School of Transportation Sciences

Master of Transportation Sciences

Master's thesis

Vehicle crash prediction model for signalized intersections in Ghent, Belgium

Yonas Gebreyesus Gebremeskel

Thesis presented in fulfillment of the requirements for the degree of Master of Transportation Sciences, specialization Traffic Safety

SUPERVISOR :

Prof. dr. Tom BRIJS

MENTOR :

De heer Roeland PAUL

2021
2022

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

School of Transportation Sciences

Master of Transportation Sciences

Master's thesis

Vehicle crash prediction model for signalized intersections in Ghent, Belgium

Yonas Gebreyesus Gebremeskel

Thesis presented in fulfillment of the requirements for the degree of Master of Transportation Sciences, specialization Traffic Safety

SUPERVISOR :

Prof. dr. Tom BRIJS

MENTOR :

De heer Roeland PAUL

PREFACE

Initially, I intended to work on my master thesis under the topic of ***"Safety evaluation of signalized intersections in Addis Ababa, Ethiopia."*** Unfortunately, because of an unexpected regional crisis (civil war) in my region(Which is still ongoing), it was impossible to collect data. Then I discussed this matter with my advisor Prof. Tom Brijs. And we decided to work on a new but related topic. As a result, I did my master thesis under the new title ***"Vehicle crash prediction model for Signalized intersections in Ghent, Belgium."***

The number of road traffic fatalities and injuries has increased at an alarming rate globally and remained a major public health issue. In the last decade, middle and high-income countries have made greater progress toward reducing road traffic fatalities than low-income countries. In Belgium, the total number of crashes resulting in injuries or death from 2010 to 2019 decreased by around 17.6%. While the steady progress, the country's traffic death rate remains high compared to the national target. Thus, road traffic crashes remain a threat to public health and the country's national economy, and traffic safety has become one of the highest priorities of the Belgian government. Intersections are identified among the most dangerous locations for traffic crashes. Many factors are believed to associate with the occurrence of a crash at the intersections. To ensure safety at intersections and provide effective and efficient interventions, it is crucial to identify the main risk factors associated with a crash at the intersections. Moreover, developing a tool for properly estimating the number of crashes at such locations is vital. Thus, this study primarily aimed to develop a vehicle crash prediction model for urban signalized intersections in Ghent, Belgium.

I'd like to express my gratitude to Professor Tom Brijs, Roeland Paul, and Wisal Khattak for their constructive feedback and guidance throughout the completion of my master thesis. Especially Professor Tom Brijs for his understanding of the regional situation and for letting me work on a new topic even though it wasn't easy to get the data due to the confidentiality nature. I want to thank you for your trust in me. I would also like to thank Ghent University for the data. Finally, I would like to express my gratitude to the Flemish Interuniversity Council (Vlaamse Interuniversitaire Raad/ VLIR-UOS) for fully funding my Master's degree program at Hasselt University.

I dedicated my thesis work to those 500,000+ individuals who lost their lives due to the ongoing genocidal war on the people of Tigray (The region where I come from) over the past 16 months.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUMMERY

Global road safety reports evidence that road traffic crashes remain as a major problem, both from the public health and socio-economic perspectives worldwide. The burden of a road traffic crash is higher in low-income countries than in middle and high-income countries. In the last decade, more progress has been shown in middle and high-income countries than in low-income countries toward reducing the number of road traffic fatalities. In Europe, between the years 2010-2019, the number of traffic death decreased from 67 to 51 fatalities per million inhabitants. In Belgium, the total number of crashes resulting in injuries or death from 2010 to 2019 reduced from 45,745 to 37,699. During this period, the number of road traffic death declined from 78 to 56 fatalities per million inhabitants, i.e., from 850 to 646. Despite the steady progress, the number of traffic death in the country is still significant compared to the national target (no more than 420 road deaths in 2020). In 2019 the overall annual cost of road crashes was estimated as EUR 5.7 billion, or 1.2% of Belgium's GDP. Accordingly, road traffic crashes remain a threat to public health and the country's national economy, and traffic safety has become one of the highest priorities of the Belgian government.

Various studies have shown that intersections are among the most dangerous locations for traffic crashes. Factors such as the traffic flow, geometry of the intersection, traffic control, environmental and operational characteristics are believed to have an association with the occurrence of a crash at the intersections. Thus, this study primarily aimed to develop a vehicle crash prediction model for urban signalized intersections in Ghent, Belgium. The main objective of the study is to identify and examine the influence of key risk factors contributing to traffic crashes at the signalized intersections and formulate a recommendation to improve safety at the signalized intersections. A cross-sectional study design was used in this study. Seventy-seven signalized intersections were analyzed using four-year crash data for model development. Ghent University provided crash data and traffic count, and road network data.

Previous works of literature were primarily done to identify possible risk factors contributing to a traffic crash at intersections. According to previous works of literature, factors related to traffic characteristics (traffic volume on the major and minor roads), traffic control (left-turn lane, right-turn lane, signalization, the presence of crosswalk, and post speed limit), Geometric characteristics (number of approaching legs, intersection skewness, number of lanes, the width of the median, lane width) found to have a significant relationship with the occurrence of the crash at the intersection.

The Highway Safety Manual (2010) recommended the generalized linear model (GLM) with the negative binomial distribution and logarithmic link function as a standard approach to model yearly crash frequencies. The GLM uses both power and exponential functions for exposure variables and risk factors, respectively. The power function ensures a non-zero positive crash number unless the exposure variable (traffic volume) is zero. On the other hand, the exponential function ensures a non-zero or negative crash number due to zero or negative values from the linear predictors (regression of risk factors). Thus, the generalized linear model (GLM) with the negative binomial distribution and logarithmic link function was utilized in this study to develop the crash prediction model for signalized intersections.

The Pearson correlation coefficient and the variance inflation factor (VIF) were used to examine correlation and multi-collinearity among variables prior to model development. Then, model validation was performed for the developed model. Based on the estimated values found from the model, major risk factors were identified with their respective influence on the number of crashes at the signalized intersections. And the results were interpreted, discussed` and compared with similar previous studies. Following the identification of the main risk factors, the relevant countermeasures and recommendations to improve traffic safety at signalized intersections were forwarded.

Depending on the data availability and previous works of literature, ten variables were considered for model development. A significant correlation was found between the traffic volume on the major and minor approach, between the left turn lane on the major approach and the right turn lane on the minor approach, and between the right turn lane on the minor approach and the number of approaches. Considering the Pearson's correlation and collinear analysis results, all variables except the right turn lane on the minor approach were selected for the model development.

First, simple models incorporating only traffic volume were developed using four functional forms. Almost similar model performance results were observed among the simple models. Thus, the researcher decided to select all the functional forms to model a fully specified crash prediction model and tried to identify a single model with the best fit based on the results from model Goodness of Fit and other model performance evaluation criteria. Only 80% (i.e., 61) of signalized intersections were used to develop the fully specified crash prediction models. And the remaining 20% (i.e., 16) signalized intersections were used for model validation.

According to the best fit model results, only five variables, including the sum of the traffic volume on the major and minor approaches, the ratio of the traffic volume on the minor approach to traffic volume on the the major approach, the Left-turn lane on the major approach, the Presence of crosswalks on the minor approach, and the number of legs/approaches were found to be significant predictor of the total number of traffic crash at signalized intersections in Ghent, Belgium. The sum of traffic volume on the major and minor approaches, the Presence of crosswalks on the minor approach, and the number of approaching legs were positively associated with the number of crashes. Whereas the ratio of the traffic volume on the minor approach to the traffic volume on the major approach, and the left-turn lane on the major approach were negatively associated to the number of crashes.

1. BACKGROUND

1.1 Introduction

Road traffic crashes remain as a major problem globally, both from the public health and socio-economic perspectives (WHO, 2018). According to the WHO (2018), road traffic deaths significantly increased to 1.35 million in 2016. Road traffic death is the 8th leading cause of death for all age groups, especially children & young adults aged 5 to 29 years (WHO, 2018). The report also indicates a strong association between the risk of road traffic death and a country's income level. With an average rate of 27.5 deaths per 100,000 population, the road traffic fatality rate in low-income countries is three times greater than in high-income countries. The average rate is 8.3 fatalities per 100,000 people (WHO, 2018). Compared to all other continents, countries in Africa and southeast Asia have regional rates of road traffic death higher than the global rate, with 26.6 and 20.7 deaths per 100,000 population, respectively (WHO, 2018). Countries in the Americas and Europe have the lowest regional rates of 15.6 and 9.3 deaths per 100,000 population, respectively (WHO, 2018).

According to WHO (2018) report, there has been more progress in reducing the number of road traffic deaths in middle and high-income countries than in low-income countries. Between the years 2013 to 2016, no reductions in the number of road traffic deaths were observed in any low-income countries, whereas some reductions were observed in 48 middle and high-income countries (WHO, 2018). In Europe, between the years 2010-2019, the number of traffic death decreased from 67 to 51 fatalities per million inhabitants, i.e., from 29,611 to 22,700 (European Commission, 2021). Yet, this number is very significant compared to the European commission goal, i.e., 15,750 by 2020. However, compared to the other continents in the year 2016 in Europe, only 50 road traffic death per one million inhabitants were sustained, while 174 fatalities per million inhabitants occurred globally(European Commission, 2018).

In Belgium, the total number of crashes resulting in injuries or death from 2010 to 2019 reduced from 45,745 to 37,699 (European Commission, 2021). During this period, the number of road traffic death declined from 78 to 56 fatalities per million inhabitants, i.e., from 850 to 646 (European Commission, 2021). Despite the steady progress, the number of traffic death in the country is still significant compared to the national target, i.e., no more than 420 road deaths in 2020(International Transportation Forum, 2020). In 2019 the overall annual cost of road crashes was estimated as EUR 5.7 billion, or 1.2% of Belgium's GDP (International Transportation Forum, 2020). Accordingly, road traffic crashes remain a threat to public health and the country's national economy, and traffic safety has become one of the highest priorities of the Belgian government (Geurts et al., 2003).

According to the Highway Safety Manual (2010), *" a crash is defined as a set of events that results in injury or property damage, due to the collision of at least one motorized vehicle and may involve a collision with another motorized vehicle, a bicyclist, a pedestrian or an object."* Crashes as rare and random events; their occurrence is influenced by several factors such as human factors, vehicle factors, and roadway/environmental factors (AASHTO, 2010). Thus, identifying and understanding the different contributing factors that influence crash occurrence helps to reduce and eliminate traffic fatalities and serious injuries (Wang & Zhang, 2017). It is also crucial to examine the nature of the relationship between the roadway/environmental factors, operational factors and crashes to understand the causal mechanism involved in the traffic crash and better predict their occurrence (Nambuusi et al., 2008). Crash prediction models (CPM) are one mechanism used to gain these insights (Eenink et al., 2008).

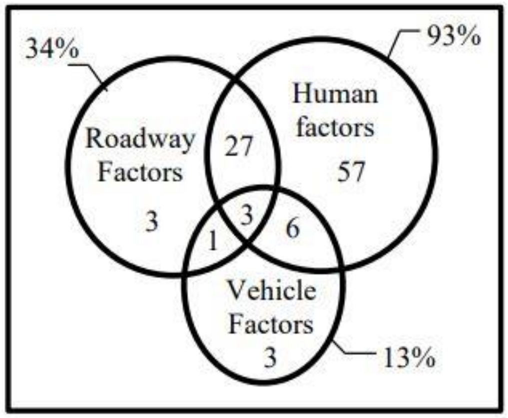

FIGURE 1: Contributing factors to vehicle crashes (AASHTO, 2010).

Crash prediction models (CPM) are used for various purposes, such as estimating the expected crash frequencies from different roadway segments (highways, intersections, interstates, etc.) and identifying geometric, environmental, and operational factors that are associated with the occurrence of traffic crashes(Nambuusi et al., 2008). Different types of roads meet at an intersection; as a result, there is a chance for different crash types to occur at such locations (Nambuusi et al., 2008). Thus, it is vital to differentiate models to examine factors associated with the different crash types. In addition, it is also important to examine the nature of the relationship between the geometric, environmental, and operational factors and the traffic crashes to understand and have a good prediction of the traffic crash. Several approaches have been developed to identify factors and elements that affect the safety of the intersections. These include the multiple linear regression models, the multiple logistic regression models, Poisson regression models, negative binomial regression models, zero-inflated Poisson and negative binomial model, Gamma model, random-effects models, and the classification and regression tree (CART) technique. Thus, This study mainly aims to develop crash prediction models for signalized intersections that are found in Ghent, Belgium.

1.2 Problem Statement

Intersections are common places for traffic crashes as a result of several conflict movements and the different geometric characteristics of the intersection(Nambuusi et al., 2008). The same author also indicates that intersections tend to experience severe traffic crashes resulting from angle and left-turn collisions. According to AASHTO (2010), an intersection is defined as "*the general area in the road network where two or more roads join or cross, and the area includes the roadway and roadside facilities for traffic movements.*" Several studies have shown that intersections are among the most dangerous locations of a roadway segment. Factors such as traffic flow, the geometry of the intersection, traffic control, environmental and operational characteristics are believed to be associated with the occurrence of a crash at such location (Abdel-Aty & Keller, 2005; Nambuusi et al., 2008). A study by Nambuusi et al. (2008) showed the significance of various variables (factors) related to crashes at intersections. According to the author, factors such as traffic flow, traffic control, geometric characteristics, driver characteristics, land use, and vehicle types and features are significantly associated with crashes at the intersection.

Identifying and understanding the relationship between factors contributing to the crash occurrence helps to propose appropriate remedial measures. A great amount of effort has been done to understand the relationship between factors contributing to crashes at the intersection over the past years; this includes developing analytical methods such as Generalized Linear Models, improved experimental designs (before-after studies, conflict analysis), and on the model functional forms (Mitra & Washington, 2012). Tools such as Crash prediction models (CPMs) are one mechanism used to examine the nature of the relationship between the factors contributing to the crash and to predict their occurrence better (Eenink et al., 2008; Nambuusi et al., 2008).

The crash prediction model (CPMs), also known as safety performance functions (SPFs), usually denotes a multivariate model fitted to crash data to estimate the statistical relationship between the number of crashes and factors that are believed to be (casually) related to crash occurrence. Over a specific period, CPMs can be used to estimate the expected average crash frequency of roadway facilities. The development of CPMs are a crucial process in which a modeler makes essential decisions(Khattak et al., 2021). Hauer and Bamfo (1997) emphasized that, in the course of modeling, the modeler will make two major decisions, i.e., *"(a) What explanatory variables to include in the model equation; and (b) What should be its functional form."*. The same author also indicates if the correct functional form has not been chosen for the model and/or if the appropriate explanatory variables are not used, the entire process on which the method for estimating rests crumbles into meaningless. Factors such as the purpose of the CPMs, and the availability, quality, and quantity of the data affect the expertise on those decisions(Khattak et al., 2021).

There are several explanatory variables, yet the selection of explanatory variables appears to depend on the availability of the data (Nambuusi et al., 2008). Nevertheless of the choice, explanatory variables should not be based on the data availability only; the explanatory variable should include variables that have been found in previous studies to have a major influence on the number of the crash and should not be highly correlated with other variables included in the model, and the variable included can be measured in a valid and reliable way (Eenink et al., 2008).

Other than the choice of the explanatory variable, the functional form is also another essential aspect of the model since it influences the accuracy and predictive performance of the model, especially in the case of the variables representing traffic flow at intersection models given the complexity of traffic interaction to be addressed (Ferreira & Couto, 2013). As described in the work by (Chin & Quddus, 2003; Lord & Mannering, 2010; Miaou & Lord, 2003), various statistical techniques, such as empirical Bayes methods and generalized linear models, have been applied to model the relationship between road crashes and traffic flow at intersections. Regardless of those techniques used, it remains a question of what functional form best addresses the essential relationship between crash occurrence and the traffic flow at the intersection, given the related complexity of the traffic movements. Miaou and Lord (2003) indicated that defining a model that produced a good fit to the data set is no longer challenging as it can be easily accomplished using many smoothing techniques and associated software, rather, criteria based on logic (such as reason, consistency, and coherency), flexibility, extensibility, and interpretability of the functional form should be taking into account in model development. Thus, this paper will present a research effort where crash prediction models (CPMs) for signalized intersections specific to Ghent, Belgium, will develop and examine to understand and identify the effects of traffic flow, the geometric/environmental, and operational characteristics that significantly contribute to the intersection and intersection-related crashes.

