

**FULL BAYESIAN MODELS TO ASSESS THE IMPACTS OF MOBILE AUTOMATED  
ENFORCEMENT ON ROAD SAFETY AND CRIME**

by

Shewkar Ibrahim

B.Sc., United Arab Emirates University, 2008

M.A.Sc., University of British Columbia, 2011

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

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES  
(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2020

© Shewkar Ibrahim, 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Full Bayesian Models to Assess the Impacts of Mobile Automated Enforcement on Road Safety and Crime

---

submitted by Shewkar Ibrahim in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Civil Engineering

---

**Examining Committee:**

Tarek Sayed, Professor, Department of Civil Engineering, UBC

---

Supervisor

Maged Senbel, Associate Professor, School of Community and Regional Planning, UBC

---

Supervisory Committee Member

David Gillen, Professor, Sauder School of Business, UBC

---

University Examiner

Pierre Bérubé, Professor, Department of Civil Engineering, UBC

---

University Examiner

**Additional Supervisory Committee Members:**

Alan Russell, Professor Emeritus, Department of Civil Engineering, UBC

---

Supervisory Committee Member

Ziad Shawwash, Assistant Professor, Department of Civil Engineering, UBC

---

Supervisory Committee Member

## **Abstract**

The success of road safety programs is highly dependent on using accurate and precise safety models. Traditionally, these safety models were developed at a micro-level and lack understanding of how safety is prioritized at a planning-level. This dissertation bridges this gap by developing macro-level models to enhance the decision-making processes by providing opportunities for planners and designers to become better informed on issues related to road safety and criminology.

The contributions of this dissertation were to develop Full Bayesian models to explore new applications for macro-level modeling, which focused on mobile automated enforcement (MAE). This type of enforcement is one of the tools that agencies use when manned enforcement is too costly or not feasible. It consists of units that are installed in vehicles that rotate between sites to improve compliance to the speed limit and to enhance safety. The first application showed that increasing the number of tickets issued for vehicles exceeding the speed limit resulted in a decrease in all collision severities. The results also showed that collision reductions were associated with an extended time enforcing a site. Decision support tools were also created to help agencies make informed decisions regarding how to optimize their enforcement strategy.

The second application explored the impact of MAE on both collisions and crime. Previous work suggested that collision and crime hotspots overlapped. It was, therefore, crucial to quantify the degree of correlation between both events. The results of the models confirmed this relationship and showed that increased MAE presence resulted in reductions in both events. This

demonstrates how a single deployment can achieve multiple objectives, and allows agencies to optimize their deployment strategy to achieve more with less. Understanding how changing the deployment strategy at a macro-level affects safety provides enforcement agencies with the opportunity to maximize the efficiency of their existing resources. Future work would include using the results in this dissertation to build an optimization tool which accounts for the safety impacts, constraints surrounding a deployment, and the cost of a deployment to allow agencies to maximize the use of their resources to achieve the highest safety benefit.

## **Lay Summary**

The main goal of this research is to arm professionals with tools that help them understand the impact of their road safety decisions. The research was divided into two components; the first component looked at developing statistical models to understand how automated speed enforcement impacts collisions. The results of the work showed that when a long time is spent enforcing a site, the safer a location is deemed. This is a novel finding that can allow road agencies to plan where they need to deploy their limited resources to yield the most safety benefits.

The second component focused on understanding the relationship between collisions and crime, as well as how mobile automated enforcement can impact both simultaneously. The results showed that both events were highly correlated. When they are modeled together, agencies can identify which neighbourhoods they need to focus their resources on to enhance both road safety and security.

## **Preface**

The dissertation chapters are published as one paper as a poster at a conference, two papers in top-tier journals, and two papers are under review and are listed below. I was the lead investigator, responsible for conceptualization, data collection and preparation, formal analysis, software coding, visualization, writing and manuscript composition. This work was all under the supervision of Prof. Tarek Sayed.

Portions of the introductory text in Chapter 1, portions of the literature review in Chapter 2, portions of the data aggregation described in Chapter 3, and a version of Chapter 4 were published in the following papers:

- Ibrahim, S. and Sayed, T. (2020). Understanding the Impact of Automated Enforcement on Collision Rates and Crime Rates using Bayesian Tobit Models. *2020 Annual meeting of the Transportation research Board, Washington DC, January.*
- Ibrahim, S. and Sayed, T. (2019). Does Automated Enforcement Presence Impact Collisions and Crime? *Transportation Research Record: Journal of the Transportation Research Board.*

Portions of the introductory text in Chapter 1, portions of the literature review in Chapter 2, portions of the data aggregation described in Chapter 3, and a version of Chapter 5 were published in

- Ibrahim, S. and Sayed, T. (2019). Use of Objective Safety Evidence to Deploy Automated Enforcement Resources. *Transportation Research Record: Journal of the Transportation Research Board*.

Portions of the literature review in Chapter 2, portions of the data aggregation described in Chapter 3, and versions of Chapters 4 and 5 are currently under review in two submitted papers:

- Using Bayesian Tobit Models to Understand the Impact of Mobile Automated Enforcement on Collision and Crime Rates.
- Identification of Hazardous Locations Based on Collisions and Crime.

## Table of Contents

<b>Abstract</b> .....	<b>iii</b>
<b>Lay Summary</b> .....	<b>v</b>
<b>Preface</b> .....	<b>vi</b>
<b>Table of Contents</b> .....	<b>viii</b>
<b>List of Tables</b> .....	<b>xii</b>
<b>List of Figures</b> .....	<b>xiii</b>
<b>List of Abbreviations</b> .....	<b>xvi</b>
<b>Acknowledgments</b> .....	<b>xvii</b>
<b>Dedication</b> .....	<b>xviii</b>
<b>Chapter 1: Introduction</b> .....	<b>1</b>
1.1    Background .....	1
1.2    Managing Road Safety.....	3
1.3    Thesis Motivation .....	4
1.4    Research Objectives.....	6
1.5    Thesis Structure .....	11
<b>Chapter 2: Literature Review</b> .....	<b>12</b>
2.1    Impact of Automated Enforcement on Collisions .....	12
2.2    Impact of Automated Enforcement on Collisions & Crime .....	16
2.3    Collision Modeling .....	20
2.3.1    Poisson-based Collision Modeling .....	21
2.3.1.1    Univariate Poisson Lognormal Models .....	22



2.3.1.2	Multivariate Poisson Lognormal.....	24
2.3.2	Tobit Collision Modeling.....	28
2.3.2.1	Traditional Tobit.....	30
2.3.2.2	Grouped Random Parameter Tobit.....	31
2.3.2.3	Random Intercept Tobit.....	32
2.3.2.4	Multivariate Random Parameter Tobit.....	32
2.3.3	Model Comparison and Goodness of Fit.....	33
2.3.4	Posterior Distribution Estimation.....	34
<b>Chapter 3:</b>	<b>Data Description.....</b>	<b>36</b>
3.1	Geographic Scope.....	36
3.2	Collision Data.....	37
3.2.1	Data Overview & Source.....	37
3.2.2	Data Quality & Aggregation.....	37
3.2.3	Data Summary.....	42
3.3	Exposure Data.....	42
3.3.1	Data Overview & Source.....	42
3.3.2	Data Quality & Aggregation.....	43
3.3.3	Data Summary.....	47
3.4	Automated Enforcement Data.....	47
3.4.1	Data Overview & Source.....	47
3.4.2	Data Quality & Aggregation.....	52
3.4.3	Data Summary.....	52

3.5	Crime Data .....	53
3.5.1	Data Overview & Source .....	53
3.5.2	Data Quality & Aggregation.....	54
3.5.3	Data Summary .....	55
<b>Chapter 4: Automated Enforcement &amp; Traffic Safety.....</b>		<b>56</b>
4.1	Introduction.....	56
4.2	Impact of Automated Enforcement on Macro-level Collisions.....	59
4.2.1	Model Development.....	59
4.2.2	Deployment Consultation Charts.....	62
4.2.3	Case Study: Comparison of Deployment Strategies.....	66
4.3	Impact of Changing Deployment Intensity on Macro-Level Collisions.....	87
4.3.1	Model Development.....	87
4.3.2	Results & Analysis.....	90
4.4	Summary.....	94
<b>Chapter 5: Crime &amp; Traffic Safety .....</b>		<b>96</b>
5.1	Introduction.....	96
5.2	Collisions & Crime .....	98
5.3	Impact of Automated Enforcement on Collisions and Crime.....	102
5.3.1	Model Development.....	102
5.3.2	Results & Discussion .....	108
5.4	Identification of Hazardous Locations.....	111
5.4.1	Methodology.....	111

5.4.2	Univariate Results .....	114
5.4.3	MVPLN vs. PLN Identification of Hazardous Locations.....	115
5.5	Impact of Changing Deployment Intensity on Collisions & Crime .....	120
5.5.1	Model Development.....	121
5.5.2	Results & Discussion .....	121
<b>Chapter 6: Conclusion .....</b>		<b>124</b>
6.1	Summary of Research Findings .....	124
6.2	Research Contributions.....	129
6.3	Limitations & Future Research.....	131
<b>References .....</b>		<b>133</b>

## List of Tables

Table 1.1 Difference between micro and macro-level analyses .....	4
Table 3.1 Descriptive Statistics Collision (2013-2015) Data by TAZ.....	42
Table 3.2 Descriptive Statistics for Exposure Data .....	47
Table 3.3 Summary of data used for Mobile Automated Enforcement analysis .....	53
Table 3.4 Summary of data for Crime and Collision Analysis.....	55
Table 4.1 Overview of Model 1 Results .....	64
Table 4.2 Overview of Model 2 Results .....	65
Table 4.3 Tobit Model Results - Parameter Estimates and 95% Confidence Intervals .....	93
Table 5.1 MVPLN Model's Statistics.....	110
Table 5.2 Univariate PLN results for Collisions.....	115
Table 5.3 Univariate PLN results for Crime.....	115
Table 5.4 Number of hotspots selected under the different PLN models.....	116
Table 5.5 Parameter Estimates and 95% Confidence Intervals .....	123

## List of Figures

Figure 1.1 Top 10 global causes of deaths ( <i>adapted from the World Health Organization, 2018</i> )	2
Figure 3.1 Difference between TAZ Boundaries (purple) and the Road Network (gray) .....	39
Figure 3.2 Difference between the Road Network (grey) and the Emme/2 Model (blue) .....	40
Figure 3.3 Example of a Link Completely within One TAZ .....	44
Figure 3.4 Example of a Link Crossing Two Zones .....	45
Figure 3.5 Example of a Link Overlapping with the Boundary of two TAZs .....	45
Figure 3.6 Example of a Link Classified as Inside Overlap .....	46
Figure 4.1 Charts demonstrating the relationship between VKT, Tickets issued per site, and the expected Total collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart .....	72
Figure 4.2 Design charts demonstrating the relationship between VKT, the ratio of hours of enforcement per site to the average number of visits per site, and the expected Total Midblock Speed Related collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart .....	74
Figure 4.3 Charts demonstrating the relationship between VKT, Tickets issued per site, and the expected Total Midblock Speed Related collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart....	76
Figure 4.4 Design charts demonstrating the relationship between VKT, the ratio of hours of enforcement per site to the average number of visits per site, and the expected Total collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart .....	78

Figure 4.5 Charts demonstrating the relationship between VKT, Tickets issued per site, and the expected $I+F$ collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart .....	80
Figure 4.6 Design charts demonstrating the relationship between VKT, the ratio of hours of enforcement per site to the average number of visits per site, and the expected $I+F$ collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart .....	82
Figure 4.7 Charts demonstrating the relationship between VKT, Tickets issued per site, and the expected $PDO$ collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart .....	84
Figure 4.8 Design charts demonstrating the relationship between VKT, the ratio of hours of enforcement per site to the average number of visits per site, and the expected $PDO$ collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart .....	86
Figure 5.1 Collision Heatmap by Neighborhood in Edmonton .....	99
Figure 5.2 Crime Heatmap by Neighborhood in Edmonton.....	100
Figure 5.3 Overlap of Neighbourhoods with high incidences of crime and collisions frequencies .....	101
Figure 5.4 Neighbourhoods that are crime & collision prone as identified by modeling both incidences independently .....	118

Figure 5.5 Neighbourhoods that are problem prone as identified by MVPLN (in teal) and neighbourhoods that are crime-prone as identified by the individual PLN model for incidents of crime ..... 119

Figure 5.6 Neighbourhoods that are problem prone as identified by MVPLN (in teal) and neighbourhoods that are collision prone as identified by the individual PLN model for collisions ..... 120

## List of Abbreviations

<b>COE</b>	<i>City of Edmonton</i>
<b>DDACTS</b>	<i>Data-Driven Approaches to Crime and Traffic Safety</i>
<b>I+F</b>	<i>Injury and Fatality</i>
<b>MAE</b>	<i>Mobile Automated Enforcement</i>
<b>MVPLN</b>	<i>Multivariate Poisson Lognormal</i>
<b>PDO</b>	<i>Property Damage Only</i>
<b>PLN</b>	<i>Poisson Lognormal Model</i>
<b>SPF</b>	<i>Safety Performance Functions</i>
<b>TAZ</b>	<i>Traffic Analysis Zones</i>



## **Acknowledgments**

I would like to extend my gratitude to my thesis supervisor, Prof. Tarek Sayed, at the University of British Columbia. His continuous support, encouragement, and patience have made this an enjoyable academic experience. His feedback from our discussions has inspired me to continue giving my best effort and have helped me recognize my true potential.

I owe sincere thanks to my professors at the undergraduate level (at the United Arab Emirates University) and at the graduate level (University of British Columbia) as well as my coworkers and colleagues at the City of Edmonton for all of their help, advice and support throughout my years of studying. They have enriched my knowledge base, and their love of their topic is what motivated me to start my career in academia so that one day I would be able to inspire new generations.

I am forever grateful to my parents and brother, whose support has helped me both morally and financially throughout my years of education. Their prayers, help, love, and presence in my life have always helped me aspire to become a better person and achieve greater things in life. I would also like to thank my lovely friend and sister-in-law, Nahla Sherif, for her continuous encouragement and her uplifting attitude throughout our years together. Special thanks are for Laura Contini for having been readily available to assist me whenever needed.

*To my loving parents and brother*

This thesis is dedicated to my late father, Prof. Mohamed Yahia El-Bassiouni, who taught me that the best kind of knowledge is that which is learned for its own sake. It is dedicated to my late mother, Esmat Fawzi, whose incredible support taught me that even the largest of tasks could be completed if tackled one step at a time. It is also dedicated to my brother, Dr. Karim El-Basyouny, without him this dissertation would not have been possible.

## Chapter 1: Introduction

This chapter consists of five sections; the first section presents background information that contextualizes the importance of traffic safety and the burden of traffic collisions worldwide. The second section discusses the traditional approach to managing road safety. This information also lays the groundwork for understanding the gaps in the existing research and the motivation behind this research, which is outlined in the third section. The fourth section states the research objectives, and finally, the chapter concludes with an outline of the thesis structure.

### 1.1 Background

The considerable increase in the world's population and economic activity over the years has resulted in an increase in the number of vehicles as well as the vehicle-kms traveled on roads worldwide. While motorized travel provides many benefits in terms of mobility, it also poses a safety concern in the form of road-related collisions. In 2010, deaths resulting from collisions was the *tenth* leading cause of death (World Health Organization, 2018). In 2016, deaths resulting from collisions increased to the *eighth* leading cause of death and accounted for 1.4 million deaths, as shown in Figure 1.1.

Rank	Leading Cause of Death	2010	Rank	Leading Cause of Death	2016
1	Ischaemic heart disease	7,029	1	Ischaemic heart disease	9,433
2	Stroke	5,170	2	Stroke	5,781
3	Lower respiratory infections	3,325	3	Chronic obstructive pulmonary disease	3,041
4	Chronic obstructive pulmonary disease	2,972	4	Lower respiratory infections	2,957
5	Diarrhoeal diseases	2,246	5	Alzheimer disease and other dementias	1,992
6	Tuberculosis	1,684	6	Trachea, bronchus, lung cancers	1,708
7	HIV/AIDS	1,469	7	Diabetes mellitus	1,599
8	Preterm birth complications	1,382	8	Road injury	1,402
9	Trachea, bronchus, lung cancers	1,257	9	Diarrhoeal diseases	1,383
10	Road injury	1,136	10	Tuberculosis	1,293

Figure 1.1 Top 10 global causes of deaths (*adapted from the World Health Organization, 2018*)

In Canada, the national and provincial trends are both consistent with the global trend. At a national level, between 2000 and 2009, road collisions were ranked as the *fifth* leading cause of death in Canada. In 2017, it increased in rank to the *third* leading cause of death (Statistics Canada, 2019). Finally, in Alberta, road collisions are the *third* leading cause of death and have claimed the lives of 273 individuals in 2016 (Alberta Government, 2019).

The societal and economic burden of these road collisions is enormous, and these collisions impact not only the victims and their families but also societies and governments. Annually, road collisions cost countries between one to three percent of their Gross National Product (World Health Organization, 2013). Road collisions incur costs related to property damage, hospital care, traffic delays, and emergency response amounting to \$37 billion in Canada (Canadian Association of Chiefs of Police, 2018).

The previous statistics solidify the motivation behind mitigating this epidemic of road collisions and justify the immediate need for developing and implementing robust road safety programs. These programs should employ innovative techniques to find practical and feasible solutions to eliminate deaths and serious injuries. Therefore, it is not surprising that road safety is currently one of the main areas of focus for governments, states, and researchers. Unless drastic measures are taken, road collisions will continue to claim the lives of people locally, nationally, and globally.

## **1.2 Managing Road Safety**

Traditional road safety improvement programs focus on the identification, diagnosis, and improvement of collision prone locations. A collision prone location is defined as any location that exhibits a higher potential for collision occurrence than an established norm. This is usually expressed in terms of typical collision measures such as frequency, rate, severity, or a combination of them. The main assumption of these programs is that road design plays a contributory role in the occurrence of many traffic collisions (Sawalha, 2006). Therefore, improving the engineering elements of collision prone locations can significantly reduce the severity and frequency of collisions.

These traditional safety programs are reactive in nature since they target problems that are identified based on current collision information. Targeting problem locations and developing plans to reduce collision potential is essential and has been proven to be highly successful. However, a significant collision history must exist before taking any action, which requires waiting for periods of 2-3 years. Therefore, transportation professionals should also adopt a proactive approach that addresses road safety concerns before they are allowed to emerge (de Leur, 2003). This proactive approach attempts to prevent unsafe road conditions by implementing modifications and changes early at the planning and design stages. When road safety is included explicitly at these early stages, safety concerns can be mitigated in a more cost-effective method.

The success of both the reactive and proactive approaches in reducing collision occurrences, however, hinges upon the existence of consistent methods that provide reliable estimates of road

safety. Several methods were developed and used in the literature to quantify safety, but the most common approach utilized is the use of Safety Performance Functions. SPFs are mathematical models with inherent statistical characteristics that attempt to relate collisions frequencies at a specific location and that location’s specific characteristics such as traffic volume or horizontal/vertical alignment. Most of the existing SPFs have been developed to analyze collision patterns, trends, and to identify statistical associations with explanatory variables at a micro-level (e.g., location-specific levels). These models provide designers with information regarding elements that may impact safety at an intersection or a road segment.

### 1.3 Thesis Motivation

To complete the shift from a reactive to a proactive approach, road safety needs to be considered at a planning-level as well. Generally, planning-level analysis is undertaken at the macro-level, and its analysis requires considering the “big picture” and investigating the traffic patterns and traffic safety at a scale larger than just a single intersection or road segment. Table 1.1 summarizes a few of the differences between the methods of analysis.

	Micro-Level	Macro-Level
<b>Unit of Analysis</b>	Segment or Intersection	Area: Zone or Neighbourhood
<b>Collisions</b>	Collision history at the intersection or road segment	Collisions within a geographical area
<b>Exposure</b>	Number of vehicles on the road (e.g., average annual daily traffic)	Zone based exposure measure such as the total vehicle km traveled within an area of interest
<b>Explanatory Variables (examples)</b>	Radius of the curve Length of the segment Number of lanes	Population Signalized intersection density Employment density

Table 1.1 Difference between micro and macro-level analyses

Many researchers have recognized the importance of implementing proactive, planning level, safety evaluations and investigated the necessary supporting empirical tools. Previous research attempted to quantify the macro-level safety impact of a planned improvement using micro-level SPFs. The earliest literature on empirical tools developed for planning-level analysis was in the 1990s; however, the issue of forecasting future volumes at a network level proved to be problematic. These preliminary models used the traditional four-step transportation planning process (i.e., trip generation, trip distribution, modal split, and trip assignment). The model output includes exposure measures, travel times, shortest paths, and congestion level, amongst other variables, and is provided for each link within the network (Ho and Guamaschelli, 1998). These values would then be aggregated mode, zone and region to allow for area-level analysis.

To develop reliable micro-level SPFs that can accurately relate the safety (i.e., collisions) of a specific location to other variables (e.g., traffic and road characteristics), data is required for each intersection and road segment in the entire region. Even if this data was not computationally expensive and cost and time prohibitive to extract, the accuracy of this data is subject to discussion. For instance, the traffic volume (i.e., the exposure measure), which is the minimum requirement for SPF development, needs to be known or at least estimated for each intersection and road segment. Therefore, this research direction was abandoned in favour of creating a road safety planning framework.

Researchers in Canada developed a proactive road safety planning framework that provided the groundwork towards quantifying a planning level predictive relationship (e.g., Lovegrove and Sayed, 2006). Early macro-level SPFs research appeared to be promising and proved the

importance of developing empirical tools for proactive network-level planning. However, the lack of reliable macro-level SPFs and guidelines for their use still hindered any progress on this research front.

Research is needed on macro-level models to address this empirical tool gap related to the inability of micro-level SPFs to do proactive planning level safety evaluations. If successful, macro-level SPFs will allow engineers and planners to target safety in a proactive manner and will ultimately improve traffic safety for all road users. It is hoped that the models will lead to significant reductions in the level of road collision frequencies beyond what has been achieved to date using reactive techniques. Moreover, proactive tools may be able to complement and enhance conventional reactive programs.

#### **1.4 Research Objectives**

The main goal of this research is to develop reliable macro-level SPFs that would assist planners and designers by quantifying safety to facilitate the decision-making process. In order to develop macro-level SPFs, several issues need to be investigated.

##### **I) Applications of Macro-Level Models**

Additional applications of macro-level SPFs are investigated to further expand the use of these models. The following two applications were used to provide more input to planners on the versatility of the macro-level models.



### ***First Application: Automated Enforcement & Traffic Safety***

The causal link between speed and traffic safety is well established in the literature (Elvik, 2005). Driving at a high speed greatly increases the risk of being involved in a collision as there is less time and distance available for the vehicle to come to a complete and safe stop. Unfortunately, speed violation is a widespread phenomenon. On a recent survey, over 50% of the respondents reported that they drive above the speed limit (Corporate Research Associates, 2019). Mobile automated speed enforcement is one of the methods undertaken to reduce speeds on the roads.

There are generally two types of speed enforcement: manned enforcement and automated enforcement. Conventional manned enforcement is conducted by police officers with speed measurement devices. However, due to a lack of resources and budget constraints, conventional enforcement is not considered to be a permanent solution. Therefore, automated enforcement is typically used as a more sustainable alternative to conventional enforcement.

The impact of mobile automated speed enforcement on traffic safety has been examined and been validated by many studies but either on a micro-level (e.g., road segment) or a jurisdictional level (e.g., country or city level). While both evaluations are needed to help determine the effectiveness of mobile automated speed enforcement, neither unit of analysis can help inform a deployment strategy.

One of the main contributions of this dissertation includes determining whether mobile automated speed enforcement has an impact on traffic safety at a macro-level. This information

will be used to develop an empirical tool that allows road agencies to understand how their deployment strategy can affect safety and how to plan their resources efficiently and effectively.

### ***Second Application: Crime & Traffic Safety***

In the past few decades, several studies have established a link between traffic collisions and crime. Research has shown that most locations with a high frequency of traffic crashes were highly correlated with crime rates (e.g., auto theft, robbery). Consequently, police departments have been adopting a new data-driven approach that integrates location-based traffic crash and crime data to optimize their deployment strategy and resources. This new approach is referred to as the Data-Driven Approach to Crime and Traffic Safety (DDACTS). Research has shown that cities with the highest rate of traffic citations per officer experienced the lowest rates of crime as well as a reduced frequency and severity in traffic crashes. The underlying premise is that the presence of more traffic officers provided more opportunities for them to identify high-risk drivers as well as discourage the presence of criminals in both targeted and surrounding areas.

While the DDACTS approach may be providing law enforcement personnel with information regarding locations that experienced a high frequency in both crashes and crime, the methodology is still quite simplistic. Additionally, although there seems to be a high correlation between traffic enforcement and crime/collision reduction, this does not necessarily infer causation. This area is relatively new, and a more evidence-based approach, such as the development of SPFs, needs to be investigated to determine and accurately quantify the relationship between crime and traffic safety.

Firstly, the correlation between collisions and crime has not been quantified. This research validates and confirms this relationship. This is important since it means that jurisdictions may no longer model collisions and criminal incidents separately. It also suggests that factors that affect collisions can also affect criminal incidents. When attempting to understand the effectiveness of a specific measure (i.e., in the case of mobile automated speed enforcement), both events need to be modeled together.

Another significant contribution of this research is determining the impact of mobile automated speed enforcement on both collisions and criminal events. This is an area that has not been studied previously but can help enforcement agencies expand their DDACTS toolkit to also include mobile automated speed enforcement. Finally, understanding how varying the enforcement agencies' deployment intensity (i.e., spending a more extended time enforcing a site) can help them plan the use of their resources more efficiently and effectively.

## **II) Enhanced Statistical Modeling of Macro-Level Safety Models**

Previous research was dedicated to improving the predictive power of micro-level SPFs and to introduce effective methods for traffic safety assessment. However, this approach only ensures that safety is accounted for at the design stage. In order to effectively bring safety to the forefront, road agencies need empirical tools that allow them to understand the impact of their decisions at a more aggregate level.

One of the contributions of this dissertation is related to developing macro-level SPFs for a jurisdiction that suffers from urban sprawl. Prior research was focused on developing these SPFs

for high-density areas; however, these models are not easily transferable to jurisdictions that have a different density (population per square km). This is particularly apparent in the transportation mode split; cities with a high density (higher population per square km) typically have more vulnerable road users (e.g., cyclists, pedestrians). Therefore, macro-level SPFs needed to be developed to provide jurisdictions with urban sprawl.

Several factors are critical to the development of any SPF: data quality and the statistical modeling technique. In developing the macro-level SPFs for the city of Edmonton, five different datasets were used. These datasets existed in various formats (e.g., spatial, Microsoft Excel spreadsheets or Microsoft Access databases). This required substantial processing to ensure that the data has been cleaned, aggregated, and ready to be used for model development. A data catalogue was created for this research, which linked all the datasets together. Indexing of the data was also completed, so if it is refreshed with more recent information, the aggregation can be completed with minimal manual intervention.

The next step that is essential to the development of any SPF is the modeling technique. More advanced statistical models need to be developed to account for the various characteristics of different datasets from both a theoretical and practical perspective. Improving the predictive power of SPFs requires examining various statistical issues, as discussed in the second chapter of this thesis. Researchers have devised enhanced regression models to address the shortcomings of previously used techniques as well as to improve the predictive accuracy of SPFs. These enhanced modeling techniques include random parameter Tobit models, multivariate PLN, and Tobit models. These models will be applied to both applications.

## **1.5 Thesis Structure**

The thesis is divided into six chapters that summarize the content and work of this dissertation; together, they provide a full view of how macro-level SPFs can be used to improve safety. Chapter Two summarizes the literature review of the development of Safety Performance Functions (SPFs) and previous work on automated enforcement. The following chapter outlines the data used in this thesis. A comprehensive database comprising exposure, collisions, automated enforcement, and crime variables was used. The data was aggregated by Traffic Analysis Zones (TAZs) and neighbourhoods for the COE (City of Edmonton). The data sources for each group, as well as the aggregation method, are summarized in Chapter Three. The following two chapters focus on the two different applications that were selected for this research. Chapter Four highlights the impact of mobile automated enforcement on collisions at a macro-level and provides additional statistical analysis to provide further insight into the deployment strategy. Chapter Five extends the models developed in Chapter Four to investigate the impact of mobile automated enforcement not only on collisions but also on crime. Finally, Chapter Six summarizes research conclusions, contributions, and suggestions for future research.

## **Chapter 2: Literature Review**

This chapter includes three sections. A summary of previous research that studied the impact of automated enforcement on collisions is provided in the first section. The second section summarizes previous research on the topic of using enforcement to reduce collisions as well as incidences of crime. Lastly, the final section provides an overview of the various statistical issues related to collision modeling..

### **2.1 Impact of Automated Enforcement on Collisions**

The causal link between speed and traffic safety is well established in the literature (Elvik, 2005). Driving at a high speed greatly increases the risk of being involved in a collision since the driver has less time and distance available for the vehicle to come to a safe and complete stop. Unfortunately, speeding is a widespread phenomenon. On a recent survey, over 50% of the respondents reported that they exceed the speed limit (Corporate Research Associates, 2019). Road agencies typically use speed enforcement as one of the methods to manage and reduce speeding.

There are generally two types of speed enforcement: manned enforcement and automated enforcement. Conventional manned enforcement is conducted by police officers with speed measurement devices. However, due to the lack of resources and budget constraints, conventional enforcement is not considered to be a permanent solution. Additionally, manned enforcement is constrained by operational considerations, for instance, it is not preferred at locations where there is a high volume of traffic, which may cause risk to personnel during

deployment. Therefore, automated enforcement is typically used as a more sustainable alternative to conventional enforcement.

Automated enforcement can either be conducted using stationary devices or using mobile devices installed in vehicles. The stationary devices are generally installed at the approaches of an intersection and either monitor red-light running vehicles (commonly referred to as red-light cameras), or they can monitor both red-light running vehicles as well as vehicles which exceed the speed limit during the green phase (commonly referred to as intersection safety devices). These devices are typically installed on approaches to intersections with high-volume roads (i.e., arterial-arterial intersections). These stationary devices have been shown to increase safety by reducing collisions and speeding through intersections [insert references for research on this topic]. However, over time drivers become more familiar with the locations of these fixed devices, and behaviors start to normalize. One of the main components of an enforcement program is specific to general deterrence, which results from the public's perception that traffic laws are enforced, and the risk of detection and apprehension (i.e., ticket/violation issued) exists if the traffic law is broken. Mobile automated enforcement allows agencies to achieve the general deterrence theory since the list of locations and schedules are not typically released, so drivers always need to obey the rules of the road to avoid the risk of receiving a traffic violation.

The impact of mobile automated speed enforcement on traffic safety has been examined and validated by many studies. These studies either focused their evaluations on either a highly aggregated level (i.e., national, state, or municipal level) or a micro-level (i.e., road segment or

intersection level). On a state level, an evaluation of mobile speed cameras in Victoria, Australia, both in urban and rural settings revealed a 20% reduction in daytime injury and fatal crashes and a 27.9% reduction in collision severity statewide (*Cameron et al., 1992*). This program also included an educational campaign which the authors indicated did not contribute to the reduction of the injury collisions but was attributed to a reduction in the overall frequency of collisions. The analysis included before and after assessments using time trend analyses and Autoregressive Integrated Moving Average (ARIMA) models were used to account for time and trend effects. At the same time, Victoria also introduced an alcohol enforcement program, which made it challenging to isolate the impact of the speed enforcement program on collision reductions (*Decina et al., 2007*).

