	478	555	392	530	Constant=-3.3, VKT=17, AUTOP=-9, κ=16
	4.2.1	S3 = 0.0443 $VKT^{0.8219}$ exp [0.0678 BUSP]	1.28	469	540	400	521	Constant=-7.4, VKT=16, BUSP=13, κ=16
	4.3.0	PDO3 = 0.7820 $VKT^{0.7812}$ exp [0.0269 PEDP + 0.0988 BIKP - 0.0063 AUTOP]	2.04	470	513	504	522	Constant=-0.6, VKT=18, PEDP=5, BIKP=4, AUTOP=-2, κ=16
	4.3.1	PDO3 = 1.4817 $VKT^{0.5845}$ exp [0.0522 BUSP]	1.96	468	510	474	519	Constant=1.1, VKT=14, BUSP=12, κ=16

As expected, VKT, BUSP, PEDP and BIKP showed positive associations with total, severe and PDO collisions for transit and non-transit vehicles. The underlying reasons behind these positive coefficients have already been described in the previous sections.

The negative coefficient of percentage of commuters using autos (AUTOP) in model 4.1.0, 4.1.3, 4.2.3 and 4.3.0 is difficult to explain although this variable is statistically significant in all of these four models. An auto passenger usually stays inside the vehicle from origin to destination and thus might be less susceptible to collision exposure than the transit passenger. Another reason may be that the less amount of AUTOP resulted in less number of autos on the road which yielded in reduced congestion in the transportation network as well as increased speed of the vehicle stream. Previous studies found a positive correlation with speed and all types of collision frequencies. Elvik et al (Elvik, Christensen, & Amundsen, 2004) developed a power model and found a strong statistical relationship between speed and road safety. They observed that when the mean speed of traffic was reduced, the number of collisions and the severity of injuries almost always went down and when the mean speed of traffic increased, the number of collisions and the severity of injuries usually increased. However, the negative association of AUTOP with collision occurrences should further be examined to ensure the validity of such outcome. While AUTOP is negatively associated with collision frequencies, at the same time an increase in AUTOP will be reflected by an increase in the total number of autos on the road network. This increase in the number of autos will likely cause an increase in VKT. Ultimately the safety effect of increased AUTOP (i.e. less collisions) will be largely offset by the safety effect of increased VKT (i.e. more collisions) as the exponent of VKT is higher than that of AUTOP. So, the negative association of AUTOP with collision frequencies does not necessarily indicate a safer zonal transportation network.

All of the independent variables shown in Table 8 are significant at the 5% level of significance. The dispersion parameters are also highly significant. The variables show reasonable signs with respect to the association with collision frequencies. Two CPMs (4.2.0 and 4.2.1) have scaled deviance values greater than corresponding χ^2 statistics, although these two SD values are within 5 percent of the target χ^2 statistics. The Pearson χ^2 values are found to be less than the χ^2 statistics and therefore satisfy the condition. However, the constant terms leading four CPMs (4.1.1, 4.1.2, 4.1.3, 4.3.0 and 4.3.1) may not significantly differ from 1.0 as those four models have intercepts with t-values less than the t-statistic (< 1.96). But these do not pose significant difficulty as similar results, with respect to SD greater than χ^2 statistics and non-significant intercept, were also found for Toronto (Hadayeghi et al., 2003) and Metro Vancouver (Lovegrove & Sayed, 2006a) data.

Changes in the existing network alignment (especially routes for transit) are sometimes occurred in the zonal transportation network. These changes in network may happen due to the change in transportation demand, land use pattern, or other factors including change in job-housing density. Table 9 shows the developed CPMs that dealt with the changes in network alignment with respect to the number and types of intersections, presence of bends and curves, etc.

Table 9 Application based CPMs: change in network alignment

Model Name	Model ID	Model Form	κ	df	SD	Pearson χ^2	$\chi^2_{(0.05,df)}$	t-statistics
Change in Network Alignment	5.1.0	T3 = 10.7855 VKT $^{0.6059}$ exp [- 0.0130 I3WP - 0.0190 SBNCP - 0.0270 HBNCP]	2.44	467	505	450	518	Constant=8.0, VKT=17, I3WP=-7, SBNCP=-5, HBNCP=-6, κ=16
	5.2.0	S3 = 0.4553 VKT$^{0.8270}$ exp [- 0.014 I3WP - 0.0217 SBNCP - 0.0387 HBNCP]	1.49	469	537	438	520	Constant=-1.94, VKT=17, I3WP=-6, SBNCP=-4 HBNCP=-7, κ=16
	5.3.0	PDO3 = 12.7 VKT $^{0.5522}$ exp [- 0.0131 I3WP - 0.0171 SBNCP - 0.0254 HBNCP]	2.42	467	505	450	518	Constant=8.4, VKT=15, I3WP=-7, SBNCP=-4, HBNCP=-6, κ=16

The coefficients of VKT in all of these three models are found to be positive, implying positive associations with total, severe and PDO collisions. Similar results have also been found in the previous models of this study.

The negative coefficients of ratio of number of soft curves (degree of curvature < 45^0) to number of conflicts (SBNCP) and ratio of number of hard curves (degree of curvature > 45^0) to number of conflicts (HBNCP) in the models satisfied intuitive expectations as the presence of curves or bends in network eventually decreased the traffic stream speed and hence reduced the number of all types of collisions (numbers of conflicts were the summation of number of intersections, soft curves, and hard curves). The positive association of speed with collision frequencies has already been described in the previous section (Elvik et al., 2004). Besides, the coefficient of HBNCP is greater than that of SBNCP as the speed reduction in hard curves will likely be more than that in soft curves. In a previous study, Milton and Mannering (Milton & Mannering, 1998) analyzed the collision frequencies on the principal arterials of the Washington state and argued that if there was a number of sharp horizontal curve sections, motorists might be more likely to drive cautiously which resulted in less collision frequencies.

The coefficient of percentage of three-way intersections (I3WP) was found negative and the reason behind this negative coefficient was due to the less number of potential vehicle to vehicle conflict points in three-way intersections in comparison with traditional four-legged intersections. Figure 11 shows that there exist only 22 potential vehicle to vehicle conflict points in two offset three-way intersections where as the traditional four-legged intersection has 32 potential vehicle to vehicle conflict points (Figure 12). This finding also supported the earlier findings. Lovegrove and Sayed (Lovegrove & Sayed, 2006a) developed macro-level (i.e., neighbourhood or traffic zone level) collision prediction models using data from 577 neighbourhoods across the Greater Vancouver Regional District and observed the negative influence of I3WP on the occurrence of total and severe collisions. According to Cinneidy and Troutbeck (Cinneide & Troutbeck, 1995), similar results were found in a Swedish (Hedman, 1990) and a TRB (Transportation Research Board (TRB), 1984) studies.

Figure 11 Points of conflicts in two three-way intersections

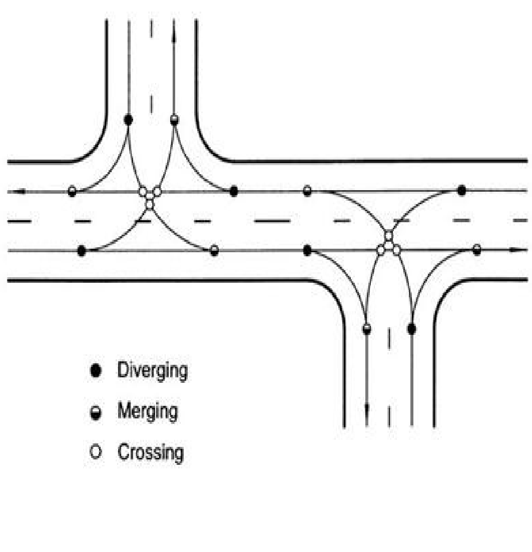
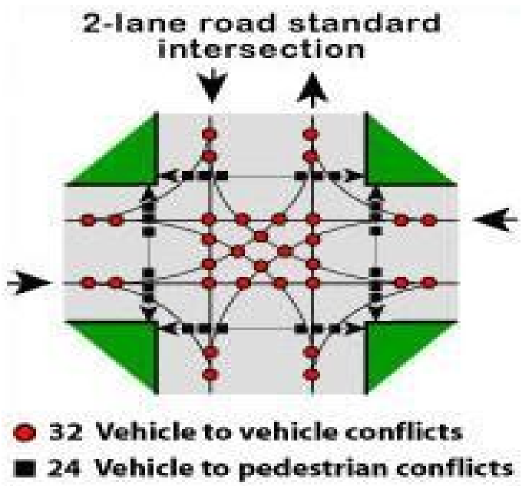


Figure 12 Points of conflicts in a four-legged intersection



All of the variables presented in Table 9 are significant at the 5% level of significance. The dispersion parameters also seem to be highly significant, similar to other models presented before. The intercepts are also significant. The variables show reasonable signs with respect to the association with collision frequencies. One CPM (5.2.0) has scaled deviance value larger than the corresponding χ^2 statistic by 3% only, which is within the acceptable limit. However, all other scaled deviances and all Pearson χ^2 values satisfy the condition.

The fifth goal of the Translink 2040 transportation plan is to have an economic growth in GVRD through efficient management of the transportation network (TransLink, 2008). Table 10 presents the CPMs showing the effect of economic growth on the collision frequencies.