1.3 Research Questions

1. What are the key variables influencing crashes at the signalized intersections?
2. How are the different variables contributing to the number of crashes at the signalized intersections?
3. Which regression model best associate the number of crash and the contributing factors?
4. How can the safety at the signalized intersections be improved?

1.4 Research Objectives

The general objective of this study is to develop a crash prediction model for signalized intersections in Gent, Belgium.
The specific objectives of the study will be:

- To identify the key factors contributing to crashes at the signalized intersections.
- To examine the influence of the exposure and explanatory variables on the number of crashes at the signalized intersections.
- Formulate a recommendation to improve safety at the signalized intersections.

1.5 Scope of The Study

This study is delimited in terms of Geographic location and road facility type. Vehicle crashes that occurred at the signalized intersection from 2014 to 2017 will be studied. Moreover, the study will be conducted in an urban region (Ghent, Belgium).

1.6 Limitation of The Study

This study shared some of the drawbacks of crash prediction modeling studies, which are omitted variable bias and controlling for confounding factors.

1.7 Conceptual FrameWork of The Study

The conceptual framework showed a detailed procedure for developing a crash prediction model for urban intersections.

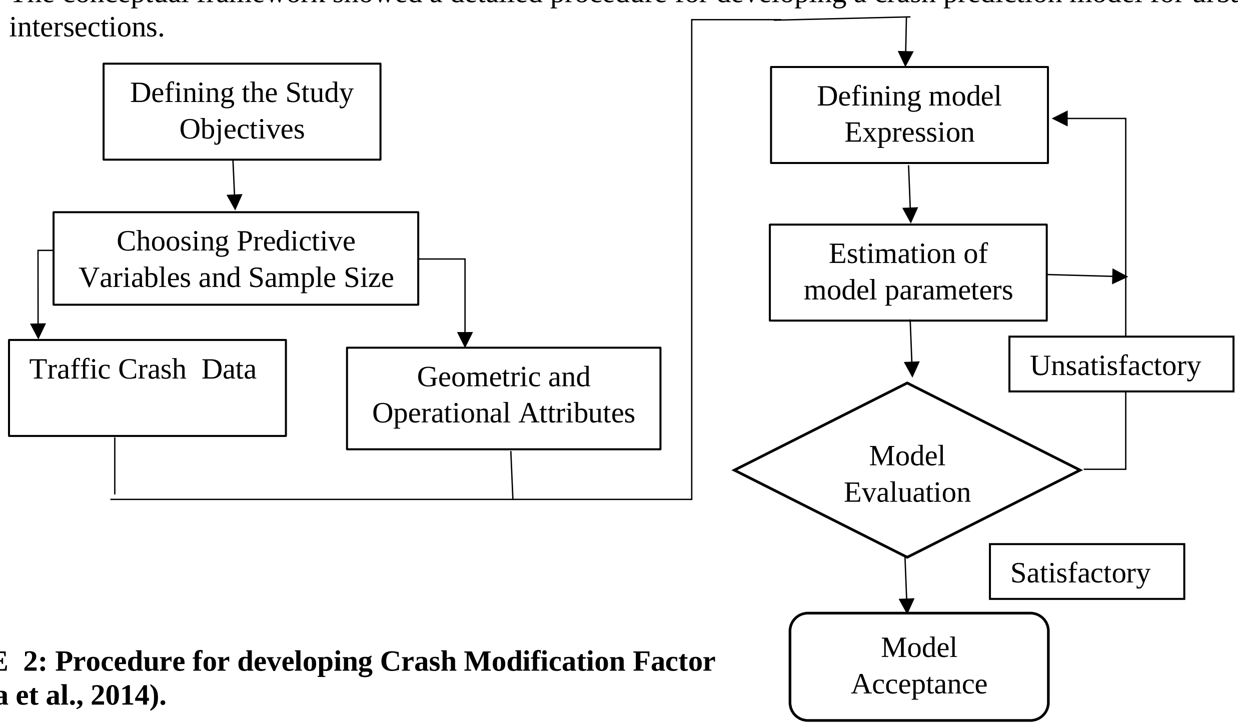

FIGURE 2: Procedure for developing Crash Modification Factor (Barbosa et al., 2014).

2. LITERATURE REVIEW

The literature review primarily aims to acquire background knowledge and insights on predictive models by exploring the different predictive variables and the method of regression used in previous studies.

2.1 Safety at Intersections

It is not surprising to consider intersections among those locations that possess the highest portion of the total crashes on the roads. This is because, at the intersection, the need for instant decision-making, the complex urban design, dense and rigorous land use, congesting, heavy traffic, vulnerable road users, and many on-and-off-vehicle distractions overload driver's attention which makes it challenging for the drivers to operate safely (Khattak et al., 2021). In addition, drivers experience various interactions with crossing and turning vehicles, pedestrians, and cyclists in such locations. In the united states, about 43% of all crashes occurred at or near intersections (Lord et al., 2005). About 40% of all traffic crash casualties in Norway occurred at intersections (Elvik & Vaa, 2004). The annual road crash statistics in Singapore showed more than one-third of crashes, i.e., 34.31%, occurred at the intersection during 1992-2002 (Tay & Rifaat, 2007).

Several major factors influence crashes at the intersection, including traffic flow, traffic control measures, geometric characteristics, and driver characteristics. Various studies have been examining the impact of traffic and geometric characteristics on the frequency of crashes at intersections, including traffic flow (Ferreira & Couto, 2013), signal timing (Bonneson & Zimmerman, 2006; Wang et al., 2006), lane arrangement (Wang et al., 2006), curvature (Savolainen & Tarko, 2005), collision type (Abdel-Aty et al., 2005; Jagannathan et al., 2006), and intersection approach conditions (Kulmala, 1998; Poch & Mannering, 1996). In addition, many studies have also been examined the influence of these factors on the severity level of crashes at intersections (Abdel-Aty, 2003; Abdel-Aty et al., 2005; Jagannathan et al., 2006). These findings are good indicators of the existence of a relationship between the contributing factors and the occurrence of crashes at intersections. Understanding the factors contributing to crashes can help improve intersections' safety by proposing an appropriate countermeasure(Khattak et al., 2021).

2.2 Vehicle-Crash Predictive Variables

2.2.1 The Effect of Traffic Volume on crash frequency

In estimating the safety performance function for intersections, traffic volume or AADT is one key factor that reflects the risk exposure at intersections (Wang et al., 2020). Even though research demonstrated the relationship between crashes and AADT as non-linear (Ivan, 2004; Jonsson et al., 2007; Qin et al., 2004), their fundamental relationship is more complex than a simple non-linear function and differ by different crash type(Ivan, 2004; Qin et al., 2004).

Khattak et al. (2021) conducted a study to estimate the safety performance function for urban intersections in Antwerp, Belgium. The analysis included a total of 760 intersections, of which 198 were signalized and 562 unsignalised. 470 three-legged and 290 four-legged intersections were analyzed. The study result showed a positive association between crash frequency and traffic volume of major and minor intersection approaches for almost every severity level and intersection type.

Alarifi et al. (2017) conducted a study to develop a crash prediction model for Orlando, Florida. The study analyzed 247 signalized intersections by including the effect of macro-level data in addition to the intersection level data. The study found that the natural log of major and minor AADTs was significant and positively associated with the crash occurrence at signalized intersections. Further, it was recognized that traffic volume on the major roads has a higher potential for crash prediction.

Wang et al. (2020) studied 4-leg intersections on urban and suburban arterials at the State of Connecticut to estimate SPF by crash type to account for crash distribution variations using the most recent five-year (2015–2019) crash data. Intersections were categorized into all-way stop-controlled and signalized intersections. 1095 stop-controlled and 1552 signalized intersections were included in the study. Further, signalized intersections were separated into two-lane and multilane intersections. The study findings revealed that major road and minor road AADT as the exposure measure performs best in estimating SPF for same direction crash, intersecting direction crash, opposite direction crash, and single-vehicle crash.

A work done by Gomes et al. (2012) to develop crash prediction models for urban intersections located in Lisbon, Portugal. Models with covariates and flow-only models were estimated using data collected at 94 intersections (44 three-legged and 50 four-legged) and crash data from 2004 to 2007. The study result showed the total traffic inflow was found to be highly significant for both three-and four-legged intersections. In addition, the results showed the ratio between the traffic flow entering in a minor direction and the total traffic flow entering the intersections have a positive effect on the safety of three-legged intersections.

Barbosa et al. (2014) did a study to develop a safety performance model for urban intersections of three Brazilian cities and investigate the transferability of models between three cities. The models were developed for 352 signalized and 132 unsignalised intersections using crash data from 2005 to 2010. The study findings showed that the AADT on the major and minor approaches were the most significant variables.

A study by Miranda-Moreno et al. (2011) was done on 519 signalized intersections in Montreal, Quebec, Canada, to develop a crash prediction model. As a base model, only traffic volume and pedestrian volume were used as a predictor, and it was found that the coefficient for traffic volume and pedestrian volume was 1.15 and 0.45, respectively. In the same study, the author developed a second model by including built environment variables and the traffic volume and pedestrian volume, and it was found a decrease in the coefficient both in traffic volume and pedestrian volume, i.e., 0.91 and 0.26, respectively. The author outlined that the slight change in coefficients (traffic volume from 1.15 to 0.91 and pedestrian volume from 0.45 to 0.26) when built environment variables are included in the model may suggest that a considerable portion of observed variability is explained by traffic and pedestrian volume. Wang et al. (2017) conducted a recent study on 279 intersections in Florida, USA. The study findings showed that the coefficient for traffic volume was 1.19 (for the model including traffic volume and road variables) and 1.15 (when macroscopic variables were included in the model). Another study by (Xie et al., 2018) in Hong Kong, based on the model fitted from 262 signalized intersections with three years of crash data, and it was founded that the coefficient for traffic volume and pedestrian volume 0.27 and 0.21, respectively.

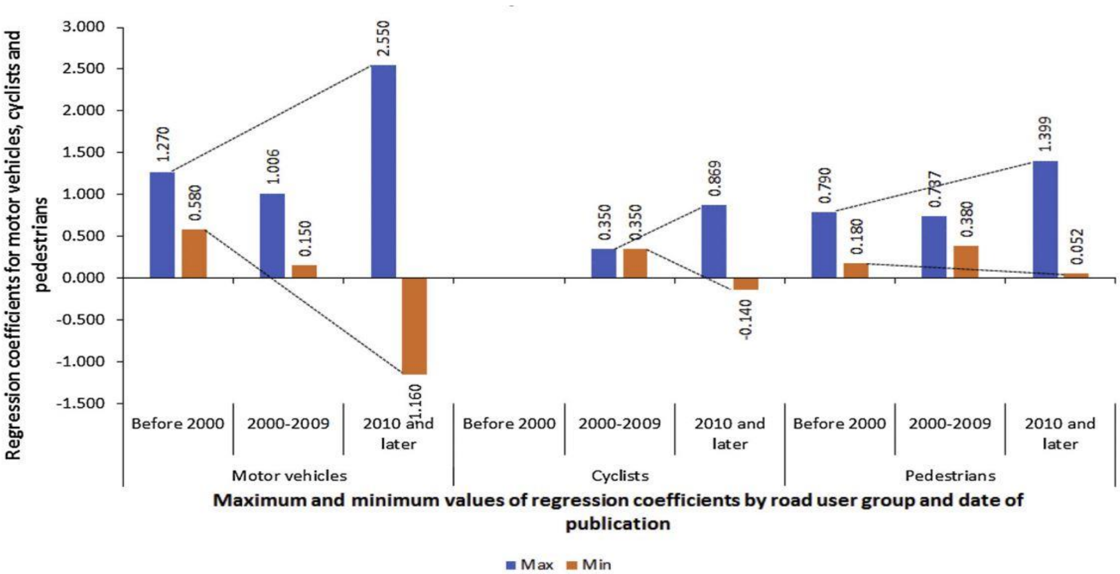

FIGURE 3:Regression coefficients for motor vehicles, cyclists, and pedestrians from previous researches (Elvik & Goel, 2019).

A summary of the trend in regression coefficient variation for traffic volume, pedestrian volume, and cyclists in previous works of literature is shown in figure 3 (Elvik & Goel, 2019). Figure 3 shows the maximum and minimum values of regression coefficients reported in studies published before 2000, between 2000 and 2009, and between 2010 and later. The tendency for the range of estimated values of the coefficients to become more extensive overtime is most evident for motor vehicles, yet it is found for cyclists and pedestrians as well. Miranda-Moreno et al. (2011) forwarded possible reasons behind the wide variation in the regression coefficient over time resulting from the difference in the number of sample size (number of intersections), type of intersections (three-legged, four-legged), number of crash data considered for analysis, quality of traffic and crash data, regression methods and many more from one study and to the other.

Finally, the recently accomplished National Cooperative Highway Research Program (NCHRP) 17-62 project (2017), as cited in (Wang et al., 2020), confirmed that the relationship between AADT and crash counts differ among different crash types and crash severities. Thus, the safety performance functions(SPFs) estimation should be disaggregated by crash type and severity level.

2.2.2 The Effect of Intersection Characteristics

Various studies showed that intersection characteristics (Traffic control and geometric characteristics) significantly affect vehicle crashes(Harwood et al., 1995; Harwood et al., 2002; Khattak et al., 2021; Kim & Washington, 2006). Accommodating the left turn lane on major roads is possibly the most challenging problem in traffic engineering because a left turn can not be made simultaneously while a vehicle is passing in the opposite direction. Left turn lanes were found to be effective in reducing total crashes at intersections (Zhou et al., 2010).

A before-after study was conducted to evaluate the safety effect of installing left-turn lanes in Connecticut. Installing left-turn lanes decreased crashes in some cases but not in all situations. The same author pointed out that adding left-turn lanes were less effective at reducing crashes under conditions associated with high traffic intensity, traffic signals, and four-legged intersections(Rimiller et al., 2003). Another study showed a crash reduction from 77% to 18% due to the installation of left-turn lanes (Gluck et al., 1999). A study by Maze et al. (1994) to assess the safety of left-turn treatment at high speed signalized intersections showed adding left-turn lanes were found to reduce crash rate by 6% with permitted phasing and by 35% with protected /permitting phasing. A study was conducted by Zhou et al. (2010) to examine the safety effects of exclusive left-turn lane installation at unsignalized intersections and driveways, the study results showed adding left-turn lanes for rural two lanes, three-leg intersections, and urban four-lane intersections creates a safer condition for the same direction crashes, including rear-end, turning, and sideswipe crashes.

A study was done by De Pauw et al. (2015) to investigate the safety effects of protected left-turn phasing at signalized intersections in Flanders, Belgium. The study included 103 signalized intersections with left-turn. The study results showed a significant decrease in the number of injury crashes by -46% and a reduction of left-turn crashes by -60%. The same study also indicates that the number of rear-end injuries did not change significantly after implementing a protected left-turn signal. However, a large effect was identified for more severe crashes, i.e., -66%.

Only a few studies have been conducted on the safety effectiveness of right-turn lanes. A study by Harwood et al. (2002) showed a 5% reduction in a crash due to right-turn lane alone, one approach in a rural stop-controlled intersection, and a 10% reduction along both major approaches. A study conducted for three-leg unsignalized intersections along rural two-lane highways showed that a right-turn lane increases intersection-related crashes by 27% (Vogt & Bared, 1998).

Kim and Washington (2006) try to justify the reason for the inconsistency in the safety effects of the left-turn lane by the potential of endogeneity of left-turn lanes that earlier researchers have not controlled. They recommended that the installation of left-turn lanes was regularly endogenous and affected by crash counts and traffic volume. Accounting endogeneity of left-turn lanes reduced angle crashes (Kim & Washington, 2006).