Chen et al., (2000) investigated the systemwide impact that increased photo-radar enforcement had on the collisions and speed in British Columbia, Canada, using an interrupted time series ARIMA model. The results of the study showed a significant collision reduction of 25% in daytime unsafe-speed related collisions, an 11% reduction in daytime injuries, and a 17% reduction in daytime fatalities. The program was also found to be successful at reducing the overall speed by 2.4 km/h. However, the traffic volume was not explicitly considered in this evaluation. Instead, gasoline sales were used as a surrogate measure of exposure.

An evaluation of a national speed camera program in France revealed a significant reduction in both fatal (21% decrease per 100,000 vehicles) and non-fatal injuries (an initial decrease of 26.2% in the first month) following the deployment of the program (*Carnis and Blais, 2013*). This program involved the use of both fixed and mobile speed cameras (which makes it



challenging to isolate the impact of each tool) to detect vehicles exceeding the speed limit and interrupted time-series analyses using ARIMA intervention time series models were adopted.

Other studies investigated the impact that the speed camera program had on specific sites where the mobile photo enforcement was used to detect speeding. Newstead and Cameron (2003) evaluated the impact of overt mobile deployment of speed cameras in Queensland, Australia. The authors used a before/after observational study with a comparison group to evaluate the success of this program. Based on the results of their study, the highest collision reductions were found within 2 km of the camera site, and those reductions included 18% in all severity collisions and 22% in collisions where hospitalization was required. However, this study did not account for the exposure measure (i.e., traffic volume) and the effectiveness was only studied within 2, 4, and 6 km area which provides jurisdictions with limited understanding of the impact of the cameras within those ranges.

Similarly, targeted effects of mobile photo enforcement were also investigated (Christie et al., 2003; Goldenbeld and van Schagen, 2005; Li et al., 2015; Li et al., 2016). Christie et al., (2003) estimated a reduction in injury collisions of 51% for road segments 500 m downstream and upstream of the location of the cameras in South Wales, United Kingdom. Goldenbeld and van Schagen (2005) reported a reduction of 21% in injury collisions at locations where the cameras were located as well as within 4 km of the location of the camera in the rural areas of Friesland, Netherlands. However, these studies did not take into consideration the impacts of traffic volume or other confounding factors (Decina et al., 2007).

More recently, two studies were conducted based on the mobile photo enforcement program in Edmonton, Alberta (*Li et al., 2015; Li et al., 2016*). *Li et. al, (2015)* used before and after Empirical Bayes method to account for regression-to-the-mean and other confounding factors that other studies did not address. The authors developed safety performance functions and calibration factors for different collision severities. The results of the assessment showed reductions between 14% and 20% in collisions. Additionally, the results also showed higher collision reductions at sites where more collisions occurred in the before period, and at sites with longer deployment hours.

Building on this assessment, further investigation of the impact of enforcement performance indicators on the program's safety outcomes was conducted in (*Li et al., 2016*). Using the monthly citywide data and a generalized linear Poisson model, the results showed that as the number of enforced sites and issued tickets increased, the number of speed-related collisions decreased.

While understanding the relationship between enforcement indicators (e.g., number enforced sites) or enforcement outcomes (e.g., number of tickets issued) on collisions is essential, additional information is still required for agencies to better plan their deployment strategy.

## **2.2 Impact of Automated Enforcement on Collisions & Crime**

In the past few decades, several studies have established a link between traffic collisions and crime. Research has shown that most locations with a high frequency of traffic crashes were highly correlated with crime rates (e.g., murder, auto theft, robbery). Consequently, police

departments have been adopting a new data-driven approach that integrates location-based traffic crash and crime data to optimize their deployment strategy and resources. This new approach is referred to as the Data-Driven Approach to Crime and Traffic Safety (DDACTS).

Research has shown that cities with the highest rate of traffic citations per officer experienced the lowest rates of crime as well as a reduced frequency and severity in traffic crashes. The underlying premise is that the presence of more traffic officers provided more opportunities for them to identify high-risk drivers as well as discourage the presence of criminals in both targeted and surrounding areas.

Although the relationship between collisions and criminal behavior was not quantified, it was investigated by several different approaches related to psychology and criminology (*Junger et al., 1995; Brace et al., 2009*). This theoretical hypothesis was then tested by generating collision and crime hotspots and overlaying the results to determine how many of the locations exhibit a high frequency of both events. Traditionally, collision hotspots were identified by using statistical tests to determine whether they are experiencing a higher frequency of collisions than is expected. However, this method is not always suitable for two reasons, firstly, the data requirements for this approach is extensive, and secondly, this approach does not account for the spatial relationship between collisions and other variables (*Kuo et al., 2013*). To address these limitations, GIS programs are identified as the preferred approach (*Harkey et al., 1999; Petch and Henson, 2000; Schneider et al., 2004; Loo, 2006; Li et al., 2007*). While the identification of crime hotspots is quite complex, there are commonly used software that can be used for the

identification of locations with a high prevalence of crime (Eck et al., 2005; Bernasco and Elffers, 2010; Ned, 2015).

The current commonly used approach to mapping DDACTS locations (defined as locations that are designated as both collision and crime hotspots) uses kernel density analysis. This approach is quite simple as it uses the observed number of criminal incidents and observed number of collisions and would not accurately identify a location as a hotspot due to confounding factors such as the regression-to-the-mean bias (Barnett, 2004). To address this limitation, Takyi et al. (2018) developed two negative binomial models to predict collisions and crime separately. Based on the results of the models, hotspots for collisions and crime were developed at a zonal level. The DDACTS zones were then identified as a priority for the enforcement agency to reduce crime and collisions, concurrently.

Evaluations of the DDACTS programs have shown significant reductions in crime and collisions (Hardy, 2010). In Baltimore, Maryland, there was a 16.6% reduction in burglaries, a 33.5% reduction in robberies, and a 40.9% reduction in vehicle thefts. Collisions were also decreased but only by 0.2% in injuries and 1.2% in total collisions. Similarly, other jurisdictions within the United States also saw a reduction in collisions and crime. In St. Albans, Vermont, there was a 27-35% reduction in crime and a 19-21% reduction in collisions. The use of DDACTS in Nashville, Tennessee, resulted in a reduction of 14% in crime and a 72.3% reduction in arrests of offenders who were driving while they were under the influence (DUI). Injury and fatal collisions also decreased by 31% and 15.6%, respectively.

These evaluations were related to the overall DDACTS strategy that each jurisdiction followed. Individual programs within this model were not investigated for their impact on collision and crime. However, understanding this impact can assist enforcement agencies in their selection of tools at the time of deployment. One of the objectives of this dissertation is to determine the impact of one of the most common traffic enforcement programs, mobile automated speed enforcement, on collisions and crime. Due to the impact speed has on the likelihood of the occurrence of a severe collision, agencies use manned and mobile automated enforcement to mitigate these safety concerns. Since manned enforcement is very resource-intensive and costly, mobile automated enforcement (e.g., photo radar) is often used to supplement the overall enforcement program.

At a macro-level, this information does not provide agencies with enough understanding of the impact enforcing speed would have on their resources. Alternatively, for larger-scale evaluations, changes in traffic volume, overall collision trends, and other confounding factors are not captured, and this can also negatively influence the quality of the results (*Thomas et al., 2008*). Therefore, a more robust unit of analysis is necessary to better investigate the impact automated enforcement activity has on collisions. Additionally, the impact of mobile automated enforcement on crime has not been investigated in previous work. Since this is part of the deployment strategy for enforcement agencies, understanding the impact of mobile automated speed enforcement on crime would assist them in the planning of their resources.

While the DDACTS approach may be providing law enforcement personnel with information regarding locations that experienced a high frequency in both crashes and crime, the

methodology is still quite simplistic. Additionally, although there seems to be a high correlation between traffic enforcement and crime/collision reduction, this does not necessarily infer causation. This area is relatively new, and a more evidence-based approach, such as the development of SPFs, needs to be investigated to determine and accurately quantify the relationship between crime and traffic safety.

### **2.3 Collision Modeling**

Due to the magnitude of the road safety problem, analysts and researchers have focused their attention on investigating the factors that affect the probability of a collision involvement. To understand the relationship the effect certain variables have on collisions, researchers have framed analytical approaches to establish this relationship through the development of SPFs. These models are used to provide an estimate of collision frequency due to any treatments carried out on existing facilities. Additionally, they can also be used to estimate predicted collision frequency on a planned facility (Shen, 2007).

The mathematical form used for the collision prediction model should generally satisfy two conditions. First, it must yield logical results meaning that it should not lead to the prediction of a negative number of collisions and should ensure a prediction of zero collision frequency for zero values of the exposure variables (Sawalha & Sayed, 2006).

To improve the predictive ability of these SPFs, researchers have used several collision modeling techniques. The model development is strongly affected by choice of the regression technique used (El-Basyouny, 2011). Earlier models were developed using linear regression, which

assumes a linear relationship between the dependent and independent variables. These models have been criticized by researchers for several reasons since road collisions are defined to be discrete, non-negative, and rare events (Jovanis and Chang, 1986; Hauer et al., 1988; Miaou and Lum, 1993).

Therefore, it is imperative that the correct collision modeling technique is used when developing SPFs. The next sections describe some of the commonly used regression models which will be used throughout this thesis.

### **2.3.1 Poisson-based Collision Modeling**

There are different types of regression models associated with various probability distributions for collision frequency. A commonly applied statistical model is the Poisson model, which recognizes that collisions are random, discrete, non-negative, and sporadic events (Hauer, 1988; Lord et al., 2005). As such, it was the first distribution used to model collisions. The main advantage of using a Poisson model lies in the ease of calculating its error structure. However, studies have shown that most collision data tends to be dispersed (i.e., the variance is greater than the mean) therefore making the Poisson distribution less likely to adequately represent the actual collision characteristics (Kulmala and Roine, 1988; Kulmala, 1995; Cameron and Trivedi, 1998; Winkelmann, 2003).

The Poisson-Lognormal model is another alternative to the Poisson model, and several researchers have advocated the use of this model for its ability to address over-dispersion (Miaou et al., 2003; Lord and Miranda-Moreno, 2008; Aquero-Valverde and Jovanis, 2008; El-Basyouny

and Sayed, 2009a, b, c, 2010a, b, c). Under this model, the error term is assumed to follow a lognormal distribution. If a dataset contains outliers, this model is a suitable modeling choice as its tails are asymptotically heavier than those of the Gamma distribution (Kim et al., 2002; Lord and Miranda-Moreno, 2008). However, this model has not been adopted as frequently as other models as it requires more computation and does not admit a closed-form posterior distribution.

There other various techniques developed to improve the prediction power of SPFs, such as the zero-inflated regression models. Collision data can include a high proportion of zero counts, which are more prevalent when modeling fatal collisions or modeling collisions occurring in rural settings. This is problematic when the observed zero counts outnumber the zero counts tolerated by the Poisson models. Zero-Inflated Poisson and Zero Inflated Negative Binomial models are characterized by having a low sample mean (Miranda-Moreno, 2006) and were developed to circumvent this phenomenon.

However, the most applied regression model for the development of SPFs is the Negative Binomial model (Poisson-Gamma). These types of models account for the over-dispersion issue that occurs with the Poisson models as well as more computationally flexible compared to the Poisson-Lognormal models.

### **2.3.1.1 Univariate Poisson Lognormal Models**

Recently, several researchers have proposed the use of the Poisson-Lognormal (PLN) model as an alternative to the Poisson-Gamma model to model collision data (Miaou et al., 2003; Lord and Miranda-Moreno, 2008; Aquero-Valverde and Jovanis, 2008). The PLN model is commonly



used to address over-dispersion for unobserved or unmeasured heterogeneity. According to this model, it is assumed that  $Y_i$  denotes the number of collisions at site  $i$  ( $i = 1, \dots, n$ ) and that collisions at the  $n$  sites are independent and that

$$Y_i | \theta_i \sim \text{Poisson}(\theta_i). \quad (2.1)$$

To address over-dispersion for unobserved or unmeasured heterogeneity, it is assumed that

$$\theta_i = \mu_i \exp(u_i), \quad (2.2)$$

where  $\mu_i$  is determined by a set of covariates representing site-specific attributes and a corresponding set of unknown regression parameters, whereas the term  $\exp(u_i)$  represents a multiplicative random effect.

The PLN regression model is obtained by the assumption

$$\exp(u_i) | \sigma_u^2 \sim \text{Lognormal}(0, \sigma_u^2) \quad \text{or} \quad u_i | \sigma_u^2 \sim \text{Normal}(0, \sigma_u^2) \quad (2.3)$$

where  $\sigma_u^2$  denotes the extra Poisson variance. Note that in the PLN model,  $\exp(u_i)$  follow a Lognormal distribution whereas in the Poisson-Gamma model  $\exp(u_i)$  followed a gamma distribution. The PLN is a good candidate for modeling collision occurrence in the presence of outliers, since its tails are known to be asymptotically heavier than those of the Gamma distribution (Kim et al., 2002). Other empirical evidence and advantages of the PLN model are discussed in Winkelmann (2003). Under the PLN model

$$E(Y_i) = \mu_i \exp(0.5 \sigma_u^2), \quad \text{Var}(Y_i) = E(Y_i) + [E(Y_i)]^2 (\exp(\sigma_u^2) - 1). \quad (2.4)$$

Prior distributions are meant to reflect prior knowledge about the parameters of interest. If such prior information is available, it should be used to formulate the so-called informative priors. The specification of informative priors for generalized linear models was dealt with by Bedrick et al. (1996). They considered conditional means priors as well as data augmentation priors of the same form as the likelihood and showed that such priors result in intractable posteriors.

The most common priors are diffused normal distributions where the regression parameters have zero mean and a large variance and  $\sigma_u^{-2}$  is typically assumed to follow a Gamma( $\varepsilon, \varepsilon$ ) or Gamma(1,  $\varepsilon$ ) where  $\varepsilon$  is a small number, e.g., 0.01 or 0.001.

A significant limitation for the application of the Poisson-Lognormal (PLN) regression model is that its' marginal distribution does not have a closed-form as the Poisson-Gamma model. As a result, the PLN has been less popular since it demands more computational effort, and few statistical programs are available for their calibration. However, recent years have witnessed enormous progress in numerical methods combined with the availability of free and easy to use the software, have permitted the application of the PLN model to analyze collision data.

### **2.3.1.2 Multivariate Poisson Lognormal**

Data on the number of collisions at a particular site are usually available where the collisions are classified by severity (e.g., fatal, minor injury, major injury or property damage only), by the number of vehicles involved (e.g., single or multiple), and/or by the type of collision (e.g., angle, head-on, rear-end, sideswipe or pedestrian-involved), etc. Despite the multivariate nature of such

data sets, they have been mostly analyzed by modeling each category separately, without taking into account the correlations that probably exist among the different levels. These correlations may be caused by omitted variables, which can influence collision occurrence at all levels of classification, or from ignoring shared information in unobserved error terms. Such univariate treatment of correlated counts as independent can lead to an imprecise analysis of road safety.

Several studies have applied multivariate models for estimating collision frequency under various collision severity levels and indicated their superiority to univariate models (Ma and Kockelman, 1950; Maher et al., 1990; Chib and Winklmann, 2001; Tunaru, 2002; Bijleveld, 2005; Ma and Kockelman, 2006; Park and Lord, 2007; Brijs et al., 2007; Ma et al., 2008; Agüero-Valverde and Jovanis, 2008; Ye et al., 2009; Agüero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009; Wang et al., 2011; Anastasopoulos et al., 2012; El-Basyouny et al., 2014a; El-Basyouny et al., 2014b).

For a dataset with two categories ( $K=1$  for collisions, and  $K=2$  for crime incidents) and  $n$  locations, define the vector  $y_i = (y_{i1} \ y_{i2} \ \dots \ y_{iK})'$ , where  $y_{ik}$  denote the number of collisions by severity, type, or any other incident at the  $i^{th}$  location in category  $k$ . It is assumed that the  $y_i$  are independently distributed and that the Poisson distribution of  $y_{ik}$ , given  $\lambda_{ik}$ , is

$$f(y_{ik}|\lambda_{ik}) = \lambda_{ik}^{y_{ik}} e^{-\lambda_{ik}} / y_{ik}!, \quad i = 1, 2, \dots, n, \quad k = 1, 2, \dots, K \quad (2.5)$$

To model extra variation, assume further that  $\ln(\lambda_{ik}) = \ln(\mu_{ik}) + \epsilon_{ik}$ , where

$$\ln(\mu_{ik}) = \beta_{k0} + \beta_{k1} X_{i1} + \dots + \beta_{kj} X_{ij}, \quad (2.6)$$

Let  $\epsilon_{ik}$  denote multivariate normal errors distributed as  $\epsilon_i \sim N_K(0, \Sigma)$ , where

$$\epsilon_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \dots \\ \epsilon_{iK} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2K} \\ \dots & \dots & \dots & \dots \\ \sigma_{K1} & \sigma_{K2} & \dots & \sigma_{KK} \end{pmatrix}$$

Let  $X_{ij}$  and  $\beta_k$  denote the matrix of covariates and the vector of regression coefficients, respectively, and let  $\beta$  denote the set  $\{\beta_1, \beta_2, \dots, \beta_K\}$ . Since the mixed approach allows for different covariates to be used in each model. Thus, given  $(X, \beta, \Sigma)$ , the  $\lambda_i$  are independently distributed as

$$f(\lambda_i | X, \beta, \Sigma) = \frac{\exp\{-0.5(\lambda_i^* - \mu_i^*)' \Sigma^{-1} (\lambda_i^* - \mu_i^*)\}}{(2\pi)^{K/2} (\prod_{k=1}^K \lambda_{ik}) |\Sigma|^{1/2}} \quad (2.7)$$

which is a  $K$ -dimensional log-normal distribution, where

$$\lambda_i = (\lambda_{i1} \quad \lambda_{i2} \quad \dots \quad \lambda_{iK}), \quad \lambda_i^* = (\ln(\lambda_{i1}) \quad \ln(\lambda_{i2}) \quad \dots \quad \ln(\lambda_{iK}))$$

$$\mu_i = (\mu_{i1} \quad \mu_{i2} \quad \dots \quad \mu_{iK}), \quad \mu_i^* = (\ln(\mu_{i1}) \quad \ln(\mu_{i2}) \quad \dots \quad \ln(\mu_{iK}))$$

Let  $\lambda$  denote the set  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ . The prior distributions for the hyperparameters  $(\beta, \Sigma)$  need to be identified first to obtain the full Bayes estimates of  $(\lambda, \beta, \Sigma)$ .

In the absence of sufficient prior knowledge of the distributions for individual parameters, uninformative proper prior distributions are usually specified. The most commonly used priors are diffused normal distributions (with zero mean and large variance) for the regression

parameters and a  $Wishart(P, r)$  prior for  $\Sigma^{-1}$ , where  $P$  and  $r \geq K$  represent the prior guess at the order of magnitude of the precision matrix  $\Sigma^{-1}$  and the degrees of freedom, respectively. The parameterization of the Wishart probability density function is

$$f(\Sigma^{-1} | P, r) = |P|^{K/2} |\Sigma^{-1}|^{(r-K-1)/2} \exp\{-0.5\text{Tr}(P \Sigma^{-1})\}. \quad (2.8)$$

Choosing  $r = K$  as the degrees of freedom corresponds to vague prior knowledge (*Spiegelhalter et al., 1996; Tunaru, 2002*).

Let  $f(\beta) \equiv \prod_{k,j} N(0, 10000)$  and  $f(\Sigma^{-1}) \equiv Wishart(P, K)$  denote the hyper-prior distributions for the regression parameters and the precision matrix, respectively. Further, let  $y$  denote the set  $\{y_1, y_2, \dots, y_n\}$ . Given  $X$ , the joint distribution of  $(y, \lambda, \beta, \Sigma)$  is

$$f(y, \lambda, \beta, \Sigma | X) = \frac{f(\beta)f(\Sigma^{-1})}{(2\pi)^{nK/2} |\Sigma|^{n/2} (\prod_{i=1}^n \prod_{k=1}^K y_{ik}^{-1})} g(y, \lambda, X, \beta, \Sigma),$$

where

$$g(y, \lambda, X, \beta, \Sigma) = \exp\left\{-0.5 \sum_{i=1}^n \left( (\lambda_i^* - \mu_i^*)' \Sigma^{-1} (\lambda_i^* - \mu_i^*) + \sum_{k=1}^K [(y_{ik} - 1) \ln(\lambda_{ik}) - \lambda_{ik}] \right)\right\}.$$

The posterior distributions are given by

$$f(\lambda_i | y, X, \beta, \Sigma) = \frac{g_i(y_i, \lambda_i, X, \beta, \Sigma)}{\int_0^\infty \int_0^\infty \dots \int_0^\infty g_i(y_i, \lambda_i, X, \beta, \Sigma) d\lambda_{i1} d\lambda_{i2} \dots d\lambda_{iK}}, \quad (2.9)$$

where

$$g_i(y_i, \lambda_i, X, \beta, \Sigma) = \exp\left\{-0.5 (\lambda_i^* - \mu_i^*)' \Sigma^{-1} (\lambda_i^* - \mu_i^*) + \sum_{k=1}^K [(y_{ik} - 1) \ln(\lambda_{ik}) - \lambda_{ik}] \right\},$$

$$f(\beta | y, \lambda, X, \Sigma) = \frac{g(y, \lambda, X, \beta, \Sigma) f(\beta)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y, \lambda, X, \beta, \Sigma) f(\beta) d\beta_{10} d\beta_{11} \dots d\beta_{KJ}}, \quad (2.10)$$

and

$$f(\Sigma^{-1} | y, \lambda, X, \beta) \equiv \text{Wishart}(P + \sum_{i=1}^n (\lambda_i^* - \mu_i^*)(\lambda_i^* - \mu_i^*)', K + n). \quad (2.11)$$

### 2.3.2 Tobit Collision Modeling

Collision rate analysis is an emerging technique with a growing body of literature (*Zeng et al., 2017*). Compared to using collision count data, rates neutralize the effect of the exposure variable and measure the risk of collision involvement (*Schultz et al., 2007; Xu et al., 2014; Ma et al., 2015*). The most common approach to analyzing collision rates is the Tobit model and *Anastasopoulos et al. (2008)* were among the first studies to introduce this new modeling technique. Since then, other studies applied the Tobit model to understanding the influencing factors of collision rates (*Xu et al., 2013; Xu et al., 2014*).

A critical issue in developing safety models such as traditional Tobit models is accounting for unobserved heterogeneity that is present across observations. For this reason, researchers propose using random parameter Tobit models to analyze collision rates (*Anastasopoulos et al., 2012; Xu et al., 2013; Ma et al., 2015; Yu et al., 2015; Bin Islam et al., 2015; Caliendo et al., 2015; Zeng et al., 2017; Anderson et al., 2017; Guo et al., 2019*). Ignoring this inherent variability could lead to biased parameters and incorrect inferences (*Washington et al., 2011*).

*Ma et al., (2015)* developed a random parameter Tobit model to accommodate serial correlation across observations. Similarly, *Yu et al., (2015)* developed a correlated random parameter Tobit

model to monitor the interactions between independent variables related to the weather and geometric characteristics. Guo et al., (2019) developed four Tobit models under the Bayesian framework to evaluate the impact of various risk factors on collision rates at freeway diverge areas.

Other research developed a random parameter Tobit model to analyze fatality rates for collisions involving heavy vehicles (Anderson et al., 2017; Zeng et al., 2017). More recently, multivariate Tobit models were also developed to model collision rates but by accounting for the injury severity (Anastasopoulos et al., 2016; Sarwar et al., 2017; Zeng et al., 2017; Zeng et al., 2018).

While the literature demonstrates the effectiveness of the use of the random parameter Tobit model, there have been limited studies on using this statistical technique, particularly in the area of automated enforcement. An evaluation of collision rates and the changes in the deployment parameter could enhance and provide a more comprehensive understanding of how best to deploy resources to minimize collision risk at a zonal level.

While more research has been focused on modeling collision frequencies, recent research has shown that Tobit regression models provide some advantages (Anastasopoulos et al., 2008; Anastasopoulos et al., 2012). Instead of using collision counts, this technique considers collision rates which neutralize the effect of the exposure variable and measures the risk of collision involvement (Schultz et al., 2007; Xu et al., 2014; Ma et al., 2015) and was first used in the traffic safety field by Anastasopoulos et al., (2008). Since then, other studies applied the Tobit model to understanding the influencing factors of collision rates (Xu et al., 2013; Xu et al., 2014).

However, traditional Tobit models do not account for the unobserved heterogeneity that is present across observations in the data. Ignoring this inherent variability could lead to incorrect inferences and may lead to biased parameter estimates. To address this, researchers have proposed developing random parameter Tobit models to analyze collision rates and to overcome this limitation (*Anastasopoulos et al., 2008; Yu et al., 2015; Caliendo et al., 2016; Bin Islam, 2016; Zeng et al., 2017; Anderson and Hernandez, 2017; Guo et al., 2019*). More recent studies have also attempted to account for the correlation between collision severities by developing multivariate Tobit models. Anastasopoulos (2016) developed a random parameter multivariate Tobit model to address unobserved heterogeneity of collision injury severity rates. Zeng et al., (2018) and Ulak et al. (2018) also developed a random parameter multivariate Tobit model to analyze crash rate by injury severity. They found significant heterogeneous effects of estimates which varied across observations. Guo et al., (2019) modeled the correlation and heterogeneity in crash rates by different collision types by developing random parameters multivariate Tobit model.

### 2.3.2.1 Traditional Tobit

Tobin (1958) first proposed the Tobit approach to model continuous dependent variables which can be left-censored, right-censored, or both. Given that collision rates are usually left-censored at zero (meaning that no collisions could occur on some roadway segments during an observation period), Anastopoulos et al. (2008) applied the Tobit model in the area of road safety evaluations. The traditional Tobit model for fitting collision rates can be given as

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (2.12)$$



$$Y_i = \begin{cases} Y_i^*, & \text{if } Y_i^* > 0 \\ 0, & \text{if } Y_i^* \leq 0 \end{cases}, \quad i = 1, 2, \dots, N \quad (2.13)$$

where  $Y_i$  is the dependent variable (observed values of collision rates),  $X_{ik}$  is the  $k$ th explanatory variable for observation  $i$ ,  $\beta_0$  is the model intercept,  $\beta_1, \beta_2, \dots, \beta_k$  are the model parameters,  $Y_i^*$  is a latent variable observed only when positive,  $N$  is the number of observations, and  $\varepsilon_i$  denotes the unstructured error which is assumed to be normally and independently distributed with zero mean and variance  $\sigma^2$  as given in

$$\varepsilon_i \sim N(0, \sigma^2) \quad (2.14)$$

### 2.3.2.2 Grouped Random Parameter Tobit

Typically, statistical models fit one regression model to the dataset. This approach does not take into consideration the effect of an explanatory variable on collisions that could vary for different observations. In the context of the deployment of mobile automated enforcement, not all enforcement sites are visited with the same frequency, nor is the duration of enforcement always consistent. For this reason, constraining the parameters can lead to an underestimation of standard errors and inconsistent inference. As such, heterogeneity may exist among different types of freeway diverge areas. As well, collisions may have some commonly shared features at a particular traffic analysis zones (TAZs). Therefore, instead of having a parameter for each observation, the grouping of observations which may share many of the same unobserved effects considers both possible heterogeneity and panel effects (Wu et al., 2013; Sarwar et al., 2017a; Sarwar et al., 2017b; Cai et al., 2018; Fountas et al., 2018). As such, observations in the same

TAZ were grouped as panel data. If the  $i^{\text{th}}$  observation belongs to group  $g(i) \in \{1, 2, 3, 4\}$ , the GRP-Tobit model is given as

$$Y_i^* = \beta_0 + \beta_{g(i),1}X_{i1} + \beta_{g(i),2}X_{i2} + \dots + \beta_{g(i),k}X_{ik} + \varepsilon_i \quad (2.15)$$

where the intercept is set to be a random parameter that follows a normal distribution as following

$$\beta_{g(i),j} \sim N(\beta_j, \sigma_j^2), j = 1, 2, \dots, k \quad (2.16)$$

### 2.3.2.3 Random Intercept Tobit

Another way to account for the within-group correlations and the heterogeneity across different enforcement deployment strategies is to allow only the intercept to vary, which leads to the RI-Tobit model. The RI-Tobit model is given as

$$Y_i^* = \beta_{m,0} + \beta_1X_{i1} + \beta_2X_{i2} + \dots + \beta_kX_{ik} + \varepsilon_i \quad (2.17)$$

where the intercept is set to be a random parameter that follows a normal distribution as following

$$\beta_{m,0} \sim N(\beta_0, \sigma_0^2), m = 1, 2, 3, 4 \quad (2.18)$$

### 2.3.2.4 Multivariate Random Parameter Tobit

The random parameter multivariate Tobit model for fitting collision and crime rates is given as:

$$Y_i^{k*} = \beta_{i0}^k + \beta_{i1}^kX_{i1}^k + \beta_{i2}^kX_{i2}^k + \dots + \beta_{ij}^kX_{ij}^k + \epsilon_i^k$$

where  $Y_i^k$  is the dependent variable for the kth event (k=1, 2 for collision rate and 2 for crime rate),  $X_i^k$  is the explanatory variable for observation  $i$ , and  $\beta_i^k$  is the coefficient corresponding to the kth event. The random parameters  $(\beta_{i1}^k, \beta_{i2}^k, \dots, \beta_{ij}^k)$  are assumed to be multnormally distributed as  $\beta^{ij} \sim N_k(B_j, \phi_j)$ , where,

$$\beta^{ij} = \begin{pmatrix} \beta_{ij}^1 \\ \beta_{ij}^2 \\ \dots \\ \beta_{ij}^k \end{pmatrix}, \beta^j = \begin{pmatrix} \beta_j^1 \\ \beta_j^2 \\ \dots \\ \beta_j^k \end{pmatrix}, \Phi_j = \begin{pmatrix} \varphi_{11}^{j2} & \varphi_{12}^{j2} & \dots & \varphi_{1k}^{j2} \\ \varphi_{21}^{j2} & \varphi_{22}^{j2} & \dots & \varphi_{2k}^{j2} \\ \dots & \dots & \dots & \dots \\ \varphi_{k1}^{j2} & \varphi_{k2}^{j2} & \dots & \varphi_{kk}^{j2} \end{pmatrix} \quad (2.19)$$

### 2.3.3 Model Comparison and Goodness of Fit

The Deviance Information Criteria (DIC) is used for model comparison. The Bayesian estimated models can be compared using the DIC, which is a measure of model complexity as follows

$$DIC = \bar{D} + p_D; p_D = \bar{D} - \hat{D} \quad (2.20)$$

where  $D$  is the un-standardized deviance of the postulated model,  $\bar{D}$  is the posterior mean of  $D$ ,  $\hat{D}$  is the point estimate obtained by substituting the posterior means of the model's parameters in  $D$ , and  $p_D$  is a measure of model complexity estimating the effective number of parameters.

Generally, the model with a smaller DIC outperforms the model with a larger DIC. According to Spiegelhalter et al. (2002), the models with DIC values' difference lower than two are considered equally well, while models with DIC values' difference between 2-7 show considerably less support to the higher DIC model.

A posterior predictive approach (Gelman et al., 1996; Stern and Cressie, 2000; Li et al., 2008) can be used to assess the goodness-of-fit (adequacy) of the model. Such procedures involve

generating replicates under the postulated model and comparing the distribution of a certain discrepancy measure such as the chi-square statistic to the value of chi-square obtained using observed data. A model does not fit the data if the observed value of chi-square is far from the predictive distribution; the discrepancy cannot reasonably be explained by chance if the  $p$ -values are close to 0 or 1 (*Gelman et al., 1996*).

The replicates are best obtained simultaneously with model estimation in WinBUGS to account for all uncertainties in model parameters, as reflected by the estimated distributions. The chi-square statistic is computed from

$$\chi^2 = \sum_{i=1}^n [y_i - E(Y_i)]^2 / Var(Y_i), \quad (2.21)$$

where the  $y_i$  denotes either the observed or the replicated collision or crime frequencies.