Table 10 Application based CPMs: economic growth

Model Name	Model ID	Model Form	κ	df	SD	Pears on χ^2	$\chi^2_{(0.05,df)}$	t-statistics
Economic Growth	6.1.0	T3 = 49.1610 $VKT^{0.8411}$ exp [- 0.0508 EMP + 0.3027 WKGD]	1.81	468	516	502	519	Constant=4.5, VKT=19, EMP=-6, WKGD=5, κ=16
	6.1.1	T3 = 2.6290 $VKT^{0.8215}$ exp [- 0.0257 PARP + 0.3662 WKGD]	1.84	467	515	511	518	Constant=2.3, VKT=20, PARP=-6, WKGD=6, κ=16
	6.2.1	S3 = 0.4795 $VKT^{0.8460}$ exp [- 0.0250 PARP + 0.2530 WKGD]	1.00	479	547	479	531	Constant=-1.2, VKT=13, PARP=-4, WKGD=4, κ=30
	6.3.0	PDO3 = 30.0498 $VKT^{0.7576}$ exp [- 0.0427 EMP + 0.3219 WKGD]	1.79	469	518	520	520	Constant=4.0, VKT=18, EMP=-5, WKGD=5, κ=17
	6.3.1	PDO3 = 2.8016 $VKT^{0.7744}$ exp [- 0.0245 PARP + 0.3771 WKGD]	1.87	467	514	518	518	Constant=2.5, VKT=19, PARP=-6, WKGD=6, κ=16

The reason behind positive coefficient of VKT is described before. The coefficient of number of workers per resident (WKGD) in all models was found to be positive which satisfied intuitive expectations. The underlying reason behind this might be more workers per resident yielded in more commuters and more economic activities outside the home, and which eventually increased the exposure to collision. Lovegrove and Sayed (Lovegrove & Sayed, 2006a) also found the similar result in a previous study on the data of Metro Vancouver.

The negative signs of participation in labour force (PARP) and employed workers in total labour force (EMP) are difficult to explain. Intuitively, PARP or EMP should be positively associated with total, severe and PDO collision frequencies. Based on the data set from Tucson, de Guevara et al (De Guevara et al., 2004) showed that the number of injury collisions and PDO collisions increased with the increase in number of employees. But this model suggested that the relationship may be otherwise. Lovegrove and Sayed (Lovegrove & Sayed, 2006a) also found similar results for Metro Vancouver data and they argued that the reason of this negative signs might be related to school drop-off trips. These trips were made by stay-at-home parent drivers and (or) postsecondary student drivers, both of who were unemployed but classed as "employable" persons, and both occurring in peak periods. Besides, employed people pass a significant portion of time of the day in their offices and as a result they might be less exposed to collision. Moreover, the carpooling and car-sharing of the employed people might reduce the number of autos which eventually reduce the amount of VKT. It has been described before that VKT has the dominance effect on the occurrence of total, severe and PDO collisions. However, further studies are needed to verify the relationship of PARP or EMP with collision frequencies.

Except for only one intercept (Model 6.2.1), all intercepts along with all independent variables in Table 10 are significant at the 5% level of significance. Besides, the dispersion parameters are also highly significant, with t-values ranging from 16 to 30. In Model 6.2.1, the SD exceeds the χ^2 statistic by 3%, although the Pearson χ^2 value satisfies the condition. All other SD and Pearson χ^2 values are within the limit.

Table 11 presents the collision prediction models with respect to the change in accessibility and mobility of a zonal transportation network.

Table 11 Application based CPMs: change in accessibility-mobility

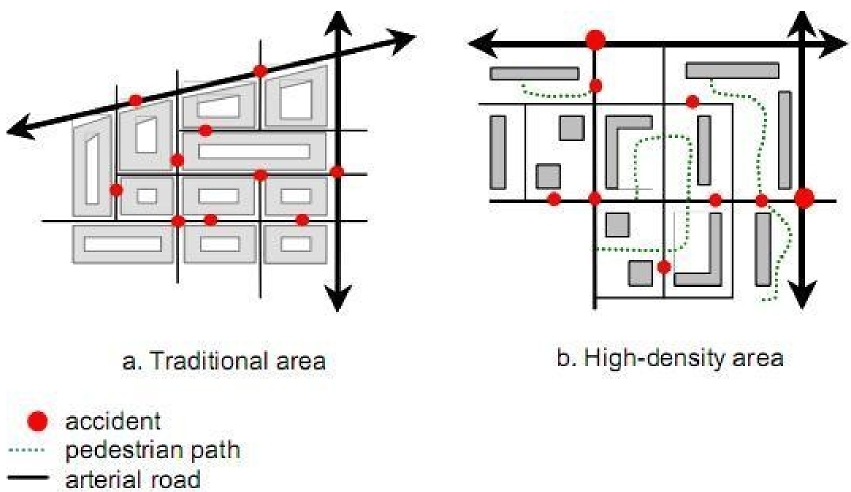
Model Name	Model ID	Model Form	κ	df	SD	Pearson χ^2	$\chi^2_{(0.05,df)}$	t-statistics
Change in Accessibility / Mobility	7.1.0	T3 = 2.4727 $VKT^{0.7502}$ exp [- 0.0168 I3WP + 5.6008 SIGD - 0.0067 ALKP]	2.46	466	504	434	517	Constant=2.6, VKT=19, I3WP=-11, SIGD=7, ALKP=-3, κ=16
	7.2.0	S3 = 0.1117 $VKT^{0.9456}$ exp [- 0.0207 I3WP + 4.0410 SIGD]	1.34	474	545	442	526	Constant=-4.6, VKT=17, I3WP=-9, SIGD=4, κ=16
	7.3.0	PDO3 = 3.0560 $VKT^{0.7021}$ exp [- 0.0169 I3WP + 6.3494 SIGD - 0.0097 ALKP]	2.55	461	498	433	512	Constant=3.2, VKT=18, I3WP=-11, SIGD=8, ALKP=-4, κ=16

Accessibility and mobility of a roadway depend on the functional type of that road. Arterial road has higher mobility than collector or local road as arterial enjoys higher speed. But it has the least accessibility among these three. Intersections are introduced to increase accessibility in the arterial road. Usually, intersections situated on the arterials are signalized. Local-arterial intersections and percentage of 3-way intersections also determines the level of accessibility and mobility in the zonal transportation network. As expected, VKT and I3WP showed positive associations with total, severe and PDO collisions for transit and non-transit vehicles. The underlying reasons behind these positive coefficients have already been described in the previous sections.

The positive coefficients of number of signalized intersection per unit area (SIGD) have large values. These large values and positive sign supported the previous study based on Metro Vancouver (Lovegrove & Sayed, 2006a). Another study on Toronto area also found positive coefficient, although the value was comparatively small (Hadayeghi et al., 2007). Vancouver study found positive coefficient for number of signalized intersection per unit area, whereas Toronto study observed positive coefficient for the total number of signalized intersections. The possible reason behind this positive sign was due to the conflicting points at signalized intersections that might be caused by left or right turning vehicles with pedestrians. Signalized intersections require focused attention, sound judgment, and quick decisions by drivers. Besides, traffic signals may be visually obstructed, distracted, or even willingly disobeyed. A speeding driver might run through a red light or make a bad decision whether to continue driving in amber signal. Combining all this information, signalized intersection environment itself makes matters worse and eventually pose more susceptibility to collision occurrences. So, Lovegrove et al. (Lovegrove & Sayed, 2006a) argued that the association of increasing signal density (SIGD) with an increasing number of collisions indicated the fact that more signals might not be safer.

The level of safety of an arterial road is immediately below that of a highway. Arterial has a moderate to high capacity and it carries large volumes of traffic between urban centres. This type of road enjoys limited accessibility as, except in older or denser communities, residential entrances directly onto the road are restricted here. In Metro Vancouver, 32% roads are arterial, 30% roads are collector and the rest of the 38% roads are local (Lovegrove, 2006). Arterial lane km per total lane km (ALKP) was found to be statistically significant for total (Model 7.1.0) and PDO collisions (Model 7.3.0) but insignificant for severe collisions (Model 7.2.0). The negative coefficient of ALKP indicated that an increase in the amount of arterial road in the total road network would cause a decrease in total and PDO collision frequencies. But this negative coefficient is really hard to explain as arterial road usually enjoys higher speed than collector or local road. Positive signs of arterial road were observed by Lovegrove and Sayed (Lovegrove & Sayed, 2006a) for ALKP for Metro Vancouver and by Hadayeghi et al (Hadayeghi et al., 2007) for Toronto data set. But this study suggested that the relationship may be otherwise. The underlying reason of negative sign might be due to the fact that arterials had better quality than collectors and local roads which reduced the odd of occurring collisions. In a previous study, Millot (Millot, 2004) showed that collisions on arterial roads were concentrated at a few 'specific' points and many tools were available to improve safety conditions of those 'specific' points. But on local and collector roads, road safety problems were dispersed. They involved car users as well as other vulnerable road users such as bike users or pedestrians. Obviously it was more difficult to manage and maintain those areas as the layout of local and collector roads were very favourable to autos with high speed. Wider roads, large green spaces, and houses that set back from the right of way – all of the factors promoted high speed and the use of autos in local and collector roads. Ultimately collector and local roads became less safe than arterial roads. Figure 13 shows the vulnerable points on arterial and local roads where collisions may occur (Millot, 2004). The figure shows that comparatively fewer number of potential conflict points exist on arterial road than on local or collector road. However, more research should be performed to verify the safety consequences of arterials.

Figure 13 Vulnerable points for occurrence of collisions



(Millot, 2004)

All of the intercepts and independent variables in Table 11 are significant at the 5% level of significance. Besides, the dispersion parameters are also highly significant. The variables show reasonable signs with respect to the association with collision frequencies. The CPM for severe collision (Model 7.2.0) has slightly larger scaled deviance value than the corresponding χ^2 statistic (exceeded by 4%). However, SD values of other models and Pearson χ^2 values of all of these models satisfy the condition.

The target of urban containment is to increase the number of inhabitants in a zone keeping the zonal boundary unchanged. Cities are adopting this initiative to have a greater density of people especially around the arterial or transit routes. Urban containment will likely to change the number and type of intersections. Besides, some unsignalized intersections may be signalized to meet the increased demand of vehicle. Table 12 shows the CPMs that deal with predicting the number of collisions due to change in the variables those are related to urban containment.