Signalizations are often installed in urban intersections to improve traffic operations, signal coordination, safety, and perceived risks. Various safety studies of intersection signalization have been conducted worldwide. A study was undertaken by McGee et al. (2003) in North America to evaluate the safety effect of signalization of 22 three-armed and 100 four-armed intersections in urban areas. The study results showed that the number of injury crashes was reduced by 14% at three-armed intersections and 23% at four-armed intersections. Many studies indicated signalization of intersections decreased the number of right-angle crashes while it was observed an increased in rear-end crashes; recent studies in the U.S. found a reduction in right-angle crashes by 34-77 % and an increment of rear-end crashes by 38-58 %(Elvik et al., 2009; Harkey, 2008; McGee et al., 2003). Jensen and ApS (2010) conducted a before-after study to evaluate the safety effects of signalization of 54 intersections in the central part of Copenhagen, Denmark. The study analyzed 35 four-armed, 18 three-armed, and 1 five-armed intersections using data from 1997-1999. Intersections were converted from yield control to signal control. The study found that signalization of intersections significantly reduced total crashes by 39% and injuries by 23% for four-armed intersections. The reduction estimated for three-armed intersections was insignificant, i.e., 21% and 17% for crashes and injuries, respectively. In the estimation for the five-armed intersection, the number of crashes decreases significantly, safety at the five-armed intersection has most probably improved after the installation of signalization.

Crosswalks are provided to guide pedestrians along roadways and to increase drivers' awareness of pedestrians. Various study results provide conflicting conclusions regarding whether the provision of marked crosswalks improves pedestrians' safety at the intersection. Several studies showed that marked crosswalks reduce crash rates in some cases by as much as 50% (Smith & Knoblauch, 1987; Wilson, 1967) as cited in (Harwood et al., 2002). On the other hand, early studies on crosswalks conducted by (Herms, 1970; Smith & Knoblauch, 1987), as cited in Harwood et al. (2002), found approximately an increase of twice as many pedestrian crashes occurred in marked crosswalks as in unmarked crosswalks and an increase of 86% of a pedestrian crash after crosswalks were marked. Khattak et al. (2021) conducted a study to estimate the safety performance function for urban intersections in Antwerp, Belgium. The study findings showed that in signalized intersections, the presence of a sidewalk on the minor approaches had a significant positive association with the crash rate only when it was present on both approaches, but it was not significant when crosswalks provided only on one of the minor approaches. The author also found that the estimated coefficients were often more than double for intersections with crosswalks on both minor approaches than intersections on one approach only. Herms highlighted the increase in crash rates resulting from marked crosswalks may "not be due to the crosswalk being marked as much as it is a reflection on the pedestrian's attitude and behavior when using the marked crosswalk." Other factors which might influence the safety of marked crosswalks incorporate visibility, intersection type, and signal timing.

A study was conducted by Zegeer et al. (2001) on 1000 marked and 1000 unmarked crosswalks at unsignalized intersections & mid-block locations in 30 U.S. cities. The study findings showed that at uncontrolled locations with more than two-lane where the average daily traffic is low, i.e., ADT<12,000, a marked crosswalk alone didn't induce a statistically significant difference in pedestrian crash rate. Whereas, on multilane roads with high ADT, i.e., ADT>12,000, a marked crosswalk without any other enhancement showed statistically significant higher pedestrian crash rates than the pedestrian crash rates on unmarked crosswalks. The same author also found that multilane roads with raised median in either marked or unmarked crosswalks induced lower crash rates compared to roads with no raised median. Even though the provision of crosswalks mainly affects pedestrian safety, it is worth noticing that the vehicle crash rate may also be affected. Heurn (1988), as cited in (Harwood et al., 2002), indicated that rear-end collisions increase after crosswalks are marked. Accordingly, the need for crosswalk provision should be analyzed from pedestrians and vehicular safety perspectives.

Most of the time, median refuge islands are found at the center of roadways, along with crosswalks to provide pedestrians with a safe place to wait for gaps in traffic while crossing a wide roadway. Several studies showed favorable benefits of median refuge islands (Bacquie & Egan, 2001; Zegeer et al., 2005) found statistically significant lower pedestrian crash rates. A recent study by Pulugurtha et al. (2012) found a statistically significant increase in the proportion of drivers yielding to pedestrians & the distance drivers yielded to pedestrians. The presence of median refuge islands results in a statistically significant reduction in mean speed (King et al., 2003).

It is reasonable to assume that as the posted speed limit on an intersection approach increases, the likelihood and severity of crashes also increase. Higher post speed limits are usually related to higher approach speeds, as it takes longer to bring an approaching vehicle to a complete stop. As a result, drivers must respond more swiftly to potential conflicts at intersections. The relationship between the speed at intersections and safety is speculated frequently, but few studies exist to clearly identify what that relationship might be. Salifu (2004) developed a crash prediction model to study the relationship between the number of crashes, traffic flow, traffic control, and geometric characteristics for urban intersections in Ghana. The study analyzed 91 intersections, of which 57 were three-arms and 34 were four-arm unsignalized intersections using three years of crash data. The study results showed that the high speed of vehicles approaching the intersections along the major road increases vehicle crashes.

In previous studies, there is broad agreement that four-leg intersections have more crashes than equivalent three-leg intersections. This finding is logical since four-leg intersections have more conflict points than three-leg intersections, which means there are more chances for a crash to happen in four-leg intersections. A study by (Bauer & Harwood, 1996), as cited in Harwood et al. (2002), found that four-leg stop-controlled intersections in rural and urban areas experienced approximately twice as many crashes as in three-leg intersections. The same author stated explicitly rural four-leg stop-controlled intersections accommodated an average of 1.1 crashes yearly, while three-leg intersections experienced 0.6 crashes per year; similarly, four-leg stop-controlled intersections in urban areas experienced 2.2 crashes yearly, while 1.3 crashes were experienced at three-leg intersections per year. Another study by Harwood et al. (1995) showed the variation in the rate of a crash as a function of the median width, i.e., four-leg intersections with narrow medians had about twice as many crashes as three-leg intersections and more than five times as many crashes as three-leg intersections with wide medians. An early study by David and Norman (1975) Showed that in urban areas, a stop-controlled intersection with a total traffic volume less than 20,000 veh/day at entering, the crash frequency for both three and four-leg intersections varied similarly. But for intersections with total traffic volume at entering above 20,000 veh/day, the crash frequency at the four-leg intersection were twice as many as three-leg intersections.

In relation to the number of intersection legs, the angle between the intersection legs has long been considered to affect the safety performance of the intersection. Intersecting roadways should be oriented at a 90-degree angle as much as possible. Intersection design, on the other hand, can diverge from this desirable configuration and resulting in a skewed intersection. A study was conducted by Nightingale et al. (2017) to determine the effect of intersection skewness on crash frequency for rural stop-controlled intersections on high-speed two-lane highways in the state of Lowa. The study separately analyzed three and four-leg intersections to estimate crash frequency as a function of AADT, skew angle, and other geometric characteristics. The study results showed that crash frequency consistently increase with skew angle, i.e., a 10-degree deviation from 90 degrees resulted in 3% more crashes at three-leg intersections and 4% more crashes at the four-leg intersection. A study by Khattak et al. (2021) found that intersection skewness was statistically significant for the total crash, injury, and fatal crashes in the case of signalized intersections. The study findings also indicated more crashes were expected on intersections with high skewness level than intersections with no or low level of skewness. Another study by (Kumfer et al., 2019) examined the effects of intersection angle on intersection safety. The study analyzed three-leg and four-leg stop-controlled intersections with two-lane and multilane major legs using crash data of seven years and five years from Minnesota and Ohio. The study results showed that more than half of the intersection types experience the highest number of predicted crashes when the intersection angle lays between 50 to 60 dgrees.

The width and number of approach lanes of an intersection are other crucial geometric characteristics that affect the safety performance of an intersection. According to a study by Khattak et al. (2021), the estimated crash frequency was shown to be influenced by the number of approach lanes of unsignalized intersections only, especially in the case of property damage only (PDO) crashes and total crashes. The same author also found that the association between the number of through lanes and the expected crash was significant and positive, i.e., an increase in the number of through lanes resulted in more crashes. Bauer and Harwood

(1996) found that for urban four-leg signalized intersections, the crash frequency tends to be higher with an increase in the number of approach lanes; whereas, for unsignalized intersections in both rural and urban areas, the rate of crash tends to be higher for intersections with one approach lane and lower at intersections with two and more approach lanes. (Poch & Mannering, 1996; Wang et al., 2006) showed that as the number of approach lanes increases, the number of crashes also increases at intersections. It is difficult to determine if any of the observed safety effects are due to the number of lanes or the traffic volume on the approach when using a demand-related design parameter like the number of lanes(Harwood et al., 2002). Similar to the number of lanes, the width of the approach lane also affects intersection safety. A study by (Harwood et al., 1995; Harwood et al., 2002) showed that an increase in approach width to an intersection reduces the crash rate along with the approach.

To summarize, exposure and explanatory variables that have been found to have significant relation with crash risks and frequency are summarized in the table below. The variables are also proposed for use in the model development for this study.

TABLE 1:Identified important exposure and explanatory variables for vehicle crash risk and frequency estimation from previous research

Category	Variables	Unit	Reference
Traffic characteristics	Traffic volume	AADT	(Khattak et al., 2021; Nambuusi et al., 2008; Wang et al., 2020)
Traffic control	Left turn lane	Yes/no	(Harwood et al., 2002; Nambuusi et al., 2008; Zhou et al., 2010)
	Right turn lane	Yes/no	(Harwood et al., 2002; Nambuusi et al., 2008)
	Signalization	Yes/no	(Elvik et al., 2009; Harkey, 2008; McGee et al., 2003)
	Presence of crosswalk	Yes/no	(Harwood et al., 2002; Khattak et al., 2021)
	Post speed limit	Km/hr	(Nambuusi et al., 2008)
Geometric characteristics	Number of legs	3 to 5	(Harwood et al., 1995; Harwood et al., 2002)
	Intersection skewness	Yes/no	(Khattak et al., 2021; Kumfer et al., 2019)
	Number of approach lanes	1 to 5	(Harwood et al., 2002; Nambuusi et al., 2008)
	Lane width	Meter	(Harwood et al., 1995; Harwood et al., 2002; Nambuusi et al., 2008)

2.3 Vehicle Crashes and The Regression Methods

2.3.1 Data and Methodological Approaches

Researchers have studied and identified issues related to data and methodology in crash frequency modeling as primary sources of error in terms of incorrectly specifying statistical models, which results in an erroneous crash frequency and wrong inference of explanatory variables (Lord & Mannering, 2010). According to the review and explanation done by Lord and Mannering (2010), these methodological and data-related issues are summarized below.

Over-dispersion: It is worth noticing that the variance of crash-frequency data exceeds the mean of the crash counts. This is problematic since the properties of most common count-data modeling approach (the Poisson regression model) are limited to the mean and variance to be equal. Estimating a common Poisson model with overdispersed data can result in biased and inconsistent parameter estimates, leading to erroneous inferences about the variables determining crash frequencies(Lord & Bonneson, 2007).

Under-dispersion: most of the time, it is not common characteristics of crash data to be under-dispersion, i.e., the sample mean value will be greater than the variance. Previous studies have shown that traditional count-data models result in incorrect parameter estimation in the presence of under-dispersed data (Oh et al., 2006).

Time-varying explanatory variables: Explanatory variables may vary dramatically over time, but the fact that crash frequency is considered only over some period of time might result in ignoring the potential within-period variation in explanatory variables, which leads to loss of potentially crucial explanatory information. This can induce error in the model estimation because of unobserved heterogeneity (Washington et al., 2020).

Under-reporting: There is a potentially serious issue with the under-reporting of crashes since less severe crashes are less likely to appear in crash databases. Although the degree of under-reporting for each severity level is often unknown, recent research has demonstrated that count-data models are prone to producing biased estimates when under-reporting is not taken into account during the model-estimation process(Ma, 2009).

Omitted-variables bias:

It might be too easy to develop a model with few explanatory variables (for example, using traffic flow as the only explanatory variable in the model). However, ignoring significant explanatory variables, with other traditional statistical estimating approaches, leads to biased parameter estimates, resulting in erroneous inferences and crash-frequency predictions. This is especially true if the omitted variable is correlated with other variables in the specification (Lord & Mannering, 2010).

Functional form: Functional forms of a model establish a link between dependent and explanatory variables and are an essential part of the modeling process. Most count-data models assume that explanatory variables have a linear effect on dependent variables. Non-linear functions, on the other hand, appear better to characterize the relationship between crash frequencies and explanatory variables (Miaou & Lord, 2003).

2.3.2 Modeling Approaches

Previously, a wide variety of methods have been applied over the years to deal with the data and methodological issues associated with crash frequency. This includes Poisson, Negative binomial, Poisson-lognormal, Zero-inflated Poisson, Negative Binomial, Gamma, Generalized estimating, Random-effects, Random-parameters equation, and Bivariate/multivariate models(Lord & Mannering, 2010). The table below summarizes model approaches used by recent researchers for crash predicting models.

TABLE 2:Summary of previous research analyzing crash-frequency data

Models	Previous researches
Poisson	Daniels et al. (2011), Wali et al. (2018)
Negative binomial	Torbic et al. (2010),(Daniels et al., 2011; Miranda-Moreno et al., 2011; Pulugurtha et al., 2013), Elvik et al. (2009), Strauss et al. (2014)
Poisson-lognormal	Siddiqui et al. (2012), Xie et al. (2018)
Zero-inflated Poisson and negative binomial	Lee et al. (2019), Cai et al. (2018)
Gamma model	Daniels et al. (2011)
Random Parameter Negative Binomial	Wali et al. (2018),Wang et al. (2017)
Random Parameter Poisson	Wali et al. (2018)
Mixed-effects negative binomial	Lee et al. (2019)
Multilevel Poissonlognormal (MPLN) joint	(Alarifi et al., 2017)

<u>Poisson regression model</u>

Since crash-frequency data are non-negative integers, applying standard ordinary least-squares regression (which assumes a continuous dependent variable) is not appropriate. The dependent variables are non-negative integers, so most recent research has used the Poisson regression model as an initial starting point (Lord & Mannering, 2010). In the Poisson regression model, the probability of a given road entity ***i*** (Segment/ intersection) having y_i (a non-negative integer) crashes per some period is given by:

$$P(y_i) = \frac{EXP(-\lambda_i)\lambda_i^{y_i}}{y_i!} \qquad (1)$$

Where, $P(y_i)$ is the probability of road segment *i* having y_i crashes over time, and λ_i is the Poisson parameter of roadway entity *i*. λ_i is equal to the expected number of crashes per year $E[y_i]$ of roadway entity *i*. The most common factional form of $\boldsymbol{\lambda_i}$ is $\boldsymbol{\lambda_i = EXP(\beta X_i)}$where X_i is a vector of an explanatory variable of roadway *i* and $\boldsymbol{\beta}$ is an estimable vector parameter(Lord & Mannering, 2010). Although the model was used as an initial starting point for crash analysis in the past, recent studies found that the Poisson regression approach encountered potential problems. One constraint is that the mean must be equal with the variance, i.e., this model can not handle over and under-dispersion crash data(Lord & Mannering, 2010).