#### **2.3.4 Posterior Distribution Estimation**

The posterior distributions needed in the full Bayes approach can be obtained using MCMC sampling. The techniques generate sequences (chains) of random points, whose distributions converge to the target posterior distributions. A sub-sample is used to monitor convergence and then excluded as a burn-in sample. The remaining iterations are used for parameter estimation, performance evaluation, and inference.

Monitoring convergence is important because it ensures that the posterior distribution has been “found”. Thereby indicating when parameter sampling should begin. To check convergence, two or more parallel chains with diverse starting values are tracked to ensure full coverage of the

sample space. Convergence of multiple chains is assessed using the Brooks-Gelman-Rubin (BGR) statistic (*Brooks and Gelman, 1998*). A value under 1.2 of the BGR statistic indicates convergence. Convergence is also assessed by visual inspection of the MCMC trace plots for the model parameters as well as by monitoring the ratios of the Monte Carlo errors relative to the respective standard deviations of the estimates; as a rule of thumb, these ratios should be less than 0.05.

## **Chapter 3: Data Description**

The ability of SPFs to accurately estimate the safety performance of traffic zones is highly dependent on the quality of the available raw data. Therefore, data extraction and aggregation processes are essential in ensuring that these models are reliable and accurate. The objective of this chapter is to provide an overview of the compiled database.

### **3.1 Geographic Scope**

The first step in successfully developing a reliable SPF is to identify the unit of analysis. For micro-level SPFs, the unit of analysis is a specific intersection or approach, or a roadway segment. For macro-level SPFs, the unit of analysis is typically Traffic Analysis Zones (TAZs), or census tracts. The most preferred unit of analysis is TAZ since it is based on scientific principles and is evidence-based. The three main principles for creating a TAZ is for it to have a homogenous land-use, be aligned with jurisdictional boundaries, and be based on trip assignments and population/employment densities. These zones are generated as part of the travel demand modeling system for urban transportation forecasting, such as the EMME/2 model (INRO, 2013). While the majority of the work in this dissertation was based on TAZ as a unit of analysis, due to data availability, there was one application that was developed used census tracts. More details will be discussed further in the final section of this chapter.

The work in this dissertation required different data sources: collision data, exposure data, automated enforcement data, and crime data. The subsequent sections will investigate the data source for each of the different datasets, the aggregation method, and a summary of the data that was extracted for each of the different datasets.

## **3.2 Collision Data**

### **3.2.1 Data Overview & Source**

The City of Edmonton's collision database, the Motor Vehicle Collision Information System (MVCIS), was used to extract the collision data. This database is based on the police record of collisions occurring on public roads within the COE boundaries. For this thesis, the time period spanned from January 2013 to December 2015, inclusive. Generally, three years is used in the literature for model development and hotspot identification processes. This time span was found to be sufficient to account for randomness and regression-to-the-mean bias in the data and was the time frame used since all datasets were available for this three year period.

### **3.2.2 Data Quality & Aggregation**

One of the significant concerns regarding any dataset is location accuracy. This is especially problematic for collision data, which form the basis of any safety review. Exact locations of collisions are usually not known or provided to the police at the time of a collision. However, a more generalized approach to identifying the collision location was adopted by the City of Edmonton's Traffic Safety section. Since most of the collisions are self-reported, they generally have significant horizontal and/or vertical errors along and across the roads. The COE has a mechanism in place to deal with these reporting errors where collisions are generally geocoded to have occurred either at a mid-block location or at the intersection and were geocoded *directly* on road centerlines.

Some of the collisions received by the COE may be missing the roadway portion (e.g., intersection) or maybe missing the collision location entirely. The process to deal with unknown roadway portions or unknown location was as follows:

- Scenario #1: the collision location is known, but the roadway portion is unknown. In some cases, the collisions received by the COE might be missing the roadway portion, or it might not be clearly explained on the police file. These collisions are geocoded to the north-west quadrant of the intersection and not on any specific road centerline. For data aggregation, these collisions were included in the dataset as they occurred within a specific TAZ, and their exact location was not pertinent to the type of analysis conducted in this thesis.
- Scenario #2: the collision location is unknown. In some cases, the collisions received by the COE might be missing the location. These collisions are geocoded to the north-west quadrant of the COE network and would not lie within any TAZ or neighbourhood or community; therefore, they were not included in the analysis.

Since the practice at the City was to directly geocode collisions on the road centerline, there was no concern regarding the collision associations with the road network.

Generally, the boundaries of the TAZs are chosen to coincide closely with boundaries of neighbourhoods (the aggregation unit for the COE municipal census tracts), the existing road network, and the Emme/2 model output (e.g., Vehicle KM Traveled). This facilitates the data extraction and aggregation processes by eliminating the data quality concerns that may arise due to the non-coinciding boundaries. While the boundaries for both the TAZs and neighbourhoods



completely overlapped; the boundaries of the TAZs, the road network, and the Emme/2 links did not; the differences are shown in Figure 3.1 and Figure 3.2. These differences pose a significant concern when aggregating the data; in some instances, the boundary of two TAZs may incorrectly cross a road segment rather than overlap. If undetected, the data corresponding to this road segment would be incorrectly assigned to the TAZs. This was especially challenging for the aggregation of collisions that occurred on the boundary. Therefore, it was imperative to adequately define which collisions were considered to be *boundary collisions*.



Figure 3.1 Difference between TAZ Boundaries (purple) and the Road Network (gray)

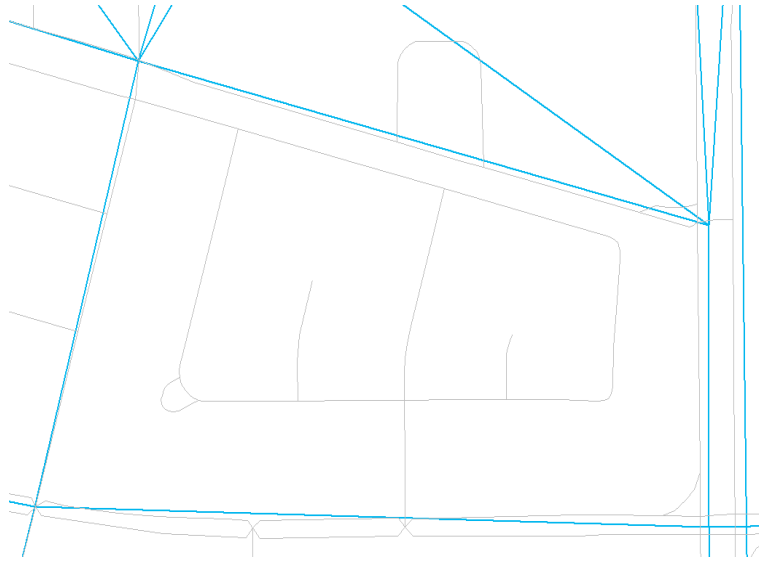


Figure 3.2 Difference between the Road Network (grey) and the Emme/2 Model (blue)

Ideally, collisions geo-coded to the road centerline which divides two TAZs are defined as boundary collisions. However, due to the coarseness of the TAZ boundaries and since they do not entirely overlap with the road network, these collisions will not be automatically identified as boundary collisions. Therefore, creating a buffer was necessary to differentiate between collisions that occurred entirely within each TAZ and collisions that need to be manually snapped to the correct boundary for a TAZ. Sensitivity analysis was conducted to identify the appropriate buffer size, and after several iterations, the most appropriate buffer size was found to be 15 m in width. Therefore, all collisions that occurred within this 15 m buffer were referred to as boundary collisions. Manual data assignment was completed for each of the collisions that occurred within that buffer size.

Based on this definition, the collision dataset was spatially divided into two subsets, and the first contained all collisions that were considered to have occurred within the TAZ and did not lie in

close proximity of any TAZ boundary ( $> 15$  m). The collisions in this subset were aggregated purely based on the frequency of collisions that occurred within the TAZ. The second subset contained all the collisions that occurred on the TAZ boundaries (i.e., within the 15 m buffer). The aggregation process for boundary collisions was more challenging and required additional analysis.

A review of the literature showed that the boundary collisions generally amounted to less than 5% of the total collisions and did not significantly impact the results. Therefore, the common practice in previous research was to ignore these collisions entirely. Based on observation of COE data, most of the TAZ boundaries coincide with arterial roads with heavy traffic volume and where collisions occur more frequently. Therefore, collisions occurring on the boundaries of TAZs accounted for a relatively more substantial proportion of the total collisions. Since more than 5% of collisions occurred on the boundaries, the standard rule of ignoring those collisions was considered inappropriate.

The aggregation method used in this project followed the conventional rule of data assignment. The collisions were assigned to the TAZ based on the direction of the travel for each record. For example, if an East-West road was on the boundary of two TAZs (one north of the road and the other south of the road). Then all collisions occurring in the west direction were associated with the TAZ to the north. Conversely, all collisions that occurred in the east direction were associated with the TAZ to the south. After the buffer was created and all the collisions were included, the collisions were manually associated according to the closest TAZ depending on the boundary.

### 3.2.3 Data Summary

The collision data were categorized by severity: fatal, injury, and property damage only. Due to the low frequency of fatal collisions, they were grouped with the injury collisions. Therefore, in this thesis, the collisions were provided by Property Damage Only (*PDO*), Injury and Fatal (*I+F*), and Total (*TOT*) collisions.

Variable	Symbol	MIN	MAX	MEAN	STDEV
Property Damage Only Collisions	PDO	1	749	142	141
Injury and Fatal Collisions	I+F	0	125	22	23
Total Collisions	Total	2	855	165	164

Table 3.1 Descriptive Statistics Collision (2013-2015) Data by TAZ

## 3.3 Exposure Data

### 3.3.1 Data Overview & Source

The City of Edmonton's Transportation Planning Branch provided Emme/2 model output files for the exposure variable. The Emme/2 is a classic four-stage gravity-based model: trip generation, trip distribution, trip mode, and trip assignment. The model output includes traffic volumes and travel times, which can be aggregated by link, mode, node, and zone (INRO. 2013). The files contained aggregated data on travel forecasts for a typical morning rush hour across the COE, including zonal totals for vehicle-kilometers-travelled, average zonal speed, and average zonal congestion (i.e., volume-to-capacity ratio). The data was based on the 2014 model developed for the COE, which was validated by the end of 2015 and was provided in a raw spatial format.

### **3.3.2 Data Quality & Aggregation**

As discussed previously, the TAZ boundaries did not overlap entirely with neither the road network nor the output from the Emme/2 model. The gap between the boundary and the road layer was not consistent and varied in displacement (between 5 – 15 m). Therefore, before extracting the data, an analysis of the different scenarios was conducted to ensure that the exposure data was extracted carefully and accurately. There was a total of 23,792 links output from the Emme/2 model. However, these links showed the connectivity within the COE and between the COE and neighboring municipalities. The links used for analysis in this research were the links confined within the boundaries of the COE and those amounted to 12,933 links. The primary variable that was taken to account for exposure was the Vehicle-KM-Travelled (VKT). For the aggregation process, the VKT value of each link was summed to achieve a representative value for each TAZ.

Upon careful review of the exposure data, four scenarios were identified to facilitate the data extraction and aggregation processes.

#### **Scenario #1: Inside**

Explanation: this scenario includes all the links that lied completely within one TAZ. Figure 3.3 shows an example of a link that lies entirely within one zone, TAZ 303. This scenario accounts for 63% (8,134 links) of the total links.

Aggregation: in this scenario, the exposure variables were included (summed for the VKT and averaged for the VC and the average speed) for each link that is included entirely within one zone. An example:  $VKT_{TAZ=303} = VKT_{link}$

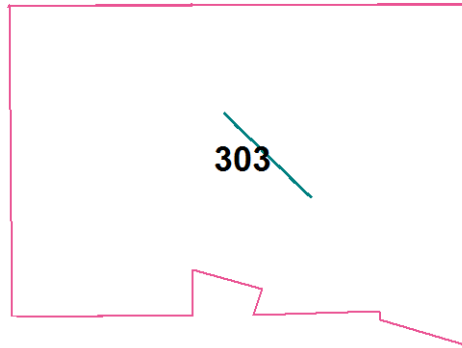


Figure 3.3 Example of a Link Completely within One TAZ

### Scenario #2: Crossing

Explanation: this scenario includes all the links which were considered to be crossing two or more zones. There were a total of 12,933 links output from the Emme/2 model. The “Crossing” links amounted to 1,508 (~12% of the total links). An example of this scenario is shown in Figure 3.4. The link (in blue) is shown crossing two zones, 2908 and 2925.

Aggregation: in this scenario, the exposure variables (e.g., VKT) were divided between the two TAZ according to the length of the link. This is similar to breaking the link into two separate sub-links, each lying entirely within one TAZ. As an example:

$$VKT_{TAZ=2908} = VKT_{link} \frac{350}{(350+178)}; \text{ and } VKT_{TAZ=2925} = VKT_{link} \frac{178}{(350+178)}; \text{ similar calculations}$$

were done for the remaining links.

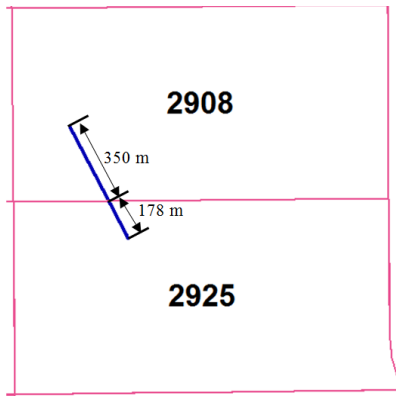


Figure 3.4 Example of a Link Crossing Two Zones

### Scenario #3: Overlap

Explanation: this scenario includes all the links which were found to entirely lie on the boundary of two or more TAZs. An example of a link that falls in this scenario is shown in Figure 3.5. There was a total of 3,226 links in this scenario, amounting to ~ 25% of the total links.

Aggregation: In this type of scenario, the exposure variables were divided between the two TAZs according to the population. An example:

$$VKT_{TAZ=2927} = VKT_{link} \times \frac{Population_{TAZ=2927}}{Population_{TAZ=2927} + Population_{TAZ=2938}}$$

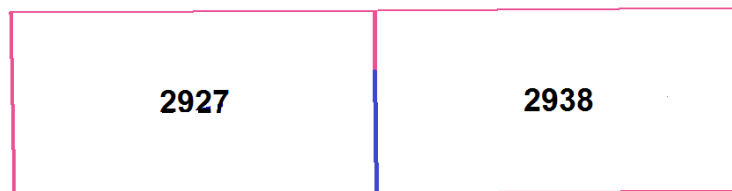


Figure 3.5 Example of a Link Overlapping with the Boundary of two TAZs

#### **Scenario #4: Inside Overlap**

Explanation: this scenario includes all the links which were partially included in one zone and partially on the boundary of two or more zones. Figure 3.6 illustrates an example of the type of link included in this scenario. This scenario comprises 0.5% of the links (61 out of 12,933 links).

Aggregation: for the purpose of calculating the exposure variables for each of the zones, the link was divided into two parts: the VKT calculation for part 1 follows that of the Scenario #3 and the VKT calculation for part 2 follows that of Scenario #1. Similarly, for the remaining two exposure variables.

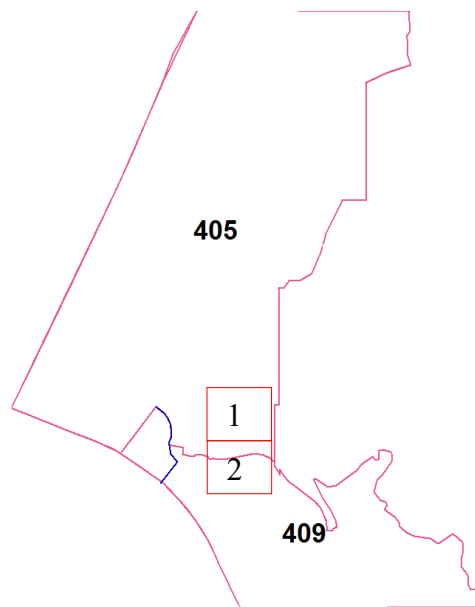


Figure 3.6 Example of a Link Classified as Inside Overlap



### 3.3.3 Data Summary

Table 3.2 provides a summary of the data extracted for the exposure theme for the data that was cleaned and aggregated.

Variable	Symbol	MIN	MAX	MEAN	STDEV
Vehicle-KM Travelled	VKT	10	20,365	3,857	3,495

Table 3.2 Descriptive Statistics for Exposure Data

## 3.4 Automated Enforcement Data

### 3.4.1 Data Overview & Source

Edmonton had two types of mobile speed enforcement technologies: photo laser detection systems and Photo Radar (PR) detectors. The photo laser detection system utilizes LiDAR technology and is capable of capturing vehicles' violating the speed limit of the road that is being enforced. However, this system is not as widely used as the photo-radar due to weather and location limitations and for this reason was not included within the scope of this research.

At the time of the research, Edmonton had 13 units that were fitted with photo radar equipment: 10 were installed in unmarked vehicles and 3 were installed in a marked vehicle. Additionally, there were a total of four photo-radar enforcement squads which were responsible for the operation of the photo radar equipment. Each day, there were two shifts of PR enforcement conducted by the squads. For the purpose of this thesis, all PR enforcement will be referred to as MAE.

A single enforcement unit consists of an operator who is a trained and qualified peace officer, a sergeant (squad leader), a vehicle, and a piece of enforcement equipment. The operator is always present in the vehicle with the equipment to observe violating vehicles and make notes accordingly. The devices monitor and capture any speed infractions committed by vehicles exceeding the speed limit.

The MAE sites are locations that are enforced using photo radar and are reflected in the spatial data as a road segment with a singular directional attribute, for example, a site enforcing traffic traveling westbound. Typically, the schedulers are provided with a list of sites to visit, 50 to 70 of these sites are included in the weekly schedule which is created for the operators based on information about speed violation rates, site priority, complaint history, high-risk time of day and day of the week time periods, and weather conditions, as well as safety.

Schedulers typically assign operators a vehicle to use each shift unless a vehicle is taken out of the circulation (e.g., for a service appointment). Typically, the operators arrive early to have a brief meeting with their squad members, perform their start-up tests, and then travel to the first enforcement site. They setup at a site and complete their deployment which typically ranges from 1 to 3.8 hours. The minimum duration can be reduced for a number of reasons which could include the weather, other field conditions, or if few speed violations were recorded during the deployment. Operators normally take breaks between their first and second deployment on a given shift and then continue to their next site.

Squads typically work four days on and then have four days off and operators who worked for 10 hours must rest at least eight hours before their next shift. Enforcement is typically deployed on all days around the year with the a few exceptions of some statutory holidays.

At the time of this research, there was no specific requirement on when a deployment occurs, and this is mainly decided upon by the operators. However, some locations are assigned or avoided during specific times of the day based on speed survey information. For example, it is typical to avoid deployment on road segments during rush hour as the road is too congested and all vehicles travel well below the speed limit. However, some sites experience high volumes of commuter speeders during rush hours or after major events, so these sites can be highlight for deployment specifically at times where higher speed violation rate occurs. Site are typically visited by different operators more than once in a scheduling cycle. Normally, there is no specific requirement on the deployment frequency for one site, however, some sites, specifically newly created ones, might require a minimum number of visits.

In Edmonton, there are five police divisions: North, West, Downtown, Southeast, and Southwest. As per the scheduling decision, the smaller Downtown area is combined with the larger North area to efficiently distribute enforcement resources; therefore, there are four enforcement quadrants. There is neither a requirement on how many sites from each quadrant should be included in the shift schedule nor a requirement on how many sites from each quadrant must be visited. The sergeants assign operations to different police divisions when scheduling. Each squad assigns one or two units to each of the quadrants each shift.

The schedule is adjusted weekly and can vary from week to week as some sites are visited more frequently than others. Site identification is based on public complains, construction complains, and is in keeping with the provincial automated enforcement guidelines. When identifying a new enforcement site to include in the monthly schedule, if the site was previously active and exists in the database, it only needs to be reactivated. If the site has never been enforced before and there is no data supporting the existence of a speeding problem at the location, a speed count is ordered. A speed count is completed by installing equipment on the surface of the road which detects the speed of the vehicle traveling; this data cannot be collected during the winter or if there is snow on the ground. The data is typically then collected in the spring after the Roadway Maintenance group in the City of Edmonton completes their annual spring sweep.

Once a speed survey can be collected, the results are analyzed and if the results support a speeding problem, the location's collision records from the past three years are then reviewed. Then the area will be validated for feasibility of enforcement and for operator safety. After meeting the previous requirements, a request is then sent to the Edmonton Police Service, the city's governing enforcement body, to obtain approval for the new enforcement site. Excluding the time needed for the speed survey results, the approval process which starts from site assessment and ends with entering the site information into the system, usually takes five to 11 days. It is important to note that there is currently a moratorium on the creation of new enforcement sites in Edmonton and that the process described above was in place for the time period of the data used for this research (2013-2015).

Changes may occur at approved sites which can affect the sites' suitability for enforcement. These sites are then identified and validated to determine whether they need to be deactivated either permanently or temporarily, depending on the nature of the underlying issues. A formal policy for site deactivation was developed. At a minimum, the sites are reviewed annually by the supervisor of the mobile enforcement program. The enforcement effectiveness at sites is based on not only a comparison of the speed data from year to year (whenever available) but also based on community feedback and the site's own speed violation counts. However, given the limited resources and strict budget, this emphasizes the need for a comprehensive evaluation framework to optimize the use of a single deployment to maximize the safety impact.

This data was slightly different from the previously extracted datasets. Firstly, all the automated enforcement data was displayed geographically within the COE boundaries. Immediately it was apparent that the deployment of the automated enforcement data occurred within 315 TAZ. This is intuitive since Edmonton suffers from urban sprawl, and the suburban zones along the edge of the City boundaries are still being developed. Therefore, only information for the 315 TAZs was extracted for the 2013-2015 time periods.

The COE Traffic Safety Section provided information regarding mobile automated enforcement data. Raw data was provided in a GIS format (which included a map of all of the locations where automated enforcement was conducted) and an accompanying spreadsheet which included data related to the deployment strategy (i.e., the number of enforced sites per zone, the number of enforcement hours spent at each site, and the number of enforcement visits for each site) as well as variables related to the infraction (i.e., the number of issued tickets). This data was extracted

for the same time period used for collisions (i.e., from 2013 to 2015) and was aggregated by TAZ.

### **3.4.2 Data Quality & Aggregation**

Since the COE periodically monitors their enforcement data, there were no challenges with the quality of the data. However, the same issue was apparent in the aggregation of enforcement data by TAZ, as was previously discussed in the Collision Data section. A similar approach of associating the data by zone was followed in that if an enforcement site was located on the boundary of two zones, the assignment of the enforcement indicators was conducted according to the direction of travel of the vehicles that were enforced.

### **3.4.3 Data Summary**

Two variables were selected as a proxy for the deployment strategy: *hpersite* (average number of hours spent at each site), and *vpersite* (average frequency of visits for each enforcement site). However, since these two variables are highly correlated, the ratio of the hours of enforcement per visit of these variables (*ratio*) was used in the model development instead. Since the data used here was a subset of that extracted for this dissertation, a summary of the collision data for the relevant TAZ is shown below in Table 3.3.

<b>Variable</b>	<b>Symbol</b>	<b>MIN</b>	<b>MAX</b>	<b>MEAN</b>	<b>STDEV</b>
<b>Property Damage Only Collisions</b>	<i>PDO</i>	1	749	208	143
<b>Injury and Fatal Collisions</b>	<i>I+F</i>	0	125	32	25
<b>Total Collisions</b>	<i>Total</i>	2	855	241	167
<b>Total Mid-block Collisions (Speed Related)</b>	<i>Total-ms</i>	2	428	92	61
<b>Vehicle KM Traveled</b>	<i>VKT</i>	457	20365	5550	3558
<b>Number of tickets issued per enforcement site</b>	<i>tickets</i>	1	31934	809	2451
<b>Average hours spent at each enforcement site</b>	<i>hpersite</i>	14	1198	61	129
<b>Average frequency of visits for each enforcement site</b>	<i>vpersite</i>	1	378	22	42
<b>Hours of enforcement per visit</b>	<i>ratio</i>	1	3.78	2.41	0.54

Table 3.3 Summary of data used for Mobile Automated Enforcement analysis

### 3.5 Crime Data

One of the goals of the dissertation is to develop models to determine the effectiveness of mobile automated enforcement on collisions and crime. As such, it was necessary to develop macro-level SPFs at locations where there was automated enforcement present. Therefore, a subset of the data was used for this application – more information is provided below.

#### 3.5.1 Data Overview & Source

The crime data was extracted from the City of Edmonton’s (COE) online Open Data Catalogue. In this catalogue, there are seven indicators of crime that are monitored and captured by neighbourhood. They include the number of assaults, number of break and enters, number of

robberies, number of sexual assaults, number of reported incidences of thefts from a vehicle, number of reported incidences of vehicle thefts, and number of homicides. Incidences of crime are defined into two categories: personal crimes (i.e., crimes that are committed against an individual), or property crimes (i.e., crimes that are related to property). Property crimes (incidents of thefts from vehicles, vehicle theft, and break and enter) were considered in this study.

### **3.5.2 Data Quality & Aggregation**

Unlike other datasets used in this dissertation, raw data was not available for individual criminal activities. The only information that was provided was the type of criminal activity that was committed. No further data was available for analysis. Additionally, the data was already aggregated by neighbourhood (which is based on the census tract in Edmonton). It was not possible to match each of the neighbourhoods to a TAZ, so for this application, all the datasets that were used were aggregated according to neighbourhoods. This included: collision data, exposure data, automated enforcement data, and population data. The reason population data was included for analysis is based on previous research. Research on the topic of developing models to predict crime included population as an exposure measure. This is similar to the use of an exposure measure (e.g., vehicle-km traveled) when modeling collisions. The results of the 2014 Municipal Census were used in this study, and since the population was already provided by neighbourhood, there was no need for any aggregation.

The other datasets (i.e., collision, exposure, and automated enforcement data) were aggregated by neighbourhoods by following the same principles described in previous sections. This was



done for the 105 neighbourhoods which had mobile automated enforcement and criminal activity.

### 3.5.3 Data Summary

Since the data for this theme is a subset of the overall data that was extracted for this dissertation, Table 3.4 summarizes the data used in the development of the SPFs.

<b>Variable</b>	<b>Symbol</b>	<b>MIN</b>	<b>MEAN</b>	<b>MAX</b>	<b>STDEV</b>
<b>Criminal Incidences (3 yrs)</b>	<i>crime</i>	8	229	1371	224
<b>Total Collisions (3yrs)</b>	<i>collisions</i>	31	378	1629	266
<b>Vehicle KM Traveled</b>	<i>VKT</i>	1168	7211	26998	4815
<b>Population</b>	<i>pop</i>	345	3365	15038	1923
<b>Average hours spent at each enforcement site</b>	<i>hpersite</i>	13	82	495	84
<b>Average frequency of visits for each enforcement site</b>	<i>vpersite</i>	5	30	174	30
<b>The ratio of hours of enforcement per site to the frequency of visits per site</b>	<i>ratio</i>	1.33	2.58	3.64	0.42

Table 3.4 Summary of data for Crime and Collision Analysis

## **Chapter 4: Automated Enforcement & Traffic Safety**

This chapter investigates the application of macro-level models as it related to automated enforcement and collisions.

### **4.1 Introduction**

Collisions have been a leading cause of death and injury for many years, and at the core of the traffic injury and fatality, the problem is speed (World Health Organization, 2017). The relationship between speed and the severity of a collision has been well established and documented in the literature (Elvik, 2005) as the average speed increases, not only does the likelihood of a collision occurring increase, but also the severity of the outcome. The Nilsson Power model suggests that for a 1 km/h increase in the average speed of the vehicle, there is a 4-5% increase in fatal collisions (Nilsson, 2004). To curb this safety concern, agencies have devised different strategies for managing speed, ranging from physical changes to the roadway to educational campaigns to enforcement programs.

Making physical changes to the roadway in the form of engineering changes is one of the most effective ways to improve safety since road users will model their behaviors based on the road environment. However, these design changes are costly, time-consuming, and cannot typically be completed on a large scale. Another strategy that road agencies often use is creating educational campaigns. An example of this would be a campaign that targets speeding and explains the risk and danger of traveling over the speed limit. One main challenge that impacts this is road users can understand this concept and believe that exceeding the speed limit is dangerous but still choose to behave that way. People's beliefs, attitudes, and behavior do not

often align, which is why educational campaigns can resonate with people but still do little to impact how they behave on the roads (Corporate Research Associates, 2019). For this reason, road safety agencies often rely on enforcement. By enforcing the rules of the road, road users' behaviors can be changed as there is a direct consequence as a result of their actions. Therefore, all traffic safety strategies typically include an enforcement component.

Enforcement programs generally include both manned and automated enforcement. Manned enforcement is conducted by police officers who use speed measurement devices to detect vehicles exceeding the speed limit. While this type of enforcement is effective at changing behavior due to the immediate action of being apprehended and issued a ticket, there are several challenges associated with only using manned enforcement. Firstly, it is very resource-intensive and costly; this is mainly a concern when compared to the number of speeding motorists. Therefore, the use of manned enforcement would result in a low perceived risk of being apprehended (Axup, 1990; Zaal, 1994). Additionally, specific locations such as sites with very high volume may present a risk to the enforcement personnel (Li et al., 2015). For this reason, automated enforcement, such as the use of red-light cameras and photo-radar, are typically used to supplement the overall enforcement program.

Mobile automated enforcement has been used to improve safety, and previous work provides strong evidence supporting the impact that these enforcement programs have on improving safety on a national or a municipal level (Axup, 1990; Cameron et al., 1992; Elvik, 1997; Chen et al., 2000; Delaney et al., 2003; Tay and de Barros; 2011; Luoma et al., 2012). However, at such a high level of analysis, changes in traffic volume, overall collision trends, and other

confounding factors are not usually captured. They can negatively affect the quality of the results (Thomas et al., 2008).

Other research on this topic aimed to quantify the impact that these programs have on safety but on a road segment (Li et al., 2015; Li et al., 2016). At this micro-level, the impact of mobile automated enforcement on collisions and safety has been quantified. Still, there is very little insight as to how this relationship can be used to inform an agency's deployment strategy. To better assist enforcement agencies in the planning of their deployment strategy, a more appropriate unit of analysis that addresses gaps in both the systemwide level and the micro-level is necessary.

The objective of this chapter is threefold. Firstly, to bridge the gap in the existing literature and to determine the impact that mobile automated enforcement activity has on collisions but at a macro-level, a unit that is large enough to inform the deployment strategy but also precise enough to capture any changes in the traffic volume. This unit of analysis will need to allow for the investigation of how parameters such as the duration of enforcement at a site and the frequency of visits in a year impact resourcing. The second objective is to develop models that would provide enforcement authorities with an empirical tool to help plan their deployment strategy and understand the impact of that deployment strategy on collisions. Finally, the third objective is to better understand how varying the number of hours of enforcement or varying the number of times a site was visited impacted safety. To explore these objectives, different modeling techniques were adopted, and they are described in the subsequent sections.

## 4.2 Impact of Automated Enforcement on Macro-level Collisions

As explained previously, the impact of automated enforcement on a micro-level (i.e., road segment or intersection) has already been studied. This section will further explore this relationship but on a macro-level. To do this, the selected unit of analysis for macro-level modeling is the TAZ, and that data needed to explore this relationship was collected and aggregated by zone as outlined in Chapter Three. This included information regarding collisions, exposure (i.e., vehicle-km-travelled), and enforcement related indicators. The indicators were classified into two categories: a deployment-related indicator, *ratio*, (i.e.,  $h_{persite}/v_{persite}$  - the ratio of the number of hours spent enforcing a site to the number of times a site is visited per year), and an outcome related indicator, *tickets* (i.e., the average number of tickets issued per site per year). Using this data, different macro-level SPFs, as well as deployment consultation charts, were developed to help road agencies understand how automated enforcement can improve safety.