Table 12 Application based CPMs: urban sprawl

Model Name	Model ID	Model Form	κ	df	SD	Pears on χ^2	$\chi^2_{(0.05,df)}$	t-statistics
Urban Sprawl	8.1.0	T3 = 0.2549 $VKT^{0.8206}$ exp [1.0391 INTD + 4.8375 SIGD + 0.0061 IALP + 0.0079 NHD]	2.27	472	516	502	524	Constant=-4.3, VKT=21, INTD=7, SIGD=6, IALP=2.4, NHD=3, κ=16
	8.1.1	T3 = 0.2558 $VKT^{0.8164}$ exp [1.0612 INTD + 4.4975 SIGD + 0.0065 IALP + 0.0033 POPD]	2.31	470	514	495	522	Constant=-4.4, VKT=21, INTD=6, SIGD=6, IALP=3, POPD=3, κ=16
	8.2.0	S3 = 0.0084 $VKT^{0.9957}$ exp [1.6695 INTD + 3.2263 SIGD + 0.0153 IALP]	1.40	465	534	422	516	Constant=-11, VKT=19, INTD=9, SIGD=3, IALP=5, κ=16
	8.3.0	PDO3 = 0.2902 $VKT^{0.7747}$ exp [1.0038 INTD + 4.6926 SIGD + 0.0054 IALP + 0.0035 POPD]	2.33	469	512	503	520	Constant=-4.0, VKT=20, INTD=6, SIGD=6, IALP=2.3, POPD=3, κ=18

Similar to the previous models, VKT and SIGD are positively associated with total, severe and PDO collisions. The positive signs of the coefficients of number of intersections per unit area (INTD) indicated that an increase in the number of intersections would cause increased number of collisions. This positive coefficient might be due to the fact that collisions are more likely to happen at intersections than on the link section of a road. Two reports of Transport Canada (Transport Canada, 2008; Transport Canada, 2008) showed that almost 30 percent of deaths from vehicle collisions and more than 40 percent of serious injuries from vehicle collisions involved an intersection. In another study of non-freeway motor vehicle collisions in four urban areas of the United states, Retting et al (Retting, 1996) found that almost 56% of the total collisions occurred at intersections. So, more intersections per unit area resulted in more collision frequencies. There might be various reasons behind the occurrences of more collisions at the intersections. Typically all types of vehicles –autos, trucks, bike, and motorcycles travel in various directions by making turns in an intersection. Besides, designated pedestrians crossing zones are found adjacent to the intersection. Moreover, it has been discussed before that in a signalized intersection, traffic signs and signals might be visually obstructed, distracted, or even willingly disobeyed. Combining all of these pieces of information, intersection environment itself makes matters worse and eventually intersections pose more susceptibility to collision occurrences. This outcome also supported some of the previous studies. Lovegrove and Sayed (Lovegrove & Sayed, 2006a) studied the Metro Vancouver area and found a positive relationship between number of intersections per unit area with total and severe collision frequencies. Using the data set from Toronto area, Hadayeghi et al (Hadayeghi et al., 2003) also found the same sign for total and severe collisions. De Guevara et al (De Guevara et al., 2004) conducted another study using the data from Tucson, Arizona, and observed similar result for PDO collision model but different results (negative coefficient of intersection density) for fatal and injury collision models. According to Lovegrove and Sayed (Lovegrove & Sayed, 2006a) this difference in results requires closer attention when planning neo-traditional communities with redundant street (grid road) patterns that may foster higher intersection densities. This requirement is also supported by a study on neighbourhood road pattern influences on road safety (Poppe, 1997).

The coefficient of percentage of arterial-local intersections to all types of intersections (IALP) was found positive. This association between total, severe and PDO collisions and arterial–local road intersections is difficult to explain. The underlying reason might be the different characteristics of arterial and local road. Arterials usually enjoy higher speed and less accessibility, where as local roads have lower speed and more accessibility. Besides, most of the time arterial-local intersections are two-way stop controlled (TWSC) intersections and stop/yield sign is placed on local road. The drivers coming from local roads need to make a judgment of ‘minimum acceptable gap’ to merge with the through traffic stream on the arterial roads. Any poor judgment in such a case would likely to cause collisions. Moreover, any driver disregarding a stop sign has the potential to be a severe collision since it typically involves a right angle type of impact. Yuan et al (Yuan et al., 2010) also showed that in an arterial-local TWSC intersection, collisions might occur due to any obstruction of the view of the stop sign (usually from vegetation or tree branches), or lack of responsibility for the placement and maintenance of the signs. In a previous study on the data of Metro Vancouver area, Lovegrove and Sayed (Lovegrove & Sayed, 2006a) also observed positive association of IALP with total and severe collisions in rural area.

The positive coefficient of population density (POPD) seems intuitive. POPD was found to be statistically significant for total and PDO collisions. Usually more population density refers to more travel activity and hence more exposure to collisions. De Guevara et al (De Guevara et al., 2004) showed that POPD was related to exposure to risk and reflected the degree of interaction among people - with higher densities implying greater interaction and potential conflicts. Eventually an increase in POPD caused increase in fatal, injury and PDO collisions. They also argued that number of population or population density served as a superior measure of exposure to risk compared to VKT because of several reasons. First, the aggregate nature of the population data and the general lack of volume data on lower classified roads made POPD a better measure of exposure to collision. Second, population-based exposure was more readily available for at the zonal level. The positive exponent of POPD also satisfied previous studies conducted by Lovegrove et al. (Lovegrove & Sayed, 2006a) and Hadayeghi et al (Hadayeghi et al., 2007) who checked the effect of POPD on total and severe collisions.

The positive coefficient of number of households per unit area (NHD) in models 8.1.0, 8.1.1 and 8.3.0 indicated that any increase in number of households per unit area would also cause increase in total and PDO collision frequencies. This positive sign was intuitive because more household density resulted in more activity and more exposure to collisions. This finding also satisfied previous studies. Lovegrove et al. (Lovegrove & Sayed, 2006a) found similar result from the data set from Metro Vancouver. Hadayeghi et al (Hadayeghi et al., 2003) found that number of households was positively associated with total and severe yearly collisions, and severe yearly collisions for morning peak period. In another study the later authors also found similar results (Hadayeghi et al., 2007).

All of the independent variables in all types of application based collision prediction models are highly significant at the 5% level. The variables selected for each type of models are of reasonable signs in terms of their expected tendency to increase or decrease collision frequency.

These variables are also found to be significant determinants of collision frequency. The value of the dispersion parameters in all of the developed models confirms that the data are over dispersed relative to the Poisson distribution and that the negative binomial error structure is therefore justified. The ratios of deviance to degree of freedom also fall within the acceptable range of 0.8 to 1.2 (Hadayeghi et al., 2003), which is another indication that the choice of the NB error structure appears to be appropriate for these models. Several CPMs have scaled deviance values slightly exceeding their χ^2 statistics. However, all SD values are within 5 percent of the target χ^2 statistics. Only a few models have intercept values less than the t statistic (<1.96), indicating that the constant term leading the CPM may not significantly differ from 1.0. But this does not pose significant difficulty as the significance of these intercepts is marginal. Similar results were also noted by Hadayeghi et al. (Hadayeghi et al., 2003) and Lovegrove and Sayed (Lovegrove & Sayed, 2006a).

4.4 Spatial CPMs

The transit reliant and application based collision prediction models with spatial correlations were developed using the software WinBUGS. The convergences of the developed models were tested by the trace plots of the parameter estimated, the BGR statistics, and ratios of Monte Carlo errors relative to the standard deviations of the estimates.

Table 13 presents the group of transit reliant collision prediction models with and without spatial correlations. The table shows that the incorporation of the spatial correlation affected the parameter estimates, the values of dispersion parameters and intercepts, and also the t-statistics. In all cases, the intercepts of spatial models are greater than those of the negative binomial models. Wang and Abdel-Aty (Wang & Abdel-Aty, 2006) dealt with 476 signalized intersections in Florida and found similar results. However, the intercept terms of total and PDO negative binomial models (1.1.0 and 1.3.0) were not significant, but the intercepts became significant in spatial models (1.1.a and 1.3.a). Besides, the effect of VKT on the number of total, severe and PDO collisions were smaller under spatial models. The smaller values of the exponents of VKT indicated the fact that not considering the spatial variables may yield a bias associated with model misspecifications. This finding also supports previous studies on the data from 281 urban road segments in Metro Vancouver (El-Basyouny & Sayed, 2009), and data from 865 rural two-segments of Centre County in central Pennsylvania (Aguero-Valverde & Jovanis, 2008). The proportions of total variability due to the spatial effects (ψ_s) were 0.406, 0.231 and 0.430 in total, severe and PDO models respectively. The values of ψ_s were less than 0.5, which indicated that the spatial effects in these models were not that much significant.

Table 13 Transit reliant CPM

Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
1.1.0	$T3 = 0.5885\ VKT^{0.7537} \exp[2.6308\ BSKD + 0.0055\ RTKT - 0.0227\ BTKT]$	2.17	Constant=-1.3, VKT=14, BSKD=9, RTKT=3, BTKT=-3, κ=16	1.1.a	$T3 = 3.3703\ VKT^{0.558} \exp[1.492\ BSKD + 0.007\ RTKT - 0.019\ BTKT]$	1.59	Constant=3.4, VKT=8, BSKD=8, RTKT=5, BTKT=-3, κ=15	0.406
1.2.0	$S3 = 0.0219\ VKT^{0.9873} \exp[3.0285\ BSKD - 0.0584\ BTKT]$	1.16	Constant=-8, VKT=18, BSKD=9, BTKT=-4, κ=16	1.2.a	$S3 = 0.0964\ VKT^{0.825} \exp[2.086\ BSKD - 0.032\ BTKT]$	1.14	Constant=-5, VKT=14, BSKD=11, BTKT=-6, κ=16	0.231
1.3.0	$PDO3 = 1.0297\ VKT^{0.6599} \exp[2.2423\ BSKD + 0.0066\ RTKT - 0.0205\ BTKT]$	2.03	Constant=0.1, VKT=12, BSKD=8, RTKT=4, BTKT=-3, κ=16	1.3.a	$PDO3 = 4.0149\ VKT^{0.505} \exp[1.417\ BSKD + 0.008\ RTKT - 0.017\ BTKT]$	1.58	Constant=2.6, VKT=7, BSKD=8, RTKT=6, BTKT=-3, κ=16	0.430

Table 14 shows the application based collision prediction models related to the change in exposure. Similar to the results of transit reliant models, the parameter estimates, the values of dispersion parameters, the values of intercepts, and the t-statistics were affected by the incorporation of the spatial correlations. The effect of VKT on the number of total, severe and PDO collisions were comparatively less in spatial models, although VKT remained to be the most dominant variable. El-Basyouny and Sayed (El-Basyouny & Sayed, 2009), and Aguero-Valverde and Jovanis (Aguero-Valverde & Jovanis, 2008) also found similar results. The proportions of spatial effects (ψ_s) were 0.558, 0.387 and 0.566 in total, severe and PDO models respectively. So, the spatial effects in total and PDO collisions were significant indicating that the neighbours of zones with high predicted total and PDO collisions tend to have higher predicted total and PDO collisions respectively, and vice versa (Aguero-Valverde & Jovanis, 2008; El-Basyouny & Sayed, 2009). There was also moderate spatial effect on the occurrence of severe collisions.