<u>The negative binomial (Poisson-gamma) regression model</u>

A new model called the negative binomial (Poisson-gamma) regression model was introduced to overcome possible over-dispersion in the data. The model assumes that the Poisson parameter follows a gamma probability distribution. Rewriting the Poisson parameter as $\boldsymbol{\lambda_i = EXP(\beta X_i + \varepsilon_i)}$**,** where the $\mathbf{EXP(\varepsilon_i)}$ is a gamma-distributed error with mean one and variance α yields the negative binomial model. The added term allow the variance to differ from the mean as $VAR[y_i] = E[y_i][1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2$. Thus, the Poisson model is a special form of the negative binomial regression model when α approaches zero. The parameter α is the over-dispersion parameter(Miaou & Lord, 2003). The negative binomial model is used widely in crash-frequency modeling. However, the model possesses some limitations in handling under-dispersion data and problems regarding dispersion-parameter-estimation when data have a low sample-mean value and small sample size(Lord & Mannering, 2010)

Poisson-lognormal model

Recent researchers used the Poisson-lognormal model as an alternative to the negative binomial model for modeling crash data (Miaou & Lord, 2003). Both models are similar, but the term **EXP(ε_i)** used to compute the Poisson parameter in the Poisson-lognormal model is lognormal rather than gamma-distribution(Lord & Mannering, 2010). Poisson-lognormal model is more flexible than negative binomial; however, model estimation is more complex because Poisson lognormal does not have a closed-form and can still be affected by small sample sizes and small sample-mean values (Miaou & Lord, 2003).

Zero-inflated Poisson and negative binomial

Zero-inflated models were developed to handle data with a considerable amount of zeros or more zeros than a typical Poisson or negative binomial/ Poisson-gamma model would predict. The model operates by splitting the dataset into two, one a crash-free (that handles for the excess zero data which Poisson/ negative binomial models cannot handle to model), and the second a crash-prone propensity of roadway facility. The binary logit or probit model can be used to determine the probability of roadway facilities being in zero or non-zero states (Washington et al., 2010). Despite its broad application to handle data characterized by significant excess zeros, (Lord & Bonneson, 2007) argued this model could not accurately describe the crash-data generation process since the zero or safe state has a long-term mean of zero.

Gamma model

The Gamma model was proposed to analyze crash data exhibited under dispersion (Oh et al., 2006). This model can handle both over-and under-dispersion, and when the variance seems to be equal to the mean of the number of crashes, it reduces to the Poisson model(Lord & Mannering, 2010). This model can perform well statistically. However, it is still a dual-state model with one state having a long-term mean equal to zero(Lord & Mannering, 2010).

2.4 Conclusion

To summarize, as evidenced by previous studies, it was shown that factors related to traffic characteristics (traffic volume on the major and minor roads), traffic control (including major and minor left-turn lanes, right-turn lanes, signalization, the presence of crosswalks, and post speed limit), Geometric characteristics (such as number of approaching legs, intersection skewness, number of lanes both on major and minor roads, the width of the median, lane width on major and minor roads) believed to contribute to the crash at the intersection. Depending on the crash data characteristics and the assumed functional form, a wide range of modeling approaches, (including the Poisson regression model, negative binomial, Poisson lognormal, Zero-inflated negative binomial, and gamma model) have been fitted to crash data to estimate the statistical relationship between the number of crashes and factors that are believed to be (casually) related to crash occurrence. The negative binomial gamma distribution model is a widely used modeling approach that overcomes mainly the problem of over-dispersion in crash data.

3. METHODOLOGY

3.1 Study Area

The study was conducted on 77 signalized intersections in Ghent, Belgium. Ghent is a port city in the northwest part of Belgium, with 469,000 inhabitants in 2021 and an average population density of 1655 h/km² (Statbel,2021). The city economy is highly determined by its port (i.e., the third-largest port in Belgium) and tourism as a result of medieval monuments and architecture (Marlinde Koopmans et al., 2012). Ghent is a vibrant city accommodating local residents, students, commuters, and tourists. Every day, thousands of individuals are on the move, resulting in a significant deal of traffic; this induces an increase in all traffic flow putting greater pressure on public spaces (Stad Gent, 2021). In order to maintain the city's accessibility and livability in the future, the city administration is heavily investing in sustainable mobility to reduce the impact of air pollution and improve road safety (Stad Gent, 2021).

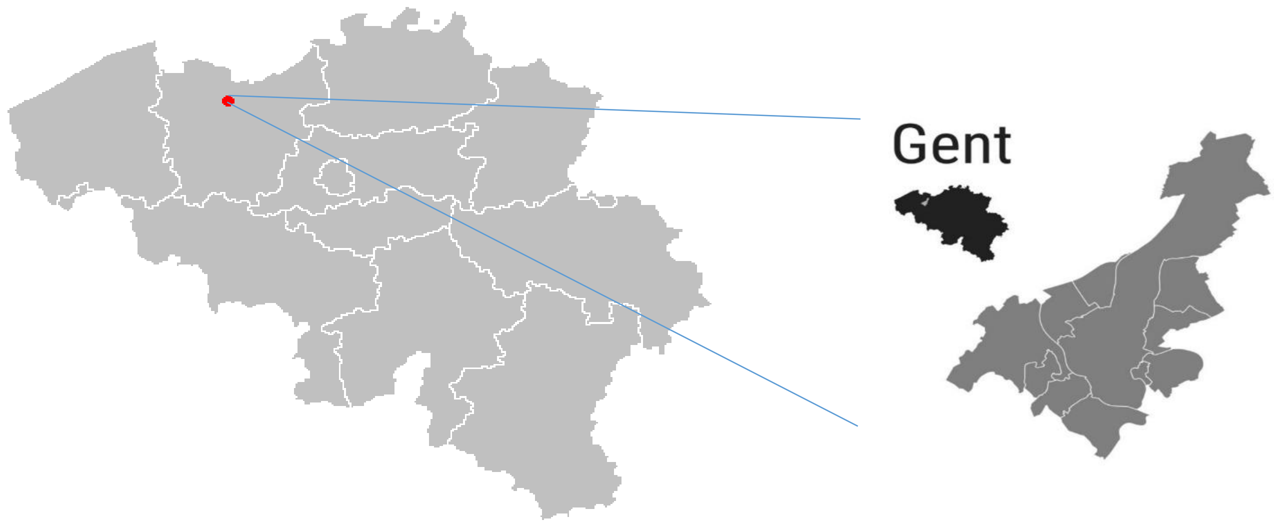

FIGURE 4: Map of Ghent, Belgium (Vectorstock).

3.2 Study Design

Most crash frequency models have traditionally used a cross-sectional data format. Because this format overlooks the long-term relationship between crashes and their contributing factors (Ambros et al., 2018). A cross-sectional study design was used to conduct this study.

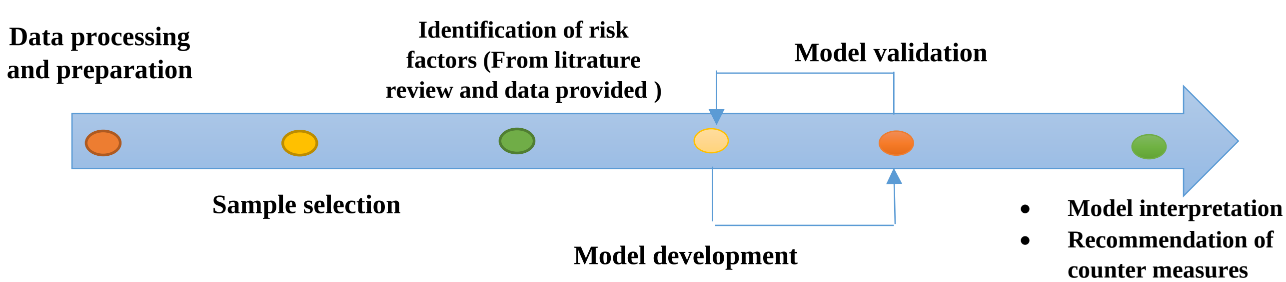

FIGURE 5:Study design flow chart (Author).

3.3 Data Collection and Preparation

This study utilized data provided by Ghent University. The provided dataset consists of four years of crash data (2014-2017), a road network map, and a traffic flow network map. The crash data included information about crash severity, geographical coordinates of the crash location, time and date of the crash, type of the vehicle involved and their number, and driver data. The Crash data categorized traffic crashes into three categories, i.e., (1) traffic crash with injuries (including slight and severe injuries), (2) traffic crash with fatalities, and (3) traffic crash with property damage only (PDO). Between the years 2014 to 2017, in the city, a total of 19,973 crashes were reported. The majority of the reported traffic crashes were Property damage only, i.e., 14,588, followed by 5364 traffic crashes with injury, and 21 traffic crashes with a fatality. The road network map shows all the road networks in the city, including information such as road type, the status of the road, owner of the road, and other road-related information. The traffic flow network map shows the number of vehicles flowing through the road network but not for the entire road network. The traffic flow count was based on three types of vehicles, i.e., passengers car, light trucks, and heavy trucks, either in one or two directions. Thus, as the first step, data processing and preparation have to be done to prepare the data required for analysis.

<u>Step 1: Data processing and preparation</u>

The entire data processing step was done using QGIS software. Since traffic flow volume is identified as a major predictor of crashes, as shown in chapter two, the traffic flow map was used as a base map to start with data processing. Then, by overlaying the traffic flow network map on the road network map, the entire road network was filtered into a new layer by excluding roads that do not have traffic flow data. Then, the new road network map with only traffic volume data was used for further process. Initially, the road network map doesn't contain any information regarding the location of intersections and intersection related characteristics. Thus, the researcher has to look for a method to identify the intersection on the road network map (i.e., the new road network map with traffic volume data). The researcher considered the AASHTO highway safety manual definition for intersections which defines an intersection as "*the general area in the road network where two or more roads join or cross and the area includes the roadway and roadside facilities for traffic movements.*" as initial thought to identify intersections on the road network, i.e., when two lines are intersecting or crossing there is a chance of that location to be an intersection on the road network, then with the help of vector analysis line intersecting tool, lines intersecting each other were identified and marked on the road network map. The road network map was made by joining line segments; as a result, the software identified two lines joining each other as intersecting lines. Thus, the researcher tries to locate and mark only lines that intersect or cross each other. Then, With the help of Google satellite plugin in QGIS, the point of intersection was checked if it is an actual intersection or not; this was done by overlaying the road network map with the marked point of intersection on the google satellite image layer. Then, after once points were identified as intersections, the researcher further checked again if the identified intersections had complete traffic flow data in all legs or not, and those intersections with no traffic flow data at least in one leg were excluded. Then with this process, the researcher managed to locate 266 intersections on the road network. Then, the next step was to identify intersections with signalization. Google map and google road with Google 3D view were used to identify signalized intersections. Then, with this process, 77 signalized intersections out of the 266 intersections were identified as signalized intersections.

Once signalized intersections were identified, the next data processing step was assigning crashes into an intersection and intersection-related crashes. Although there is no clear definition of an intersection-related crash, different researchers use different criteria to define intersection-related crashes based on intersection influence area. Harwood et al. (2002) considered all intersection-related crashes within 250 ft of each intersection to evaluate the safety effectiveness of providing dedicated left- and right-turn lanes for at-grade intersections. A study by Joksch and Kostyniuk (1998) to examine the relationship between crash counts and traffic volume at intersections in the states of Michigan, California, and Minnesota using a maximum of 350 ft length of influence area. Mitra et al. (2007) study to examine spatial variables' effect on intersection crash occurrence considered intersection-related crashes by including the intersection's influence area within 250 ft long from the center of intersection along any leg of the intersection. Thus, this study decided to use the HSM guidelines to define intersection-related crashes by using an intersection influence area of 70m length from the center of the intersection along all the legs. Accordingly, all crashes that have occurred within the boundary of an intersection area (A) are designated as intersection crashes. On the other hand, crashes on the road segment (B) were considered only as intersection related crashes if it falls within the influence area of the intersection and when there was no other road facility within the influence area (for example, there was a situation where a neighboring unsignalized intersection located within the influence area of a signalized intersection, so in that case crashes located nearby to the unsignalized intersection were excluded from consideration as intersection related crashes for the signalized intersection under consideration). In a case where two signalized intersections were situated closely, and overlapping of intersection influence area occurred, intersection related crashes were assigned to the nearby intersection.

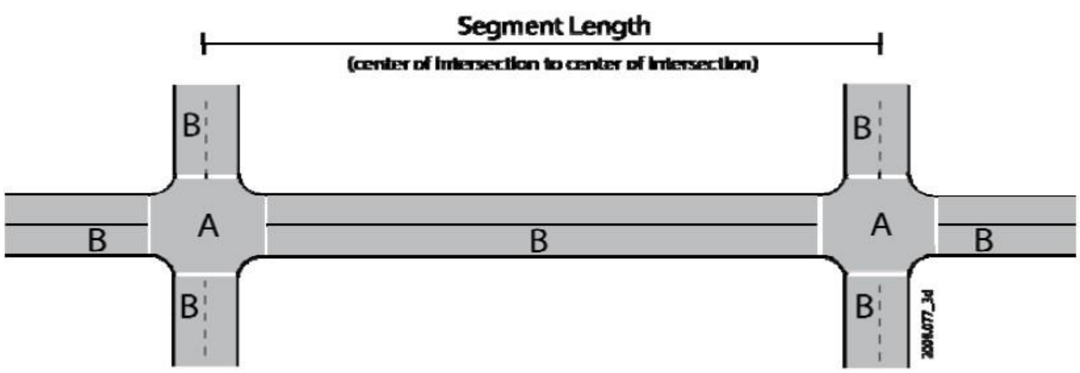

FIGURE 6:Definition of Roadway Segments and Intersections (AASHTO, 2010).

Thus, with the help of QGIS software, a 70m buffer zone was drawn as an intersection influence area in each signalized intersection on the road network map. Then, crash data was overlaid on the road network map with intersection buffer zone, and crashes located outside of the buffer zone were excluded. Only crashes situated within the buffer zone were designated as intersection and intersection-related crashes. With this process, the researcher examined each signalized intersection to see how crashes situated inside the intersection influence area. In some signalized intersections, crashes were situated neither in boundary (A) nor region (B); even in some cases, crashes were situated inside resident buildings. Thus, in cases like this, crashes were excluded from consideration. A total of 1338 crashes were assigned as intersection and intersection-related crashes as a result of the above definitions and considerations. Of which 748 crashes were property only damage, 586 were traffic crashes with injuries, and 4 traffic crashes with fatalities.

Traffic volume data

In this stage, the traffic flow network map was processed. Only mid-block traffic flow counts along all the intersection legs were considered. Traffic counts on the rest of the road network were excluded. Traffic flow counts were available in one or two directions of the road network, depending on the type of road. The traffic count data consisted of counts of three types of vehicles, i.e., passenger cars, light trucks, and heavy trucks. As a first step, the total number of vehicles was determined by summing up the three types of vehicles (total number of vehicle=passenger cars + light truck vehicles + heavy truck vehicles). And then, the total number of vehicles count was divided by 365 in order to convert the traffic count into Average Annual Daily Traffic (AADT). The calculated AADT was only in one direction. For the roads with two directions flow, the total AADT was calculated by adding the AADT in the respective directions. Then the calculated AADTs were assigned to the corresponding intersections. In cases where the AADTs on the two major and minor road legs of a four-leg intersection and the two major road legs of a three-leg intersection are different, the AASHTO's highway safety manual recommendations were used. According to the HSM (2010), if the AADTs on the two major-road legs of an intersection differ, the larger of the two AADT value is used. If the AADTs on the two minor-road legs of a four-leg intersection differ, then the larger of the two AADT value is used. For three-leg intersections, the AADT of the single minor-road leg is used.

3.4 Study Population

Step 2: Sample selection

The current study purposely selected all signalized intersections identified in previous steps. A total of 77 signalized intersections were identified, of which 39 were four-leg and 38 three-leg.