### 4.2.1 Model Development

When trying to understand the impact of automated enforcement on collisions, two models were developed using Poisson Lognormal technique. It is assumed that  $Y_i$  denotes the number of collisions at TAZ  $i$  ( $i = 1, \dots, n$ ) and that collisions at the  $n$  TAZs are independent and that

$$Y_i | \theta_i \sim \text{Poisson}(\theta_i). \quad (4.1)$$

To address over-dispersion for unobserved or unmeasured heterogeneity, it is assumed that

$$\theta_i = \mu_i \exp(u_i), \quad (4.2)$$

where  $\mu_i$  is determined by the covariates such as the VKT and the MAE deployment parameters (i.e., tickets or ratio of hours of enforcement per visit) which are specific to each TAZ and a corresponding set of unknown regression parameters, whereas the term  $exp(u_i)$  represents a multiplicative random effect. The PLN regression model is obtained by the assumption

$$exp(u_i) | \sigma_u^2 \sim \text{Lognormal}(0, \sigma_u^2) \quad \text{or} \quad u_i | \sigma_u^2 \sim \text{Normal}(0, \sigma_u^2) \quad (4.3)$$

where  $\sigma_u^2$  denotes the extra Poisson variance. Note that in the PLN model  $exp(u_i)$  follow a Lognormal distribution. Under the PLN model

$$E(Y_i) = \mu_i exp(0.5 \sigma_u^2), \quad \text{Var}(Y_i) = E(Y_i) + [E(Y_i)]^2 (exp(\sigma_u^2) - 1). \quad (4.4)$$

A diffused normal distribution with zero mean and a large variance is the most commonly used prior to estimating the regression parameters (El-Basyouny and Sayed, 2009). The diffused gamma distribution  $\text{Gamma}(0.001, 0.001)$  was used as the prior of precisions for  $\sigma_u^2$  (El-Basyouny and Sayed, 2009).

The posterior mean and standard deviation were sampled using the Markov chain Monte Carlo (MCMC) technique using WinBUGS software. The MCMC method uses the sampling technique to generate chains of random points, the distribution of which converges to the target posterior distributions. The convergence was monitored in several ways. First, two parallel chains with diverse starting values were tracked so that the full coverage of the sample space is ensured. Brooks-Gelman-Rubin (BGR) statistic was also used, where convergence occurred if the value

of the BGR statistic was less than 1.2. Typically, convergence is achieved with the Monte Carlo errors relative to the respective standard deviations of the estimates are less than 0.05 (*El-Basyouny and Sayed, 2009*). In this study, the posterior summaries were obtained using two chains with 100,000 iterations, 10,000 of which were excluded as a burn-in sample. Examination of the BGR statistics, ratios of the Monte Carlo errors relative to the standard deviations of the estimates and trace plots for all model parameters indicated convergence.

The first set of models (Models 1) included the number of tickets issued per site within each TAZ. The coefficients of the variables and the  $p$ -values are shown in Table 5.1. The regression coefficients of the exposure variable are significant and positive (as expected), indicating a positive relationship between predicted collisions and the vehicle-km traveled. The results also show a significant negative relationship between tickets issued per site and predicted collisions. This relationship is intuitive since the purpose of the automated enforcement program is to deter speeding. By reducing the incidences of drivers exceeding the speed limit, the frequency of collisions is also expected to decrease, which is similar to the results found in the research on a micro-level.

The second set of models (Model 2) identifies the relationship between the enforcement activity or deployment strategy and the frequency and severity of collisions. The coefficients of the variables and the  $p$ -values are shown in Table 5.2. The regression coefficients of the exposure variable are significant and positive (as expected), indicating a positive relationship between predicted collisions and the vehicle-km traveled. There is a negative relationship between predicted collisions and the *ratio* parameter. This indicates that an overall collision reduction is

associated with spending a longer duration, enforcing a site for each visit throughout the year. This relationship is investigated further by comparing two different deployment strategies and the effect that has on the frequency of collisions.

A tool was developed based on Models 1 and 2 to quantify and visualize the impact of a given deployment on the predicted collisions. This will allow agencies to better understand and manage their programs and resources more efficiently and effectively. This tool can be used in two ways by using an SPF and/or a deployment consultation chart. The use of this tool is further illustrated by using case studies, examples of which are provided in the subsequent section.

#### **4.2.2 Deployment Consultation Charts**

A set of deployment consultation charts (monographs) were created based on Models 1 and 2 to provide road agencies with an empirical tool to easily quantify the impact their deployment strategy has on collisions. To estimate the predicted *Total* collision frequency based on the number of tickets issued per site, the enforcement agencies would follow these steps:

- 1- Determine the vehicle-km traveled for the TAZ
- 2- Use Figure 5.1(a) to estimate the expected *Total* collision frequency (which is plotted on the *y*-axis) using the expected VKT (on the *x*-axis) along with the number of tickets issued per site (per year)

Alternatively, the equation below can be used to quantify the impact of the change in the number of tickets issued per site per year:

$$\mu = 0.407VKT^{0.648}exp^{-0.0001tickets} \quad (4.5)$$



where  $\mu$  is the expected *Total* collision frequency per year, VKT is the vehicle-km traveled for the specific TAZ that is being investigated, and *tickets* are the number of tickets issued per site per year for the same TAZ.

To estimate the predicted *Total* collision frequency based on the enforcement activity, the enforcement agencies would follow these steps:

- 1- Determine the vehicle-km traveled for the TAZ
- 2- Determine the number of hours of enforcement planned per site and the number of visits per year
- 3- Use Figure 5.2(a) to estimate the expected Total collision frequency based on the VKT and the *ratio* of *hpersite* to *vpersite*

Alternatively, the equation below can be used to quantify the impact of the change in the deployment strategy:

$$\mu = 0.431VKT^{0.683}exp^{-0.456ratio} \quad (4.6)$$

where  $\mu$  is the expected *Total* collision frequency per year, VKT is the vehicle-km traveled for the specific TAZ that is being investigated, and the *ratio* is the portion of the *hpersite* to *vpersite*.

Variable	<i>Total</i>			<i>Total-ms</i>		
	Coefficient	Confidence Interval		Coefficient	Confidence Interval	
<b>a<sub>0</sub></b>	0.407	-1.305	-1.255	-0.267	-1.310	-0.030
<b>VKT</b>	0.648	0.474	0.780	0.547	0.479	0.674
<b>tickets</b>	-0.0001	-0.0001	-0.001	0.000	0.000	0.000
<b>p.chi</b>	0.697	0	1	0.663	0.000	1.000
<b>Goodness-of-fit Criteria</b>						
<b>DIC</b>	2852			2567		

Variable	<i>I+F</i>			<i>PDO</i>		
	Coefficient	Confidence Interval		Coefficient	Confidence Interval	
<b>a<sub>0</sub></b>	0.431	-2.557	-0.437	-0.163	-1.005	-0.056
<b>VKT</b>	0.558	0.436	0.687	0.627	0.530	0.730
<b>tickets</b>	-0.0001	-0.0001	-0.001	-0.0001	-0.0001	-0.001
<b>p.chi</b>	0.756	0.000	1.000	0.717	0.000	1.000
<b>Goodness-of-fit Criteria</b>						
<b>DIC</b>	2189			2810		

Table 4.1 Overview of Model 1 Results

Variable	<i>Total</i>			<i>Total-ms</i>		
	Coefficient	Confidence Interval		Coefficient	Confidence Interval	
<b>a<sub>0</sub></b>	0.4306	0.3305	1.485	0.5985	0.2885	1.284

<b>VKT</b>	0.6827	0.5747	0.7977	0.543	0.457	0.646
<b>ratio</b>	-0.4558	-0.6061	-0.0995	-0.3558	-0.4761	-0.2396
<b>p.chi</b>	0.6405	0	1	0.5955	0	1

<b>Goodness-of-fit Criteria</b>	<b>Total</b>	<b>Total-ms</b>
<b>DIC</b>	2852	2567

<b>Variable</b>	<b>I+F</b>			<b>PDO</b>		
	<b>Coefficient</b>	<b>Confidence Interval</b>		<b>Coefficient</b>	<b>Confidence Interval</b>	
<b>a<sub>0</sub></b>	0.9254	1.957	0.2037	0.7156	0.2659	1.26
<b>VKT</b>	0.6066	0.4545	0.7438	0.6527	0.5876	0.7183
<b>ratio</b>	-0.4149	-0.5646	-0.2687	-0.475	-0.5694	-0.2493
<b>p.chi</b>	0.6745	0	1	0.6433	0	

<b>Goodness-of-fit Criteria</b>	<b>I+F</b>	<b>PDO</b>
<b>DIC</b>	2187	2809

Table 4.2 Overview of Model 2 Results

### 4.2.3 Case Study: Comparison of Deployment Strategies

To better understand how to use these charts and to test the feasibility of these models, case studies are presented to determine how best to proceed with the deployment strategy. Using Figure 4.1(a) or Equation (22), for a TAZ with 5,000 VKT, visiting a site and issuing 1,000 tickets per year to vehicles exceeding the speed limit would result in approximately 90 collisions per year. Alternatively, visiting a site and issuing 5,000 tickets per year would yield 60 collisions per year. The safety effect of this change can be quantified using the equation below:

$$\text{Overall Safety Effect (\%)} = \frac{N_a - N_b}{N_b} \times 100 \quad (4.7)$$

where  $N_b$  is the expected collision frequency of the base strategy (i.e., 90 collisions in the case study), and  $N_a$  is the expected collision frequency of the alternative strategy (i.e., 60 collisions in the case study). According to Eq'n (22), the safety effect of the alternative strategy of increased hours of enforcement per site and less frequent visits to each site is -33%, with the negative sign indicating a reduction in the expected collision frequency.

For Model 2, two examples are presented to demonstrate how these charts can impact an agency's deployment strategy i) a case study representing the impacts of a less concentrated enforcement effort by spending fewer hours per visit, and ii) a case study representing a concentrated enforcement effort by spending a more extended time per visit throughout the year.

To select realistic deployment parameters, an overview of the City of Edmonton's program is necessary. Enforcement sites are selected per Alberta Justice and Solicitor General Guidelines as well as local expertise. These sites are created for any of the following reasons, the prevalence of

i) high frequency of collisions, ii) high frequency of speeding (as evidenced by the results of speed studies), iii) proximity to a school or playground zone or iv) high frequency of speeding complaints made by members of the public.

Enforcement shifts are ten hours in duration and include time spent on transportation to and from the site, equipment setup, breaks, as well as the actual enforcement time spent at each site. The maximum continuous amount of time that an operator can spend enforcing a site per visit is 3.8 hours. However, the duration of enforcement at each site and the frequency of visits of these sites is left to the discretion of the photo radar operator. Therefore, the developed models would help provide the enforcement agencies with an empirical tool to quantify the impact their deployment strategy has on safety.

Using these parameters, the first example compares the differences in two deployment strategies based on an average deployment in a year. For a TAZ with 10,000 VKT, for a deployment strategy that represents a more dispersed enforcement effort, for example, an average of 1.5 hours of enforcement per visit per year, results in approximately 117 collisions. However, increasing this ratio to 3.5 hours of enforcement per visit per year results in approximately 47 collisions. This is a reduction of 60% in the expected *Total* collisions per year.

For a more detailed application of how this ratio affects the deployment strategy within the parameters of the City's enforcement program, more detailed case-studies are discussed. First, the baseline of average deployment is presented. For a TAZ with 10,000 VKT, an average deployment in a year of 65 hours of enforcement that were completed by visiting a site twice a

month (i.e.,  $ratio = 2.5$ ), the average number of expected collisions using Figure 4.2(a) or Equation (23) is 74 collisions per year. An example of a dispersed enforcement effort could include spending 18 hours enforcing a site per year but visiting this site once a month each year (i.e.,  $ratio = 1.5$ ). The expected number of collisions using this strategy yields 117 collisions per year (a 37% increase in collisions from an average deployment).

An example of a more concentrated enforcement effort in a year could include spending 182 hours enforcing a site and visiting this site once a week (i.e.,  $ratio = 3.5$ ). The expected number of collisions using this strategy yields 47 collisions per year (a 36% reduction in collisions from an average deployment). Therefore, to achieve a reduction in collisions, this model indicates that the number of hours of enforcement for each visit should be increased.

To further investigate the impact of changing one parameter at a time, two examples are provided. Over a year, for a site that is visited once a week, increasing the number of hours of enforcement results in a reduction in collisions. For example, for the same TAZ (of 10,000 VKT), if the site was enforced for a total of 104 hours each week (i.e.,  $ratio = 2$ ), the expected *Total* collision frequency per year is 93 collisions. Increasing the total hours of enforcement by 50% to 156 hours of enforcement (i.e.,  $ratio = 3$ ), the expected collision frequency is 59 collisions per year, resulting in a 37% collision reduction.

Alternatively, for the same TAZ, over a year for the same number of enforcement hours, reducing the number of visits per year results in a reduction in collisions. For example, if 100 hours of enforcement were completed by visiting a site 50 times (i.e.,  $ratio = 2$ ), the expected

collision frequency is 93 collisions per year. If the number of visits is decreased by 20% to 40 visits per year (i.e., *ratio* = 2.5), the expected collision frequency is 74 resulting in a 20% collision reduction.

The previous examples showcased how the change in the deployment strategy for existing TAZ could impact the collision frequency. A higher ratio of hours of enforcement per visit (i.e., the more time spent at a site for each unique visit) is associated with the lower collision frequency. This is useful to quantify the impact of operational changes on safety and provides agencies with a tool that can assist in the decision-making process of their deployment strategy. However, to further refine the use of these models, additional deployment charts were developed to assist planners when knowledge of parameters such as the VKT is unknown.

The impact that speed has on collisions has been investigated quite extensively. Even with the studies quantifying the impact that road design has on speed, agencies are still finding challenges with regards to achieving compliance with the speed limit. Therefore, even with new developments, an understanding of how enforcement programs can assist in achieving compliance with the speed limit is still vital. Since the impact of the deployment strategy has already been studied using Models 1 and 2, it is just as essential to be able to provide guidelines for how these tools can be used proactively to plan for resourcing needs. Figure 4.1(a) allows planners to understand the impact the frequency of *tickets* issued has on TAZ with different VKT. For the same number of tickets issued (e.g., 5,000 tickets issued per site), the expected *Total* collisions for a TAZ with a higher VKT (e.g., 20,000) is approximately 150 collisions per

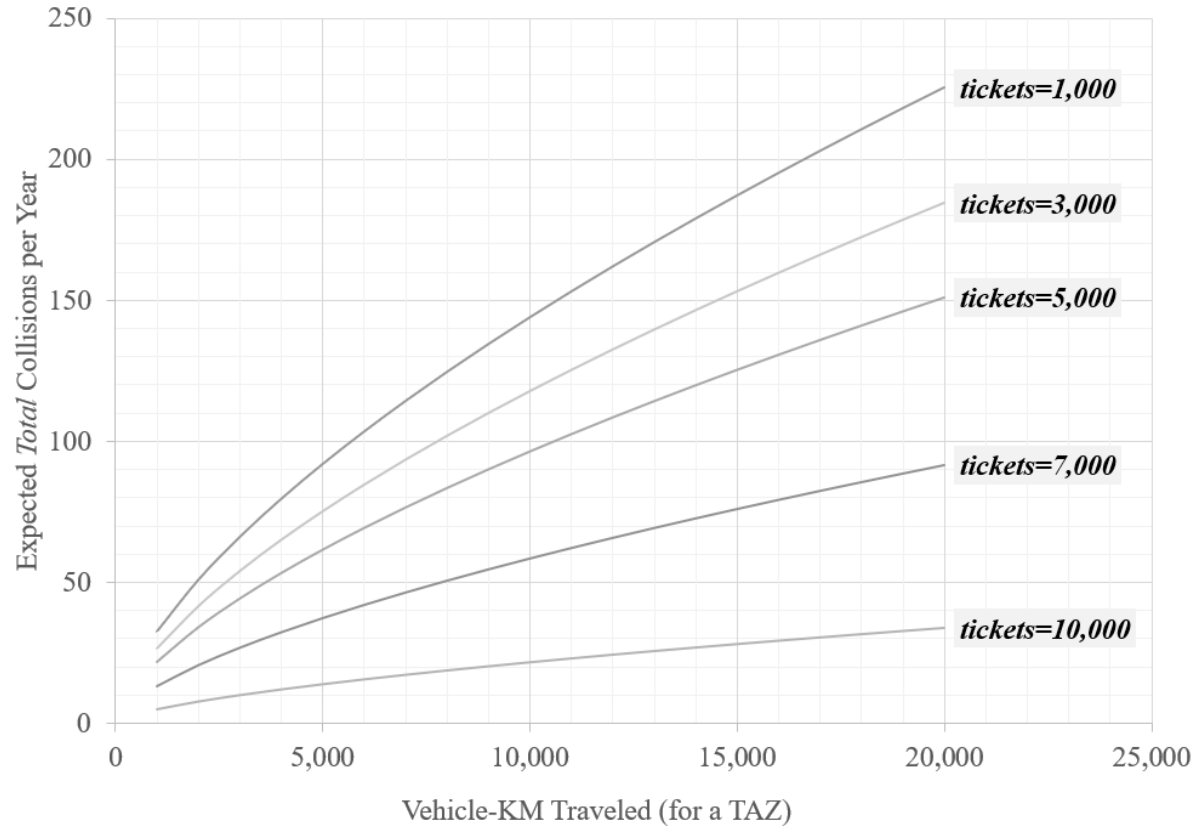
year. However, for a TAZ with a lower VKT (e.g., 5,000), the expected *Total* collisions is approximately 60 collisions per year.

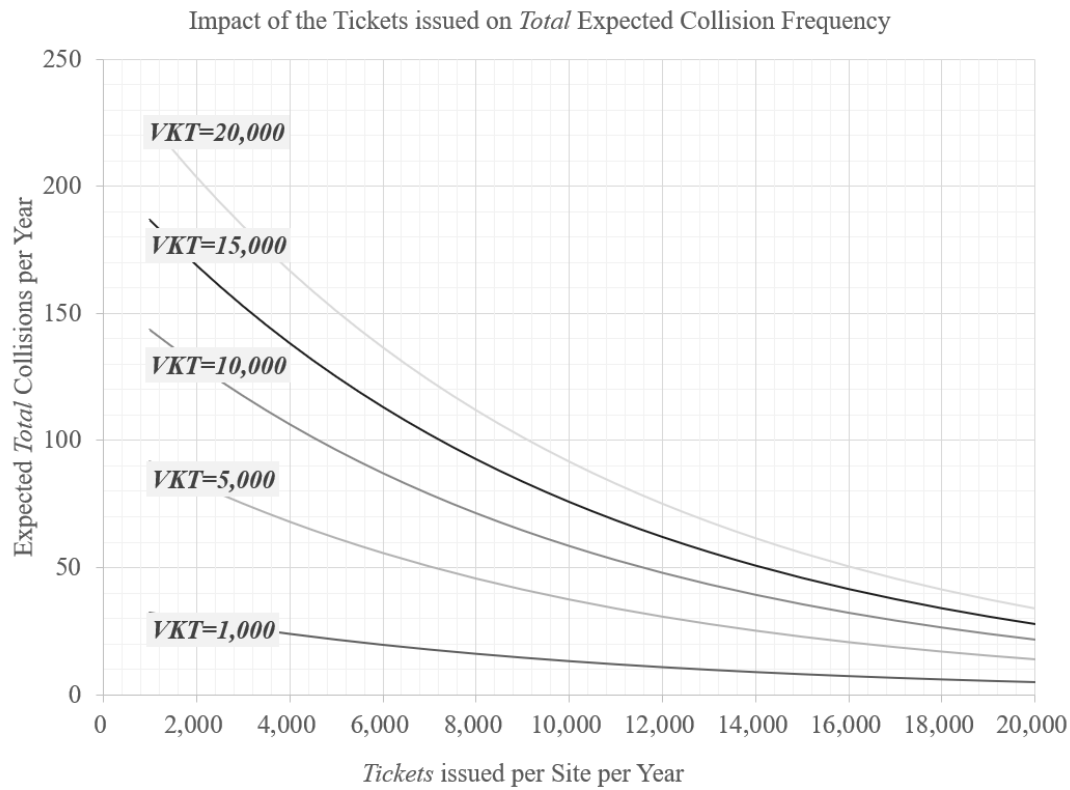
Figure 4.2 (b) demonstrates the impact that the *ratio* of hours of enforcement per visit has for TAZ values with varying VKT. Spending an hour enforcing a site per visit each year (*ratio=1*) has a significant impact on TAZ with different VKT. A TAZ with expected VKT of 5,000, the expected *Total* collision is 92 collision per year. For the same deployment strategy at a TAZ with expected VKT of 15,000, the expected *Total* collisions is 193 collisions per year. This demonstrates that a TAZ with a higher volume would benefit more from spending a longer time enforcing a site compared to a TAZ with a lower volume. This is intuitive since increasing the time spent enforcing a site increases the driver's expectation of receiving a ticket, and as a result, they reduce their speed. However, these charts quantify this relationship and provide an evidence-based approach for enforcement agencies to better plan their strategy to ensure that safety is prioritized.



Impact of the Tickets issued on *Total* Expected Collision Frequency

(a)



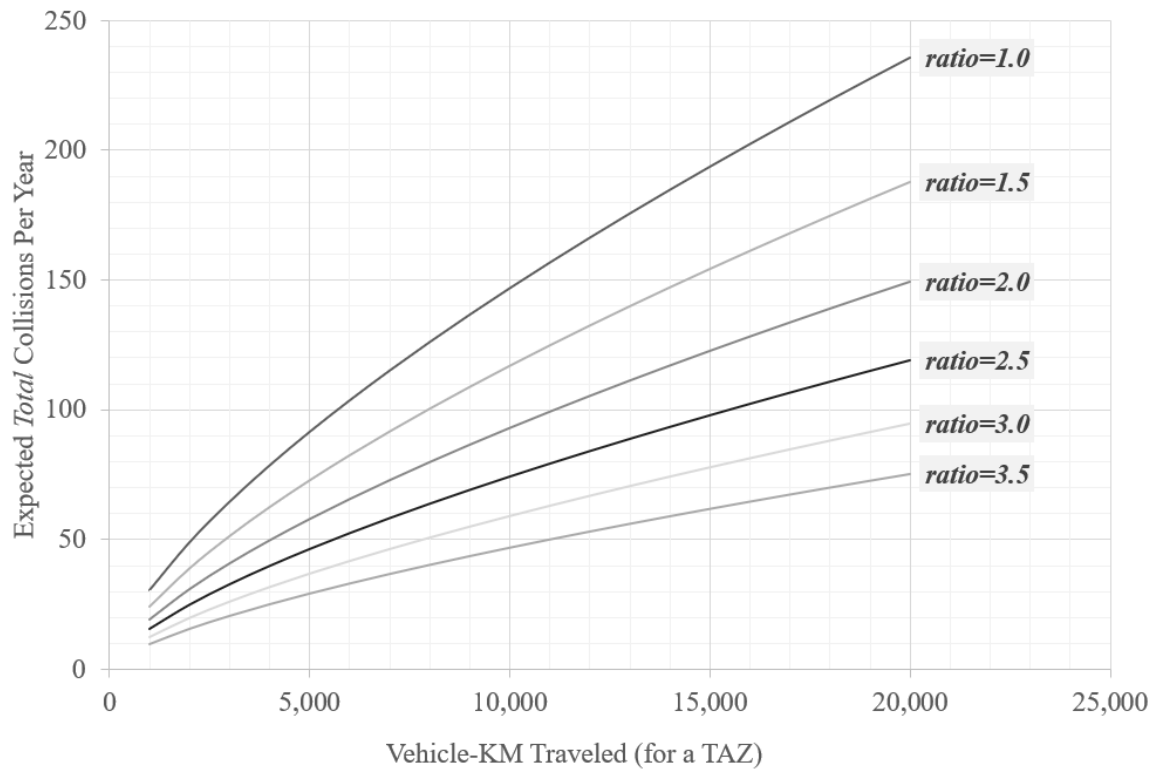


(b)

Figure 4.1 Charts demonstrating the relationship between VKT, Tickets issued per site, and the expected Total collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart

Impact of the Ratio of Hours of Enforcement to Average Number of Visits per Site on *Total* Expected Collision Frequency

(a)



Impact of the *Ratio* of Hours of Enforcement to Average Number of Visits on  
*Total Expected Collision Frequency*

(b)

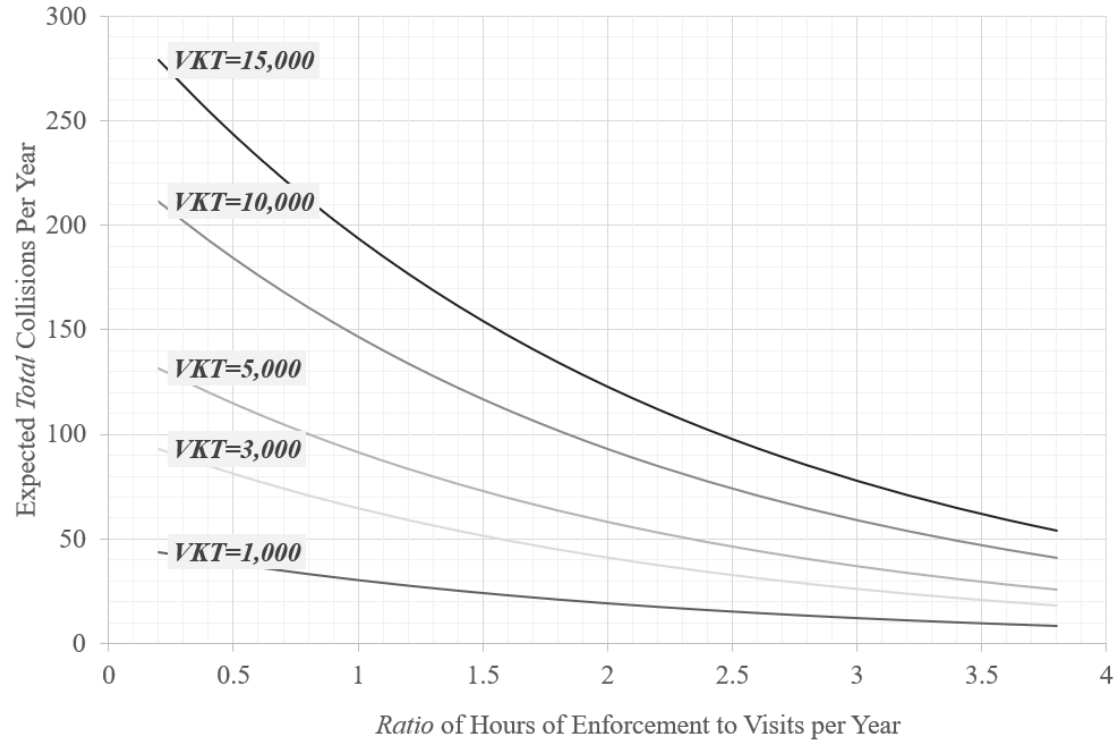
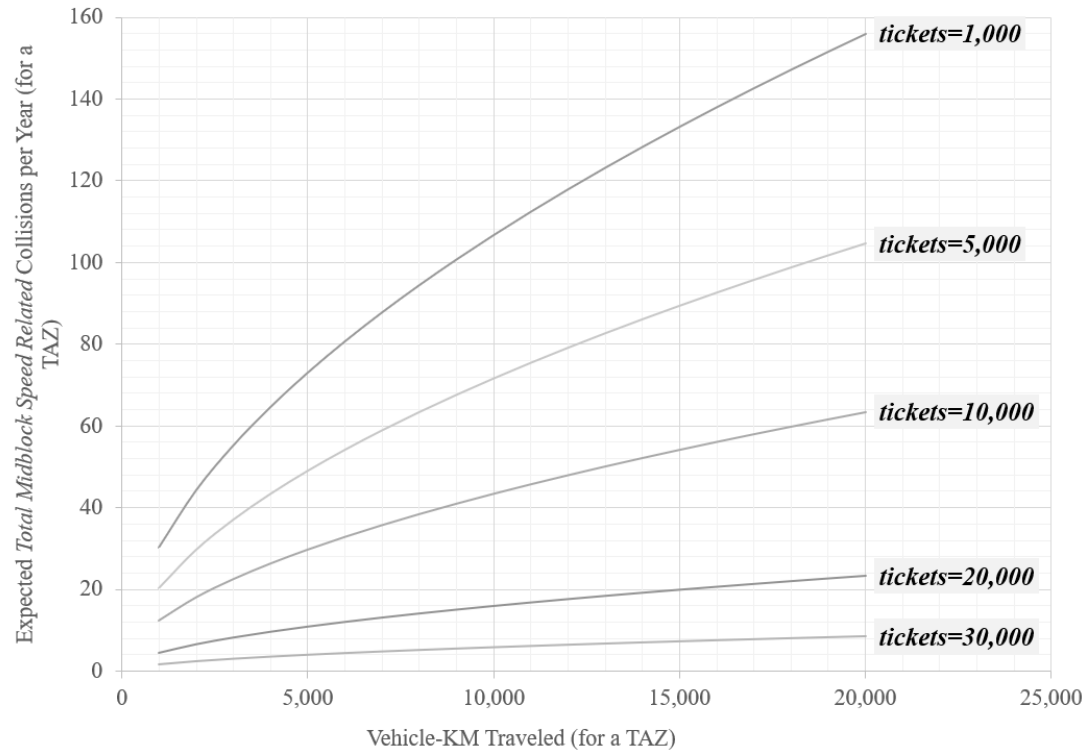
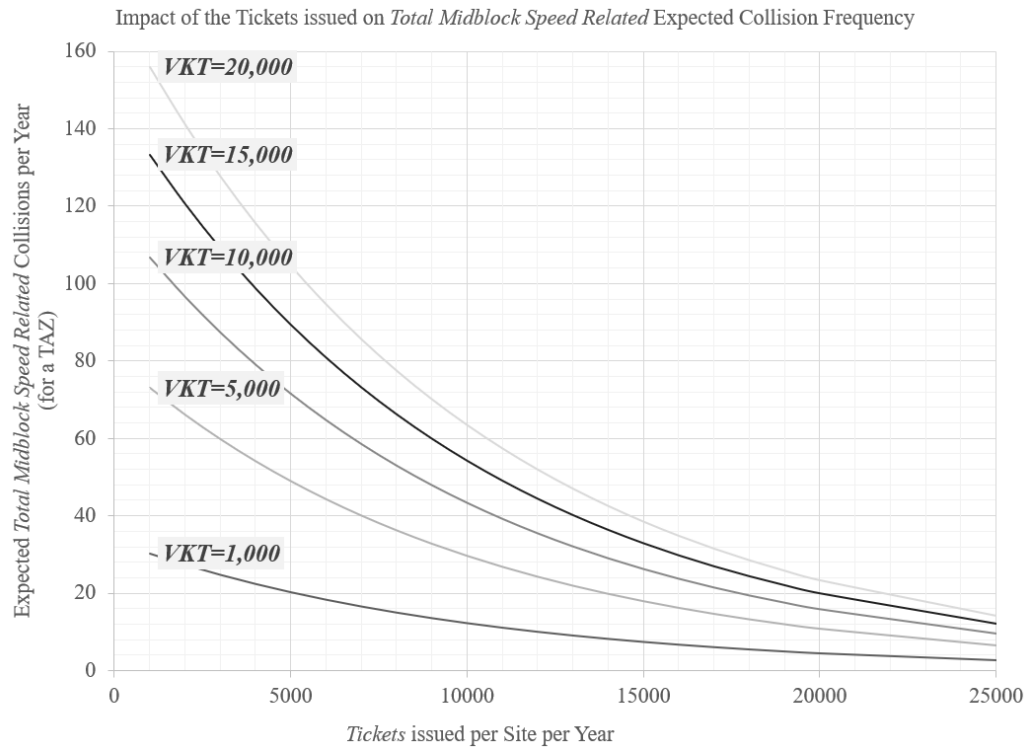