Table 14 Application based CPMs: change in exposure

Model Name	Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
Change in Exposure	2.1.0	$T3 = 3.8023\ VKT^{0.607}$	1.42	Constant=3.9, VKT=14, κ =17	2.1.a	$T3 = 3.6656\ VKT^{0.592}$	1.27	Constant=3.8, VKT=14, κ =18	0.558
	2.2.0	$S3 = 0.0448\ VKT^{0.9555}$	1.02	Constant=-7, VKT=17, κ =16	2.2.a	$S3 = 0.1969\ VKT^{0.783}$	1.00	Constant=-4, VKT=16, κ =15	0.387
	2.3.0	$PDO3 = 4.0338\ VKT^{0.5676}$	1.47	Constant=4.2, VKT=14, κ =17	2.3.a	$PDO3 = 3.8884\ VKT^{0.555}$	1.26	Constant=4, VKT=13, κ =17	0.566

CPMs based on the green house gas reduction/sustainable environment were presented in Table 15. Incorporation of spatial effect made the intercepts of total and PDO collision models not significant, and lowered the exponent values of VKT in all models. Significant effect of spatial correlation was not found in any model, although moderate spatial effect was observed in model 3.1.a.

Table 15 Application based CPMs: GHG emission reduction / sustainable environment

Model Name	Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
GHG Emission Reduction / Sustainable Environment	3.1.0	T3 = 0.3260 $VKT^{0.8581}$ exp [0.3959 TKTP]	1.72	Constant=-3, VKT=19, TKTP=9, κ=17	3.1.a	T3 = 1.1688 $VKT^{0.703}$ exp [0.394 TKTP]	1.49	Constant=0.23, VKT=9, TKTP=10, κ=16	0.411
	3.1.1	T3 = 0.3577 $VKT^{0.8376}$ exp [0.0271 PEDP + 0.1608 TKTP + 0.0599 BIKP]	1.83	Constant=-2.8, VKT=19, PEDP=6, TKTP=4, BIKP=3, κ=17	3.1.b	T3 = 0.8976 $VKT^{0.727}$ exp [0.022 PEDP + 0.16 TKTP + 0.07 BIKP]	1.67	Constant=-0.2, VKT=9, PEDP=6, TKTP=3, BIKP=3, κ=15	0.249
	3.2.0	S3 = 0.0143 $VKT^{0.9934}$ exp [0.3046 TKTP]	1.05	Constant=-8.3, VKT=16, TKTP=6, κ=16	3.2.a	S3 = 0.06614 $VKT^{0.88}$ exp [0.296 TKTP]	1.11	Constant=-5.5, VKT=15, TKTP=8, κ=16	0.315
	3.3.0	PDO3 = 0.3597 $VKT^{0.8026}$ exp [0.1806 TKTP + 0.026 PEDP + 0.0729 BIKP]	1.90	Constant=-2.9, VKT=18, TKTP=4, PEDP=6, BIKP=3, κ=16	3.3.a	PDO3 = 0.9277 $VKT^{0.689}$ exp [0.178 TKTP + 0.021 PEDP + 0.083 BIKP]	1.90	Constant=-0.12, VKT=9, TKTP=4, PEDP=5, BIKP=43, κ=16	0.235

Table 16 shows the application based collision prediction models based on the change in modal share. As expected, the effect of VKT on the number of total, severe and PDO collisions were comparatively less in spatial models, although VKT remained to be the most dominant variable. Models 4.1.b, 4.1.c and 4.3.b showed significant effects of spatial correlations on the occurrence of collisions as their ψ_s were 0.659, 0.571 and 0.688 respectively. Model 4.2.b also showed considerable amount of spatial effects.

Table 16 Application based CPMs: change in modal share

Model Name	Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
Change in Modal Share	4.1.0	T3 = 6.0279 VKT $^{0.7858}$ exp [- 0.0264 AUTOP]	2.04	Constant=5.3, VKT=19, AUTOP=-13, κ=16	4.1.a	T3 = 7.9725 VKT$^{0.699}$ exp [- 0.021 AUTOP]	1.78	Constant=5, VKT=12, AUTOP=-12, κ=16	0.348
	4.1.1	T3 = 1.5869 VKT $^{0.6049}$ exp [0.0536 BUSP]	1.93	Constant=1.39, VKT=15, BUSP=12, κ=16	4.1.b	T3 = 2.3467 VKT$^{0.564}$ exp [0.051 BUSP]	1.64	Constant=2.1, VKT=12, BUSP=11, κ=17	0.659
	4.1.2	T3 = 0.5642 VKT $^{0.739}$ exp [0.0391 BUSP + 0.0251 PEDP]	1.99	Constant=-1.64, VKT=17, BUSP=8, PEDP=7, κ=16	4.1.c	T3 = 1.3192 VKT$^{0.642}$ exp [0.036 BUSP + 0.022 PEDP]	1.75	Constant=0.49, VKT=9, BUSP=8, PEDP=6, κ=14	0.571
	4.1.3	T3 = 2.0643 VKT $^{0.7983}$ exp [0.0146 PEDP - 0.0164 AUTOP + 0.0645 BIKP]	2.19	Constant= 1.5, VKT=18, PEDP=3, AUTOP=-5, BIKP=3, κ=16	4.1.d	T3 = 2.1340 VKT$^{0.715}$ exp [0.022 PEDP - 0.008 AUTOP + 0.06 BIKP]	1.72	Constant= 1.3, VKT=10, PEDP=5, AUTOP=-3, BIKP=3, κ=16	0.296
	4.2.0	S3 = 0.2284 VKT $^{0.9687}$ exp [- 0.0233 AUTOP]	1.10	Constant=-3.3, VKT=17, AUTOP=-9, κ=16	4.2.a	S3 = 0.4369 VKT$^{0.891}$ exp [- 0.023 AUTOP]	1.23	Constant=-2.3, VKT=19, AUTOP=-13, κ=17	0.262
	4.2.1	S3 = 0.0443 VKT $^{0.8219}$ exp [0.0678 BUSP]	1.28	Constant=-7.4, VKT=16, BUSP=13, κ=16	4.2.b	S3 = 0.1011 VKT$^{0.733}$ exp [0.0678 BUSP]	1.29	Constant=-5.5, VKT=13, BUSP=13, κ=16	0.477

Model Name	Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
Change in Modal Share	4.3.0	PDO3 = 0.7820 $VKT^{0.7812}$ exp [0.0269 PEDP + 0.0988 BIKP - 0.0063 AUTOP]	2.04	Constant=-0.6, VKT=18, PEDP=5, BIKP=4, AUTOP=-2, κ=16	4.3.a	PDO3 = 2.0897 $VKT^{0.672}$ exp [0.023 PEDP + 0.077 BIKP - 0.007 AUTOP]	1.71	Constant=1.2, VKT=10, PEDP=5, BIKP=4, AUTOP=-3, κ=16	0.319
	4.3.1	PDO3 = 1.4817 $VKT^{0.5845}$ exp [0.0522 BUSP]	1.96	Constant=1.1, VKT=14, BUSP=12, κ=16	4.3.b	PDO3 = 2.6939 $VKT^{0.522}$ exp [0.049 BUSP]	1.56	Constant=2.6, VKT=11, BUSP=11, κ=14	0.688

Collision prediction models due to the change in network alignment are presented in Table 17. Apart from comparatively smaller values of exponents of VKT, the spatial models showed that total and PDO collision occurrences were significantly affected by the spatial correlations.

Table 17 Application based CPMs: change in network alignment

Model Name	Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
Change in Network Alignment	5.1.0	$T3 = 10.7855\ VKT^{0.6059} \exp[-0.0130\ I3WP - 0.0190\ SBNCP - 0.0270\ HBNCP]$	2.44	Constant=8.0, VKT=17, I3WP=-7, SBNCP=-5, HBNCP=-6, κ=16	5.1.a	$T3 = 14.9693\ VKT^{0.566} \exp[-0.0130\ I3WP - 0.0190\ SBNCP - 0.02\ HBNCP]$	1.75	Constant=5.4, VKT=9, I3WP=-8, SBNCP=-5, HBNCP=-3, κ=14	0.629
	5.2.0	$S3 = 0.4553\ VKT^{0.8270} \exp[-0.014\ I3WP - 0.0217\ SBNCP - 0.0387\ HBNCP]$	1.49	Constant=-1.94, VKT=17, I3WP=-6, SBNCP=-4 HBNCP=-7, κ=16	5.2.a	$S3 = 0.8728\ VKT^{0.752} \exp[-0.016\ I3WP - 0.023\ SBNCP - 0.026\ HBNCP]$	1.18	Constant=-0.3, VKT=16, I3WP=-7, SBNCP=-6 HBNCP=-3, κ=16	0.397
	5.3.0	$PDO3 = 12.7\ VKT^{0.5522} \exp[-0.0131\ I3WP - 0.0171\ SBNCP - 0.0254\ HBNCP]$	2.42	Constant=8.4, VKT=15, I3WP=-7, SBNCP=-4, HBNCP=-6, κ=16	5.3.a	$PDO3 = 14.6436\ VKT^{0.534} \exp[-0.012\ I3WP - 0.019\ SBNCP - 0.022\ HBNCP]$	1.69	Constant=5.3, VKT=8, I3WP=-6, SBNCP=-4, HBNCP=-3, κ=15	0.675

Table 18 shows the negative binomial and spatial models related to the change in total, severe and PDO collisions due to economic growth. Although, the parameter estimates, the values of dispersion parameters and intercepts, and the t-statistics were affected by the incorporation of the spatial correlations, the proportions of spatial effects (ψ_s) were significantly smaller than 0.5. So, the spatial effects in total, severe and PDO collisions were not significant in these models.