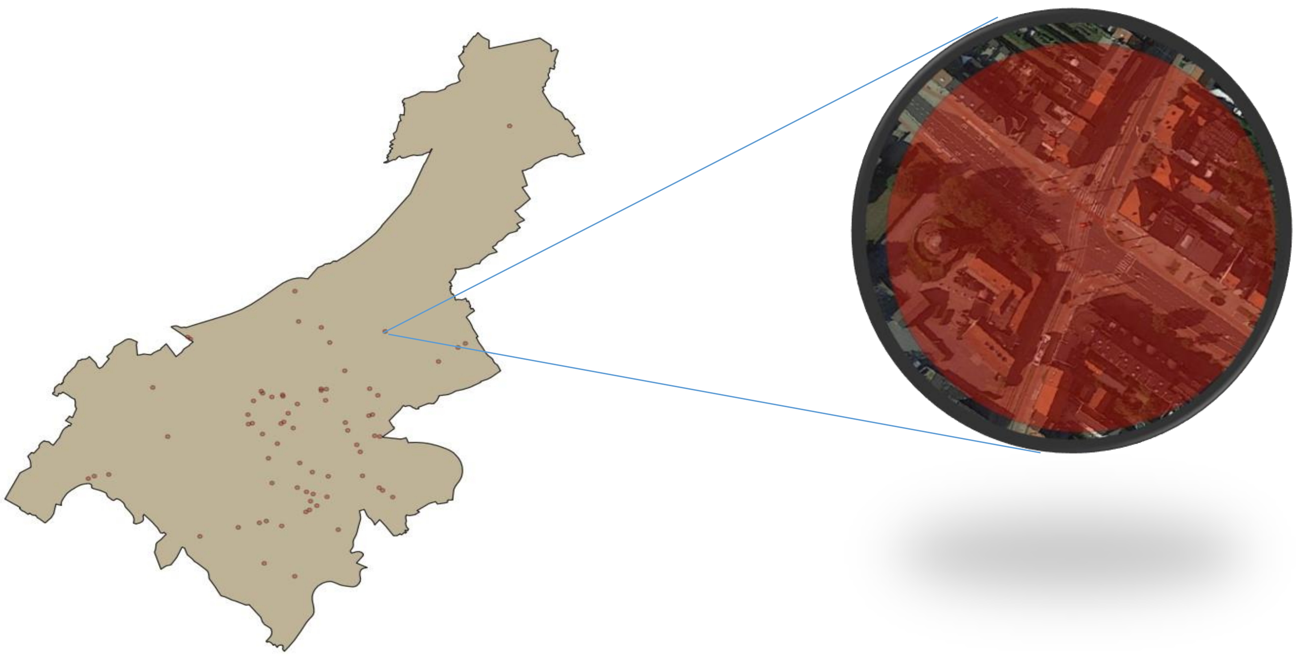

FIGURE 7: Signalized intersections (Author).

3.5 Identification of Risk Factors

Step 3: Identifying possible risk factors/variables

Three main criteria were considered to select variables based on the highway safety manual (2010).

1. Variables that have found in previous studies to wield a major influence on the number of crashes.
2. The variable included can be measured validly and reliably.
3. The variables should not be highly correlated with other variables included in the model. (This requirement will be checked in step 4).

Based on the previous literature review and the data given, ten variables that qualified the first two requirements were selected under two main categories; exposure variables (traffic flow) and intersection characteristics variables (explanatory variables, traffic control variables). The exposure variable includes the AADT on major and minor approaches, and eight variables describing the characteristics of the intersection were selected. The variables include; average lane width on major and minor approaches, left-turn lane on major approaches, right-turn lane on minor approaches, the presence of crosswalks on major and minor approaches, number of approaches/legs, and intersection skewness. Intersections were divided into skewed and un-skewed intersections depending on the angle formed when two or more roads meet. Intersections with an angle other than 90 degrees were considered as skewed intersections.
Intersection characteristics data (i.e., explanatory variables, traffic control variables) were collected extensively using Google satellite image, Google Maps and road, and google earth pro software. A summary of the selected variables are shown in the table below.

TABLE 3: Summary of variables description

Category	Variables	Unit
Exposure variables	AADT on major approaches	AADT
	AADT on minor approaches	AADT
Intersection characteristics variables	Average lane width on major approaches	Meter
	Average lane width on minor approaches	Meter
	Left-turn lane on the major approach	0 to 2
	Right-turn lane on the minor approach	0 to 2
	Presence of crosswalks on major approaches	0 to 2
	Presence of crosswalks on minor approaches	0 to 2
	Number of legs	3 to 4
	Intersection skewness	Yes/no

A summary and description of the variables (Level) selected for the vehicle crash prediction model for signalized intersections in this study are presented in appendix A.

3.6 Model Development

Step 4: Model development
Correlation

The Highway safety manual (2010) states that variables exhibiting high correlation within explanatory variables should be avoided from the model. Thus, it is essential to check for correlation among explanatory variables to identify important contributing factors (Miranda-Moreno et al., 2011). The Pearson coefficient of correlation and the variance inflation factor (VIF) were utilized to avoid the problem of multicollinearity. The Pearson coefficient of correlation determines whether two variables are correlated or not. If the Pearson correlation coefficient is between -0.3 and +0.3, the variables have a weak correlation(Pulugurtha et al., 2012). When two variables are identified to have a high correlation, the one with the weakest significance for crash involvement was excluded from the model. VIF, on the other hand, was utilized to identify collinearity among variables in the fitted model. A VIF value greater than ten implies significant multicollinearity problems among variables (Chen et al., 2016). The variable with the highest value was eliminated from the model if the VIF value was greater. SPSS version 28 software was used to conduct the analysis.

Modeling approach and model development

In this stage, the predictive model was developed. Crash prediction models are multivariate models that relate the number of crashes with exposure (Traffic flow) and explanatory variables (road geometry, land use, traffic control) (AASHTO, 2010). Since crash occurrence is described as a random, discrete and non-negative event, the Poisson regression model appears to be more suitable. Still, this model has a strict limitation, i.e., the mean must be equal to the variance, which is rare in the case of traffic crash data. Previous studies identified that crash data has a significant over-dispersion, i.e., the variance of the data is usually greater than the mean (Lord & Mannering, 2010). The Highway Safety Manual (2010) recommends the generalized linear model (GLM) with the negative binomial distribution and logarithmic link function as a standard approach to model yearly crash frequencies. The GLM with the negative binomial distribution provides a relatively reliable distribution for studying random, discrete, and non-negative events such as crashes(Khattak et al., 2021). In NB (negative binomial) regression model, the probability of roadway entity *i* having y_i crashes per period of time is defined as;

$$P(y_i) = \frac{EXP(-\lambda_i)\lambda_i^{y_i}}{y_i!} \quad (2)$$

$$\lambda_i = EXP(\beta X_i + \varepsilon_i) \quad (3)$$

Where;

y_i = the number of crashes for segment *i* in year t

β = a vector of the estimable parameters

Xi = a vector of the explanatory variables

EXP(ε_i) = a gamma-distributed error term with mean 1 and variance α.

The GLM uses both power and exponential functions for exposure variables and risk factors, respectively. The power function ensures a non-zero positive crash number unless the exposure variable (traffic volume) is zero. On the other hand, the exponential function ensures a non-zero or negative crash number due to zero or negative values from the linear predictors (regression of risk factors), even if the sum of the linear function results in a zero or negative value, it will not result in a zero or negative crash value since the linear function is the exponent of the exponential function, (i.e., $e^{\sum y_i x_i}$) with value e^0 being one and e^{-x} Non-negative number. Thus, the generalized linear model (GLM) with the negative binomial distribution and logarithmic link function (equation 3) was used in this study to develop the crash prediction model for signalized intersections. SPSS software version 28 was used to fit the model.

$$\mu_i = \beta_0 * Q_{MA}^{\beta_1} * Q_{MI}^{\beta_2} * e^{\sum \beta_i x_i} \tag{3}$$

Where μ_i = predicted number of crashes at intersection type *i*

Q_{MA}= number of vehicles entering an intersection from the major road

Q_{MI}= number of vehicles entering an intersection from the minor road

x_i= vector value of risk factor *i* other than the number of vehicles

β_0= intercept

β_1 *and* β_2= The effect of traffic volume on the expected number of crashes (elasticity)

β_i= parameter to be estimated and represent the effect of risk factor *i*

Generally, separate models for different intersection types and crash types are often suggested than one model for all intersection types since they provide a better fit and description of the data. It is recommended to fit disaggregated models than aggregated models if data on intersection types and crash types are available(Eenink et al., 2008). Therefore, this study intended to develop separate models for total injuries (injuries and fatality) and property damage only (PDO) for three and four-legged signalized intersections.

3.7 Model Validation and Goodness of Fit

Step 5: Model evaluation

Since the main objective of the study is to develop a crash prediction model that will realistically estimate the crash frequency at the signalized intersection, the model's goodness of fit and its statistical adequacy has to be checked. Different methods were used to assess the goodness of fit for predictive models. In this study, Mean absolute deviation (MAD), Mean squared prediction error (MSPE), and Mean prediction bias (MPB) was used to validate the goodness of fit of the developed model see (Oh et al., 2003). Mean absolute deviation (MAD) provides a measure of the average misprediction of the model, a value close to zero suggests that, on average, the model predicts the observed data well. On the other hand, Mean squared prediction error (MSPE) was used to assess the error associated with divation. MAD is given by;

$$MAD = \frac{\sum_{i=1}^{n} |\hat{\mu} - y_i|}{n} \tag{4}$$

Where: $\hat{\mu}$ = Predicted number of crashes per year for the site *i*

y_i= observed number of crashes per year for the site *i*

n= number of sites

In addition to MAD and MSPE, the mean prediction bise (MPB) was used to determine the magnitude and the direction of average model bias. A value close to zero suggests that, on average, the model predicts the observed data well. On the other hand, a positive value indicates that the model overestimates the observed number of crashes and a negative value indicates an underestimation of the observed number of crashes. MPB is given by;

$$MPB = \frac{\sum_{i=1}^{n}(\hat{\mu} - y_i)}{n} \tag{5}$$

Where: $\hat{\mu}$ is the fitted value of y_i With sample size n.

Step 6: Model interpretation and recommendation

Based on the estimated values found from the model, major risk factors were identified with their respective influence on the number of crashes at the signalized intersections. And finally, the results were interpreted, and the coefficients of explanatory variables were compared with similar previous studies on crash prediction models at intersections. Following the identification of the main risk factors, the relevant countermeasures and recommendations to improve traffic safety at signalized intersections were forwarded.

4. DATA ANALYSIS

This chapter discusses about variables considered in this study, the correlation between the variables, and model development.

4.2 Descriptive Analysis

The mean, variance, minimum, maximum, and standard deviation values of the identified variables are summarized in tables (4) below. A total of 77 signalized intersections were considered in the study.

TABLE 4: Descriptive statistics of the Response variable, Exposure variables and Intersection characteristics variables (n=77)

Category	Variables	Minimum	Maximum	Mean	Std. Deviation
A. Response variables	Total crash	0	15	5.04	3.401
B. Exposure variables	AADT on the major approach	5093	35,738	14,802	6566.132
	AADT on the minor approach	6	31,578	8880	6307.587
C. Intersection characteristics variables	Average lane width on the major approach	2.78	3.81	3.107	.171
	Average lane width on the minor approach	1.98	3.39	2.974	.229
	Left turn lane on the major approach	On each side: 37 On one side: 15 No dedicated lane: 25			
	Right turn lane on the minor approach	On each side: 23 On one side: 10 No dedicated lane: 44			
	Presence of crosswalks on the major approach	On each side: 49 On one side: 15 No crosswalk: 13			
	Presence of crosswalks on the minor approach	On each side: 35 On one side:29 No crosswalk: 13			
	Number of legs/approaches	Four legs:39 Three legs: 38			
	Intersection skewness	yes: 35 No: 42			

4.2 Exploratory Data Analysis

4.2.1 Correlation

A Pearson correlation analysis was done to check for correlation between the independent variables. A Pearson coefficient value of 0.5 (moderate relationship) were used as a threshold. Thus, for variables with a Pearson coefficient value above 0.5, one variable with a better significance in predicting vehicle crashes was selected. A summary of the Pearson correlation results between the independent variables are shown in table 5.

Based on the Pearson correlation results, a significant correlation was observed between the Annual Average Daily Traffic (AADT) on the major and minor approaches with a Pearson coefficient value of 0.798. A significant correlation was also observed between the left turn lane on major approach and the right turn lane on the minor approach with a Pearson coefficient value of 0.820. And a significant correlation was observed between the right turn lane on the minor approach and the number of approaches with a Pearson coefficient value of -0.613.

In addition to the Pearson coefficient, the variance inflation factor (VIF) was used to check for multicollinearity within the independent variables. The collinearity analysis showed that a maximum variance inflation factor (VIF) value of 4.121, which is less than 10, implies no significant multicollinearity problems among the variables (Chen et al., 2016). A summary of the collinearity analysis is shown in appendix B. Finally, based on Pearson's correlation and collinearity analysis results, all variables except the right turn lane on the minor approach were selected for the model development.

TABLE 5: Pearson correlation

Correlations											
		AADT_maj	**AADT_min**	**ALW_maj**	**ALW_min**	**LTL_maj**	**RTL_min**	**CW_maj**	**CW_min**	**Nº approaches**	**Skewness**
AADT_maj	Pearson C.	--									
AADT_min	Pearson C.	**.794****	--								
	Sig.	<.001									
ALW_maj	Pearson C.	.056	.099	--							
	Sig.	.631	.390								
ALW_min	Pearson C.	.177	.176	.221	--						
	Sig.	.123	.125	.053							
LTL_maj	Pearson C.	-.161	-.218	-.010	.122	--					
	Sig.	.162	.056	.929	.292						
RTL_min	Pearson C.	-.131	-.112	-.025	.135	**.820****	--				
	Sig.	.257	.332	.830	.242	<.001					
CW_maj	Pearson C.	-.022	.033	-.071	-.136	-.084	-.050	--			
	Sig.	.851	.775	.538	.237	.466	.665				
CW_min	Pearson C.	-.271*	-.124	-.091	-.135	-.190	-.117	.475**	--		
	Sig.	.017	.283	.430	.240	.097	.310	<.001			
Nº approaches	Pearson C	.256*	.278*	-.002	-.132	-.070	**-.613****	.162	.117	--	
	Sig.	.024	.014	.986	.254	<.001	<.001	.160	.311		
Skewness	Pearson C.	.196	.161	-.052	.022	-.016	-.020	.158	.214	.119	--
	Sig.	.087	.161	.655	.852	.889	.863	.170	.062	.304	

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

4.3 Negative Binomial Model Development

In this section, a summary of the results from the Negative Binomial GLM were discussed. As explained in the methodology, the researcher intended to develop two separate models for the two types of crashes, i.e., Property Damage Only crashes (PDO) and Injury and Fatality crashes (IF). But unfortunately, model results for Injury and Fatality crashes showed statistically insignificant results. Both Poisson and Negative binomial GLM were used separately to develop a model for the Injury and Fatality crashes, but the model results showed statistically insignificant results in both cases. The SPSS model output results are summarized in Appendix C. As a result, instead of disregarding injury and fatality crash data from the dataset, the researcher decided to combine the two types of crashes, i.e., (Total crash= Property Damage Only crashes+ Injury and Fatality crashes) and tried to fit a single model for the total crash. Only 80% (i.e., 61) of signalized intersections were used to develop the fully specified crash prediction models. And the remaining 20% (i.e., 16) signalized intersections were used for model validation.

4.3.1 Negative Binomial Model Development For Total Crash

Before starting model development, the distribution of the response variable data was first checked. This helps in deciding to choose which model to use. The analysis result showed that the response variable (Total crash) has a higher value of variance (11.564) than the mean (5.04); this implies there is an overdispersion in the response variable. Various literature recommended to assume negative binomial distribution for count data with overdispersion. Thus, the negative binomial GLM model with loglink was used for the model fitting.

The modeling process was started by recalling the correlation and collinearity analysis results for the variables. As indicated in table 5, a high correlation was found between AADT on the major approach and AADT on the minor approach, between the left-turn lane on the major approach and right-turn lane on the minor approach, and between right turn lane and the number of approaches. Thus, considering the significance of the AADT on the minor approach for the response variable, the researcher agreed to combine the AADT on the major and minor approaches using various functional forms rather than discarding this variable, and the variable right turn lane on the minor approach was excluded. Since there were no significant correlations and collinearity problems between other variables, all the remaining variables were considered for model development.