Figure 4.2 Design charts demonstrating the relationship between VKT, the ratio of hours of enforcement per site to the average number of visits per site, and the expected Total Midblock Speed Related collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart

Impact of the Tickets issued on *Total Midblock Speed Related* Expected Collision Frequency

(a)



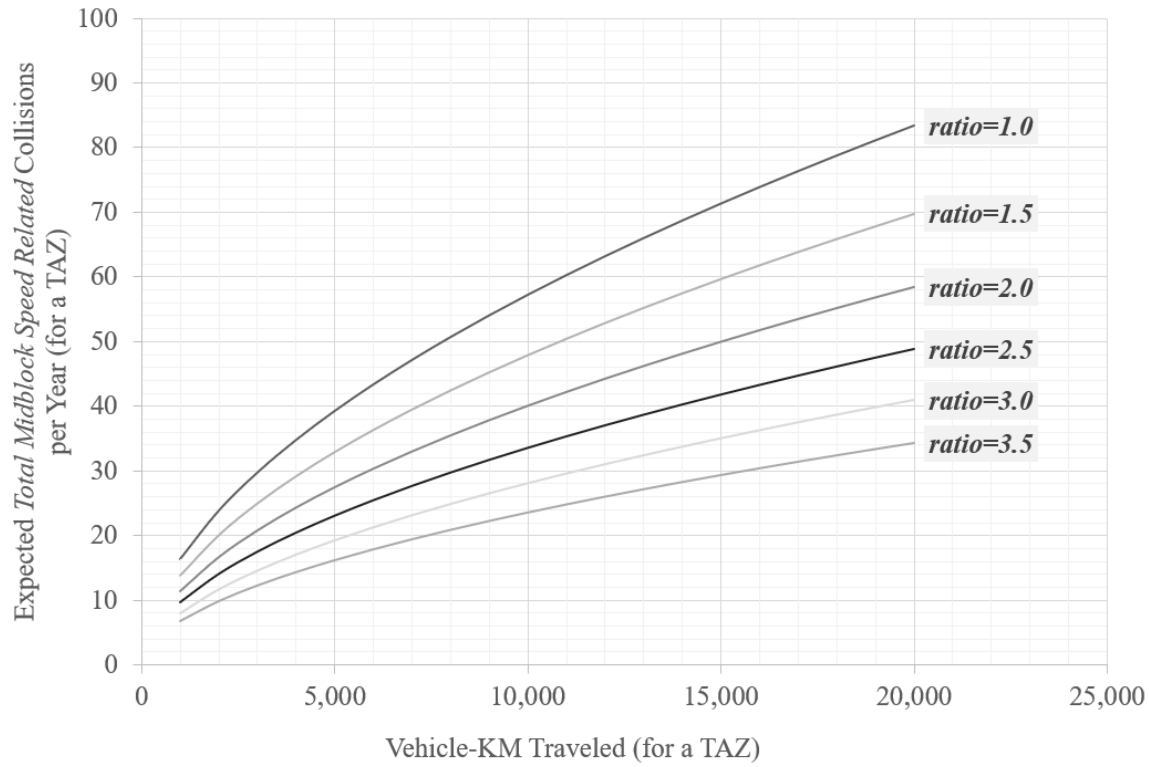


(b)

Figure 4.3 Charts demonstrating the relationship between VKT, Tickets issued per site, and the expected Total Midblock Speed Related collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart

Impact of the Ratio of Hours of Enforcement to Average Number of Visits per Site on *Total Midblock Speed Related Expected Collision Frequency*

(a)



(b)

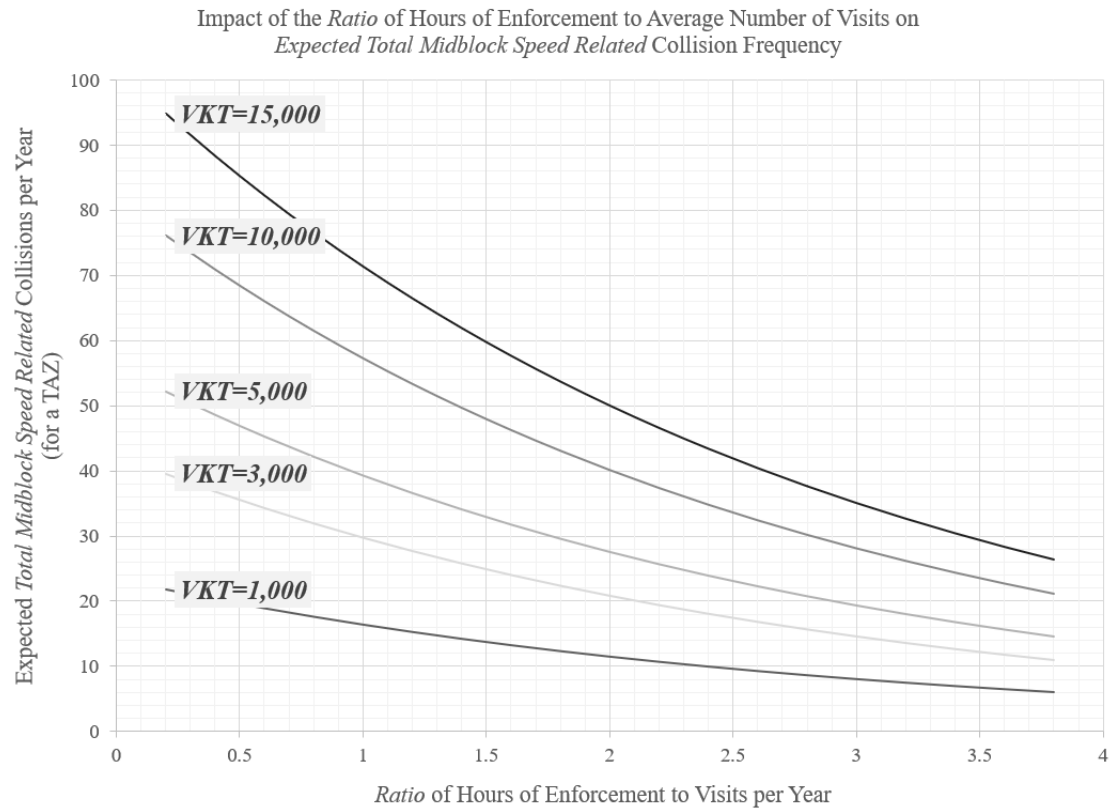
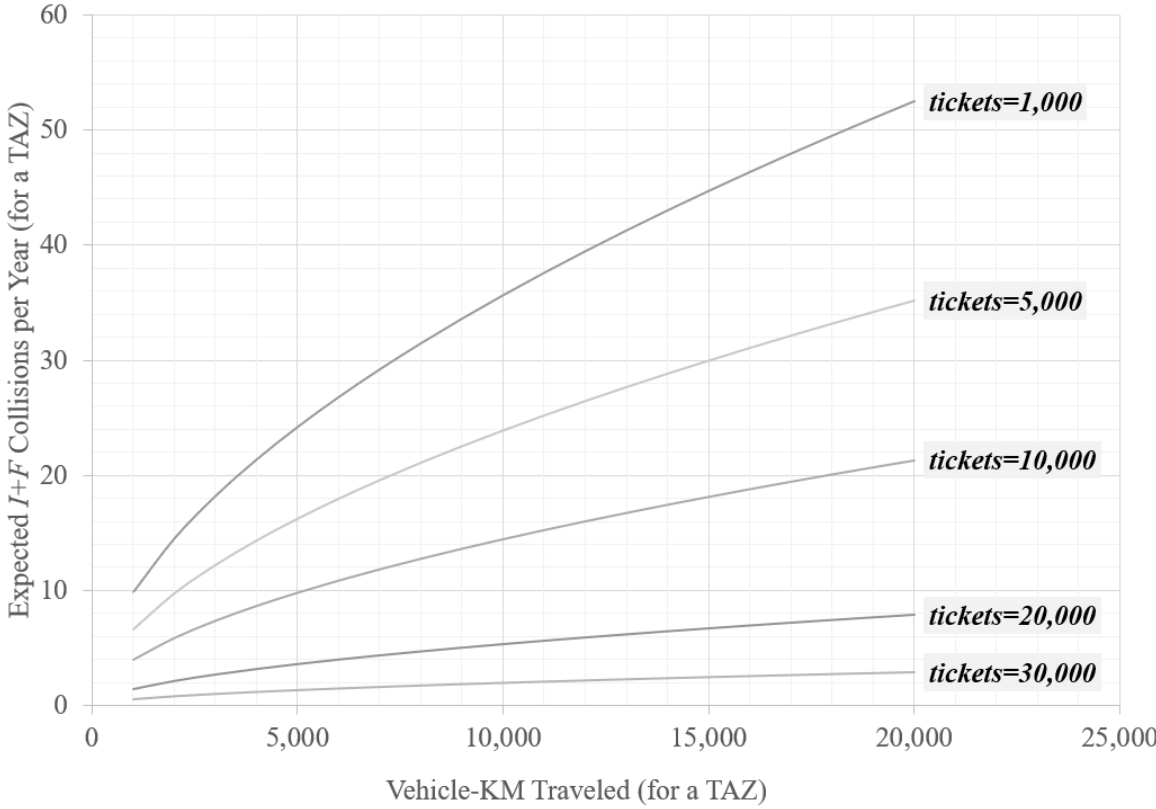


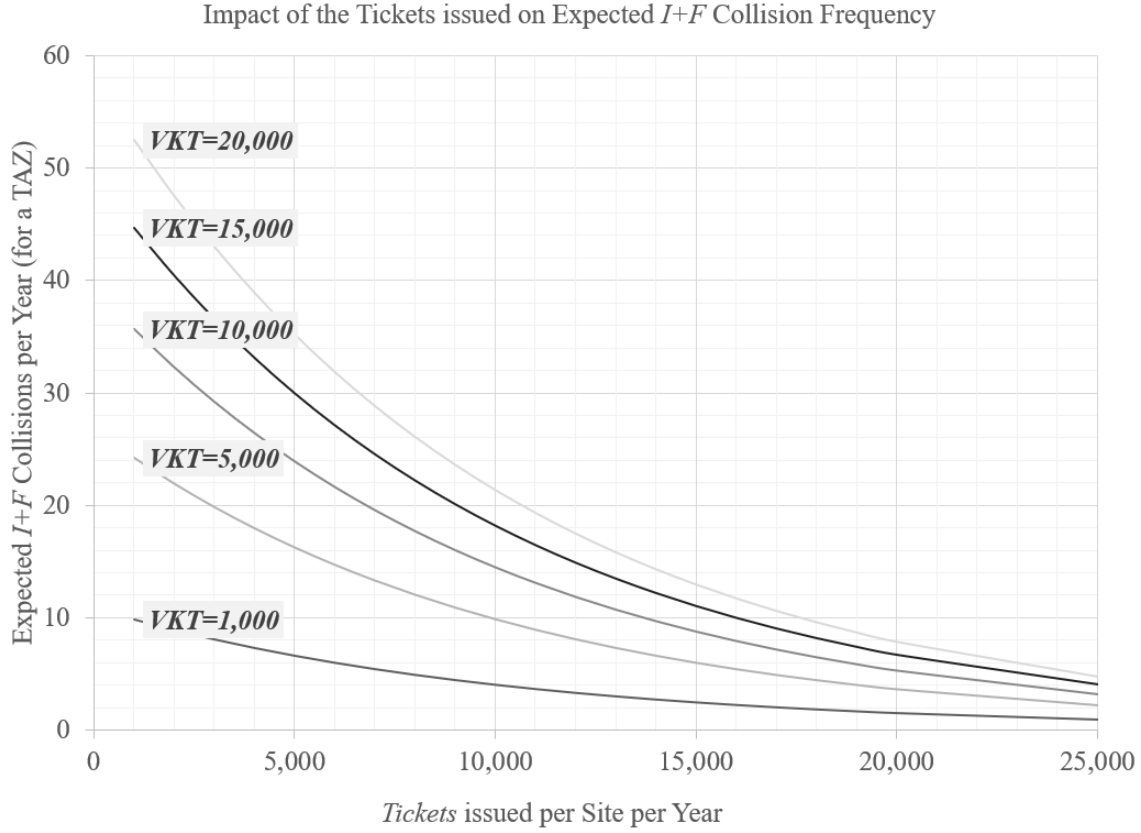
Figure 4.4 Design charts demonstrating the relationship between VKT, the ratio of hours of enforcement per site to the average number of visits per site, and the expected Total collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart



Impact of the Tickets issued on Expected  $I+F$  Collision Frequency

(a)



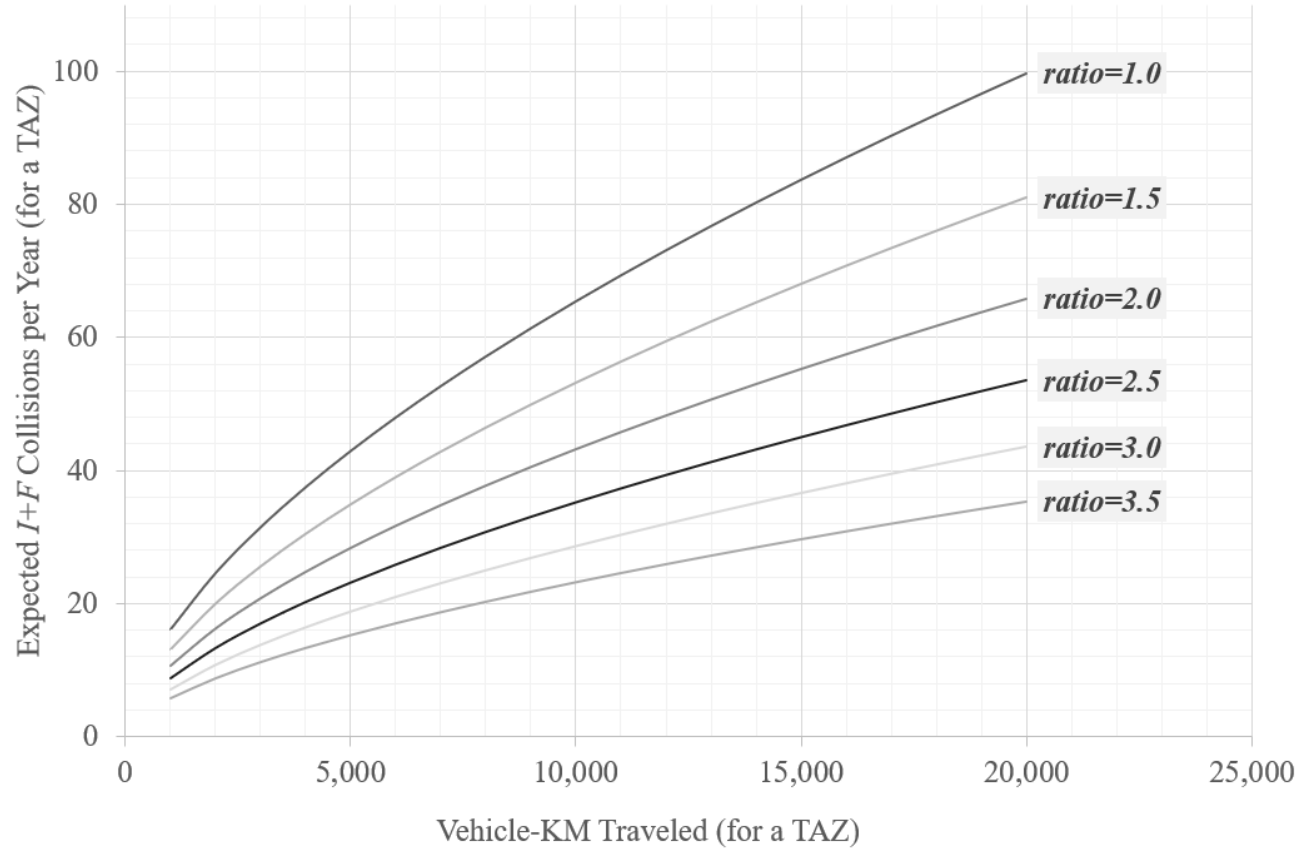


(b)

Figure 4.5 Charts demonstrating the relationship between VKT, Tickets issued per site, and the expected *I+F* collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart

Impact of the Ratio of Hours of Enforcement to Average Number of Visits per Site on  $I+F$  Expected Collision Frequency

(a)



Impact of the *Ratio* of Hours of Enforcement to Average Number of Visits on *I+F* Expected Collision Frequency

(b)

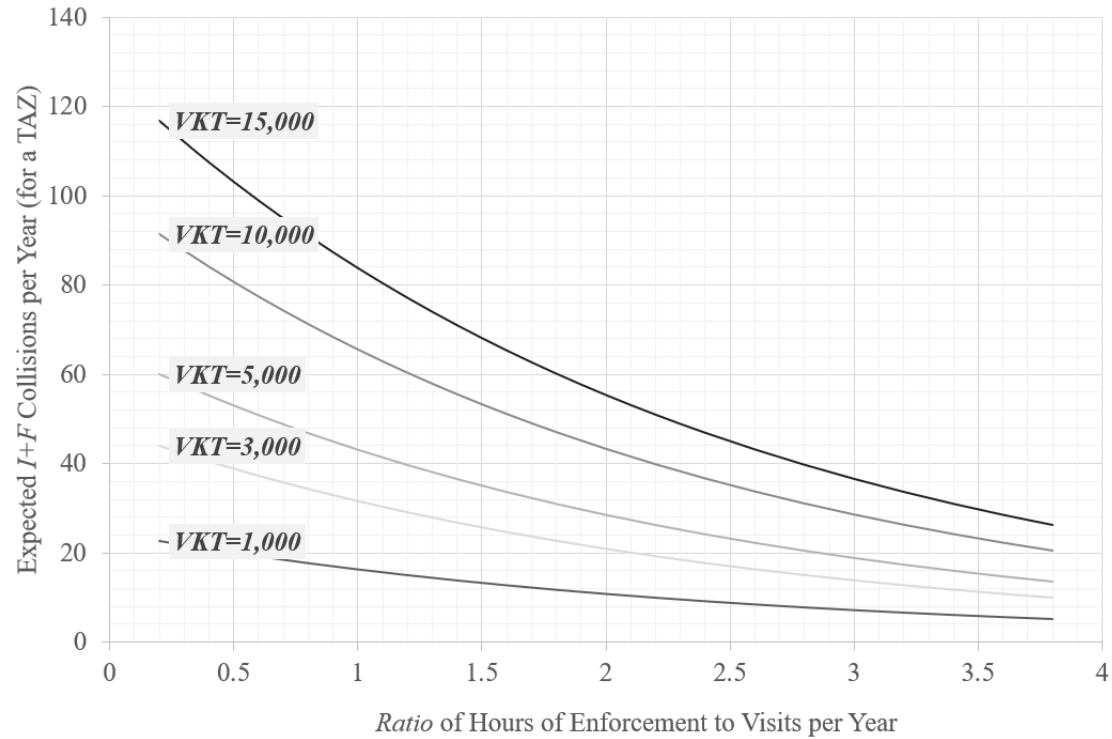
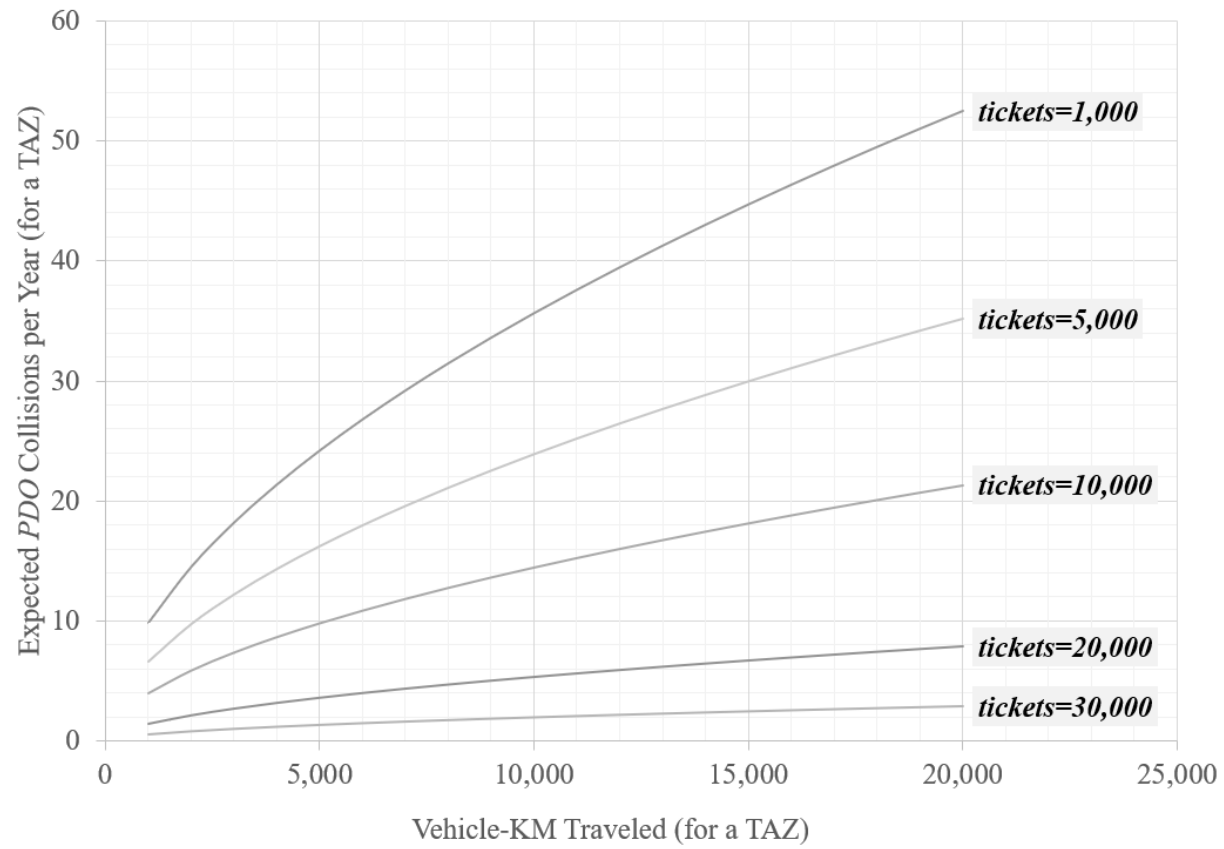
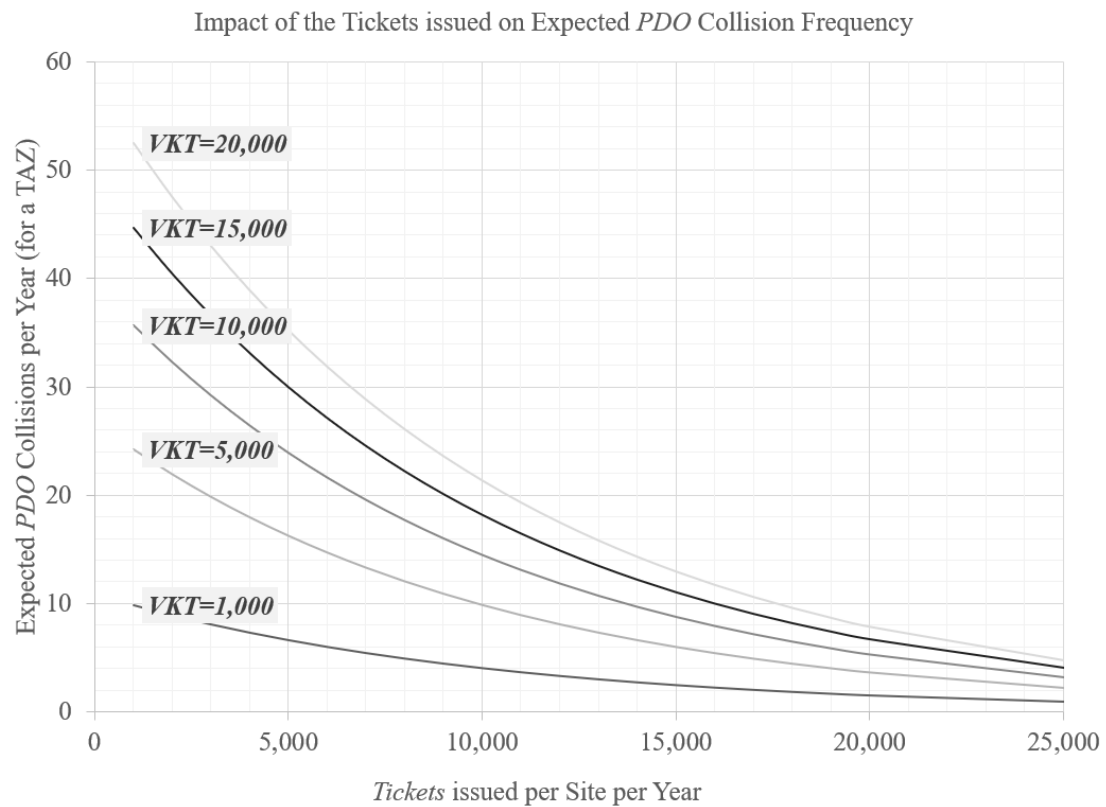


Figure 4.6 Design charts demonstrating the relationship between VKT, the ratio of hours of enforcement per site to the average number of visits per site, and the expected *I+F* collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart

Impact of the Tickets issued on Expected *PDO* Collision Frequency

(a)



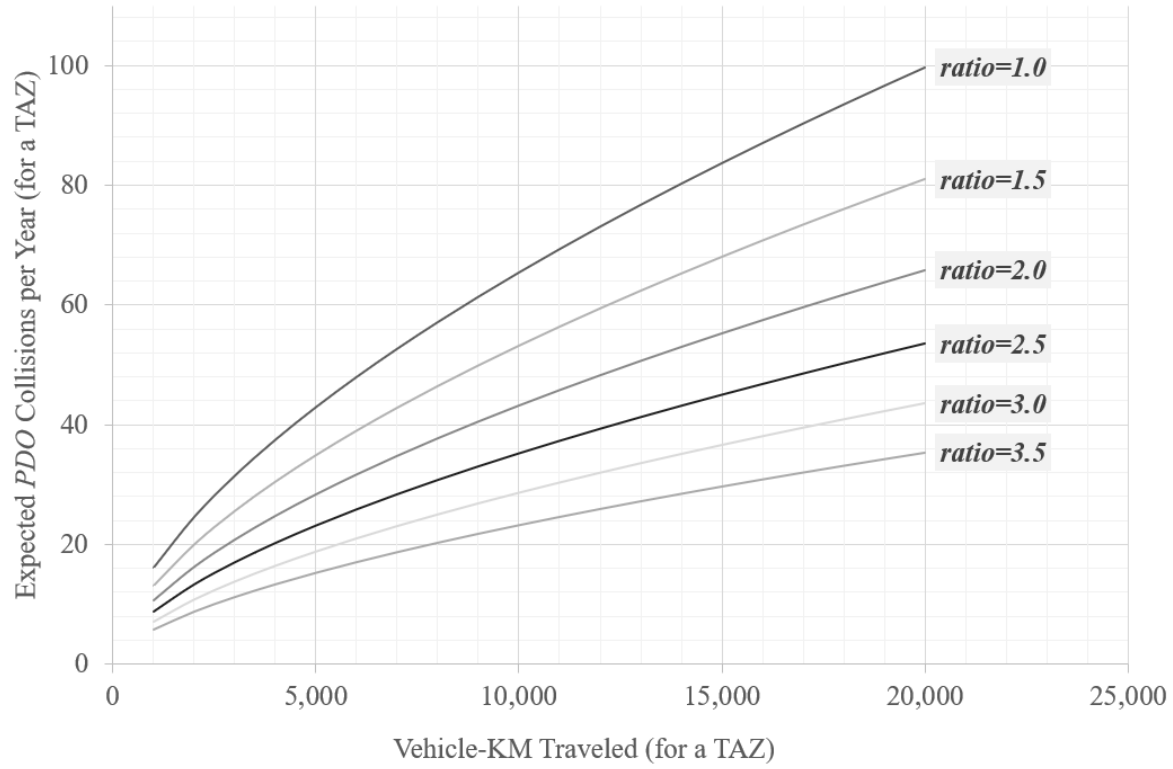


(b)

Figure 4.7 Charts demonstrating the relationship between VKT, Tickets issued per site, and the expected *PDO* collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart

Impact of the Ratio of Hours of Enforcement to Average Number of Visits per Site on *PDO* Expected Collision Frequency

(a)



Impact of the *Ratio* of Hours of Enforcement to Average Number of Visits on *PDO* Expected Collision Frequency

(b)

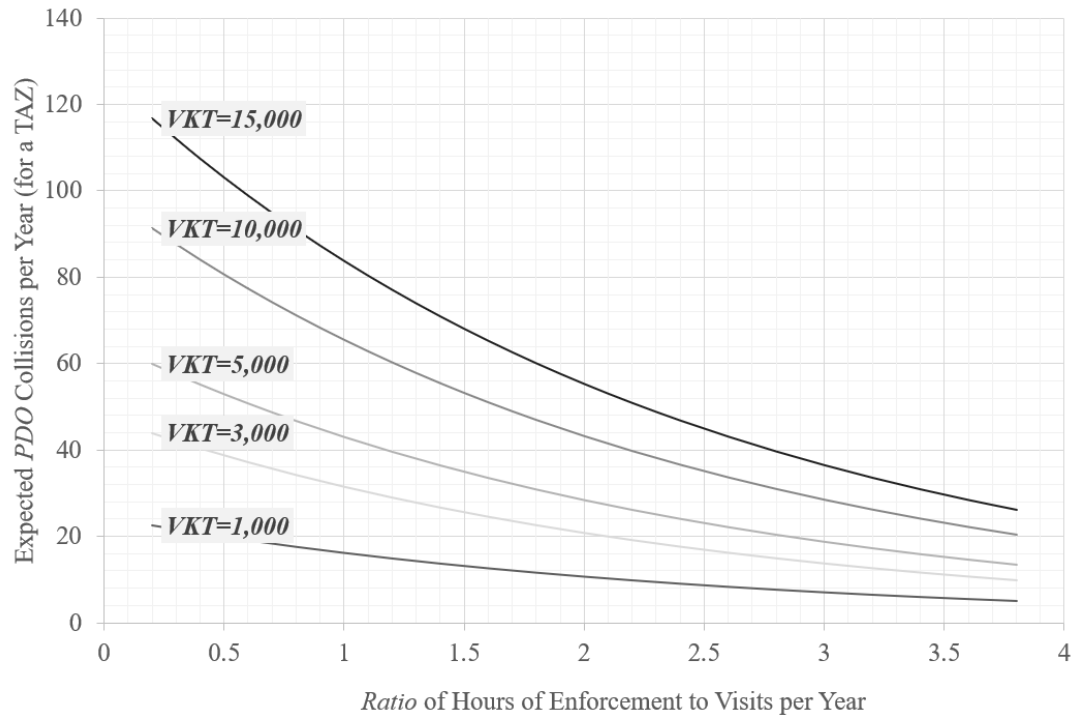


Figure 4.8 Design charts demonstrating the relationship between VKT, the ratio of hours of enforcement per site to the average number of visits per site, and the expected PDO collision frequency (a) showing the operational deployment consultation chart, and (b) showing the planning deployment consultation chart



Figures 4.3 to 4.8 provide the deployment charts for other collisions types: total midblock speed related collisions, I+F collisions, and PDO collisions.