Table 18 Application based CPMs: economic growth

Model Name	Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
Economic Growth	6.1.0	T3 = 49.1610 $VKT^{0.8411}$ exp [- 0.0508 EMP + 0.3027 WKGD]	1.81	Constant=4.5, VKT=19, EMP=-6, WKGD=5, κ=16	6.1.a	T3 = 159.3336 $VKT^{0.664}$ exp [- 0.047 EMP +0.213 WKGD]	1.37	Constant=7, VKT=12, EMP=-6, WKGD=5, κ=16	0.279
	6.1.1	T3 = 2.6290 $VKT^{0.8215}$ exp [- 0.0257 PARP + 0.3662 WKGD]	1.84	Constant=2.3, VKT=20, PARP=-6, WKGD=6, κ=16	6.1.b	T3 = 5.9299 $VKT^{0.663}$ exp [- 0.016 PARP +0.291 WKGD]	1.31	Constant=3.4, VKT=12, PARP=-3, WKGD=8, κ=16	0.199
	6.2.1	S3 = 0.4795 $VKT^{0.8460}$ exp [- 0.0250 PARP + 0.2530 WKGD]	1.00	Constant=-1.2, VKT=13, PARP=-4, WKGD=4, κ=30	6.2.a	S3 = 0.5081 $VKT^{0.84}$ exp [- 0.025 PARP +0.255 WKGD]	1.00	Constant=-1.1, VKT=14, PARP=-4, WKGD=6, κ=28	0.171
	6.3.0	PDO3 = 30.0498 $VKT^{0.7576}$ exp [- 0.0427 EMP + 0.3219 WKGD]	1.79	Constant=4.0, VKT=18, EMP=-5, WKGD=5, κ=17	6.3.a	PDO3 = 73.1126 $VKT^{0.631}$ exp [- 0.038 EMP +0.238 WKGD]	1.36	Constant=5.9, VKT=11, EMP=-5, WKGD=6, κ=17	0.301
	6.3.1	PDO3 = 2.8016 $VKT^{0.7744}$ exp [- 0.0245 PARP + 0.3771 WKGD]	1.87	Constant=2.5, VKT=19, PARP=-6, WKGD=6, κ=16	6.3.b	PDO3 = 5.4466 $VKT^{0.628}$ exp [- 0.015 PARP +0.302 WKGD]	1.31	Constant=3.2, VKT=11, PARP=-3, WKGD=8, κ=14	0.234

Table 19 shows the application based collision prediction models related to the change in accessibility and mobility. Similar to the results of the previous models, the parameter estimates and corresponding t-statistics were changed due to the incorporation of the spatial correlations. The proportions of spatial effects (ψ_s) were 0.535, 0.332 and 0.555 in total, severe and PDO spatial models respectively. So, the spatial effects in total and PDO collisions were significant indicating that the neighbours of zones with high predicted total and PDO collisions tend to have higher predicted total and PDO collisions respectively, and vice versa (Aguero-Valverde & Jovanis, 2008; El-Basyouny & Sayed, 2009). The spatial effect on the occurrence of severe collisions was comparatively small.

Table 19 Application based CPMs: change in accessibility-mobility

Model Name	Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
Change in Accessibility / Mobility	7.1.0	T3 = 2.4727 $VKT^{0.7502}$ exp [- 0.0168 I3WP + 5.6008 SIGD - 0.0067 ALKP]	2.46	Constant=2.6, VKT=19, I3WP=-11, SIGD=7, ALKP=-3, κ=16	7.1.a	T3 = 5.3016 $VKT^{0.662}$ exp [- 0.016 I3WP + 4.632 SIGD - 0.007 ALKP]	1.87	Constant=2.3, VKT=7, I3WP=-10, SIGD=7, ALKP=-3, κ=15	0.535
	7.2.0	S3 = 0.1117 $VKT^{0.9456}$ exp [- 0.0207 I3WP + 4.0410 SIGD]	1.34	Constant=-4.6, VKT=17, I3WP=-9, SIGD=4, κ=16	7.2.a	S3 = 0.2690 $VKT^{0.842}$ exp [- 0.02 I3WP + 3.03 SIGD]	1.23	Constant=-2.4, VKT=13, I3WP=-12, SIGD=5, κ=16	0.332
	7.3.0	PDO3 = 3.0560 $VKT^{0.7021}$ exp [- 0.0169 I3WP + 6.3494 SIGD - 0.0097 ALKP]	2.55	Constant=3.2, VKT=18, I3WP=-11, SIGD=8, ALKP=-4, κ=16	7.3.a	PDO3 = 5.2646 $VKT^{0.633}$ exp [- 0.015 I3WP + 5.097 SIGD - 0.008 ALKP]	1.84	Constant=2.2, VKT=7, I3WP=-8, SIGD=8, ALKP=-3, κ=15	0.555

Negative binomial and spatial collision prediction models related to urban sprawl were shown in Table 20. Unlike the previous models, inclusion of spatial correlations hardly changed the negative binomial model for total collisions. However, severe and PDO negative binomial models changed when spatial effects were considered. PDO spatial model showed that the neighbours of zones with high predicted PDO collisions tend to have higher predicted PDO collisions, and vice versa, as the value of the proportions of spatial effects (ψ_s) was found to be 0.743.

Table 20 Application based CPMs: urban sprawl

Model Name	Model ID	Model Form	κ	t-statistics	Model ID	Model Form	κ	t-statistics	Ψ_s
Urban Sprawl	8.1.0	$T3 = 0.2549\ VKT^{0.8206} \exp[1.0391\ INTD + 4.8375\ SIGD + 0.0061\ IALP + 0.0079\ NHD]$	2.27	Constant=-4.3, VKT=21, INTD=7, SIGD=6, IALP=2.4, NHD=3, κ=16	8.1.a	$T3 = 0.2549\ VKT^{0.8106} \exp[1.0391\ INTD + 4.8375\ SIGD + 0.0061\ IALP + 0.0079\ NHD]$	1.59	Constant=-4.3, VKT=21, INTD=7, SIGD=6, IALP=2.4, NHD=3, κ=16	0.406
	8.1.1	$T3 = 0.2558\ VKT^{0.8164} \exp[1.0612\ INTD + 4.4975\ SIGD + 0.0065\ IALP + 0.0033\ POPD]$	2.31	Constant=-4.4, VKT=21, INTD=6, SIGD=6, IALP=3, POPD=3, κ=16	8.1.b	$T3 = 0.2558\ VKT^{0.8064} \exp[1.0612\ INTD + 4.4975\ SIGD + 0.0065\ IALP + 0.0033\ POPD]$	1.59	Constant=-4.4, VKT=21, INTD=6, SIGD=6, IALP=3, POPD=3, κ=15	0.403
	8.2.0	$S3 = 0.0084\ VKT^{0.9957} \exp[1.6695\ INTD + 3.2263\ SIGD + 0.0153\ IALP]$	1.40	Constant=-11, VKT=19, INTD=9, SIGD=3, IALP=5, κ=16	8.2.a	$S3 = 0.026\ VKT^{0.877} \exp[1.62\ INTD + 2.427\ SIGD + 0.016\ IALP]$	1.29	Constant=-8, VKT=15, INTD=10, SIGD=3, IALP=5, κ=15	0.397
	8.3.0	$PDO3 = 0.2902\ VKT^{0.7747} \exp[1.0038\ INTD + 4.6926\ SIGD + 0.0054\ IALP + 0.0035\ POPD]$	2.33	Constant=-4.0, VKT=20, INTD=6, SIGD=6, IALP=2.3, POPD=3, κ=18	8.3.a	$PDO3 = 0.6984\ VKT^{0.673} \exp[1.016\ INTD + 3.762\ SIGD + 0.005\ IALP + 0.004\ POPD]$	1.91	Constant=-0.5, VKT=8, INTD=6, SIGD=6, IALP=2.5, POPD=4, κ=16	0.743

4.5 Summary

This chapter has presented the developed transit reliant and application based negative binomial and spatial collision prediction models. The developed models are classified in different groups for total, severe and property damage only collisions with corresponding goodness of fit statistics. The CPMs deal with the number of collision frequencies associated with the changes in exposure, green house gas (GHG) emissions, percentages of people using different modes of a transportation system, network alignment, economic growth, accessibility and mobility of a zonal transportation network, and urban sprawl. Analyses of possible casual mechanisms with respect to the sign and magnitude of the coefficient of each variable have also been presented with explanations and comparisons with previous studies.

The negative binomial models indicate that the collision frequencies increase with an increase in vehicle-km traveled, number of workers per resident, number of signalized intersections per unit area, number of intersections per unit area, population density, percentage of arterial-local intersections, and bus stop density. However, fewer collisions are associated with longer average transit trip length, and percentage of three-way intersections. This study also indicates that percentage of transit-km traveled with regard to total vehicle-km traveled (TKTP), percentage of commuters walking (PEDP), percentage of commuters biking (BIKP) and percentage of commuters using transit (BUSP) have positive exponents. However, an increase in any of these variables (TKTP, PEDP, BIKP, and BUSP) will likely to cause a decrease in percentage of commuters using autos which eventually will yield a decrease in vehicle-km traveled (VKT). VKT also has a positive exponent but the value of this exponent is much higher than the coefficients TKTP, PEDP, BIKP and BUSP. Ultimately an increase in TKTP, PEDP, BIKP and BUSP do not necessarily refer to a less safe condition of the zonal transportation network.