Before directly going to a multivariate model fit, the researcher agreed to develop simple models incorporating only the exposure variable (i.e., AADT on major and minor approaches). Simple models have wide applications considering the transferability of crash prediction models for other regions. Miaou and Lord (2003) used different functional forms to combine the average annual daily traffic on the major and minor approaches. According to Miaou and Lord (2003), four functional forms are considered as the most popular forms to combine the average annual daily traffic on the major and minor approaches. These are;

FF_1: $\ln\mu = \beta_0 + \beta_1 \ln(Q_{MA} + Q_{MI})$ (6)

FF_2: $\ln\mu = \beta_0 + \beta_1 \ln(Q_{MA}) + \beta_2 \ln(Q_{MI})$ (7)

FF_3: $\ln\mu = \beta_0 + \beta_1 \ln(Q_{MA} * Q_{MI})$ (8)

FF_4: $\ln\mu = \beta_0 + \beta_1 \ln(Q_{MA} + Q_{MI}) + \beta_2 \ln(\frac{Q_{MI}}{Q_{MA}})$ (9)

Where: Q_{MA}= number of vehicles entering an intersection from the major road

Q_{MI} = number of vehicles entering an intersection from the minor road

Thus, simple models were developed using the four functional forms (FF) as a first step toward model development. All signalized intersections were included for the development of the simple models. The SPSS model estimation results for the four functional forms are shown in the table below.

TABEL 6: Model estimation results for the four functional forms (simple models, n=77)

	FF_1		FF_2		FF_3		FF_4	
	β	***p-value***	**β**	***p-value***	**β**	***p-value***	**β**	***p-value***
Intercept	-1.782	0.172	-2.891	0.065	-0.442	0.626	-2.146	0.153
$\ln(Q_{MA} + Q_{MI})$	0.340	0.009					0.374	0.011
$\ln(Q_{MA})$			0.468	0.016				
$\ln(Q_{MI})$			0.004	0.960				
$\ln(Q_{MA} * Q_{MI})$					0.112	0.023		
$\ln(\frac{Q_{MI}}{Q_{MA}})$							-0.040	0.627
Dispersion	0.198		0.187		0.207		0.197	
Log Likelihood	-191.043		**-190.020**		-191.757		-190.924	
AIC	388.086		**388.040**		389.513		389.848	
BIC	**395.117**		397.416		396.545		399.223	

From the four simple models, to select a functional form with better performance, measures of Goodness of Fit and other model performance evaluation measures were checked for the simple models—table 7 below shows the model performance evaluation measures used in this study. The bolded values are the desirable values.

TABEL 7: Measures of model performance (Goodness of Fit) for the simple models

	FF_1	FF_2	FF_3	FF_4
Deviance	86.248	86.296	**86.040**	86.164
MAD	2.4285	**2.4025**	2.4805	2.4155
MSPE	10.3246	**10.1428**	10.7662	10.2077
MPB	-0.0129	0.0389	**-0.0129**	0.0259
Calibration Factor	0.9948	0.9923	1.0026	**1.0025**

Based on the Goodness of fit and other model performance evaluation measures, almost similar results were observed between the four functional forms**.** Thus, the researcher decided to select all the functional forms for further modeling process. The researcher also believes choosing all the four functional forms instead of a single form will help to examine whether the functional forms would continue to perform similarly in the presence of other covariates or if a single best model would be revealed.

Only 80% (i.e., 61) of signalized intersections were used to develop the fully specified crash prediction models. And the remaining 20% (i.e., 16) signalized intersections were used for model validation. Initially, all the variables(covariates) except the right turn lane on the minor approach were used simultaneously to develop a Negative Binomial Generalized Linear Model (NBGLM). However, most of the explanatory variables were found to be insignificant. This could be because of the variation in variable level between the independent variables or because of the small size of the dataset. As a result, the forward selection process was used to fit the Generalized Linear Models. As a first step, covariates were separately used for model development. From this step, only covariates such as (AADT, left turn lane on the major approach, the presence of crosswalk on both approaches, the number of legs/approaches, and skewness) were found to be significant. And then, as a second step, only those significant covariates from step one were used simultaneously. From this step, only covariates such as AADT, left turn lane on the major approach, the presence of crosswalk on the minor approach, and the number of legs/approaches were found to be significant. Then, in step three, only significant covariates from step two were combined with the insignificant covariates from step one. And then, in step four, only significant covariates from step three were incorporated again with insignificant covariates from step two. This process continued until the best model fit was found. Then, in the end, only five variables were found significant for signalized intersections at 5% confidence levels.

A summary of parameter estimation for the variables (covariates) by the four functional forms are shown in table (8) below. From the table below it can be clearly shown that coefficient estimations across all the functional forms are almost similar for the explanatory variables.

TABEL 8:Model estimation results for the four functional forms (full models, n=61)

		FF_1		**FF_2**		**FF_3**		**FF_4**	
		β	***p-value***	**β**	***p-value***	**β**	***p-value***	**β**	***p-value***
Intercept		-2.125	0.107	-4.297	0.006	-0.276	0.784	-3.712	0.011
$\mathbf{ln}(\boldsymbol{Q_{MA}} + \boldsymbol{Q_{MI}})$		0.314	0.013					0.451	<0.001
$\mathbf{ln}(\boldsymbol{Q_{MA}})$				0.629	<0.001				
$\mathbf{ln}(\boldsymbol{Q_{MI}})$				-0.092	0.146				
$\mathbf{ln}(\boldsymbol{Q_{MA}} * \boldsymbol{Q_{MI}})$						0.073	0.162		
$\mathbf{ln}(\frac{\boldsymbol{Q_{MI}}}{\boldsymbol{Q_{MA}}})$								-0.150	0.023
LTL_maj	LTL_maj=2	-0.549	0.010	-0.569	0.007	-0.538	0.013	-0.578	0.006
Base LTL_maj=0	LTL_maj=1	0.043	0.829	0.076	0.699	0.018	0.931	0.068	0.730
CW_min	CW_min=2	0.360	0.061	0.482	0.014	0.301	0.125	0.466	0.018
Base CW_min=0	CW_Min=1	0.301	0.405	0.468	0.194	0.230	0.535	0.446	0.216
No. Of Legs Base three legs=0	No. Legs=4	0.606	0.001	0.603	0.001	0.616	0.001	0.625	<0.001
Dispersion		0.013		3.746E-9		0.027		3.613E-8	
Log-Likelihood		-131.049		**-128.258**		-132.920		-128.935	
AIC		278.099		**274.516**		281.938		275.869	
BIC		294.986		**293.514**		298.726		294.867	

In order to choose the fully specified crash prediction model that best fits the data, Goodness of fit and other model performance evaluation measures were calculated for each model. Table 9 shows the model performance evaluation measures used in this study. Thus, according to the Goodness of fit and other model performance evaluation measures from table 8 and 9, it is clear that the model with functional form 4 (FF-4) showed the best fit over the other models. The bolded values are the desirable value.

TABEL 9: Measures of model performance (Goodness of Fit) for the fully specified models

	FF1	**FF2**	**FF3**	**FF4**
Deviance	67.858	67.458	68.002	**66.105**
MAD	**2.4590**	2.6393	2.4918	2.6065
MSPE	**10.7868**	11.9508	10.9180	11.4262
MPB	-0.6229	-0.6721	-0.6229	**-0.6065**
Calibration factor	1.1258	1.1371	1.1258	**1.1221**

Once the model with the best fit was identified, interpretation and discussions were given according to the results obtained from the best fit model (i.e., FF-4). Accordingly, the best model fit incorporates the exposure variable (AADT) and three intersection characteristics variables (i.e., left-turn lane on the major approach, Presence of crosswalk on the minor approach and the number of legs/approaches).

Based on the fitted NBGLM, a positive and significant relationship was found between the expected crash frequency and the total number of vehicles on the major and minor approaches. On the other hand, a negative and significant association was found between the expected crash frequency and the ratio of the number of vehicles on the minor approach to the number of vehicles on the major approach (i.e., Q_{mi}/Q_{ma}), but this needs interpretation. When the proportion between the traffic volume on the minor approaches to the traffic volume on the major approaches increases, the expected crashes will decrease. This means, for a given signalized intersection, when the traffic volume is highly concentrated on the major approach only, there is a higher chance of the vehicles being involved in a crash compared to the same signalized intersection with proportional distribution of traffic volume on major and minor approaches. A simulation was done for a random signalized intersection from the sample size to clarify this interpretation see *Appendix E*.

The relationship between the "left turn lane on the major approach" and the expected crash was found negative and significant when the left turn lanes are provided on each side of the major road. Compared to a signalized intersection without a left-turn lane, the logarithmic expected crash frequency in signalized intersection with Left turn lane on each side of the major approach will be less by a factor of 0.578. Signalized intersections with only one side left turn lane on the major approach were found insignificant compared to signalized intersections without a left-turn lane. Based on the research findings, providing a left-turn lane on each side of the major approach will decrease the expected crash at signalized intersections.

The relationship between the presence of crosswalks on the minor approach and expected crash frequency for signalized intersections was found positive and significant. Compared to a signalized intersection without marked crosswalks, the logarithmic expected crash frequency in signalized intersection with crosswalks along each side of the minor road will be greater by a factor of 0.466.

The association between the number of legs/approaches and the expected crash frequency was found to be positive and significant. The logarithmic expected crash frequency for four-leg signalized intersections will be greater by a factor of 0.625 than for three-leg signalized intersections.

TABLE 10: Coefficient estimates detailed (model-4, FF-4)

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-3.712	1.4530	-6.560	-.864	6.526	1	.011
Ln(Q_{maj}+Q_{min})	.451	.1352	.186	.716	11.126	1	<.001
Ln(Q_{min}/Q_{maj})	-.150	.0656	-.278	-.021	5.204	1	.023
[Left turn lane on the major approach =2]	-.578	.2107	-.991	-.165	7.531	1	.006
[Left turn lane on the major approach =1]	.068	.1968	-.318	.454	.120	1	.730
[Left turn lane on the major approach =0]	0[a]	.	.	.	.	.	.
[Presence of crosswalks on the minor approach=2]	.466	.1967	.080	.851	5.604	1	.018
[Presence of crosswalks on the minor approach=1]	.446	.3602	-.260	1.152	1.531	1	.216
[Presence of crosswalks on the minor approach=0]	0[a]	.	.	.	.	.	.
[Number of legs/approaches =1]	.625	.1864	.259	.990	11.235	1	<.001
[Number of legs/approaches =0]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						
(Negative binomial)	3.613E-8[c]	.	.	.			

The model outperformed the simple model of $\ln \mu = \beta_0 \ + \ \beta_1 \ln(Q_{MA} + Q_{MI}) + \beta_2 \ln(\frac{Q_{MI}}{Q_{MA}})$. Moreover, the likelihood ratio chi-square test indicates that the model was a significant improvement in fit over a null (no predictor) model (P<.001). The full SPSS model output results are shown in *Appendix D*. The transformed form of the model is shown as follows:

Best fit model: (Model 4)

ln(Total number of crashes)$_i$ =(-3.712)+0.451*ln(*AADT on the major approach + AADT on the minor approach*) – 0.15 *ln(*AADT on the minor approach/ AADT on the major approach*) -0.578 ∗ (*Left turn lane on the major approach*) + 466∗ (*Presence of crosswalks on the minor approach*) + 0.625∗ (*Number of legs/approaches*)

4.4 Model validation

Model validation was done using the best fit model (**Model-4**) for the remaining 16 signalized intersections. Accordingly, number of predicted crashes were calculated using the equation from model-4 for each of the 16 signalized intersections. And then, by comparing the observed crash and the calculated predicted crash of each intersection, the measures of Goodness of Fit were calculated to validate the model performance. Table 11 below summarizes the results of the Goodness of fit measures. According to the HSM, calibration factors are calculated to see the relative performance of the models. The calibration factor is defined as the ratio of the sum of the observed number of crashes to the sum of the predicted expected number of crashes for all sites. A calibration factor with the value of unity shows that the model performance is satisfactory for the given dataset.

Model 4 (FF-4)

$$\ln \mu = (-3.712) + 0.451 \ln(Q_{MA} + Q_{MI}) - 0.15 \ln(\frac{Q_{MI}}{Q_{MA}}) - 0.578(LTL_{maj}) + 0.466(CW_{min}) + 0.625(\text{No Approaches})$$

TABLE 11: Measures of model performance (Goodness of Fit)

MAD	1.8057
MSPE	6.1685
MPB	0.2299
Calibration Factor	0.9162

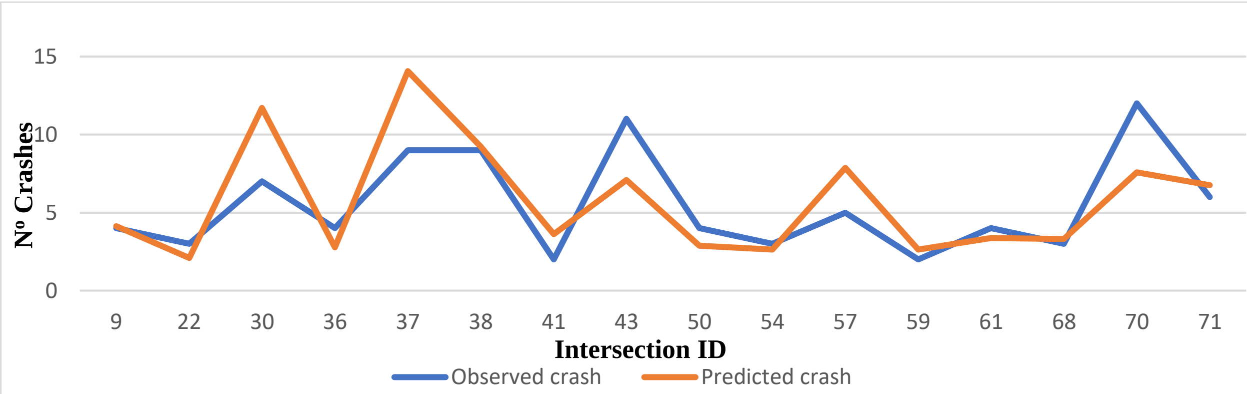

FIGURE 8: Model validation

5. DISCUSSION

This chapter discusses the key research findings of the study in the context of the research questions. Recalling the analysis result from chapter four, it was observed that there was an overdispersion in the response variable. Previous works of literature recommended to assume negative binomial distribution for count data with overdispersion. Thus, the negative binomial GLM model with loglink was used for the model fitting in this study.

According to the study results (Best model fit), factors related to traffic characteristics (the sum of traffic volume on the major and minor approaches and the ratio of traffic volume on minor approach to major approach), traffic control (left-turn lane on the major approach, and the presence of crosswalk on the major approach), Geometric characteristics (number of approaches) found to be key predictors of crashes at the signalized intersection in the city of Ghent, Belgium.

The sum of the traffic volume on the major and minor approaches have a positive association with the expected crash frequency at the signalized intersection. This means that for a one percent increase in the sum of traffic volume on both approaches, the expected crash increases by a factor of 0.629. This was expected as it was also reported in previous studies by (Alarifi et al., 2017; Barbosa et al., 2014; Gomes et al., 2012; Khattak et al., 2021; Wang et al., 2020). On the other hand, the ratio of traffic volume on minor approach to the major approach found to have a negative association with the expected crashes at the signalized intersection. This implies that for a given signalized intersection, when there is a proportional traffic flow in the major and minor approaches, the expected crash will be reduced (i.e., for a one percent increase in the ratio traffic volume on the minor approach to the major approach the expected crash frequency decrease by a factor of 0.569).

The research findings also demonstrate that when left turn lanes are provided in both directions of the major approach, the number of expected crash occurrences at the signalized intersection will be reduced. Compared to a signalized intersection without a left-turn lane, the logarithmic expected crash frequency in signalized intersection with Left turn lane on each side of the major approach will be less by a factor of 0.578. This was not a surprise since various previous studies (De Pauw et al., 2015; Gluck et al., 1999; Harwood et al., 1995; Zhou et al., 2010) also reported a negative association between Left-turn lane and the expected crash frequency.