### **4.3 Impact of Changing Deployment Intensity on Macro-Level Collisions**

Previous research has modeled collisions as counts and regards them as discrete, non-negative values. However, one of the limitations of these regression models is that they assume that any lack of observation is censored. For instance, in the collision dataset, if an observed collision frequency for a TAZ is zero, typical regression models would assume those zeros as observations. This assumption can result in a biased estimate since the lack of collision frequency should not be considered only to be dependent on the explanatory variables used. The Tobit model accounts for this by, incorporating a binary case which investigates whether a crash occurred or not, and a second case which evaluates the rate of those collisions given that the observation is not zero (Ulak et al., 2018). Since there were some TAZ with automated enforcement but no collisions, traditional statistical techniques would result in a biased estimate. Therefore, the application of the three Tobit statistical techniques is explored further in this section: traditional Tobit model, Grouped Random Parameter Tobit model, and Random Intercept Tobit model.

#### **4.3.1 Model Development**

The traditional Tobit model for fitting collision rates can be given as

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (4.8)$$

$$Y_i = \begin{cases} Y_i^*, & \text{if } Y_i^* > 0 \\ 0, & \text{if } Y_i^* \leq 0 \end{cases}, \quad i = 1, 2, \dots, N \quad (4.9)$$

where  $Y_i$  is the collision rate,  $X_{ik}$  is the  $k$ th explanatory variable (i.e., VKT and *ratio* variables) for TAZ  $i$ ,  $\beta_0$  is the model intercept,  $\beta_1, \beta_2, \dots, \beta_k$  are the model parameters,  $Y_i^*$  is a latent variable observed only when positive,  $N$  is the number of TAZs, and  $\varepsilon_i$  denotes the unstructured error which is assumed to be normally and independently distributed with zero mean and variance  $\sigma^2$  as given in

$$\varepsilon \sim N(0, \sigma^2) \quad (4.10)$$

Typically, statistical models fit one regression model to the dataset. This approach does not take into consideration the effect of an explanatory variable on collisions could vary for different observations. In the context of the deployment of mobile automated enforcement, not all enforcement sites are visited with the same frequency, nor is the duration of enforcement always consistent. For this reason, constraining the parameters can lead to an underestimation of standard errors and inconsistent inference. As such, heterogeneity may exist among different zones and may have some commonly shared features at a certain traffic analysis zones (TAZs). Therefore, instead of having a parameter for each observation, the grouping of observations which may share many of the same unobserved effects considers both possible heterogeneity and panel effects (Wu et al., 2013; Sarwar et al., 2017a; Sarwar et al., 2017b; Cai et al., 2018; Fountas et al., 2018). As such, observations in the same TAZ were grouped as panel data. If the  $i^{\text{th}}$  observation belongs to group  $g(i) \in \{1, 2, 3, 4\}$ , the GRP-Tobit model is given as

$$Y_i^* = \beta_0 + \beta_{g(i),1} X_{i1} + \beta_{g(i),2} X_{i2} + \dots + \beta_{g(i),k} X_{ik} + \varepsilon_i \quad (4.11)$$

where the intercept is set to be random parameter that follows a normal distribution as following

$$\beta_{g(i),j} \sim N(\beta_j, \sigma_j^2), j = 1, 2, \dots, k \quad (4.12)$$

The third and final model that was developed also addresses the heterogeneity across the different enforcement deployment strategies but by allowing the intercept to vary. This model is the Random Intercept Tobit model and is given by

$$Y_i^* = \beta_{m,0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (4.13)$$

where the intercept is set to be random parameter that follows a normal distribution as following

$$\beta_{m,0} \sim N(\beta_0, \sigma_0^2), m = 1, 2, 3, 4 \quad (4.14)$$

To determine the impact of varying enforcement intensity on the overall collision rates, the Full Bayes (FB) method was used for the model estimation, and the specification of the prior distribution of the model parameters is required before the FB estimates. Due to the absence of sufficient prior knowledge, non-informative priors are selected for the parameters. A diffused normal distribution with zero mean and a large variance is the most commonly used prior to estimating the regression parameters (El-Basyouny and Sayed, 2009; Guo et al., 2018a). The diffused gamma distribution  $\text{Gamma}(0.001, 0.001)$  was used as the prior of precisions for  $\sigma^{-2}$ ,  $\sigma_j^{-2}$ , and  $\sigma_0^{-2}$  (El-Basyouny and Sayed, 2009; Guo et al., 2018b).

The posterior mean and standard deviation were sampled using the Markov chain Monte Carlo (MCMC) technique using WinBUGS software. The MCMC method uses the sampling technique

to generate chains of random points, the distribution of which converges to the target posterior distributions. The convergence was monitored in several ways. First, two parallel chains with diverse starting values were tracked so that the full coverage of the sample space is ensured. Brooks-Gelman-Rubin (BGR) statistic was also used, where convergence occurred if the value of the BGR statistic was less than 1.2. Typically, convergence is achieved with the Monte Carlo errors relative to the respective standard deviations of the estimates are less than 0.05 (*El-Basyouny and Sayed, 2009*). In this study, the posterior summaries were obtained using two chains with 100,000 iterations, 10,000 of which were excluded as a burn-in sample. Examination of the BGR statistics, ratios of the Monte Carlo errors relative to the standard deviations of the estimates and trace plots for all model parameters indicated convergence.

#### **4.3.2 Results & Analysis**

Three different Tobit models were developed, and Table 5.3 shows the results of the three models; all variable estimates are statistically significant at the 95% credible interval and are bounded away from zero. The first model that was developed was the traditional Tobit model. This approach does not account for any variation in the deployment parameters (i.e., there is a single coefficient for the *ratio* variable). In this model, there were two variables, the *VKT*, which is the exposure variable, and the *ratio*, which depicts the deployment strategy. The table shows a positive coefficient for *VKT*, an intuitive result, and means that an increase in the mean of collision rates is associated with an increase in traffic volumes.

On the other hand, the coefficient of the *ratio* of hours of enforcement per visit is negative. This indicates that a higher presence of enforcement activity (i.e., longer hours of enforcement for

each visit at a specific site) is associated with a reduction in collision rates. This result is in line with previous results, which provides further evidence that mobile automated speed enforcement is successful in reducing collision rates.

The remaining two models take into consideration that not all enforcement sites are enforced for a similar number of hours or with the same frequency of visits. To account for the heterogeneity, the deployment parameter *ratio* was divided into different clusters. The clusters for the random parameters were determined using the FASTCLUS Procedure in SAS. This approach performs a disjoint cluster analysis based on distances that are computed from one or more quantitative variables. As a result, the data was separated into three different clusters, the first cluster is the *ratio* between 0.9 and 1.8 hours of enforcement per visit per site, the second cluster is the *ratio* between 1.8 and 2.8 hours of enforcement per visit per site, and lastly, the third cluster is the *ratio* between 2.8 and 3.8 hours of enforcement per visit per site.

The second model that was developed was the GRP Tobit model. This approach accounts for the within-group correlations in the data by providing different coefficients for *VKT* and the deployment parameter *ratio* for each of the three clusters. All three coefficients for *VKT* were positive and statistically significant, which is similar to the traditional Tobit model results. However, the coefficient of the *ratio* variable is different for cluster 1 in comparison to clusters 2 and 3. The results indicate that clusters 2 and 3, where there was a longer enforcement duration per visit, were associated with lower collision rates. This result indicates that for agencies to yield the highest safety benefits from a single deployment, their enforcement should be conducted for a longer duration to increase their visibility.

The third model was the RI Tobit model, which accounts for the heterogeneity in the data through changes in the intercept (as opposed to changes in the coefficients of the parameter that is being investigated). The results also suggest that there is a difference in the impact of the deployment strategy on the collision rates by showing a different coefficient for the intercept for clusters 2 and 3 compared to the first cluster.

Finally, the model performance measure is identified in the DIC results. The traditional Tobit model shows the least performance among the three models since it has the highest DIC value. The RI Tobit model provides slightly better performance than the traditional Tobit model, which suggests that accounting for the variation is essential. However, the GRP Tobit model had the best performance with the lowest DIC value, which again indicates that accommodating unobserved heterogeneity across observations is essential when capturing the variability in collision rates because of changing the deployment strategy.

Variable*		Traditional Tobit				Group Random Parameter Tobit (GRP Tobit)				Random Intercept Tobit (RI Tobit)			
		Mean	SD*	95% Confidence Interval		Mean	SD**	95% Confidence Interval		Mean	SD*	95% Confidence Interval	
Intercept	b0 [1]									3.378	0.086	3.568	3.201
	b0 [2]	5.559	0.075	5.405	5.666	-3.930	1.884	-5.909	-1.880	-2.865	0.170	-3.076	-2.554
	b0 [3]									-2.280	0.347	-2.818	-1.699
VKT	b1 [1]					1.121	0.203	0.890	1.343				
	b1 [2]	0.604	0.019	0.580	0.640	1.201	0.086	1.101	1.316	0.636	0.045	0.551	0.702
	b1 [3]					1.809	0.607	1.173	2.437				
Ratio	b2 [1]					3.038	2.055	5.184	0.870				
	b2 [2]	-4.379	0.116	-4.587	-4.240	-2.610	0.540	-3.184	-1.864	-1.032	0.229	-1.391	-0.639
	b2 [3]					-3.723	2.235	-6.089	-1.397				
						<b>Goodness of Fit</b>							
DIC		1621				1590				1605			

\* Numbers in brackets [] refer to the cluster

\*\*SD = Standard Deviation

Table 4.3 Tobit Model Results - Parameter Estimates and 95% Confidence Intervals

#### 4.4 Summary

The impact of automated enforcement on traffic safety has been investigated and validated by many studies. The results have shown that there are significant safety benefits related to the use of these programs; however, there is little knowledge of the safety implications of this program on a macro-level (or zonal) collisions. Additionally, there are no tools that can quantify the safety effects associated with changes in an agency's deployment strategy. This tool translates the proven theoretical benefits of mobile automated speed enforcement into a practical application to assist decision-makers with efficiently allocating their resources to improve safety.

To address these gaps, the analysis in this chapter included the development of two models. The first model showed a reduction in zonal collisions associated with an increase in the number of tickets issued (for vehicles exceeding the speed limit). While this relationship was expected, quantifying the impact that automated enforcement activity has on overall zonal collisions is further evidence of how successful this program is at improving safety. A chart was created to illustrate this relationship as well as to aid agencies in better understanding the impact on collisions. The results of the second model showed that collision reductions were also associated with an increase in the ratio of hours of enforcement per visit. To better understand the impact that these parameters have on the expected zonal collisions, several case studies were proposed.

While the results have shown that there are significant demonstrated safety benefits associated with increasing the number of hours of enforcement at a site per visit, there is little knowledge of how changing the deployment intensity can impact collisions. To address this gap, the analysis in



this chapter developed a new model based on the Tobit approach. Three models were developed i) traditional Tobit model to quantify the impact that the deployment parameter has on collision rates, ii) GRP Tobit model which accounts for variation in the data by categorizing the data into three distinct clusters, and iii) RI Tobit model which accounts for the variation in the intercept.

The results of all three models are in line with earlier evaluations in that increasing the number of hours of enforcement at a site per visit was associated with higher reductions in collisions, which is confirmed by the results of clusters 2 and 3. However, the additional benefit that this analysis provides is to demonstrate the impact of the deployment strategy, explicitly spending a more extended time enforcing a site, to see benefits. Both the GRP Tobit and RI Tobit model results highlight the need to use clusters to define the deployment strategy and to better understand the impact it has on collision rates.

The statistical comparison between all three models indicated that the GRP-Tobit model outperformed the other two models, which is reflected by a significant decrease in the DIC value. Additionally, this approach captures the variation in the deployment parameter *ratio*, which allows agencies to better understand how to best deploy their resources. While using the RI Tobit model is still a better approach when compared to the Traditional Tobit model, it does not clarify the nuances in the changes in the deployment parameter compared to the GRP Tobit model.

## Chapter 5: Crime & Traffic Safety

This chapter investigates the second application of macro-level models as it relates to automated enforcement and its impact on collisions *and* crime.

### 5.1 Introduction

Enforcement agencies typically operate under a strict budget and with limited resources. Therefore, if a single deployment can achieve several objectives (e.g., reducing crime and collisions), this could decrease the demand for their resources and would allow them to manage their deployments efficiently and effectively. Early research on the topic of crime and traffic enforcement found that the most effective policing style that produced the most successful results (i.e., the lowest rate of commercial robbery) was achieved in cities with the highest traffic citations (*Wilson, 1968*). The author attributed this success to the increased presence of enforcement, which meant an increased risk of being apprehended.

Based on this premise, the Data-Driven Approaches to Crime and Traffic Safety (DDACTS) was developed. There are three distinct elements of DDACTS that have been iteratively updated to maximize the program's results. The first element includes a shift in the deployment strategy away from targeting individuals and moving towards targeting locations. This change alleviates any concerns for preconceived biases and emphasizes that the approach is more evidence-based (*NHTSA, 2013*).

Secondly, the DDACTS approach recognizes that locations with a high frequency of collisions also coincide with locations with a high frequency of criminal incidents. Michalowski (*1975*)

explained this phenomenon by suggesting that individuals who exhibit aggressive tendencies and a characteristic of violence were also likely to be aggressive drivers. This relationship is intuitive and logical but was only inferred because of the overlap between locations that exhibited a high frequency of collisions and locations with a high frequency of criminal incidents. It has only been quantified recently (*Takyi et al., 2018*).

The third and final element focuses on the development of tools that assist enforcement agencies in visualizing the overlap between collisions and crime hotspots and improve their ability to optimize their deployment strategies. These tools can vary from a simplistic display of collision and crime counts to a more advanced GIS approach, which accounts for spatial autocorrelation. Based on these three elements, guidelines were developed and shared with enforcement agencies so that they could incorporate the DDACTS model in the deployment of their policing resources (*NHTSA, 2014*).

While evaluations of DDACTS programs are successful at reducing collisions and crime (*Hardy, 2010*), there were limitations associated with the methodology that was used (*Kuo et al., 2013*). The first limitation was related to the level of aggregation; most of the analysis was conducted at a city or a county level, which can negatively impact the quality of the results. At such a high aggregation level, the overall collision trends, changes in the volume of traffic, and other confounding factors are not considered. Therefore, a more robust unit of analysis is required to inform the deployment strategy better. Secondly, most approaches used a simple before and after evaluation, which does not accurately represent or capture the actual impact of the change on collision or crime and suffers from site-selection bias and regression to the mean effect (*Hauer,*

1997; Lord and Kuo, 2012). A more rigorous approach is needed to determine the effectiveness of the DDACTS programs on collisions and crime.

The first objective of this chapter is to address the gap in the literature and quantify the correlation between collisions and crime at a macro-level. The second objective is to determine the impact of one of the most common traffic enforcement programs, automated speed enforcement, on crime and collisions. While the impact of automated enforcement on collisions was previously studied, there was no research on the impact of these programs on crime incidences. The third objective is how to use this information to accurately determine which neighbourhoods are crime or collision prone. Finally, the last objective is to explore the impact of the deployment strategy by investigating the ratio of how often to visit a site per year as well as the length of time spent enforcing a site per visit, at a neighbourhood level.

## **5.2 Collisions & Crime**

As outlined previously, the relationship between collisions and crime has long been hypothesized to be highly correlated. One of the main justifications for this is due to the visualization of the data showing that locations with high incidents of crime also have a high frequency of collisions. While the rest of this chapter will look at quantifying this correlation, first, maps were created for neighbourhoods where mobile automated enforcement occurs to determine whether the visualization supports the correlation between collisions and crime.

Figure 5.1 shows a list of all the neighbourhoods in the city of Edmonton, where mobile automated enforcement was active. The map shows three categories of high collision locations

based on total numbers. It is noteworthy that some of the neighbourhoods with a higher frequency of collisions seem to be clustered around the core area in the city. Additionally, the neighbourhoods in the suburban areas in the city seemed to have fewer collisions compared to the mature areas. This is due to the traffic patterns and the higher trips that are generated towards Edmonton's core.

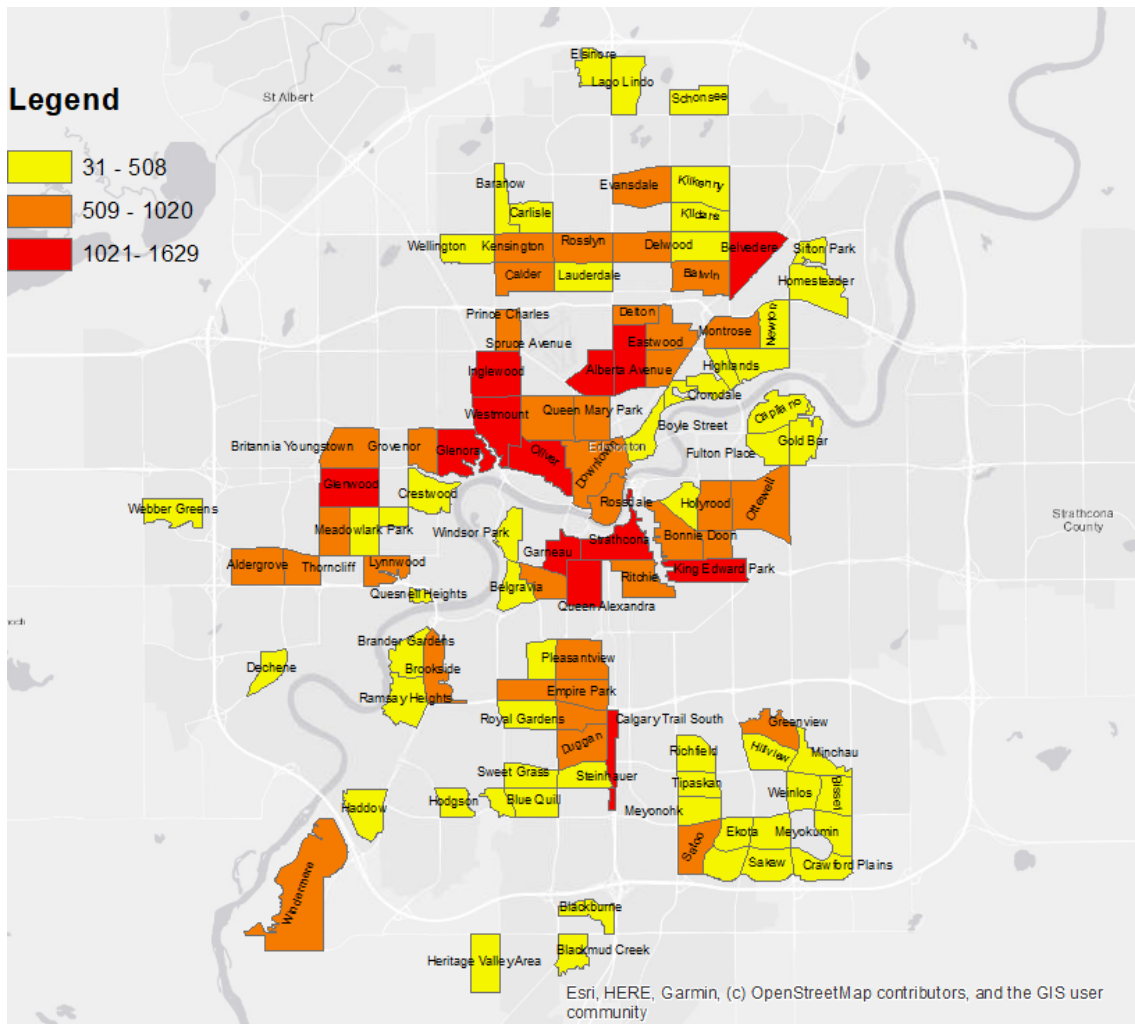


Figure 5.1 Collision Heatmap by Neighborhood in Edmonton

Similarly, Figure 5.2 shows a heat map of the incidents of property crime in neighbourhoods in Edmonton. The neighbourhoods with the highest incidents of crime again appear to be in the core of the city.

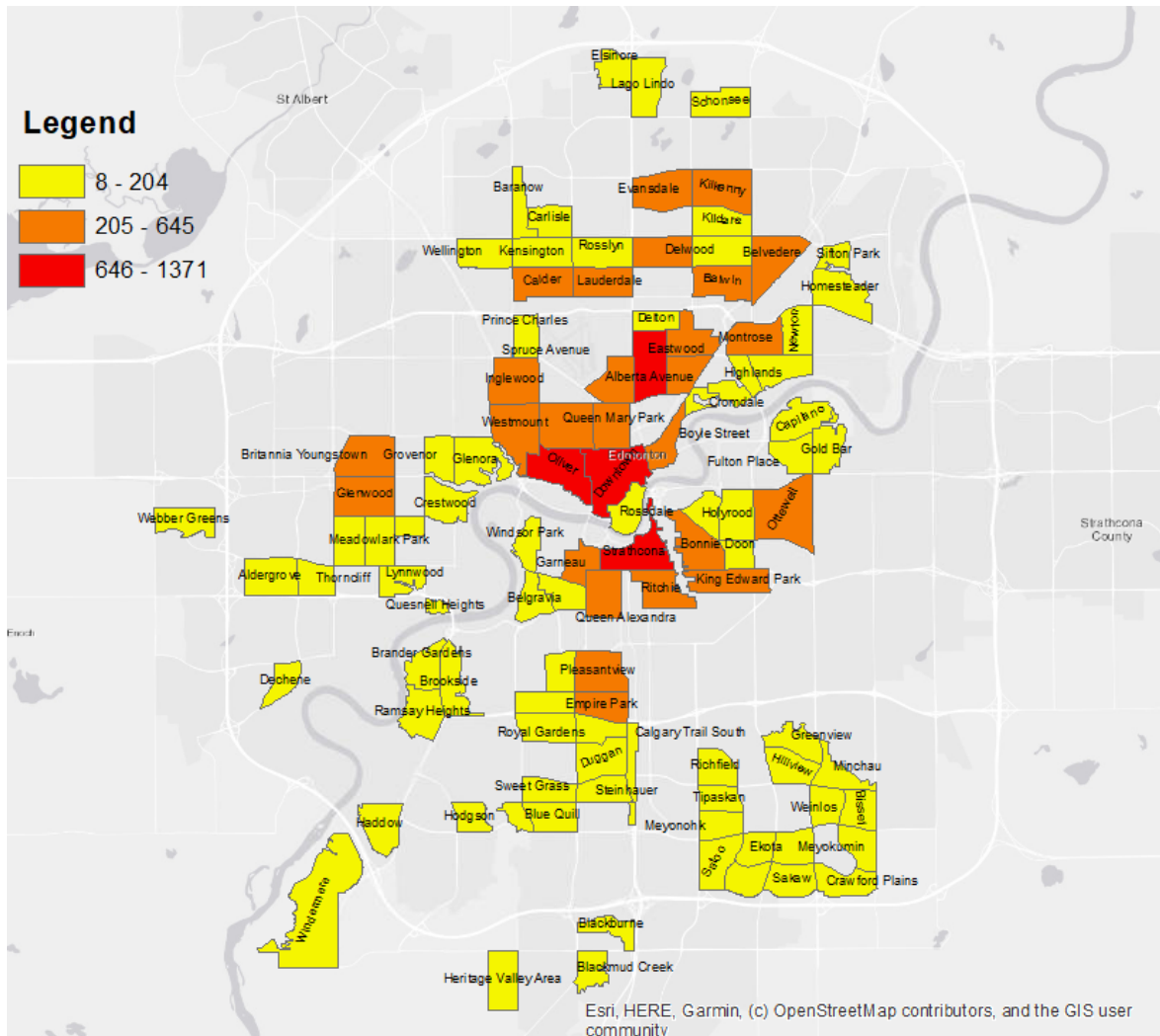


Figure 5.2 Crime Heatmap by Neighborhood in Edmonton



### **5.3 Impact of Automated Enforcement on Collisions and Crime**

Since the visualization has shown that there seems to be an overlap between collisions and crime, the next step is to quantify this relationship to provide irrefutable proof. To determine whether there is a correlation between collisions and criminal incidents in a neighbourhood and whether the presence of automated enforcement impacts both a Multivariate Poisson Lognormal (MVPLN) model was developed. Table 5.2 summarizes the parameter estimates and their associated statistics for the MVPLN model.

#### **5.3.1 Model Development**

Data on the number of collisions at a particular site are usually available where the collisions are classified by severity (e.g., fatal, minor injury, major injury or property damage only), by the number of vehicles involved (e.g., single or multiple), and/or by the type of collision (e.g., angle, head-on, rear-end, sideswipe or pedestrian-involved), etc. Despite the multivariate nature of such data sets, they have been mostly analyzed by modeling each category separately, without taking into account the correlations that probably exist among the different levels. These correlations may be caused by omitted variables, which can influence collision occurrence at all levels of classification, or from ignoring shared information in unobserved error terms. Such univariate treatment of correlated counts as independent can lead to imprecise analysis of road safety.

Several studies have applied multivariate models for estimating collision frequency under various collision severity levels and indicated their superiority to univariate models (Ma and Kockelman, 1950; Maher et al., 1990; Tunaru, 2002; Bijleveld, 2005; Ma and Kockelman, 2006; Park and Lord, 2007; Brijs et al., 2007; Ma et al., 2008; Agüero-Valverde and Jovanis, 2009; Ye



et al., 2009; Agüero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009; Wang et al., 2011; Anastasopoulos et al., 2012; El-Basyouny et al., 2014a; El-Basyouny et al., 2014b). However, in this research the two events that are modeled together are collisions and crime, rather than different severity models. While the correlation between collisions and crime is unknown, the use of the DDACTS model suggests that there is a relationship between the two events. Using this model, the correlation between collisions and crime can be quantified.

There are two approaches to developing the MVPLN models: the standard approach and the mixed approach (*Wright, 2018*). The standard approach involves using the same variables for the two modeled events (e.g., collisions and crime, or two different collision severities) or using only variables that are expected to commonly impact both events (*Osama and Sayed, 2017*). However, given that collisions and crime are impacted by different variables, the standard approach is not a suitable choice. The mixed approach accounts for this limitation by allowing the model to take on another form (e.g., if the same model form cannot be used for the two events), or by allowing the model to include a different set of variables for each event. This is the approach that is undertaken in this study.

For a set of data on collisions and crime incidents at  $n$  neighbourhoods, where the collisions and crime incidents at each neighbourhood are classified into two categories ( $K=1$  for collisions, and  $K=2$  for crime incidents), define the vector  $y_i = (y_{i1} \ y_{i2} \ \dots \ y_{iK})'$ , where  $y_{ik}$  denote the number of collisions or crime incidents at the  $i^{th}$  TAZ in category  $k$ . It is assumed that the  $y_i$  are independently distributed and that the Poisson distribution of  $y_{ik}$ , given  $\lambda_{ik}$ , is

$$f(y_{ik}|\lambda_{ik}) = \lambda_{ik}^{y_{ik}} e^{-\lambda_{ik}} / y_{ik}!, i = 1, 2, \dots, n, k = 1, 2, \dots, K \quad (5.1)$$

To model extra variation, assume further that  $\ln(\lambda_{ik}) = \ln(\mu_{ik}) + \epsilon_{ik}$ , where

$$\ln(\mu_{ik}) = \beta_{k0} + \beta_{k1} X_{i1} + \dots + \beta_{kK} X_{iK}, \quad (5.2)$$

Let  $X_{ij}$  denote relevant traffic, population and enforcement characteristics and the  $\epsilon_{ik}$  denote multivariate normal errors distributed as  $\epsilon_i \sim N_K(0, \Sigma)$ , where

$$\epsilon_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \dots \\ \epsilon_{iK} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2K} \\ \dots & \dots & \dots & \dots \\ \sigma_{K1} & \sigma_{K2} & \dots & \sigma_{KK} \end{pmatrix}$$

Let  $X_{ij}$  and  $\beta_k$  denote the matrix of covariates and the vector of regression coefficients, respectively, and let  $\beta$  denote the set  $\{\beta_1, \beta_2, \dots, \beta_K\}$ . Since the mixed approach allows for different covariates to be used in each model, the unique covariate that was used for both collisions and crime is the *ratio* of hours of enforcement per visit. This covariate represents the enforcement activity level and needs to be included in both, so that its impact can be evaluated. However, the exposure measure for collisions was taken as the vehicle-km travelled (*VKT*) which is a common covariate used in collision modeling. Using the *VKT* as a covariate when modeling crime is not an appropriate surrogate for exposure. Instead, the population (*pop*) was used as a covariate. Thus, given  $(X, \beta, \Sigma)$ , the  $\lambda_i$  are independently distributed as

$$f(\lambda_i | X, \beta, \Sigma) = \frac{\exp\{-0.5(\lambda_i^* - \mu_i^*)' \Sigma^{-1} (\lambda_i^* - \mu_i^*)\}}{(2\pi)^{K/2} (\prod_{k=1}^K \lambda_{ik}) |\Sigma|^{1/2}}, \quad (5.3)$$

which is a  $K$ -dimensional log-normal distribution, where

$$\lambda_i = (\lambda_{i1} \quad \lambda_{i2} \quad \dots \quad \lambda_{iK})', \quad \lambda_i^* = (\ln(\lambda_{i1}) \quad \ln(\lambda_{i2}) \quad \dots \quad \ln(\lambda_{iK}))'$$

$$\mu_i = (\mu_{i1} \quad \mu_{i2} \quad \dots \quad \mu_{iK})', \quad \mu_i^* = (\ln(\mu_{i1}) \quad \ln(\mu_{i2}) \quad \dots \quad \ln(\mu_{iK}))'$$

Let  $\lambda$  denote the set  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ . The prior distributions for the hyperparameters  $(\beta, \Sigma)$  need to be identified first to obtain the full Bayes estimates of  $(\lambda, \beta, \Sigma)$ . Prior distributions are meant to reflect prior knowledge about the parameters of interest. If such prior information is available, it should be used to formulate the so-called informative priors. The specification of informative priors for generalized linear models was dealt with by Bedrick et al. (1996), who considered conditional means priors as well as data augmentation priors of the same form as the likelihood and showed that such priors result in intractable posteriors.

In the absence of sufficient prior knowledge of the distributions for individual parameters, uninformative proper prior distributions are usually specified. The most commonly used priors are diffused normal distributions (with zero mean and large variance) for the regression parameters and a *Wishart*( $P, r$ ) prior for  $\Sigma^{-1}$ , where  $P$  and  $r \geq K$  represent the prior guess at the order of magnitude of the precision matrix  $\Sigma^{-1}$  and the degrees of freedom, respectively. The parameterization of the Wishart probability density function is

$$f(\Sigma^{-1} | P, r) = |P|^{K/2} |\Sigma^{-1}|^{(r-K-1)/2} \exp\{-0.5 \text{Tr}(P \Sigma^{-1})\}. \quad (5.4)$$

Choosing  $r = K$  as the degrees of freedom corresponds to vague prior knowledge (*Spiegelhalter et al., 1996; Tunaru, 2002*).