The spatial models have shown that the incorporation of the spatial correlation has affected the parameter estimates, the values of dispersion parameters and intercepts, and also the t-statistics. The effect of VKT on all of the models for total, severe and PDO collisions have been found to be smaller under spatial models. The proportions of total variability due to the spatial effects (ψ_s) range from 0.171 (Model 6.2.a) to 0.743 (Model 8.3.a). Spatial models are usually considered significant for ψ_s values greater than 0.5.

5 Conclusions and Recommendations

5.1 Introduction

This chapter consists of three main sections. Section 5.2 contains the summary of the study along with the motivation behind this research and the main research conclusions. It is followed by Section 5.3 which briefly discusses the overall significance and research contributions. Finally, Section 5.4 describes some recommendations for future research, which can strengthen the methodologies used in this study by promoting consistency in proactive evaluation of city transportation plans.

5.2 Summary and conclusions

The motivation for this research results from the necessity of developing a framework that can predict the zonal safety effect of changes in the transportation and transit network configurations, and ultimately can evaluate the safety estimates among alternatives of different transportation plans and policies. While modern transportation planning considers issues such as road congestion, pollution and mobility proactively, road safety has usually not been dealt with while choosing between the alternatives of different transportation plans and policies. Road safety is usually evaluated only after the implementation of transportation plans, and when safety problems arise. This reactive road safety approach has been effective in reality, but at the same time it is associated with significant costs on communities. Therefore, some studies have developed macro-level CPMs that can assess the road safety in a proactive manner and provide a safety planning decision support tool to community planners and engineers. However, this attempt has limitation due to lack of availability of macro-level CPMs that are solely based on the safety evaluation of different targets/goals of a city transportation plan.

The city authority of Metro Vancouver as well as TransLink want to see a change in people's travel behaviour that would result in a transit-oriented city with a transportation system meeting the needs of residents, business, and goods movers, in a manner that protects the environment and supports the economic and social objectives of the region. This approach targets at reducing the reliance on autos and making massive investments in the development of alternatives to autos, such as public transit, and active forms of transportation including walking and biking. This plan identifies six key goals and proposes four strategies to reach those goals. However, there is a need to develop tools to estimate the safety impact of these goals. But the safety Therefore, the motivation for this research thus resulted from the necessity of developing some macro-level collision prediction models that can provide one additional input to transportation planners as well as city authorities to comprehensively assess the future network scenarios and ultimately select the best alternatives among various transportation plans.

Most of the previous studies on macro level collision prediction models have considered the Poisson variations and the heterogeneity (extra-variation) on the occurrence of collisions. But the spatial effect caused by the similarity in environment and geography of the neighbouring sites on the distribution of collisions is typically not considered. However, some previous studies have observed that the incorporation of spatial effect into CPMs not only improves the goodness of fit, but also it can explain enough variation that some of the model covariates would deem non-significant. Therefore, another purpose of this research was to investigate and evaluate the spatial effects on the distribution of collisions.

The first objective of this study was to thoroughly screen the possible data sources to ensure the quality of data. The data extraction sources included geo-coded database of ICBC, TransLink's digital road map and network data, Metro Vancouver land use map, and demographic database of Census Canada for 485 TAZs of Metro Vancouver. City transportation plans of different cities were carefully analyzed and proxy variables related to each goal/target of a city transportation plan were sorted out. Around 58 explanatory variables were found that could be included in model development. Out of these 58 available variables, 30 variables were chosen based on the relevance to planning and transit, availability at the planning stage, and easiness of forecasting. The variables were then classified based on the target to be achieved and the corresponding variables for each model type were distributed with respect to the relevance with achieving that target.

The next objectives were to develop transit reliant and application-based macro-level (i.e. zone based) collision prediction models to quantify the impact of various transit network elements on the overall roadway safety as well as to evaluate the safety consequences of various goals/targets of a city transportation plan. Total, severe and property damage only collision prediction models for all traffic (transit and non transit) were developed in this study. The Generalized Linear

Modelling (GLM) approach was used in this research as it has the advantage of overcoming the limitations associated with the use of conventional linear regression. GLM process followed a forward stepwise procedure at 95% desired level of confidence. The error structure was assumed to follow a negative binomial distribution. The estimation of parameters was carried out by the maximum likelihood method in statistical analysis software package SAS. Then the outlier analysis was performed by finding and then removing the extremely influential outliers with the help of Cook's distance. The goodness of fit of the developed models was assessed by the Pearson χ^2 and scaled deviance values. Besides, significance of the t-statistics of all of the parameters was also ensured.

The developed models indicated that collision frequencies increased with an increase in vehicle-km traveled (VKT), number of workers per resident (WKGD), number of signalized intersections per unit area (SIGD), number of intersections per unit area (INTD), population density (POPD), percentage of arterial-local intersections (IALP) and bus stop density (BSKD). However, fewer collisions are associated with longer average transit trip length (BTKT) and percentage of three-way intersections (I3WP). VKT was found to be the most dominant variable in occurrence of collisions. The coefficients of percentage of transit-km traveled with regard to total vehicle-km traveled (TKTP), percentage of commuters walking (PEDP), percentage of commuters biking (BIKP) and percentage of commuters using transit (BUSP) were found to be positive. But at the same time an increase in any of these variables would likely to cause a decrease in percentage of commuters using autos (AUTOP) which eventually will yield a decrease in VKT. As the exponent of VKT is positive and comparatively larger, an increase in TKTP, PEDP, BIKP and BUSP did not necessarily refer to a less safe transportation network. However, further investigation is needed to explain the change in VKT due to modal shift from auto to transit and auto to active forms of transportation including walking and biking..

As the incorporation of spatial effect in collision prediction models was found to be highly significant in some of the previous studies, another objective of this research was to evaluate the spatial effects on the distribution of collisions and to check whether the inclusion of spatial variables can improve the goodness of fit and inference capability of the developed CPMs. First-order neighbouring structure was used in this study to develop the Spatial Poisson-Gamma CPMs. Gaussian conditional autoregressive (CAR) techniques were followed and Markov chain

Monte Carlo (MCMC) methodology was applied using WinBUGS 2.2.0, the Windows interface of Open-BUGS. The effects of spatial correlation were estimated by computing the proportion of total variation that was caused by spatial variation. CPMs with and without spatial effects were then compared based on the values of dispersion parameter (κ). The developed spatial models showed that the incorporation of the spatial correlation affected the parameter estimates, the values of dispersion parameters and intercepts, and also the t-statistics. The effect of VKT on all models for total, severe and PDO collisions were found to be smaller under spatial models. The smaller values of the exponents of VKT indicated the fact that not considering the spatial variables yielded a bias associated with model misspecifications. The proportions of total variability due to the spatial effects (ψ_s) ranged from 0.171 (Model 6.2.a) to 0.743 (Model 8.3.a) with an average value close to 0.5. Usually the spatial effects are considered significant for ψ_s values greater than 0.5. In these models significantly high proportions of the total variability could be explained by the spatial correlations. On average, the spatial effects on the occurrences of property damage only (PDO) collisions were found to be much higher than that on the occurrences of severe collisions. The reason might be that severe collisions were scarcer and it was harder to draw relationships among zones. However, further research should be performed to investigate this issue.

All of the independent variables and parameter estimates in all types of negative binomial and spatial Poisson-gamma collision prediction models were significant at the 5% level of significance. The variables selected for each type of models were of reasonable signs with respect to their expected tendency to increase or decrease with collision frequency. These variables were also found to be significant determinants of collision occurrences. The value of the dispersion parameters in all of the developed models confirmed that the data were over dispersed relative to the Poisson distribution and the negative binomial error structure was therefore justified. The ratios of deviance to degree of freedom also fell within the acceptable range of 0.8 to 1.2, which was another indication that the choice of the NB error structure appeared to be appropriate for these models. Only a few CPMs had scaled deviance values slightly exceeding their χ^2 statistics. However, all SD values were within 5 percent of the target χ^2 statistics. The t-values of the intercepts of some models were found to less than the critical t-statistic (<1.96), indicating that the constant term leading the CPM might not significantly differ from 1.0. But this phenomenon did not pose significant difficulty as the significance of these intercepts was marginal.

Therefore, this study developed a proactive tool that can address and evaluate the safety estimate at the very early stage of transportation planning. The developed models can also give a clear understanding on how the change in people's travel behaviour, more reliance on transit as well as a transit-oriented society approach affect the overall zonal roadway safety. Moreover, effect of the spatial incorporation on the occurrences of collisions were also investigated and quantified in this study. To sum up, these collision prediction models captured the interaction between the collision frequency of all traffic (transit and non-transit) of different type and severity, and the implemental aspects related to the goals of long-term transportation plans including travel demand management policy impacts and transit investment implications.

5.3 Research contributions

The overall significance and main contributions of this research are presented in the following paragraphs.

Development of transit reliant and application based macro-level collision prediction models based on city transportation plans is feasible and successful. Various cities and transportation authorities across Canada have adopted long-term transportation plans with a view to creating and sustaining a transportation system that meets the needs of residents, in a sustainable and environment friendly manner. Each plan contains different goals and targets to achieve this objective. But the safety estimates of these goals are rarely evaluated at the planning stage. Therefore, it is obvious that an effective approach should be taken where road safety will be assessed at the early stage of transportation planning process so that the safest plan among various alternatives can be sorted out. Analyzing the transportation plans of different cities, this study found out some proxy variables related to each goal/target of a city transportation plan. Those variables were also relevant to planning and transit, available at the planning stage, and easy to forecast. Micro-level CPMs have very successfully been used in reactive road safety engineering approach. But application of micro-level CPMs to evaluate safety in planning stage has been largely obstructed by some inherent empirical limitations of these CPMs. Therefore, this study developed macro-level (zone-based) transit reliant and seven types of application based collision prediction models which were solely based on the safety evaluation of different transit network configurations as well as various targets/goals of a city transportation plan. The statistical tests of these developed models provided enough evidence that the safety estimates of the first five goals of TransLink's *Transport 2040* transportation plan for Metro Vancouver could be successfully evaluated. Moreover, the t-test statistics indicated the fact that the proxy variables related to each goal were statistically significant for the occurrences of total, severe and property damage only (PDO) collisions. So, zone level CPMs can successfully be developed based on the goals of city transportation plans.