The presence of a crosswalk on the minor approach has a positive relationship with the expected crash frequency at the signalized intersection when provided in both direction of the minor approach. In comparison to a signalized intersection without marked crosswalks, the logarithmic expected crash frequency in signalized intersection with crosswalks along each side of the minor road will be greater by a factor of 0.466. Previous studies by (Harwood et al., 2002; Herms, 1970; Khattak et al., 2021; Smith & Knoblauch, 1987) also report a positive relationship between the presence of crosswalks and crash frequency at signalized intersections.

The association between the number of legs/approaches and the expected crash frequency was found to be positive. The logarithmic expected crash frequency for four-leg signalized intersections will be greater by a factor of 0.625 than for three-leg signalized intersections. According to research findings in this study, four-leg intersections can increase the expected crash frequency by around 60% compared to three-leg intersections. This finding was logical since four-leg intersections have more conflict points than three-leg intersections, which means there are more chances for a crash to happen in four-leg intersections. This was not a surprise compared with the broad agreement that four-leg intersections have more crashes than equivalent three-leg intersections. Previous research findings by (Bauer & Harwood, 1996; Harwood et al., 1995; Harwood et al., 2002) also revealed similar results. Apart from the above key predictive variables, some variables which were previously found to be significant in other studies were either insignificant or not considered in this study because of the small sample size and lack of data. Moreover, in this study, it was observed that most of the research results were conceded with the research findings of previous studies.

6. CONCLUSION AND RECOMMENDATION

In this chapter, according to the research findings, a general conclusion was drawn, and recommendations were forwarded to improve the safety of signalized intersections in Ghent, Belgium.

6.1 Conclusion

This research attempted to make an effort to develop a motorized vehicle crash prediction model for signalized intersections for the city of Ghent, Belgium. Initially, it was intended to develop two separate models for Property Damage Only crashes (PDO) and for Injuries and Fatality crashes(IF). However, due to the small sample size and the number of observations in the data set, statistically insignificant results were observed for the Injuries and fatality crashes. As a result, rather than ignoring the second model, the author decided to combine the two types of crashes and develop a single crash prediction model for a total crash using four types of functional forms.

A negative binomial generalized linear model (NBGLM) with loglink were used to fit a model. First, before directly going to a multivariate model fit, the researcher developed simple models that incorporate only the exposure variable (i.e., AADT on major and minor approaches) using four functional forms. Then, from the four simple models developed, to choose a functional form with better performance, measures of Goodness of Fit were checked for the simple models. However, similar results were observed among the four simple models. Thus, the researcher decided to select all the functional forms to model a fully specified crash prediction model. By doing this, the researcher believes choosing all the four functional forms instead of a single form will help to examine whether the functional forms would continue to perform similarly in the presence of other covariates or if a single best model would be revealed. Only 80% (i.e., 61) of signalized intersections were used to develop the fully specified crash prediction models. And the remaining 20% (i.e., 16) signalized intersections were used for model validation.

According to the best-fitted model, only five variables, including the sum of the traffic volume on the major and minor approaches, the ratio of the traffic volume on the minor approach to the major approach, the Left-turn lane on the major approach, the Presence of crosswalks on the minor approach, and the number of legs/approaches were found to be significant. The sum of traffic volume(AADTs) on the major and minor approaches was positively associated with the expected crash frequency. On the other hand, a negative and significant association was found between the expected crash and the ratio of the number of vehicles on the minor approach to the number of vehicles on the major approach (i.e., Q_{mi}/Q_{ma}).

A negative association was found between the expected crash frequency and the left-turn lane on the major approach (when provided on each side of the major approach). A positive relation was found between the presence of marked crosswalks on the minor approach and the expected crash frequency. This could be due to the number of vehicle-pedestrian interactions; for signalized intersections with marked crosswalks on each side of the minor approach, there will be more vehicle-pedestrian interaction points. Thus, when the number of interaction points increases, the risk of exposure to crashes also increases. A positive relationship was observed between the expected crash frequency and the number of legs/approaches at the signalized intersections.

6.2 Recommendations And Direction For Future Studies

It is evident that the cumulative result of the mitigation action taking on the three contributing factors (i.e., road users, vehicles, and road system) can ensure the safety of signalized intersections. According to the research findings in this study, a proportional traffic volume flow on both major and minor approaches will reduce the chance of vehicle crashes. Moreover, the provision of providing a left-turn lane on each side of the major approach will help in reducing the occurrence of crashes at the signalized intersections. Thus, incorporating a left-turn lane on each side of the major approach for new signalized intersections will have a positive impact on reducing traffic crashes.

Usually, pedestrian crosswalks are provided to create safe crossing zones for Pedestrians. But it has to be also taken into account that when crosswalks are provided, it can also increase the risk of exposure to crashes, as shown in this study. As a result, the following countermeasures should be considered when crosswalks are provided:

- Crosswalks should be provided on the narrow width of the minor road segment rather than on the intersection's neck. This helps to shorten the time taken to cross the road and reduce the exposure of pedestrians on the road.
- Adopting Leading Pedestrian Interval (LPI) in the traffic signal timing. The Leading Pedestrian Interval (LPI) allows pedestrians to enter the crosswalk at the signalized intersection 3-7 seconds before vehicles are given a green indication (Federal Highway Administration, 2021). This way, pedestrians can better establish their presence in the crosswalk before vehicles have priority to turn right or left.
- Installing Pedestrian countdown signals at signalized intersections with high pedestrian activities.
- Installing advanced pedestrian warning signs and higher visibility crosswalks.
- Installation of traffic calming devices.
- Providing raised crosswalks. This can be used to ramp up crosswalks and also used as a traffic calming technique.

In this study, it was also observed that four-leg signalized intersections increase the expected crash frequency by around 60% compared to three-leg signalized intersections. Thus, in relation to three-leg Signalized intersections, great attention should be given to incorporate proven safety countermeasures recommended in this and other studies for constructing new four-leg signalized intersections.

The current study attempted to make an effort to develop a vehicle crash prediction model based on only data provided on traffic crashes, traffic count, and road network. No data were provided regarding the location of intersections and intersection characteristics data. The author developed a methodology to locate intersections from the road network data and to collect intersection-related data. This might have an effect on the selection of the number of samples and the quality of the data. The author also tried to include most of the variables that were found significant in previous studies, but because of data unavailability, some co-founding factors(such as speeding) were not included in the study. In addition, some intersection characteristics data were collected using Google satellite images and google earth; this also affects the data quality and precision (Lane width and intersection angle of skewness). On top of this, the developed crash prediction model only considered traffic flow and intersection characteristics data; other essential contributor factors such as road user and vehicle factors were not included. Considering the above limitations, the author believed that a better vehicle crash prediction model for signalized intersections could be developed by providing additional data and by incorporating road user and vehicle factors.

REFERENCE

AASHTO. (2010). Highway Safety Manual.

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of safety research*, *34*(5), 597-603.

Abdel-Aty, M., & Keller, J. (2005). Exploring the overall and specific crash severity levels at signalized intersections. *Accident Analysis & Prevention*, *37*(3), 417-425. https://doi.org/https://doi.org/10.1016/j.aap.2004.11.002

Abdel-Aty, M., Keller, J., & Brady, P. A. (2005). Analysis of types of crashes at signalized intersections by using complete crash data and tree-based regression. *Transportation Research Record*, *1908*(1), 37-45.

Alarifi, S. A., Abdel-Aty, M. A., Lee, J., & Park, J. (2017). Crash modeling for intersections and segments along corridors: a Bayesian multilevel joint model with random parameters. *Analytic methods in accident research*, *16*, 48-59.

Ambros, J., Jurewicz, C., Turner, S., & Kieć, M. (2018). An international review of challenges and opportunities in development and use of crash prediction models. *European transport research review*, *10*(2), 1-10.

Bacquie, R., & Egan, D. (2001). Pedestrian refuge island safety audit. Improving Transportation Systems Safety and Performance. 2001 Spring Conference and ExhibitInstitute of Transportation Engineers (ITE),

Barbosa, H., Cunto, F., Bezerra, B., Nodari, C., & Jacques, M. A. (2014). Safety performance models for urban intersections in Brazil. *Accident Analysis & Prevention*, *70*, 258-266.

Bauer, K., & Harwood, D. W. (1996). *Statistical models of at-grade intersection accidents*.

Bonneson, J. A., & Zimmerman, K. (2006). Identifying Intersections with Potential for Red Light–Related Safety Improvement. *Transportation Research Record*, *1953*(1), 128-136.

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., & 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*, 1-15.

Chen, P., Wu, C., & Zhu, S. (2016). Interaction between vehicles and pedestrians at uncontrolled mid-block crosswalks. *Safety science*, *82*, 68-76.

Chin, H. C., & Quddus, M. A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention*, *35*(2), 253-259.

Daniels, S., Brijs, T., Nuyts, E., & Wets, G. (2011). Extended prediction models for crashes at roundabouts. *Safety science*, *49*(2), 198-207.

David, N., & Norman, J. (1975). *Motor Vehicle Accidents in Relation to Geometric and Traffic Features of Highway Intersections: Volume II-Research Report*.

De Pauw, E., Daniels, S., Van Herck, S., & Wets, G. (2015). Safety Effects of Protected Left-Turn Phasing at Signalized Intersections: An Empirical Analysis. *Safety*, *1*(1), 94-102.

Eenink, R., Reurings, M., Elvik, R., Cardoso, J., Wichert, S., & Stefan, C. (2008). Accident prediction models and road safety impact assessment: recommendations for using these tools. *Institute for Road Safety Research, Leidschendam*.

Elvik, R., & Goel, R. (2019). Safety-in-numbers: An updated meta-analysis of estimates. *Accident Analysis & Prevention*, *129*, 136-147.

Elvik, R., & Vaa. (2004). *The handbook of road safety measures*. Emerald Group Publishing.

Elvik, R., Vaa, T., Hoye, A., & Sorensen, M. (2009). *The handbook of road safety measures*. Emerald Group Publishing.

European Commission. (2018). Annual Accident Report 2018.

European Commission. (2021). Road safety thematic report – Fatigue. *European Road Safety Observatory. Brussels, European Commission, Directorate General for Transport.*

Federal Highway Administration. (2021). *OFFICE OF SAFETY Proven Safety Countermeasures; Leading Pedestrian*

Interval. U.S. department of transportation Federal highway administration. https://safety.fhwa.dot.gov/provencountermeasures/lead_ped_int.cfm

Ferreira, S., & Couto, A. (2013). Traffic flow-accidents relationship for urban intersections on the basis of the translog function. *Safety science*, *60*, 115-122.

Geurts, K., Wets, G., Brijs, T., & Vanhoof, K. (2003). Profiling high frequency accident locations using association rules. *Proceedings of the 82nd annual transportation research board*, 18.

Gluck, J., Levinson, H. S., & Stover, V. (1999). Impacts of Access Management Techniques, NCHRP Report 420. *Transportation Research Board*.

Gomes, S. V., Geedipally, S. R., & Lord, D. (2012). Estimating the safety performance of urban intersections in Lisbon, Portugal. *Safety science*, *50*(9), 1732-1739.

Harkey, D. L. (2008). *Accident modification factors for traffic engineering and ITS improvements* (Vol. 617). Transportation Research Board.

Harwood, D., Pietrucha, M., Wooldridge, M., Brydia, R., & Fitzpatrick, K. (1995). Median Intersection Design, NCHRP Report 375. *Transportation Research Board*.

Harwood, D. W., Bauer, K. M., Potts, I. B., Torbic, D. J., Richard, K. R., Rabbani, E., Hauer, E., & Elefteriadou, L. (2002). *Safety effectiveness of intersection left-and right-turn lanes*.

Hauer, E., & Bamfo, J. (1997). Two tools for finding what function links the dependent variable to the explanatory variables. Proceedings of the ICTCT 1997 Conference, Lund, Sweden,

Herms, B. F. (1970). *Pedestrian crosswalk study: Accidents in painted and unpainted crosswalks*. Police Department, Traffic Bureau.

Heurn, E. (1988). The safety of older persons at intersections.

International Transportation Forum. (2020). *ROAD SAFETY REPORT 2020 | BEGIUM*. OECD/ITF.

Ivan, J. N. (2004). New approach for including traffic volumes in crash rate analysis and forecasting. *Transportation Research Record*, *1897*(1), 134-141.

Jagannathan, R., Gimbel, M., Bared, J. G., Hughes, W. E., Persaud, B., & Lyon, C. (2006). Safety comparison of New Jersey jug handle intersections and conventional intersections. *Transportation Research Record*, *1953*(1), 187-200.

Joksch, H. C., & Kostyniuk, L. P. (1998). *Modeling intersection crash counts and traffic volume*.

Jonsson, T., Ivan, J. N., & Zhang, C. (2007). Crash prediction models for intersections on rural multilane highways: Differences by collision type. *Transportation Research Record*, *2019*(1), 91-98.

Khattak, M. W., Pirdavani, A., De Winne, P., Brijs, T., & De Backer, H. (2021). Estimation of safety performance functions for urban intersections using various functional forms of the negative binomial regression model and a generalized Poisson regression model. *Accident Analysis & Prevention*, *151*, 105964.

Kim, D.-G., & Washington, S. (2006). The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. *Accident Analysis & Prevention*, *38*(6), 1094-1100.

King, M. R., Carnegie, J. A., & Ewing, R. (2003). Pedestrian safety through a raised median and redesigned intersections. *Transportation Research Record*, *1828*(1), 56-66.

Kulmala, R. (1998). Safety at rural three-and four-arm junctions: Development and application of accident prediction models.

Kumfer, W., Harkey, D., Lan, B., Srinivasan, R., Carter, D., Patel Nujjetty, A., Eigen, A. M., & Tan, C. (2019). Identification of critical intersection angle through crash modification functions. *Transportation Research Record*, *2673*(2), 531-543.

Lee, J., Abdel-Aty, M., & Shah, I. (2019). Evaluation of surrogate measures for pedestrian trips at intersections and crash modeling. *Accident Analysis & Prevention*, *130*, 91-98.

Lord, D., & Bonneson, J. A. (2007). Development of accident modification factors for rural frontage road segments in Texas. *Transportation Research Record*, *2023*(1), 20-27.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation research part A: policy and practice*, *44*(5), 291-305.

Lord, D., van Schalkwyk, I., Staplin, L., & Chrysler, S. (2005). Reducing older driver injuries at intersections using more accommodating roundabout design practices. *Texas Transportation Institute and The Texas A&M University System*.

Ma, P. (2009). *Bayesian analysis of underreporting Poisson regression model with an application to traffic crashes on two-lane highways*.

Marlinde Koopmans, Evy Mettepenningen, & Guido Van Huylenbroeck. (2012). *SUPURBFOOD Towards sustainable modes of urban and peri-urban food provisioning; Case Study WP2: The Ghent City Region*.

Maze, T., Henderson, J. L., & Sankar, R. (1994). *Impacts on Safety of Left-Turn Treatment at High Speed Signalized Intersections*.

McGee, H., Taori, S., & Persaud, B. (2003). NCHRP Report 491: Crash Experience Warrant for Traffic Signals. *Transportation Research Board of the National Academies, Washington, DC*.

Miaou, S.-P., & Lord, D. (2003). Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. *Transportation Research Record*, *1840*(1), 31-40.

Miranda-Moreno, L. F., Morency, P., & El-Geneidy, A. M. (2011). The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections. *Accident Analysis & Prevention*, *43*(5), 1624-1634.

Mitra, S., & Washington, S. (2012). On the significance of omitted variables in intersection crash modeling. *Accident Analysis & Prevention*, *49*, 439-448.

Nambuusi, B., Brijs, T., & Hermans, E. (2008). *A review of accident prediction models for road intersections*. UHasselt.

Nightingale, E., Parvin, N., Seiberlich, C., Savolainen, P. T., & Pawlovich, M. (2017). Investigation of skew angle and other factors influencing crash frequency at high-speed rural intersections. *Transportation Research Record*, *2636*(1), 9-14.