Let  $f(\beta) \equiv \prod_{k,j} N(0,10000)$  and  $f(\Sigma^{-1}) \equiv \text{Wishart}(P, K)$  denote the hyper-prior distributions for the regression parameters and the precision matrix, respectively, Further, let  $y$  denote the set  $\{y_1, y_2, \dots, y_n\}$ . Given  $X$ , the joint distribution of  $(y, \lambda, \beta, \Sigma)$  is

$$f(y, \lambda, \beta, \Sigma | X) = \frac{f(\beta)f(\Sigma^{-1})}{(2\pi)^{nK/2} |\Sigma|^{n/2} (\prod_{i=1}^n \prod_{k=1}^K y_{ik}^{-1})} g(y, \lambda, X, \beta, \Sigma),$$

where

$$g(y, \lambda, X, \beta, \Sigma) = \exp\left\{-0.5 \sum_{i=1}^n (\lambda_i^* - \mu_i^*)' \Sigma^{-1} (\lambda_i^* - \mu_i^*) + \sum_{k=1}^K [(y_{ik} - 1) \ln(\lambda_{ik}) - \lambda_{ik}]\right\}.$$

The posterior distributions are given by

$$f(\lambda_i | y, X, \beta, \Sigma) = \frac{g_i(y_i, \lambda_i, X, \beta, \Sigma)}{\int_0^\infty \int_0^\infty \dots \int_0^\infty g_i(y_i, \lambda_i, X, \beta, \Sigma) d\lambda_{i1} d\lambda_{i2} \dots d\lambda_{iK}}, \quad (5.5)$$

where

$$g_i(y_i, \lambda_i, X, \beta, \Sigma) = \exp\left\{-0.5 (\lambda_i^* - \mu_i^*)' \Sigma^{-1} (\lambda_i^* - \mu_i^*) + \sum_{k=1}^K [(y_{ik} - 1) \ln(\lambda_{ik}) - \lambda_{ik}]\right\},$$

$$f(\beta | y, \lambda, X, \Sigma) = \frac{g(y, \lambda, X, \beta, \Sigma) f(\beta)}{\int_{-\infty}^\infty \int_{-\infty}^\infty \dots \int_{-\infty}^\infty g(y, \lambda, X, \beta, \Sigma) f(\beta) d\beta_{10} d\beta_{11} \dots d\beta_{KJ}}, \quad (5.6)$$

and

$$f(\Sigma^{-1} | y, \lambda, X, \beta) \equiv \text{Wishart}(P + \sum_{i=1}^n (\lambda_i^* - \mu_i^*) (\lambda_i^* - \mu_i^*)', K + n). \quad (5.7)$$

The posterior distributions needed in the full Bayes approach can be obtained using MCMC sampling. In this thesis, the posterior distributions are sampled using the MCMC techniques available in WinBUGS 2.2.0; the windows interface of OpenBUGS. The techniques generate sequences (chains) of random points, whose distributions converge to the target posterior distributions. A sub-sample is used to monitor convergence and then excluded as a burn-in sample. The remaining iterations are used for parameter estimation, performance evaluation, and inference.

Monitoring convergence is important because it ensures that the posterior distribution has been “found”. Thereby indicating when parameters sampling should begin. To check convergence, two or more parallel chains with diverse starting values are tracked to ensure full coverage of the sample space. Convergence of multiple chains is assessed using the Brooks-Gelman-Rubin (BGR) statistic (*Brooks and Gelman, 1998*). A value under 1.2 of the BGR statistic indicates convergence. Convergence is also assessed by visual inspection of the MCMC trace plots for the model parameters as well as by monitoring the ratios of the Monte Carlo errors relative to the respective standard deviations of the estimates; as a rule of thumb, these ratios should be less than 0.05.

A posterior predictive approach (*Gelman et al., 1996; Stern and Cressie, 2000; Li et al., 2008*) can be used to assess the goodness-of-fit (adequacy) of the model. Such procedures involve generating replicates under the postulated model and comparing the distribution of a certain discrepancy measure such as the chi-square statistic to the value of chi-square obtained using observed data. A model does not fit the data if the observed value of chi-square is far from the

predictive distribution; the discrepancy cannot reasonably be explained by chance if the  $p$ -values are close to 0 or 1 (*Gelman et al., 1996*).

The replicates are best obtained simultaneously with model estimation in WinBUGS to account for all uncertainties in model parameters as reflected by the estimated distributions. The chi-square statistic is computed from

$$\chi^2 = \sum_{i=1}^n [y_i - E(Y_i)]^2 / Var(Y_i), \quad (5.8)$$

where the  $y_i$  denotes either the observed or the replicated collision or crime frequencies.

### **5.3.2 Results & Discussion**

The posterior summaries in Table 5.1 were obtained via two chains with 100,000 iterations, 10,000 of which were excluded as a burn-in sample using WinBUGS. A Wishart prior with an identity scale matrix and two degrees of freedom was adopted (*Chib and Winkelmann, 2001; Congdon, 2006*). Examination of the BGR statistics, ratios of the Monte Carlo errors relative to the standard deviations of the estimates and trace plots for all model parameters indicated convergence.

The results of Table 5.1 show that the parameter estimates are all statistically significant at the 95% credible interval and are bounded away from zero. For collisions, the coefficient for the exposure measure (VKT) is positive, which indicates that there is an increase in the mean collision frequency with an increase in traffic volumes.

Additionally, the coefficient of the *ratio* of hours of enforcement per visit is negative. This indicates that a higher presence of enforcement activity (more extended hours of enforcement for each visit at a specific site) is associated with a reduction in collisions. This result is in line with previous research on this topic, which suggests that automated speed enforcement is successful in reducing collisions (*Li et al., 2015*).

For crime, the coefficient for the exposure measure (*pop*) is positive, which indicates that there is an increase in the mean crime frequency with an increase in population. The coefficient of the *ratio* of hours of enforcement per visit is also negative for crime. This indicates that increasing the presence of enforcement activity (i.e., spending more time at a site for each unique visit per site) is associated with a reduction in crime. This demonstrates that mobile automated enforcement is also successful in reducing crime. The reason this relationship is pertinent is that most of the focus of DDACT programs has been on manned enforcement, which involves the use of police officers for traffic and crime enforcement. The results of this model show that there are other tools available that road agencies can use to reduce collisions and crime.

In addition to the results mentioned above, the MVPLN model also confirms and quantifies the correlation between collisions and crime; this correlation ( $\rho$ ) is estimated at 0.720, which is highly significant. Therefore, locations that exhibit a high frequency of collisions also have a high prevalence of criminal incidences. This confirms the basis of DDACT strategies; however, this correlation suggests that modeling these two events individually (as they have been in previous studies) can result in imprecise evaluations and analysis.

Multivariate Poisson-Lognormal	Estimate	Standard Deviation	95% Credible Intervals	
			Lower Limit	Upper Limit
<b><i>Collisions</i></b>				
Intercept	3.119	0.766	1.902	4.854
<i>VKT</i>	0.431	0.122	0.135	0.595
<i>ratio</i>	-0.480	0.190	-0.775	-0.085
$\sigma_{11}$	0.324	0.063	0.229	0.478
<b><i>Crime</i></b>				
Intercept	4.203	0.785	2.971	5.430
<i>pop</i>	0.280	0.089	0.143	0.438
<i>ratio</i>	-0.569	0.238	-1.033	-0.067
$\sigma_{22}$	0.614	0.099	0.447	0.836
<b><i>Covariance</i></b>				
$\sigma_{12}$	0.321	0.064	0.216	0.466
<b><i>Correlation</i></b>				
$\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$	0.720	0.051	0.610	0.808
DIC	1932			

Table 5.1 MVPLN Model's Statistics



## 5.4 Identification of Hazardous Locations

Now that the correlations between collisions and crime has been statistically proven and quantified, the next step would be to identify which neighbourhoods are hazardous (i.e., prone to a high frequency of collisions or criminal incidents). This also presents an opportunity to demonstrate the difference between identifying these hazardous neighbourhoods individually (i.e., by collisions or by criminal incidents) and between using the multivariate approach.

### 5.4.1 Methodology

The conventional approach to select the hazardous locations is based on the posterior probability of excess (Higle and Witkowski, 1988; Sayed and Abdelwahab, 1997; Heydecker and Wu, 2001; El-Basyouny and Sayed, 2009). In this approach, the Bayesian posterior probability that a site has an excessive, mean collision frequency provides an indication of sites at which the predicted collision record is expected to be higher than what is typical. In a univariate analysis, this probability is shown by:

$$\int_{\mu_0}^{\infty} f(i|y, X, \beta, \sigma^2) d\lambda_i > 1 - \delta, i = 1, 2, \dots, n, \quad (5.9)$$

where  $\sigma^2$ ,  $\mu_0$  and  $\delta$  denote the extra-Poisson variation, an upper limit of the typical mean number of collisions or criminal incidents, and a threshold value between 0 and 1, respectively. In practice,  $\mu_0$  is typically specified as either the median or mean of the prior distribution, whereas  $\delta$  has been arbitrarily assumed to equal 0.05 (Higle and Witkowski, 1988; Sayed and Abdelwahab, 1997).

For the multivariate Poisson-lognormal (MVPLN) model, equation (21) generalizes to:

$$\int_{\mu_{10}}^{\infty} \int_{\mu_{20}}^{\infty} \dots \int_{\mu_{K0}}^{\infty} f(\lambda_i | y, X, \beta, \Sigma) d\lambda_{i1} d\lambda_{i2} \dots d\lambda_{iK} > 1 - \delta, i = 1, 2, \dots, n, \quad (5.10)$$

where  $f(\lambda_i | y, X, \beta, \Sigma)$  is given by:

$$f(\lambda_i | y, X, \beta, \Sigma) = \frac{g_i(y_i, \lambda_i, X, \beta, \Sigma)}{\int_0^{\infty} \int_0^{\infty} \dots \int_0^{\infty} g_i(y_i, \lambda_i, X, \beta, \Sigma) d\lambda_{i1} d\lambda_{i2} \dots d\lambda_{iK}} \quad (5.11)$$

where

$$g_i(y_i, \lambda_i, X, \beta, \Sigma) = \exp\{-0.5\lambda_i^* - \mu_i^* \Sigma^{-1} (\lambda_i^* - \mu_i^*) + \sum_{k=1}^K [(y_{ik} - 1) \ln(\lambda_{ik}) - \lambda_{ik}]\} \quad (5.12)$$

$\mu_{K0}$  denotes the upper limit of the typical mean number of collisions or criminal incidents in category  $k$ . It is noteworthy that the univariate equation (5.10) is the special case  $K=1$ . The evaluation of the multiple integrals in equation (5.11) is somewhat complicated, and previous researchers (Clayton and Kaldor, 1987) have proposed the following equations to simplify the calculation.

$$f(\lambda_i | y, X, \beta, \Sigma) \approx N_K(m_i, S_i) \quad (5.13)$$

where

$$m_i = S_i \left( \Sigma^{-1} \mu_i^* + \begin{pmatrix} (y_{i1} + 0.5) \ln(y_{i1} + 0.5) - 0.5 \\ \dots \\ (y_{iK} + 0.5) \ln(y_{iK} + 0.5) - 0.5 \end{pmatrix} \right), \quad (5.14)$$

and

$$S_i \left( \Sigma^{-1} + \begin{pmatrix} y_{i1} + 0.5 & 0 & \dots & 0 \\ 0 & y_{i2} + 0.5 & \dots & 0 \\ 0 & 0 & \dots & y_{iK} + 0.5 \end{pmatrix} \right)^{-1} \quad (5.15)$$

When modeling using the univariate approach, (i.e.,  $K=1$ ),  $\mu_0^*$  is the natural logarithm of the average of the prior means. Then the  $i$ th intersection would be selected as a hazardous location whenever

$$\phi(\mu_{i0}^*) < \delta, \quad (5.16)$$

where  $\phi$  denotes the univariate standard normal distribution function and

$$\mu_{i0}^* = \frac{\mu_{0-m_i}^*}{\sqrt{S_i}} \quad (5.17)$$

The standardized multivariate normal distribution can be used in a similar procedure to select hotspots under MVPLN. Therefore, using a similar notation, the  $i$ th intersection would be selected as a hazardous location whenever

$$\phi_K(\mu_{i10}^*, \mu_{i20}^*, \dots, \mu_{iK0}^*, \hat{R}) < \delta, \quad (5.18)$$

where  $\phi_K$  denotes the multivariate standard normal distribution function,

$$\mu_{ik0}^* = \frac{\mu_{k0-m_{ik}}^*}{\sqrt{S_{ikk}}}, k = 1, 2, \dots, K \quad (5.19)$$

and  $\hat{R}$  denotes the correlation matrix corresponding to  $\hat{\Sigma}$ , the Bayesian estimate of the covariance matrix  $\Sigma$ .

### 5.4.2 Univariate Results

The MVPLN results are shown in Table 5.1; however, univariate PLN models are needed for collisions and crime independently. The results of both models are shown below.

Tables 5.2 and 5.3 show the univariate PLN model results for the neighbourhoods in Edmonton, where mobile automated enforcement was active. The posterior summaries in both tables were obtained via two chains with 90,000 iterations, 10,000 of which were excluded as a burn-in sample using WinBUGS. Examination of the BGR statistics, ratios of the Monte Carlo errors relative to the standard deviations of the estimates and trace plots for all model parameters indicated convergence.

The results of both tables show that the parameter estimates are significant as the 95% credible intervals are bounded away from zero. The results of both models also show that the coefficients of both exposure parameters (i.e., VKT for collisions and Population for crime) are positive, which is intuitive and in line with previous results. Additionally, both results show that the coefficient of the *ratio* variable (i.e., the ratio of the hours of enforcement to the number of visits) is negative, which is in line with the MVPLN results that were shown previously. This further proves that mobile automated enforcement is effective in reducing both collisions and crime.

Variable	<i>Total</i>		
	Coefficient	Confidence Interval	
<b>Intercept</b>	0.6365	0.5634	0.7757
<b>VKT</b>	0.8445	0.6448	1.0640
<b>Ratio</b>	-0.4424	-0.7867	-0.1936
<b>Goodness-of-fit Criteria</b>			
<b>DIC</b>	1001		

Table 5.2 Univariate PLN results for Collisions

Variable	<i>Total</i>		
	Coefficient	Confidence Interval	
<b>Intercept</b>	-0.1909	-0.9823	-0.0092
<b>Population</b>	0.7717	0.6797	0.8672
<b>Ratio</b>	-0.3523	-0.6593	-0.0294
<b>Goodness-of-fit Criteria</b>			
<b>DIC</b>	930.8		

Table 5.3 Univariate PLN results for Crime

### 5.4.3 MVPLN vs. PLN Identification of Hazardous Locations

Hazardous locations were identified at different threshold values:  $\delta = 0.1, 0.05,$  and  $0.01.$   $K = 2,$  the standardized bivariate normal distribution function PROBBNRM (SAS, 2020), was used in the implementation of equation 20. The results are shown in Table 5.4 below.

		Multivariate PLN		Multivariate PLN		Multivariate PLN	
		$\delta = 0.1$		$\delta = 0.05$		$\delta = 0.01$	
		Hazardous?	No	Yes	No	Yes	No
Univariate PLN							
Collisions	No	46	10	47	9	49	9
	Yes	0	49	0	49	0	47
Crime	No	46	21	48	20	48	20
	Yes	0	38	0	37	0	37

Table 5.4 Number of hotspots selected under the different PLN models

The table shows the number of hotspots neighbourhoods that were identified by all models, the two univariate PLN for collision and crime independently, and the MVPLN, which modeled both events together. It is clear from the table that none of the neighbourhoods that were identified as hazardous by either the collision or crime univariate PLN was missed by the MVPLN. This proves that the multivariate model does not miss any of the hotspot neighbourhoods. As expected, the number of hotspots neighbourhoods decreased with an increase in the threshold value  $\delta$  but the change was not significant.

To better illustrate the difference in the identified locations, GIS maps were created for  $\delta = 0.05$ . Figure 5.4 shows the difference between neighbourhoods that were identified as collision or crime-prone using the univariate models. The neighbourhoods in the teal color are the neighbourhoods that were identified as collision prone using the collision PLN model, in grey were the neighbourhoods that were not identified as hotspots for either crime or collisions. The

neighbourhoods with the red boundaries are locations that have been identified as hotspots for crime incidences. This map shows a close overlap between neighbourhoods that were identified as collisions and crime-prone (i.e., neighbourhoods that are teal in color and have a red boundary).

Figure 5.5 shows the overlap between neighbourhoods that were identified as hotspot locations using the MVPLN model (teal) and neighbourhoods that were identified as hotspots using the univariate crime PLN model (those with red boundaries). Again, it is clear that all of the neighbourhoods that were identified as hazardous using the crime PLN have all been identified by the MVPLN model. However, it is quite evident that there are a few additional neighbourhoods that were identified as hotspots that were missed by the crime PLN model.

Figure 5.6 shows the overlap between neighbourhoods that were identified as hazardous using the MVPLN model (shown in teal) and neighbourhoods that were identified as collision prone (those with the red boundaries). Likewise, all of the neighbourhoods that were identified as collision prone have also been identified by the MVPLN model. On the other hand, there are a few additional neighbourhoods that were missed by the collision PLN model.

These maps and the results shown in Table 5.4 provide further evidence that collisions and crime should not be modeled independently and that the MVPLN approach provides decision-makers with a higher level of precision and accuracy when identifying where resources need to be focused for the most prominent safety return on the investment of their resources.

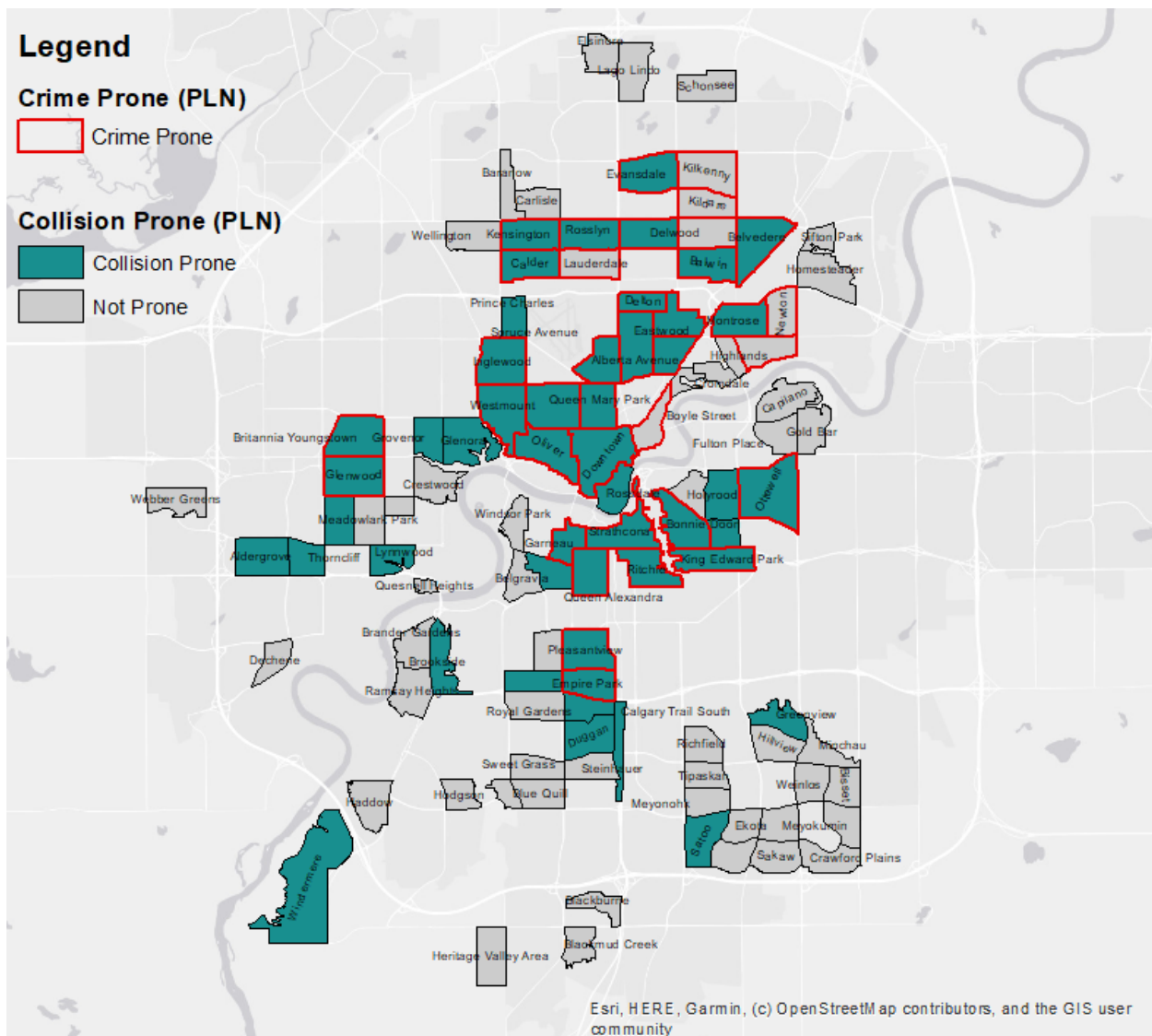


Figure 5.4 Neighbourhoods that are crime & collision prone as identified by modeling both incidences independently



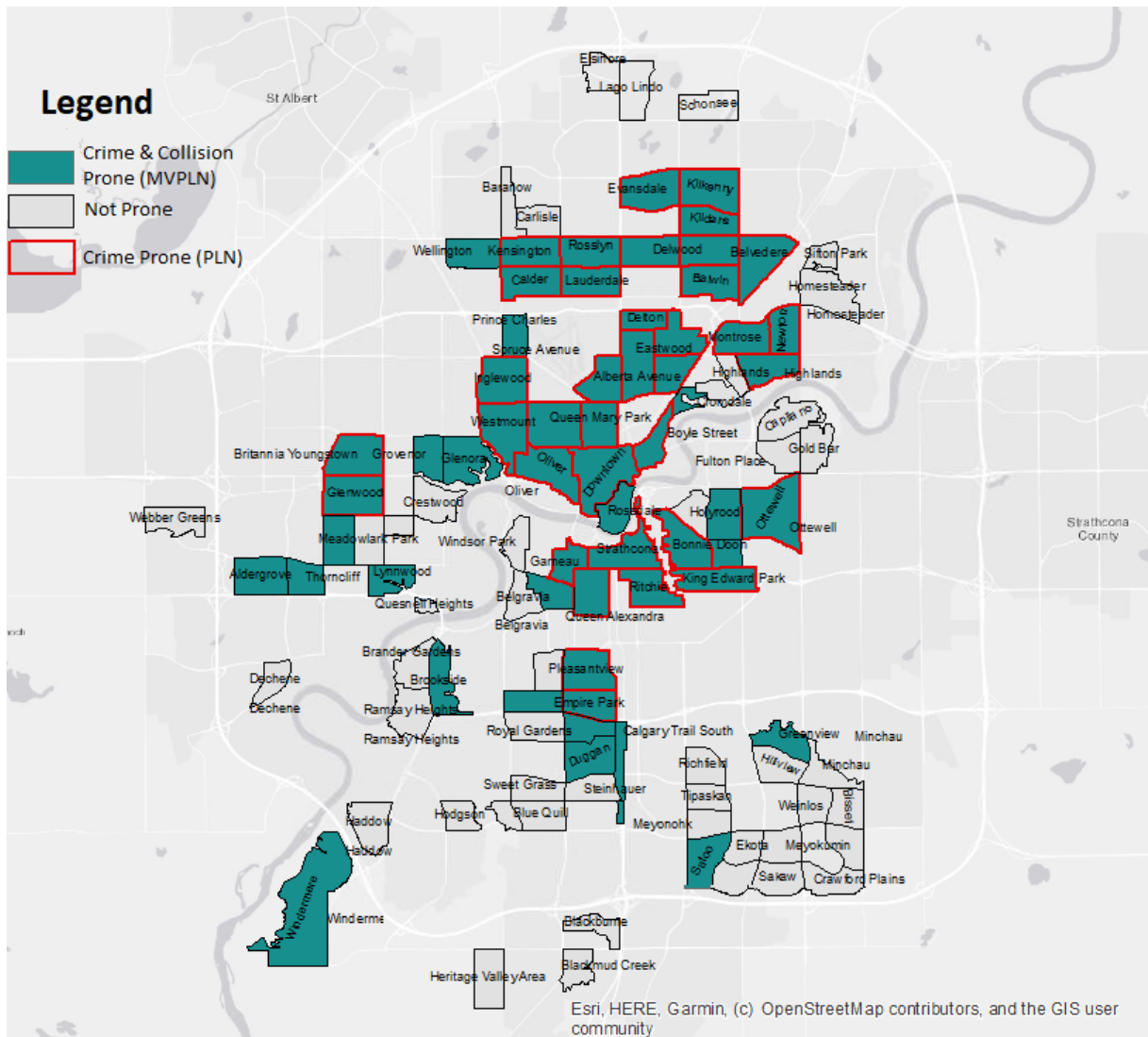


Figure 5.5 Neighbourhoods that are problem prone as identified by MVPLN (in teal) and neighbourhoods that are crime-prone as identified by the individual PLN model for incidents of crime

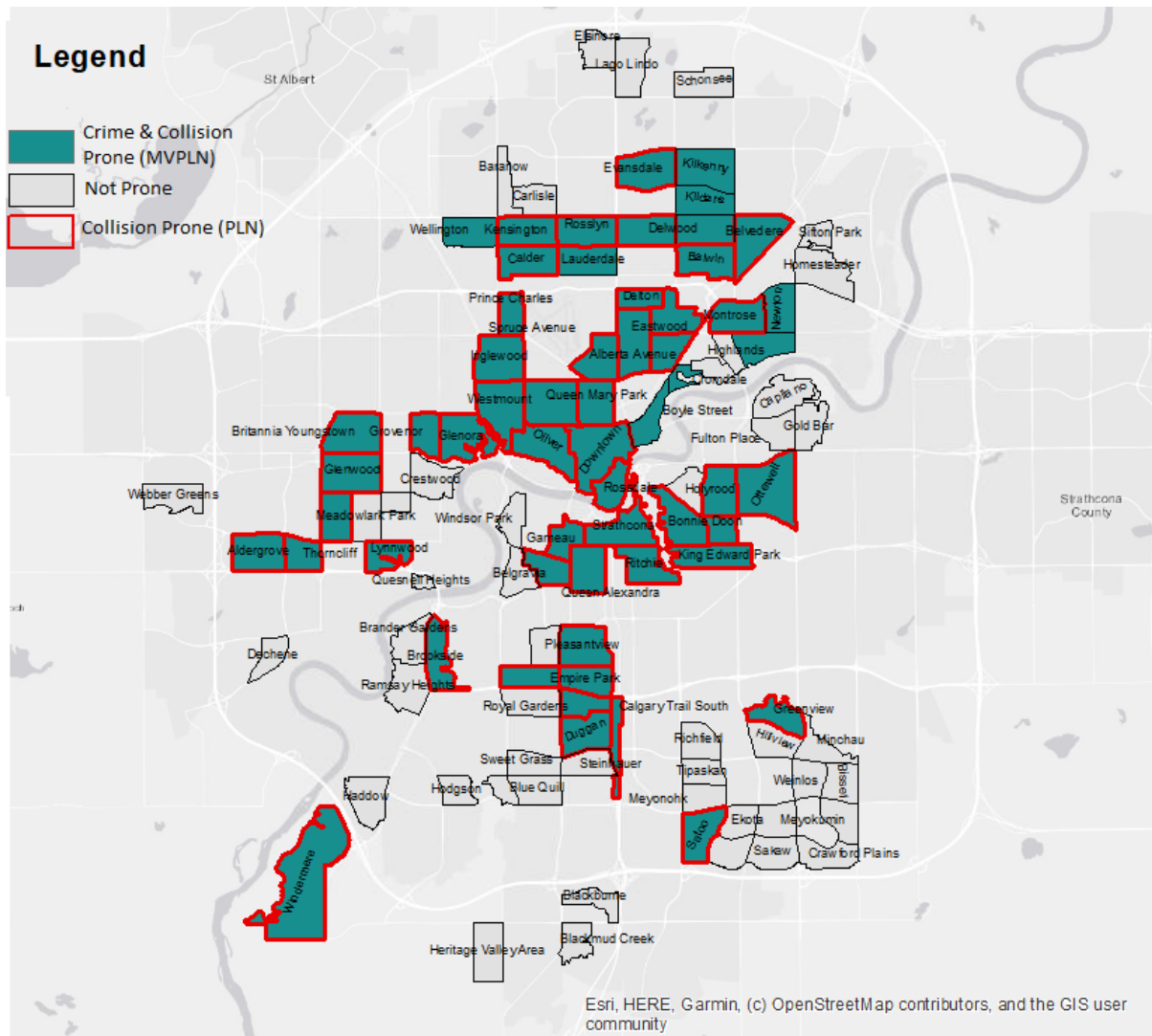


Figure 5.6 Neighbourhoods that are problem prone as identified by MVPLN (in teal) and neighbourhoods that are collision prone as identified by the individual PLN model for collisions

### 5.5 Impact of Changing Deployment Intensity on Collisions & Crime

All the previous research has shown that mobile automated enforcement is a useful tool to increase safety by reducing collisions and also incidents of crime. However, given that enforcement agencies are operating under limited resources and budget, further analysis is required to better understand how varying the MAE deployment strategy can impact collisions

and crime. For this reason, a multivariate Tobit model was developed. The Tobit model was selected to overcome the limitation of the regression models since they assume that any lack of observation is censored.

### 5.5.1 Model Development

The random parameter multivariate Tobit model for fitting collision and crime rates is given as:

$$Y_i^{k*} = \beta_{i0}^k + \beta_{i1}^k X_{i1}^k + \beta_{i2}^k X_{i2}^k + \dots + \beta_{ij}^k X_{ij}^k + \epsilon_i^k \quad (5.20)$$

where  $Y_i^k$  is the dependent variable for the  $k$ th event ( $k=2, 1$  for collision rate and 2 for crime rate),  $X_i^k$  is the explanatory variable for observation  $i$ , and  $\beta_i^k$  is the coefficient corresponding to the  $k$ th event. The random parameters  $(\beta_{i1}^k, \beta_{i2}^k, \dots, \beta_{ij}^k)$  are assumed to be multnormally distributed as  $\beta^{ij} \sim N_k(B_j, \phi_j)$ , where,

$$\beta^{ij} = \begin{pmatrix} \beta_{ij}^1 \\ \beta_{ij}^2 \\ \dots \\ \beta_{ij}^k \end{pmatrix}, \beta^j = \begin{pmatrix} \beta_j^1 \\ \beta_j^2 \\ \dots \\ \beta_j^k \end{pmatrix}, \Phi_j = \begin{pmatrix} \varphi_{11}^{j2} & \varphi_{12}^{j2} & \dots & \varphi_{1k}^{j2} \\ \varphi_{21}^{j2} & \varphi_{22}^{j2} & \dots & \varphi_{2k}^{j2} \\ \dots & \dots & \dots & \dots \\ \varphi_{k1}^{j2} & \varphi_{k2}^{j2} & \dots & \varphi_{kk}^{j2} \end{pmatrix} \quad (5.21)$$

### 5.5.2 Results & Discussion

Table 5.5 summarizes the parameter estimates and their associated statistics for the random parameter multivariate Tobit model. The posterior summaries were obtained via two chains with 100,000 iterations, 10,000 of which were excluded as a burn-in sample using WinBUGS. A Wishart prior with an identity scale matrix and two degrees of freedom was adopted (*Chib et al., 2001; Congdon, 2006*). Examination of the BGR statistics, ratios of the Monte Carlo errors

relative to the standard deviations of the estimates and trace plots for all model parameters indicated convergence.

The results show that the parameter estimates are all statistically significant at the 95% credible interval and are bounded away from zero. For collision rates, the coefficient of the *ratio* of the hours of enforcement per site is positive for cluster 1 and negative for clusters 2 and three, which suggests that a longer enforcement duration per visit was associated with lower collision rates. This indicates to agencies that the highest safety benefits can be yielded from a single deployment when enforcement is conducted for a more extended time period per shift per visit.

For crime rates, the results are similar; the coefficient of the *ratio* of hours of enforcement per visit is also negative for clusters 2 and 3. This indicates that the increasing presence of enforcement activity (i.e., spending more time at a site for each unique visit per site) is associated with a reduction in crime rates. This confirms the results of previous research in that mobile automated enforcement is also successful in reducing crime, specifically when enforcement is conducted for longer durations. This reinforces the need to include MAE as part of the DDACT approach and to expand the scope of the program, not only to include manned enforcement.