The best transportation plan with respect to safety among various alternatives can be sorted out at the planning stage. City transportation authority will largely benefit from this approach as it will help them find out the safest plan at the very early stage of the transportation planning process. Presently the city transportation authorities consider various transportation issues such as road congestion, pollution, accessibility, and mobility while choosing among the best transportation plan and policy among more than one alternative. But road safety is usually not considered at this stage because of lack of available tools that can explicitly evaluate the safety estimate of each goal and target of any plan in a proactive manner. This limitation results in significant costs on communities and societies not only across Canada, but also all over the world as the retrofitting countermeasures at collision prone locations in existing transportation network is usually very costly. Using the transit reliant and application based zone level collision prediction models developed in this study, transportation planners as well as city authorities can have a clear understanding on how the change in people's travel behaviour, more reliance on active forms of transportation including walking and biking, as well as a transit-oriented society approach affect the overall zonal roadway safety. As these CPMs captured the complex interaction between the collision frequency of all traffic (transit and non-transit) of different type and severity, and various implemental aspects related to the goals of long-term transportation plans, it is an important and effective proactive tool to the transportation planners and city authorities to sort out the best transportation plan with respect to safety among various alternatives.

Incorporation of spatial correlations significantly improves the inference capability of collision prediction models. The spatial Poisson-gamma models in this research indicated that incorporation of the spatial correlation affected the parameter estimates, the values of dispersion parameters and intercepts, and also the t-statistics. The average proportions of total variability due to the spatial effects were close to 0.5, and the spatial effects on the occurrence of PDO collisions were more significant than on severe collisions. Therefore, these models indicated that significantly high proportions of the total variability could be explained by the spatial correlations. Besides, all of the significant variables in negative binomial models remained significant in spatial models as well. However, the total, severe and PDO spatial models had comparatively smaller coefficients of VKT indicating that not considering the spatial variables yielded a bias associated with model misspecifications. Hence, with a little more extra effort, spatial models could significantly improve the inference capability of transit reliant and application based zone level collision prediction models. So, it is highly advocated that the city authority and safety researchers should try to incorporate the spatial effects in all the CPMs.

5.4 Recommendations for future research

This section proposed some recommendations on future research directions to strengthen and intensify the methodologies used in this study as well as to promote consistency in proactive evaluation of the city transportation plan.

- **Improved Data Quality and Variable Selection Method-** The goodness of fit of the developed models found to be reasonable. All of the explanatory variables seemed to be statistically significant with respect to the t-statistic. Still the models may be improved by ensuring better quality of data and including the variables that might have been omitted. In this study collision and explanatory data were extracted and then aggregated by a combination of manual, modelled (Emme/2) and automated (GIS) techniques. Manual extraction was done to identify which zones were predominantly urban or rural. All of these manual and modelled techniques can be replaced by an automated technique to maximize the quality of data. Besides, analyzing the city transportation plans and the previous studies on collision prediction models, a group of explanatory variables was sorted out and then classified based on the relevance to each goal of TransLink's transportation plan. So there remained a high possibility that this variable selection and classification process influenced the goodness of fit statistics of the developed models.

Further efforts are recommended to refine and improve this selection and stratification process, as well as to search for more potential variables.

- **Transit Collision Prediction Models -** The developed negative binomial and spatial models in this study appeared to have great potential to be applied in a proactive manner to evaluate the safety estimates of different goals of a city transportation plan. However, these models considered total, severe and PDO collisions for all traffics (transit and non-transit both). Similar models may be developed considering the transit collisions only. At the same time other groups of models may be developed for collisions regarding active forms of transportation including walking and biking.
- **Modal Shift Impact on Explanatory Variables -** This study found positive coefficients for percentage of transit-km traveled with regard to total vehicle-km traveled (TKTP), percentage of commuters walking (PEDP), percentage of commuters biking (BIKP) and percentage of commuters using transit (BUSP). It was expected that an increase in any of these variables would cause a decrease in percentage of commuters using autos (AUTOP) which eventually will yield a decrease in vehicle-km traveled (VKT). This expectation was logical, but no other studies were found to establish this relationship. So, further investigation is recommended to explain the change in VKT caused by the modal shift from auto to transit and auto to active forms of transportation including walking and biking.
- **Higher Order Neighbouring Structure for Spatial Modelling -** The developed spatial models showed that incorporation of spatial correlations affected PDO collisions more than it affected severe collisions. The underlying mechanism behind this difference in spatial effect should further be investigated. This study also considered the first-order neighbouring structure while preparing the data for spatial analysis. In a future study, the effects of considering the second-order and the third-order neighbours may be evaluated. Besides, the distance from the centroid of each zone may be used to define and sort out the neighbours with order.
- **Evaluation of Transportation Plans for Other Canadian Cities -** This study used data from 485 TAZs of Metro Vancouver to develop collision prediction models. The models were based on five goals of TransLink's *Transport 2040* transportation plan. Almost all of the big cities across Canada have their own city transportation plans. Further research may

be undertaken to assess the transferability as well as inference capability of the developed negative binomial and spatial models to other cities.

References

Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in pennsylvania. *Accident Analysis and Prevention, 38*(3), 618-625.

Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of road crash frequency with spatial models. *Transportation Research Record,* (2061), 55-63.

Bedrick, E. J., Christensen, R., & Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association, 91*, 1450-1460.

Brude, U., & Larsson, J. (1993). Models for predicting accidents at junctions where pedestrians and cyclists are involved. how well do they fit? *Accident Analysis and Prevention, 25*(5), 499-509.

Buyco, C., & Saccomanno, F. F. (1988). Analysis of truck accident rates using loglinear models. *Canadian Journal of Civil Engineering, 15*(3), 397-408.

Census Canada. (1996). *Statistics from the 1996 census.* Ottawa, Ont.: Census Canada, Government of Canada.

Cheung, C. (2007). *Models for safety analysis of road surface transit.* Department of Civil Engineering, University of Toronto, Ontario, Canada.

Cheung, C., Shalaby, A. S., Persaud, B. N., & Hadayeghi, A. (2008). Models for safety analysis of road surface transit. *Transportation Research Record,* (2063), 168-175.

Cinneide, D. O., & Troutbeck, R. J. (1995). At-grade intersections/Worldwide review. Paper presented at the *Proceedings of the International Symposium on Highway Geometric Design Practices,* 26-1.

City of Vancouver. (2005). *Downtown transportation plan.*Downtown Transportation Plan Implementation Team.

De Guevara, F. L., Washington, S. P., & Oh, J. (2004). Forecasting crashes at the planning level : Simultaneous negative binomial crash model applied in tucson, arizona. (1897) 191-199.

De Leur, P., & Sayed, T. (2000). The development of an auto insurance claim prediction model for road safety evaluation in british columbia. *2000 Annual Conference - Canadian Society for Civil Engineering, June 7, 2000 - June 10,* 302.

De Leur, P., & Sayed, T. (2002). Development of a road safety risk index. (1784) 33-42.

de Leur, P., & Sayed, T. (2003). A framework to proactively consider road safety within the road planning process. *Canadian Journal of Civil Engineering, 30*(4), 711-719.

El-Basyouny, K., & Sayed, T. (2009). Urban arterial accident prediction models with spatial effects. *Transportation Research Record,* (2102), 27-33.

Elvik, R., Christensen, P., & Amundsen, A. (2004). *Speed and road accidents: An evaluation of the power model.* No. 740/2004). Oslo: Institute of Transport Economics (TOI).

Gaspers, K. (2004). *On the road to danger: WHO call traffic injuries a "global public health problem".* Illinois, USA: Safety & Health, National Safety Council.

Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2003). Macrolevel accident prediction models for evaluating safety of urban transportation systems. (1840) 87-95.

Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2007). Safety prediction models: Proactive tool for safety evaluation in urban transportation planning applications. *Transportation Research Record,* (2019), 225-236.

Haddon, W. (1980). Advances in the epidemiology of injuries as a basis for public policy. *Landmarks in American Epidemiology, 95*(5), 411.

Hauer, E., Ng, J. C. N., & Lovell, J. (1988). Estimation of safety at signalized intersections. *Transportation Research Record,* (1185), 48-61.

Hedman, K. O. (1990). Road design and safety. Paper presented at the *Proceedings of Strategic Highway Research Program and Traffic Safety on Two Continents,*

Ho, G., Nepomuceno, J., & Zein, S. R. (1998). *Introducing road safety audits and design safety reviews.* (Draft discussion paper prepared for ICBC Vancouver, Canada:

Huang, H., Abdel-Aty, M. A., & Darwiche, A. L. (2010). County-level crash risk analysis in florida: Bayesian spatial modeling. *Transportation Research Record: Journal of the Transportation Research Board,* (2148), 27-37. doi:10.3141/2148-04

ICBC. (2003). *Crash claim statistics for 1996, 1997, 1998.* North Vancouver, B.C.: Road Safety Program, Insurance Corporation of British Columbia (ICBC).

ICBC. (2008). *Traffic collision statistics: Police attended injury and fatal collisions: British columbia 2007.* Victoria, British Columbia: Road Safety Research, Insurance Corporation of British Columbia (ICBC). Retrieved from British Columbia traffic collision statistics / Motor Vehicle Branch.

Jegede, F. J. (1988). Spatio-temporal analysis of road traffic accidents in oyo state, nigeria. *Accid. Anal. Prev., 20*(3), 227-243.

Jensen, S. U. (2008). Bicycle tracks and lanes: A before-after study. Paper presented at the *TRB 87th Annual Meeting Compendium of Papers DVD,*

Jovanis, P. P., & Chang, H. (1986). Modeling the relationship of accidents to miles traveled. *Transportation Research Record,* , 42-51.

Jovanis, P. P., Schofer, J. L., Prevedouros, P., & Tsunokawa, K. (1991). Analysis of bus transit accidents: Empirical, methodological, and policy issues. *Transportation Research Record, 1322*, 17-28.

Keeler, T. E. (1994). Highway safety, economic behavior, and driving enforcement. *The American Economic Review, 84*(3), 684-693.

Khondaker, B. (2008). *Transferability of community-based macro-level collision prediction models for use in road safety planning applications.* (Master of Applied Science, The University of British Columbia, Vancouver). Retrieved from http://hdl.handle.net/2429/2867

Kiattikomol, V., Chatterjee, A., Hummer, J. E., & Younger, M. S. (2008). Planning level regression models for prediction of crashes on interchange and noninterchange segments of urban freeways. *Journal of Transportation Engineering, 134*(3), 111-117.

Kulmala, R. (1996). Safety at rural three- and four-arm junctions. development and application of accident prediction models. *VTT Publications,* (233), X-104.

Levine, N., Kim, K. E., & Nitz, L. H. (1995a). Spatial analysis of honolulu motor vehicle crashes: I. spatial patterns. *Accident Analysis and Prevention, 27*(5), 663-663.

Levine, N., Kim, K. E., & Nitz, L. H. (1995b). Spatial analysis of honolulu motor vehicle crashes: II. zonal generators. *Accident Analysis and Prevention, 27*(5), 675-675.

Lord, D. (2000). *The prediction of accidents on digital networks: Characteristics and issues related to the application of accident prediction models.* (Unpublished

Lord, D., & Persaud, B. N. (2004). Estimating the safety performance of urban road transportation networks. *Accident Analysis and Prevention, 36*(4), 609-620.

Lovegrove, G. R. (2006). *Community-based, macro-level collision prediction models.* (Doctor of Philosophy, The University of British Columbia, Vancouver). Retrieved from http://hdl.handle.net/2429/18172

Lovegrove, G. R., Lim, C., & Sayed, T. (2010). Community-based, macrolevel collision prediction model use with a regional transportation plan. *Journal of Transportation Engineering, 136*(2), 120-128.

Lovegrove, G. R., & Sayed, T. (2006a). Macro-level collision prediction models for evaluating neighbourhood traffic safety. *Canadian Journal of Civil Engineering, 33*(5), 609-621.

Lovegrove, G. R., & Sayed, T. (2006b). Using macrolevel collision prediction models in road safety planning applications. *Statistical Methods and Crash Prediction Modeling,* (1950) 73-82.

Lovegrove, G. R., & Sayed, T. (2007). Macrolevel collision prediction models to enhance traditional reactive road safety improvement programs. *Transportation Research Record,* (2019), 65-73.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. New York: Chapman and Hall.

Miaou, S., & Lum, H. (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis and Prevention, 25*(6), 689-709.

Millot, M. (2004). The influence of urban planning on road safety. Paper presented at the *Proceedings of European Transport Conference,* Strasbourg, France.

Milton, J., & Mannering, F. (1998). The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation, 25*(4), 395-413.

Miranda-Moreno, L. F., Morency, P., & El-Geneidy, A. M. (2010). How does built environment influence pedestrian activity and pedestrian collisions at intersections? Paper presented at the *89th Annual Meeting of the Transportation Research Board,*

Moudon, P. M., Hess, P. M., & Matlick, J. M. (2004). Pedestrian safety and transit corridors. *Journal of Public Transportation, 7*(2), 74-93.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., & Mathers, C. (2004). *World report on road traffic injury prevention.* (WHO Library Cataloguing No. WA 275). Geneva, Switzerland: World Health Organization. doi:ISBN 92 4 156260 9

Persaud, B., & Dzbik, L. (1993). *Accident prediction models for freeways.*

Poppe, F. (1997). *Traffic models: Inner areas and road dangers.* No. R-97-10). Leidschendam, The Netherlands: SWOV Institute for Road Safety Research.

Retting, R. A. (1996). Urban motor vehicle crashes and potential countermeasures. *Transportation Quarterly, 50*(3), 19-31.

Saccomanno, F. F., Fu, L., & Roy, R. K. (2007). Geographic information System—Based integrated model for analysis and prediction of road accidents. *Transportation Research Record: Journal of the Transportation Research Board, 1768*(2001), 193-202.

Sawalha, Z., & Sayed, T. (1999). *Accident prediction and safety planning models for urban arterial roadways.* (Draft reportThe Insurance Corporation of British Columbia.

Sawalha, Z., & Sayed, T. (2006). Traffic accident modelling:Some statistical issues. *Canadian Journal of Civil Engineering, 33*, 1115-1124. doi:10.1139/L06-056

Sayed, T., & Rodriguez, F. (1999). Accident prediction models for urban unsignalized intersections in british columbia. *Transportation Research Record,* (1665), 93-99.

Schluter, P. J., Deely, J. J., & Nicholson, A. J. (1997). Ranking and selecting motor vehicle accident sites by using a hierarchical bayesian model. *Statistician, 46*(3), 293-316.

Shahla, F., Shalaby, A. S., Persaud, B. N., & Hadayeghi, A. (2009). Analysis of transit safety at signalized intersections in toronto, ontario, canada. *Transportation Research Record,* (2102), 108-114.

Société de Transport de Montréal. (2008). *Transportation Plan|2008.*

TransLink. (2003). *Greater vancouver regional transportation model, input/output files, scenario 1996.* Greater Vancouver Regional District, Burnaby, B.C.: Strategic Planning Department.

TransLink. (2008). *Transport 2040: A transportation strategy for metro vancouver, now and in the future.*

Transport Canada. (2008). *A quick look at intersection crashes in Canada* . Retrieved September 07, 2010, from http://www.tc.gc.ca/eng/roadsafety/tp-tp2436-rs200806-menu-839.htm

Transport Canada. (2010). *Canadian motor vehicle collision statistics: 2007.* No. TP 3322). Ottawa, Ontario: Road Safety and Motor Vehicle Regulation Directorate, Transport Canada. Retrieved from http://www.tc.gc.ca/Publications/en/TP3322/PDF/HR/TP3322E_2007.pdf

Transportation Research Board (TRB). (1984). *Designing safer roads, practices for resurfacing, restoration and rehabilitation.* No. 214). Washington: National Research Council.

Waller, P. (2000). *Introduction to safety.* (Presented at the Transportation Research Board Safety-Concious Planning Meeting TRB, Washington, D.C.:

Wang, X., & Abdel-Aty, M. (2006). Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis and Prevention, 38*(6), 1137-1150.

Washington, S., Schalkwyk, I. V., Meyer, M., Dumbaugh, E., & Zoll, M. (2006). *Incorporating safety into long-range transportation planning.* No. NCHRP 546). Washington, D.C.: Transportation Research Board of the National Academies.

WHO. (2008). *WHO global burden of disease: 2004 update.* (WHO Library Cataloguing-in-Publication Data No. ISBN 978 92 4 156371 0). World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland: WHO Press. Retrieved from http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf

Yuan, L., Yuan, H., & Huang, Z. (2010). Safety improvement at two-way stop-controlled intersection. *2010 International Conference on Intelligent Computation Technology and Automation, ICICTA 2010, May 11, 2010 - May 12, 3* 252-254.

Appendices

Appendix A: Sample data set

Table 21 Sample data set

TZ Prefix	VKT	POPD	WKGD	PARP	BUSP	RTKT	PEDP	BSKD	T3	S3
1	1198	13.4	0.12	67	5.2	8.853	3.1	0.095	87	6
2	4193	10.1	0.12	67	5.2	2.095	3.1	0.071	177	11
4	96	2.1	0.12	67	5.2	12.14	3.1	0.009	30	0
5	1401	12.4	0.06	68	8.3	3.127	2.2	0.115	160	16
6	1537	10.9	0.06	55	5.7	8.39	3	0.042	254	35
7	3347	10	0.08	58	5.2	4.59	1	0.065	200	14
8	1752	26.5	0.04	60	10.6	19.02	6.5	0.1	276	42
9	2082	14.9	0.04	60	10.6	22.15	6.5	0.025	322	70
10	1354	52.2	0.22	46	18.3	6.074	9.3	0.242	598	75
11	4152	30.1	0.56	49	17.1	4.514	9.3	0.16	1393	219
12	9319	9.4	0.56	49	17.1	1.036	9.3	0.02	22	2
13	6520	25.4	0.42	65	13.8	5.476	5.3	0.154	955	184
14	1695	4.2	0.42	65	13.8	2.093	5.3	0.042	234	52
15	6437	49.1	0.11	65	16.7	17.48	9.6	0.096	1044	200
16	1265	64.4	0.29	63	24	33.03	11.1	0.175	175	41
17	2016	63.8	0.29	63	24	9.509	11.1	0.614	536	121
18	1610	110.5	0.23	66	26.5	36.04	14.1	0.285	1184	156
19	4504	31.8	0.04	69	16.7	48.85	8.1	0.058	389	87
20	1437	39.7	0.04	66	23	22.5	10.5	0.203	190	53

Appendix B: Output in SAS

Figure 14 Output in SAS

The SAS System 15:40 Wednesday, December 15, 2010 1

The GENMOD Procedure

Model Information

Data Set	WORK.T3
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	t3
Observations Used	479
Missing Values	2

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	473	545.5569	1.1534
Scaled Deviance	473	545.5569	1.1534
Pearson Chi-Square	473	442.5083	0.9355
Scaled Pearson X2	473	442.5083	0.9355
Log Likelihood		233066.6939	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-4.5880	0.4419	-5.4542	-3.7219	107.79	<.0001
vkt	1	0.9954	0.0544	0.8888	1.1020	334.89	<.0001
intd	1	1.4377	0.2216	1.0034	1.8720	42.10	<.0001
ialp	1	0.0130	0.0033	0.0066	0.0194	16.02	<.0001
sigd	1	2.5310	0.9522	0.6648	4.3972	7.07	0.0079
popd	1	0.0026	0.0015	-0.0004	0.0056	2.86	0.0908
Dispersion	1	0.7510	0.0471	0.6586	0.8433		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.