In addition to the results mentioned above, the random parameter multivariate Tobit model also quantified the correlation between collisions and crime; this correlation is estimated at 0.86, which is highly significant. Therefore, locations that exhibit a high frequency of collisions also have a high prevalence of criminal incidences. This confirms the basis of DDACT strategies and

shows that that modeling these two events individually (as they have been in previous studies) can result in imprecise road safety evaluations and analysis.

Variable		Estimate	SD	95% Confidence Interval	
<b>Collisions</b>					
Intercept	$b_0$ [Cluster 1]	29.3	6.5	23.0	45.3
	$b_0$ [Cluster 2]	135.4	15.9	117.0	170.1
	$b_0$ [Cluster 3]	36.3	4.5	32.1	50.1
Ratio	$b_1$ [Cluster 1]	18.4	3.2	14.2	26.2
	$b_1$ [Cluster 2]	-43.5	4.8	-38.2	-54.1
	$b_1$ [Cluster 3]	-26.2	2.6	-33.8	-23.5
<b>Crime</b>					
Intercept	$b_0$ [Cluster 1]	15.8	4.0	6.6	20.5
	$b_0$ [Cluster 2]	85.0	18.7	40.9	102.7
	$b_0$ [Cluster 3]	20.5	6.6	7.1	28.8
Ratio	$b_1$ [Cluster 1]	11.8	2.0	7.2	14.1
	$b_1$ [Cluster 2]	-28.1	5.7	-14.8	-33.6
	$b_1$ [Cluster 3]	-17.0	4.0	-22.2	-8.9
<b>Correlation</b>					
0.86					
<b>DIC</b>					
387					
*SD = Standard Deviation					

Table 5.5 Parameter Estimates and 95% Confidence Intervals

## **Chapter 6: Conclusion**

This chapter is divided into three sections. The first section provides an overview of the goal of this research, the methodology, and results and presents the conclusions drawn from the work. The second discusses the research contributions resulting from this thesis. Finally, the third part identifies the limitations and opportunities for future research.

### **6.1 Summary of Research Findings**

The main goal of this research is to provide transportation and enforcement agencies with tools that can help them make informed decisions based on evidence. The study used extensive data from the city of Edmonton's traffic analysis zones and neighbourhoods to develop empirical macro-level models that incorporated variables related to automated enforcement deployment, collisions, and criminal incidents by TAZ and neighbourhood. This is the first time that these variables were used in macro-level modeling and demonstrated the value they can offer.

The first part of this dissertation built upon previous work related to macro-level models but took this further by considering a completely new application; the use of mobile automated enforcement. While previous evaluations of MAE have shown that they are an effective tool to deter speeding and reduce collisions, the unit of analysis that was used could not quantify the safety effects associated with changes to an agency's deployment strategy. This research proposed the use of TAZ to allow for a more in-depth investigation of how parameters such as the duration of enforcement at a site and the frequency of visits in a year impact resourcing.

Two regression models were developed to determine the effectiveness of MAE on collisions at a macro-level (i) Model 1 which determined the impact of mobile automated enforcement outcome (i.e., tickets issued) on collisions at a zonal level, and (ii) Model 2 which is a decision-support tool to assist enforcement agencies with planning their deployment strategy. The first model showed a reduction in zonal collisions associated with an increase in the number of tickets issued (for vehicles exceeding the speed limit). While this relationship was expected, quantifying the impact that automated enforcement activity has on overall zonal collisions is further evidence of how successful this program is at improving safety. Figure 5.1 was then created to illustrate this relationship as well as to aid agencies in better understanding the impact the number of issued tickets has on collisions.

The results of the second model showed that collision reductions were also associated with an increase in the ratio of hours of enforcement per visit. To better understand the impact that these parameters have on the expected zonal collisions, two case studies were proposed. The first case study represented a more concentrated deployment strategy, and the second case study represented a more diffused deployment strategy. The results quantified the impact these two strategies have on the expected frequency of collisions.

Figure 4.2 was also created to aid enforcement agencies with devising a plan to maximize the use of their resources to have a more significant impact on safety. Using case studies that are based on the actual constraints and parameters, bridges the gap between theory and practice. This proves its value in decision making and the planning of resources to maximize safety benefits.

The results are easily replicated and updated to reflect any changes in an agency's deployment strategy.

To account for some of the limitations of regression models in the application of MAE, a new modeling technique was used. Tobit regression models account for the censorship in the data due to the presence of zero-collision frequencies in some zones. An additional benefit of this model is its ability to allow agencies to understand how changing the deployment strategy can impact collisions. Three models were developed i) traditional Tobit model to quantify the impact that the deployment parameter has on collision rates, ii) GRP Tobit model which accounts for variation in the data by categorizing the data into three distinct clusters, and iii) RI Tobit model which accounts for the variation in the intercept.

The results of all three models are in line with earlier evaluations in that increasing the number of hours of enforcement at a site per visit was associated with higher reductions in collisions, which is confirmed by the results of clusters 2 and 3. However, the additional benefit that this analysis provides is to demonstrate the impact of the deployment strategy, explicitly spending a more extended time enforcing a site, to see benefits. Both the GRP Tobit and RI Tobit model results highlight the need to use clusters to define the deployment strategy and to better understand the impact it has on collision rates.

The statistical comparison between all three models indicated that the GRP-Tobit model outperformed the other two models, which is reflected by a significant decrease in the DIC value. Additionally, this approach captures the variation in the deployment parameter *ratio*, which



allows agencies to better understand how to best deploy their resources. While using the RI Tobit model is still a better approach when compared to the Traditional Tobit model, it does not clarify the nuances in the changes in the deployment parameter compared to the GRP Tobit model. In conclusion, the analysis proves the superiority of the Bayesian Tobit models and the importance of accounting for the heterogeneous effects of risk factors in collision rate analysis.

DDACTS approach has been shown to improve crime and collisions across jurisdictions due to the highly correlated nature of both events. However, jurisdictions have applied the DDACTS approach very differently (e.g., number of officers designated to a deployment, hours of enforcement during each time period, the type of enforcement being conducted). This makes it particularly challenging to attribute the reductions in collisions or crime to one specific tool in their deployment strategy. Additionally, it makes it difficult for the road safety agencies to understand the impact of a specific change in their enforcement strategy on collisions and crime.

Typically, traffic enforcement agencies use manned enforcement as part of their DDACTS strategy, which is resource-intensive and costly. There is no research that investigated the impact of one of the most common traffic enforcement programs, MAE, on collision, and crime. For this reason, there were four main objectives for this application 1) confirm and quantify the correlation between collisions and crime, 2) understand the impact of MAE on collisions and crime, 3) understand how to use the models to determine problem neighbourhoods based on collisions and crime, and 4) understand how different deployment parameters for MAE can affect both events.

Firstly, maps were generated to determine whether the correlation between collisions and crime was apparent visually. This is a typical approach that is followed by DDACTS, and the results based on this research confirmed this relationship visually. The next step was to develop models to statistically quantify this correlation.

The first model that was developed included an MVPLN, which was used to confirm the correlation between collisions and crime. The results confirmed the premise of the DDACTS approach since both collision and crime were highly correlated (i.e., 0.72). This indicates that hotspots based on collision frequencies are also highly likely to be crime hotspots, which suggests that continuing to model both events independently is inaccurate. The results showed that increasing MAE was associated with reduced collisions and crime rates, as confirmed by the results from both statistical techniques.

One of the important uses of these models is to understand how to use them to identify which neighbourhoods were hazardous (crime or collision-prone). Two univariate PLN models were developed for collisions and crime separately, and two maps were created to display the results visually. Using the results from the MVPLN model, a map and table were created to show how many locations were identified as hazardous by each of the three models at different threshold values. The results of the analysis further demonstrated that the MVPLN was the superior model as it had identified hazardous locations that were missed by the univariate models. This further proves that modeling both events independently leads to inaccurate and imprecise results.

An additional multivariate random parameter Tobit model was developed to study the impact of the deployment strategy to better understand how it can affect both collisions and crime. Again, the high correlation between collision and crime rates was confirmed using this model (i.e., correlation = 0.86). However, to further investigate how varying the deployment strategy can impact collision and crime rates, the MAE enforcement parameter (i.e., *ratio*) was categorized into three different clusters according to the number of hours spent enforcing a site per visit. The results showed that increasing MAE was associated with reduced collisions and crime rates, confirmed by results of clusters 2 and 3. The analysis demonstrates the impact of the deployment strategy, specifically that spending a more extended time enforcing a site is needed to see benefits in both collision and crime rates.

These overall results provide evidence to suggest that a single deployment of MAE not only results in a reduction in collisions but also in crime. For this reason, focusing deployment strategies to target both collision and crime hotspots is an effective use of the agency's resources. This allows DDACT programs to expand their toolkit to include other strategies such as mobile automated speed enforcement.

## **6.2 Research Contributions**

The research contributions are divided into different categories: the impact of MAE on collisions, and finally, the impact of MAE on crime and collisions.

The main contributions related to the use of automated enforcement and its impact on collisions are:

1. Identified an appropriate unit of analysis to account for the limitations associated with high-level analysis and micro-level analysis
2. Quantified the impact of automated enforcement on collisions at a macro-level
3. Developed a tool to quantify and visualize the impact of a given deployment on predicted collisions, this was provided in the format of:
  - An SPF relating the explanatory variables to collisions
  - A deployment consultation chart
4. Demonstrated the use of these tools through case studies
5. Developed a new model (Tobit model) to account for the presence of the censored data associated with zero collisions in zones
6. Used a new modeling technique to develop three candidate Tobit models, which were compared under the Bayesian framework to determine the impact of changing the deployment on traffic safety.

The main contributions related to the use of automated enforcement and its impact on collisions and crime are:

1. Confirmed and quantified the correlation between collisions and crime, which has previously not been demonstrated
2. Developed a multivariate poisson lognormal model which showed that automated enforcement not only impacts collisions but also incidences of crime

3. Developed a new model (Tobit model) to account for the presence of the censored data associated with zero collisions and criminal incidences in zones
4. Used a new modeling technique to develop a multivariate Tobit model to determine the impact of changing the deployment on traffic safety

### **6.3 Limitations & Future Research**

This dissertation has demonstrated the benefits of macro-level modeling and how it can be used for evidence-based decision making. However, there are several limitations that need to be acknowledged. This includes limitations in the data that is available for modeling as well as challenges that are inherent to the modeling techniques. Further, there are additional applications outside of the scope of this dissertation that is identified for future research.

One of the data limitations was the level of aggregation of criminal incidents. This data was provided at a neighbourhood level as opposed to a TAZ. Previous research has identified the TAZ as the preferred unit of analysis due to the boundary selection process. Conversely, neighbourhoods are defined based on arbitrary boundaries that may not always represent homogenous units. Additionally, due to privacy concerns, any descriptive information regarding the location or details of the crime (e.g., age/gender) was also not available.

The case studies that were selected for the application of macro-level models investigated the impact of MAE on collisions and crime. They also quantified how different deployment strategies regarding the number of hours of enforcement per visit impact safety. However, further model development could include other explanatory variables such as socio-demographic

factors (e.g., employment density, income levels...etc.). Not only would this account for what is the best deployment strategy for a zone, but it can also highlight which zones would benefit the most from this deployment.

Another important application that should be considered for future work is to use the results in this dissertation to build an optimization tool. This tool can take into consideration the safety impact as a result of deployment, the constraints around a deployment (e.g., minimum and maximum hours of enforcement at a site, the distance between sites, available resource hours, etc.), and the cost of deployment (e.g., fleet and personnel considerations) to provide agencies with the ability to maximize their resources to achieve the highest safety benefit.

Lastly, this dissertation has confirmed that macro-level models can be used effectively to better understand the impact of one tool (i.e., MAE programs) on both collisions and crime. Given these results, it is critical to replicate this process to evaluate the effectiveness of other DDACTS tools. For example, developing macro-level models based on manned enforcement deployments and studying their impact on collisions and crime together rather than model them independently (as they had been in previous work). This would give agencies a holistic understanding of the effectiveness of all the elements of their programs, which would help road safety agencies plan the use of their resources more efficiently and effectively.

## References

Aguero-Valverde, J. and Jovanis, P. (2008). Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2061(1), pp.55-63.

Aguero-Valverde, J. and Jovanis, P. (2009). Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking. *Transportation Research Record: Journal of the Transportation Research Board*, 2136(1), pp.82-91.

Alberta Transportation Office of Traffic Safety (2017). *Alberta traffic collision statistics*. Edmonton: Alberta Transportation Office of Traffic Safety.

Anastasopoulos, P. and Mannering, F. (2009). A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis & Prevention*, 41(1), pp.153-159.

Anastasopoulos, P., Mannering, F., Shankar, V. and Haddock, J. (2012). A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis & Prevention*, 45, pp.628-633.

Anastasopoulos, P., Tarko, A. and Mannering, F. (2008). Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis & Prevention*, 40(2), pp.768-775.

Anderson, J. and Hernandez, S. (2017). Heavy-Vehicle Crash Rate Analysis: Comparison of Heterogeneity Methods Using Idaho Crash Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2637(1), pp.56-66.

ArcGIS. (2020). Esri.

Axup, D. (1990). Enforcement - a review of Australian techniques. *Australian Road Research Board (ARRB) Conference*, 15(7), pp.45-60.

Barnett, A. (2004). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, 34(1), pp.215-220.

Bedrick, E., Christensen, R. and Johnson, W. (1996). A New Perspective on Priors for Generalized Linear Models. *Journal of the American Statistical Association*, 91(436), p.1450.

Bernasco, W. and Elffers, H. (2010). *Statistical Analysis of Spatial Crime Data*. New York: Springer, pp.699-724.

Bijleveld, F. (2005). The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. *Accident Analysis & Prevention*, 37(4), pp.591-600.

Bin Islam, M. and Hernandez, S. (2015). Fatality rates for crashes involving heavy vehicles on highways: A random parameter tobit regression approach. *Journal of Transportation Safety & Security*, 8(3), pp.247-265.

Brace, C., Whelan, M., Clark, B. and Oxley, J. (2009). *The Relationship between Crime and Road Safety*. Australia: Monash University Accident Research Centre.



Brijs, T., Karlis, D., Van den Bossche, F. and Wets, G. (2007). A Bayesian model for ranking hazardous road sites. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(4), pp.1001-1017.

Brooks, S. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), p.434.

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L. and Wang, X. (2018). Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. *Analytic Methods in Accident Research*, 19, pp.1-15.

Caliendo, C., De Guglielmo, M. and Guida, M. (2015). Comparison and analysis of road tunnel traffic accident frequencies and rates using random-parameter models. *Journal of Transportation Safety & Security*, 8(2), pp.177-195.

Cameron, M. and Delaney, A. (2006). *Development of strategies for best practice in speed enforcement in Western Australia: Final report*. Victoria, Australia: Monash University Accident Research Centre.

Cameron, M., Cavallo, A. and Gilbert, A. (1992). *Crash-based evaluation of the speed camera program in victoria 1990–91. Phase 1: General effects. Phase 2: Effects of program mechanisms.* Victoria, Australia: Monash University Accident Research Centre.

Canadian Association of Chiefs of Police (2018). *Canada Road Safety Week*. Canadian Association of Chiefs of Police.

Carnis, L. and Blais, E. (2013). An assessment of the safety effects of the French speed camera program. *Accident Analysis & Prevention*, 51, pp.301-309.

Chen, E. and Tarko, A. (2014). Modeling safety of highway work zones with random parameters and random effects models. *Analytic Methods in Accident Research*, 1, pp.86-95.

Chen, G., Wilson, J., Meckle, W. and Cooper, P. (2000). Evaluation of photo radar program in British Columbia. *Accident Analysis & Prevention*, 32(4), pp.517-526.

Chen, G., Wilson, J., Meckle, W. and Cooper, P. (2000). Evaluation of photo radar program in British Columbia. *Accident Analysis & Prevention*, 32(4), pp.517-526.

Chib, S. and Winkelmann, R. (2001). Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business & Economic Statistics*, 19(4), pp.428-435.

Chib, S. and Winkelmann, R. (2001). Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business & Economic Statistics*, 19(4), pp.428-435.

Christie, S., Lyons, R., Dunstan, F. and Jones, S. (2003). Are mobile speed cameras effective? A controlled before and after study. *Injury Prevention*, 9(4), pp.302-306.

Clayton, D. and Kaldor, J., 1987. Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. *Biometrics*, 43(3), p.671.

Congdon, P. (2006). *Bayesian Statistical Modeling*. 2nd ed. New York: Wiley.

Corporate Research Associates (2019). 2018 Edmonton and Area Traffic Safety Culture Survey. [online] Edmonton: City of Edmonton. Available at: [https://www.edmonton.ca/transportation/PDF/2018\\_TrafficSafetyCultureReport-web.pdf](https://www.edmonton.ca/transportation/PDF/2018_TrafficSafetyCultureReport-web.pdf) [Accessed 5 Apr. 2019].

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

Decina, I., Thomas, L., Srinivasan, R. and Staplin, L. (2007). *Automated Enforcement: A Compendium of Worldwide Evaluations of Results*. New Jersey, New York: National Highway Traffic Safety Administration (NHTSA).

Delaney, A., Diamantopoulou, K. and Cameron, M. (2003). *MUARC's speed enforcement research: Principles learnt and implications for practice*. Victoria, Australia: Monash University Accident Research Centre.

Dinu, R. and Veeraragavan, A. (2011). Random parameter models for accident prediction on two-lane undivided highways in India. *Journal of Safety Research*, 42(1), pp.39-42.

Eck, J., Chainey, S., Cameron, J., Leitner, M. and Wilson, R. (2005). Mapping crime: Understanding hot spots. *USA: National Institute of Justice*.

El-Basyouny, K. and Sayed, T. (2009). Accident prediction models with random corridor parameters. *Accident Analysis & Prevention*, 41(5), pp.1118-1123.

El-Basyouny, K. and Sayed, T. (2009). Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis & Prevention*, 41(4), pp.820-828.

El-Basyouny, K. and Sayed, T. (2009). Urban Arterial Accident Prediction Models with Spatial Effects. *Transportation Research Record: Journal of the Transportation Research Board*, 2102(1), pp.27-33.

El-Basyouny, K. and Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis & Prevention*, 41(4), pp.820-828.

El-Basyouny, K. and Sayed, T. (2010). A method to account for outliers in the development of safety performance functions. *Accident Analysis & Prevention*, 42(4), pp.1266-1272.

El-Basyouny, K. and Sayed, T. (2010). Application of generalized link functions in developing accident prediction models. *Safety Science*, 48(3), pp.410-416.

El-Basyouny, K. and Sayed, T. (2010). Safety performance functions with measurement errors in traffic volume. *Safety Science*, 48(10), pp.1339-1344.

El-Basyouny, K., Barua, S. and Islam, M. (2014). Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. *Accident Analysis & Prevention*, 73, pp.91-99.

Elvik, R. (1997). Effects on Accidents of Automatic Speed Enforcement in Norway. *Transportation Research Record: Journal of the Transportation Research Board*, 1595, pp.14-19.

Elvik, R. (2005). Speed and Road Safety: Synthesis of Evidence from Evaluation Studies. *Transportation Research Record: Journal of the Transportation Research Board*, 1908, pp.59-69.

Emme. (2015). INRO.

Ewing, R., Hamidi, S. and Grace, J. (2014). Urban sprawl as a risk factor in motor vehicle crashes. *Urban Studies*, 53(2), pp.247-266.

Fountas, G., Anastasopoulos, P. and Mannering, F. (2018). Analysis of vehicle accident-injury severities: A comparison of segment- versus accident-based latent class ordered probit models with class-probability functions. *Analytic Methods in Accident Research*, 18, pp.15-32.

Functional Manipulation Engine (FME). (2016). Safe Software.

Gelman, A., Meng, X. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, pp.733-807.

Goldenbeld, C. and van Schagen, I. (2005). The effects of speed enforcement with mobile radar on speed and accidents. *Accident Analysis & Prevention*, 37(6), pp.1135-1144.

Guo, Y., Li, Z. and Sayed, T. (2019). Analysis of Crash Rates at Freeway Diverge Areas using Bayesian Tobit Modeling Framework. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(4), pp.652-662.

Guo, Y., Li, Z., Wu, Y. and Xu, C. (2018). Evaluating factors affecting electric bike users' registration of license plate in China using Bayesian approach. *Transportation Research Part F: Traffic Psychology and Behaviour*, 59, pp.212-221.

Guo, Y., Osama, A. and Sayed, T. (2018). A cross-comparison of different techniques for modeling macro-level cyclist crashes. *Accident Analysis & Prevention*, 113, pp.38-46.

Hardy, E. (2010). Data-Driven Policing: How Geographic Analysis Can Reduce Social Harm. *Geography & Public Safety*, [online] 2(3), pp.1-20. Available at: <https://www.nij.gov/topics/technology/maps/Documents/gps-bulletin-v2i3.pdf?Redirected=true> [Accessed 2 Aug. 2018].

Harkey, D. (1999). Evaluation of Truck Crashes Using a GIS-Based Crash Referencing and Analysis System. *Transportation Research Record: Journal of the Transportation Research Board*, 1686, pp.13-21.

Hauer, E. (1997). *Observational before--after studies in road safety*. U.K.: Emerald Group Pub.

Heydecker, B. and Wu, J., 2001. Identification of sites for road accident remedial work by Bayesian statistical methods: an example of uncertain inference. *Advances in Engineering Software*, 32(10-11), pp.859-869.

Higle, J. and Witkowski, J., 1988. Bayesian identification of hazardous sites. *Transportation Research Record*, 1185, pp.24-35.

Ho, G. and Guamaschelli, M. (1998). *“Developing a Road Safety Module for the Regional Transportation Model, Technical Memorandum One: Framework.* Vancouver: Insurance Corporation of British Columbia.

Junger, M., Terlouw, G. and Van der Heijden, P. (1995). Crime and accident involvement in young road users. *Behavioural Research in Road Safety: Transportation and Road Research Laboratory*, pp.35-54.

Kim, H., Sun, D. and Tsutakawa, R. (2002). Lognormal vs. Gamma: Extra Variations. *Biometrical Journal*, 44(3), p.305.

Kuo, P., Lord, D. and Walden, T. (2013). Using geographical information systems to organize police patrol routes effectively by grouping hotspots of crash and crime data. *Journal of Transport Geography*, [online] 30, pp.138-148. Available at: <https://www-sciencedirect-com.ezproxy.library.ubc.ca/science/article/pii/S0966692313000707>.

Li, L., Zhu, L. and Sui, D. (2007). A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography*, 15(4), pp.274-285.

Li, R., El-Basyouny, K. and Kim, A. (2015). Before-and-After Empirical Bayes Evaluation of Automated Mobile Speed Enforcement on Urban Arterial Roads. *Transportation Research Record: Journal of the Transportation Research Board*, 2516(1), pp.44-52.

Li, R., El-Basyouny, K., Kim, A. and Gargoum, S. (2016). Relationship between road safety and mobile photo enforcement performance indicators: A case study of the city of Edmonton. *Journal of Transportation Safety & Security*, 9(2), pp.195-215.

Li, W., Carriquiry, A., Pawlovich, M. and Welch, T. (2008). The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. *Accident Analysis & Prevention*, 40(4), pp.1531-1542.

Loo, B. (2006). Validating crash locations for quantitative spatial analysis: A GIS-based approach. *Accident Analysis & Prevention*, 38(5), pp.879-886.

Lord, D. and Kuo, P. (2012). Examining the effects of site selection criteria for evaluating the effectiveness of traffic safety countermeasures. *Accident Analysis & Prevention*, 47, pp.52-63.

Lord, D. and Miranda-Moreno, L. (2008). Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. *Safety Science*, 46(5), pp.751-770.

Lord, D. and Park, P. (2008). Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. *Accident Analysis & Prevention*, 40(4), pp.1441-1457.

Lovegrove, G. and Sayed, T. (2006). Using Macrolevel Collision Prediction Models in Road Safety Planning Applications. *Transportation Research Record: Journal of the Transportation Research Board*, 1950(1), pp.73-82.



Luoma, J., Rajamäki, R. and Malmivuo, M. (2012). Effects of reduced threshold of automated speed enforcement on speed and safety. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(3), pp.243-248.

Ma, J. and Kockelman, K. (2006). Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity. *Transportation Research Record: Journal of the Transportation Research Board*, 1950(1), pp.24-34.

Ma, J. and Kockelman, K. (2006). Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity. *Transportation Research Record: Journal of the Transportation Research Board*, 1950(1), pp.24-34.

Ma, J., Kockelman, K. and Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention*, 40(3), pp.964-975.

Ma, J., Kockelman, K. and Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention*, 40(3), pp.964-975.

Ma, L., Yan, X. and Weng, J. (2015). Modeling traffic crash rates of road segments through a lognormal hurdle framework with flexible scale parameter. *Journal of Advanced Transportation*, 49(8), pp.928-940.

Ma, X., Chen, F. and Chen, S. (2015). Modeling Crash Rates for a Mountainous Highway by Using Refined-Scale Panel Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2515(1), pp.10-16.

Maher, M. (1990). A bivariate negative binomial model to explain traffic accident migration. *Accident Analysis & Prevention*, 22(5), pp.487-498.

Miaou, S. and Lord, D. (2003). Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods. *Transportation Research Record: Journal of the Transportation Research Board*, 1840(1), pp.31-40.

Michalowski, R. (1975). Violence in the Road: The Crime of Vehicular Homicide. *Journal of Research in Crime and Delinquency*, 12(1), pp.30-43.

Milton, J., Shankar, V. and Mannering, F. (2008). Highway accident severities and the mixed logit model: An exploratory empirical analysis. *Accident Analysis & Prevention*, 40(1), pp.260-266.

Milton, J., Shankar, V. and Mannering, F. (2008). Highway accident severities and the mixed logit model: An exploratory empirical analysis. *Accident Analysis & Prevention*, 40(1), pp.260-266.

Mitra, S. and Washington, S. (2007). On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis & Prevention*, 39(3), pp.459-468.

National Highway Traffic Safety Administration (2014). *Data-Driven Approaches to Crime and Traffic Safety: Operational Guidelines*. Washington, D.C.: NHTSA.

Newstead, S. and Cameron, M. (2003). *Evaluation of the crash effects of the Queensland speed camera program*. Victoria, Australia: Monash University Accident Research Centre.

NHTSA (2013). *Data-Driven Approaches to Crime and Traffic Safety: An Historical Overview*. Washington, D.C.: NHTSA.

Nilsson, G. (2004). *Traffic speed dimensions and the power model to describe the effect of speed on safety*. Lund, Sweden: Lund Institute of Technology.

Park, E. and Lord, D. (2007). Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record: Journal of the Transportation Research Board*, 2019(1), pp.1-6.

Park, E. and Lord, D. (2007). Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record: Journal of the Transportation Research Board*, 2019(1), pp.1-6.

Petch, R. and Henson, R. (2000). Child road safety in the urban environment. *Journal of Transport Geography*, 8(3), pp.197-211.

Sarwar, M., Anastasopoulos, P., Golshani, N. and Hulme, K. (2017). Grouped random parameters bivariate probit analysis of perceived and observed aggressive driving behavior: A driving simulation study. *Analytic Methods in Accident Research*, 13, pp.52-64.

Sarwar, M., Fountas, G. and Anastasopoulos, P. (2017). Simultaneous estimation of discrete outcome and continuous dependent variable equations: A bivariate random effects modeling approach with unrestricted instruments. *Analytic Methods in Accident Research*, 16, pp.23-34.

Sawalha, Z. and Sayed, T. (2006). Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*, 33(9), pp.1115-1124.

Sayed, T. and Abdelwahab, W., 1997. Using Accident Correctability to Identify Accident-Prone Locations. *Journal of Transportation Engineering*, 123(2), pp.107-113.

Schneider, R., Ryznar, R. and Khattak, A. (2004). An accident waiting to happen: a spatial approach to proactive pedestrian planning. *Accident Analysis & Prevention*, 36(2), pp.193-211.

Sodduqi, C. (2012). . *Macroscopic Crash Analysis and Its Implications for Transportation Safety Planning*. Ph.D. University of Central Florida.

Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), pp.583-639.

Statistics Canada. (2019). [online] Available at: <https://www150.statcan.gc.ca/t1/tb11/en/tv.action?pid=1310039401> [Accessed 18 Aug. 2019].

Stern, H. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models. *Statistica Sinica*, 19, pp.2377-2397.

Takyi, E., Oluwajana, S. and Park, P. (2018). Development of Macro-Level Crime and Collision Prediction Models to Support Data-Driven Approach to Crime and Traffic Safety (DDACTS). *Transportation Research Record: Journal of the Transportation Research Board*, p.036119811877735.

Tay, R. (2003). The efficacy of unemployment rate and leading index as predictors of speed and alcohol related crashes in Australia. *International Journal of Transport Economics*, 30(3), pp.363-375.

Tay, R. and de Barros, A. (2011). Should traffic enforcement be unpredictable? The case of red light cameras in Edmonton. *Accident Analysis & Prevention*, 43(3), pp.955-961.

Thomas, L., Srinivasan, R., Decina, L. and Staplin, L. (2008). Safety Effects of Automated Speed Enforcement Programs. *Transportation Research Record: Journal of the Transportation Research Board*, 2078(1), pp.117-126.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26(1), p.24.

Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. *Austrian Journal of Statistics*, 31(3), pp.221-229.

wagWang, C., Quddus, M. and Ison, S. (2011). Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. *Accident Analysis & Prevention*, 43(6), pp.1979-1990.

Washington, S., Karlaftis, M. and Mannering, F. (2011). *Statistical and econometric methods for transportation data analysis*. Boca Raton, Florida: CRC Press.

Wilson, J. (1968). *Varieties of Police Behavior: The Management of Law and Order in Eight Communities (A Harvard paperback)*. Harvard University Press.

World Health Organization (2004). *World report on road traffic injury prevention*. Geneva: World Health Organization.

World Health Organization (2009). *Global Status Report on Traffic Safety: Time for Action*. [online] Geneva: World Health Organization. Available at: [http://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2009](http://www.who.int/violence_injury_prevention/road_safety_status/2009) [Accessed 3 Sep. 2017].

World Health Organization (2017). *Managing Speed*. Geneva, Switzerland: World Health Organization.

World Health Organization. (2018). *The top 10 causes of death*. [online] Available at: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> [Accessed 18 Feb. 2019].

Wright, S. (2018). Multivariate analysis using the MIXED procedure. Nashville: Proceedings of the 38th Annual SAS Users Group International Conference.

Xu, X., Kouhpanejade, A. and Šarić, Ž. (2013). Analysis on Influencing Factors Identification of Crash Rates Using Tobit Model with Endogenous Variable. *PROMET - Traffic&Transportation*, 25(3), pp.217-224.

Xu, X., Wong, S. and Choi, K. (2014). A two-stage bivariate logistic-Tobit model for the safety analysis of signalized intersections. *Analytic Methods in Accident Research*, 3-4, pp.1-10.

Ye, X., Pendyala, R., Washington, S., Konduri, K. and Oh, J. (2009). A simultaneous equations model of crash frequency by collision type for rural intersections. *Safety Science*, 47(3), pp.443-452.

Yu, R., Xiong, Y. and Abdel-Aty, M. (2015). A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway. *Transportation Research Part C: Emerging Technologies*, 50, pp.68-77.

Zaal, D. (1994). *Traffic law enforcement: a review of the literature*. Netherlands: SWOV.

Zeng, Q., Wen, H., Huang, H. and Abdel-Aty, M. (2017). A Bayesian spatial random parameters Tobit model for analyzing crash rates on roadway segments. *Accident Analysis & Prevention*, 100, pp.37-43.

Zeng, Q., Wen, H., Huang, H., Pei, X. and Wong, S. (2017). A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. *Accident Analysis & Prevention*, 99, pp.184-191.

Zeng, Q., Wen, H., Huang, H., Pei, X. and Wong, S. (2018). Incorporating temporal correlation into a multivariate random parameters Tobit model for modeling crash rate by injury severity. *Transportmetrica A: Transport Science*, 14(3), pp.177-191.