Oh, J., Lyon, C., Washington, S., Persaud, B., & Bared, J. (2003). Validation of FHWA crash models for rural intersections: Lessons learned. *Transportation Research Record*, *1840*(1), 41-49.

Oh, J., Washington, S. P., & Nam, D. (2006). Accident prediction model for railway-highway interfaces. *Accident Analysis & Prevention*, *38*(2), 346-356.

Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of transportation engineering*, *122*(2), 105-113.

Pulugurtha, S. S., Duddu, V. R., & Kotagiri, Y. (2013). Traffic analysis zone level crash estimation models based on land use characteristics. *Accident Analysis & Prevention*, *50*, 678-687.

Pulugurtha, S. S., Vasudevan, V., Nambisan, S. S., & Dangeti, M. R. (2012). Evaluating effectiveness of infrastructure-based countermeasures for pedestrian safety. *Transportation Research Record*, *2299*(1), 100-109.

Qin, X., Ivan, J. N., & Ravishanker, N. (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis & Prevention*, *36*(2), 183-191.

Rimiller, J., Ivan, J., & Garrick, N. (2003). Estimating benefits from specific highway safety improvements: Phase III: Safety benefits from left turn treatment. *Connecticut Cooperative Highway Research Program, Project*, *97*(1), 3.

Salifu, M. (2004). Accident prediction models for unsignalised urban junctions in Ghana. *IATSS research*, *28*(1), 68-81.

Savolainen, P. T., & Tarko, A. P. (2005). Safety impacts at intersections on curved segments. *Transportation Research Record*, *1908*(1), 130-140.

Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, *45*, 382-391.

Smith, S. A., & Knoblauch, R. L. (1987). Guidelines for the installation of crosswalk markings (discussion and closure). *Transportation Research Record*, *1141*, 15-25.

Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2014). Multimodal injury risk analysis of road users at signalized and non-signalized intersections. *Accident Analysis & Prevention*, *71*, 201-209.

Tay, R., & Rifaat, S. M. (2007). Factors contributing to the severity of intersection crashes. *Journal of Advanced Transportation*, *41*(3), 245-265.

Torbic, D. J., Harwood, D. W., Bokenkroger, C. D., Srinivasan, R., Carter, D., Zegeer, C. V., & Lyon, C. (2010). Pedestrian safety prediction methodology for urban signalized intersections. *Transportation Research Record*, *2198*(1), 65-74.

Vogt, A., & Bared, J. (1998). Accident models for two-lane rural segments and intersections. *Transportation Research Record*, *1635*(1), 18-29.

Wali, B., Khattak, A. J., Waters, J., Chimba, D., & Li, X. (2018). Development of safety performance functions: incorporating unobserved heterogeneity and functional form analysis. *Transportation Research Record*, *2672*(30), 9-20.

Wang, J., Huang, H., & Zeng, Q. (2017). The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. *Accid Anal Prev*, *98*, 223-231. https://doi.org/10.1016/j.aap.2016.10.018

Wang, K., Zhao, S., & Jackson, E. (2020). Investigating exposure measures and functional forms in urban and suburban intersection safety performance functions using generalized negative binomial-P model. *Accident Analysis & Prevention*, *148*, 105838.

Wang, X., Abdel-Aty, M., & Brady, P. A. (2006). Crash estimation at signalized intersections: significant factors and temporal effect. *Transportation Research Record*, *1953*(1), 10-20.

Wang, Y., & Zhang, W. (2017). Analysis of Roadway and Environmental Factors Affecting Traffic Crash Severities. *Transportation Research Procedia*, *25*, 2124-2130. https://doi.org/10.1016/j.trpro.2017.05.407

Washington, S., Karlaftis, M., Mannering, F., & Anastasopoulos, P. (2010). *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC.

Washington, S., Karlaftis, M., Mannering, F., & Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC.

WHO. (2018). *Global Status Report on Road Safety 2018*.

Wilson, J. E. (1967). Simple types of intersection improvements. In *Improved Street Utilization Through Traffic Engineering* (pp. 144-159). Highway Research Board Washington, DC.

Xie, S., Dong, N., Wong, S., Huang, H., & Xu, P. (2018). Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. *Accident Analysis & Prevention*, *121*, 285-294.

Zegeer, C. V., Richard Stewart, J., Huang, H., & Lagerwey, P. (2001). Safety effects of marked versus unmarked crosswalks at uncontrolled locations: analysis of pedestrian crashes in 30 cities. *Transportation Research Record*, *1773*(1), 56-68.

Zegeer, C. V., Stewart, J. R., Huang, H. H., Lagerwey, P. A., Feaganes, J. R., & Campbell, B. (2005). *Safety effects of marked versus unmarked crosswalks at uncontrolled locations final report and recommended guidelines*.

Zhou, H., Ivan, J. N., Sadek, A. W., & Ravishanker, N. (2010). Safety effects of exclusive left-turn lanes at unsignalized intersections and driveways. *Journal of Transportation Safety & Security*, *2*(3), 221-238.

APPENDIX

Appendix-A: List and description of variables proposed for model development

Category	Variable	Description	Unit	Variable level	Buffer zone
Dependent variables	PDO_crashes	Average property damage only crashes	Count	-	70m
	IF_crashes	Average injury and fatality crashes	Count	-	70m
Exposure variables	AADT_Maj	AADT on the major approach	AADT	-	n/a
	AADT_Min	AADT on the minor approach	AADT	-	n/a
Intersection characteristics variables	LTL_Maj	Left turn lane on the major approach	0 to 2	2- on each side 1- on one side 0-no dedicated lane	n/a
	RTL_min	Right turn lane on the minor approach	0 to 2	2- on each side 1- on one side 0-no dedicated lane	n/a
	CW_Maj	Presence of crosswalks on the major approach	0 to 2	2- on each side 1- on one side 0-no crosswalk	n/a
	CW_Min	Presence of crosswalks on the minor approach	0 to 2	2- on each side 1- on one side 0-no crosswalk	n/a
	Legs	Number of legs/approaches	3 or 4	1-four leg 0-three leg	n/a
	Skewness	Intersection skewness	Yes or no	1-yes 0-no	n/a
	ALW_Maj	Average lane width on the major approach	Meter	-	n/a
	ALW_Min	Average lane width on the minor approach	Meter	-	n/a

Appendix B: A summary of collinearity analysis

Model		Unstandardized Coefficients		Collinearity Statistics	
		B	Std. Error	Tolerance	VIF
1	(Constant)	2.213	6.114		
	AADT on the major approach	.000	.000	.315	3.173
	AADT on the minor approach	-5.482E-5	.000	.337	2.966
	Average lane width on the major approach	-.242	1.716	.933	1.072
	Average lane width on the minor approach	.098	1.317	.874	1.145
	Left turn lane on the major approach	-.874	.646	.243	4.121
	Right turn lane on the minor approach	-.869	.710	.310	3.227
	Presence of crosswalks on the major approach	-.048	.428	.735	1.360
	Presence of crosswalks on the minor approach	.775	.432	.613	1.630
	Number of legs/approaches	1.612	.831	.458	2.183
	Intersection skewness	.169	.607	.864	1.158

Appendix C: SPSS Model Output results For Injury and fatality crashes

Generalized Linear Models (Poisson)

```
DATASET ACTIVATE DataSet2.
* Generalized Linear Models.
GENLIN IF_crashes BY LTL_Maj RTL_Min CW_Maj CW_Min Legs Skewness
(ORDER=DESCENDING) WITH AADT_Maj
    ALW_Maj ALW_Min
  /MODEL AADT_Maj ALW_Maj ALW_Min LTL_Maj RTL_Min CW_Maj CW_Min Legs
Skewness INTERCEPT=YES
 DISTRIBUTION=POISSON LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100
MAXSTEPHALVING=5
    PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD)
CILEVEL=95 CITYPE=WALD
    LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
```

Model Information

Dependent Variable	Average injury and fatality crashes
Probability Distribution	Poisson
Link Function	Log

Case Processing Summary

	N	Percent
Included	72	100.0%
Excluded	0	0.0%
Total	72	100.0%

Goodness of Fit[a]

	Value	df	Value/df
Deviance	38.478	58	.663
Scaled Deviance	38.478	58	
Pearson Chi-Square	37.960	58	.654
Scaled Pearson Chi-Square	37.960	58	
Log Likelihood[b]	-114.313		
Akaike's Information Criterion (AIC)	256.626		
Finite Sample Corrected AIC (AICC)	263.995		
Bayesian Information Criterion (BIC)	288.500		
Consistent AIC (CAIC)	302.500		

Dependent Variable: Average injury and fatality crashes
Model: (Intercept), AADT on the major approach , Average lane width on the major approach , Average lane width on the minor approach, Left turn lane on the major approach , Right turn lane on the minor approach, Presence of crosswalks on the major approach , Presence of crosswalks on the minor approach, Number of legs/approaches , Intersection skewness[a]

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
19.812	13	.100

Parameter Estimates

Parameter	B	Std. Error	Hypothesis Test			Exp(B)
			Wald Chi-Square	df	Sig.	
(Intercept)	.776	1.6861	.212	1	.645	2.174
AADT on the major approach	1.132E-5	1.5000E-5	.569	1	.450	1.000
Average lane width on the major approach	.248	.4385	.319	1	.572	1.281
Average lane width on the minor approach	-.399	.3992	1.000	1	.317	.671
[Left turn lane on the major approach =2]	.236	.2153	1.198	1	.274	1.266
[Left turn lane on the major approach =1]	.493	.2669	3.405	1	.065	1.636
[Left turn lane on the major approach =0]	0[a]	.	.	.	.	1
[Right turn lane on the minor approach=2]	-1.072E-5	.2520	.000	1	1.000	1.000
[Right turn lane on the minor approach=1]	.085	.2241	.144	1	.705	1.089
[Right turn lane on the minor approach=0]	0[a]	.	.	.	.	1
[Presence of crosswalks on the major approach =2]	-.297	.3117	.907	1	.341	.743
[Presence of crosswalks on the major approach =1]	.175	.2880	.368	1	.544	1.191
[Presence of crosswalks on the major approach =0]	0[a]	.	.	.	.	1
[Presence of crosswalks on the minor approach=2]	.169	.2567	.436	1	.509	1.185
[Presence of crosswalks on the minor approach=1]	-.936	.6027	2.413	1	.120	.392
[Presence of crosswalks on the minor approach=0]	0[a]	.	.	.	.	1
[Number of legs/approaches =1]	.320	.2366	1.826	1	.177	1.377

[Number of legs/approaches =0]	0[a]	.	.	.	.	1
[Intersection skewness=1]	-.041	.1751	.056	1	.814	.960
[Intersection skewness=0]	0[a]	.	.	.	.	1
(Scale)	1[b]					

Dependent Variable: Average injury and fatality crashes

Model: (Intercept), AADT on the major approach , Average lane width on the major approach , Average lane width on the minor approach, Left turn lane on the major approach , Right turn lane on the minor approach, Presence of crosswalks on the major approach , Presence of crosswalks on the minor approach, Number of legs/approaches , Intersection skewness

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Generalized Linear Models (Negative binomial)

```
* Generalized Linear Models.

GENLIN IF_crashes BY LTL_Maj RTL_Min CW_Maj CW_Min Legs Skewness
(ORDER=DESCENDING) WITH AADT_Maj

  ALW_Maj ALW_Min

  /MODEL AADT_Maj ALW_Maj ALW_Min LTL_Maj RTL_Min CW_Maj CW_Min Legs
Skewness INTERCEPT=YES

 DISTRIBUTION=NEGBIN(MLE) LINK=LOG

  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100
MAXSTEPHALVING=5

    PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD)
CILEVEL=95 CITYPE=WALD

    LIKELIHOOD=FULL

  /MISSING CLASSMISSING=EXCLUDE

  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION.
```

Model Information

Dependent Variable	Average injury and fatality crashes
Probability Distribution	Negative binomial (1)
Link Function	Log

Case Processing Summary

	N	Percent
Included	72	100.0%
Excluded	0	0.0%
Total	72	100.0%

Goodness of Fit[a]

	Value	df	Value/df
Deviance	11.539	58	.199
Scaled Deviance	11.539	58	
Pearson Chi-Square	10.946	58	.189
Scaled Pearson Chi-Square	10.946	58	
Log Likelihood[b]	-145.910		
Akaike's Information Criterion (AIC)	319.821		
Finite Sample Corrected AIC (AICC)	327.189		
Bayesian Information Criterion (BIC)	351.694		
Consistent AIC (CAIC)	365.694		

Dependent Variable: Average injury and fatality crashes

Model: (Intercept), AADT on the major approach , Average lane width on the major approach , Average lane width on the minor approach, Left turn lane on the major approach , Right turn lane on the minor approach, Presence of crosswalks on the major approach , Presence of crosswalks on the minor approach, Number of legs/approaches , Intersection skewness[a]

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
5.611	13	.959

Parameter Estimates

Parameter	B	Std. Error	Hypothesis Test		
			Wald Chi-Square	df	Sig.
(Intercept)	.509	3.2660	.024	1	.876
AADT on the major approach	8.612E-6	2.7618E-5	.097	1	.755
Average lane width on the major approach	.374	.9267	.163	1	.686
Average lane width on the minor approach	-.425	.6841	.385	1	.535
[Left turn lane on the major approach =2]	.241	.3942	.372	1	.542
[Left turn lane on the major approach =1]	.544	.5118	1.128	1	.288
[Left turn lane on the major approach =0]	0[a]	.	.	.	.
[Right turn lane on the minor approach=2]	.008	.4467	.000	1	.986
[Right turn lane on the minor approach=1]	.131	.4313	.093	1	.761
[Right turn lane on the minor approach=0]	0[a]	.	.	.	.
[Presence of crosswalks on the major approach =2]	-.321	.5311	.365	1	.546
[Presence of crosswalks on the major approach =1]	.108	.5270	.042	1	.837
[Presence of crosswalks on the major approach =0]	0[a]	.	.	.	.
[Presence of crosswalks on the minor approach=2]	.198	.4451	.198	1	.656

[Presence of crosswalks on the minor approach=1]	-.967	.9954	.943	1	.331
[Presence of crosswalks on the minor approach=0]	0[a]	.	.	.	.
[Number of legs/approaches =1]	.300	.4298	.487	1	.485
[Number of legs/approaches =0]	0[a]	.	.	.	.
[Intersection skewness=1]	-.066	.3331	.039	1	.843
[Intersection skewness=0]	0[a]	.	.	.	.
(Scale)	1[b]				
(Negative binomial)	1[b]				

Dependent Variable: Average injury and fatality crashes

Model: (Intercept), AADT on the major approach , Average lane width on the major approach , Average lane width on the minor approach, Left turn lane on the major approach , Right turn lane on the minor approach, Presence of crosswalks on the major approach , Presence of crosswalks on the minor approach, Number of legs/approaches , Intersection skewness

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Appendix D: SPSS Model Output result For Total Crash (FF-4)

```
DATASET ACTIVATE DataSet1.
* Generalized Linear Models.
GENLIN Total_crash BY LTL_Maj CW_Min Legs (ORDER=DESCENDING) WITH
Ln_Q_sum Ln_Q_proportion
  /MODEL Ln_Q_sum Ln_Q_proportion LTL_Maj CW_Min Legs INTERCEPT=YES
 DISTRIBUTION=NEGBIN(MLE) LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100
MAXSTEPHALVING=5
    PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD)
CILEVEL=95 CITYPE=WALD
    LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION.
```

Model Information

Dependent Variable	Average total crash
Probability Distribution	Negative binomial (MLE)
Link Function	Log

Case Processing Summary

	N	Percent
Included	61	100.0%
Excluded	0	0.0%
Total	61	100.0%

Categorical Variable Information

			N	Percent
Factor	Left turn lane on the major approach	On each side	30	49.2%
		On one side	11	18.0%
		No dedicated lane	20	32.8%
		Total	61	100.0%
	Presence of crosswalks on the minor approach	On each side	47	77.0%
		On one side	3	4.9%
		No crosswalk	11	18.0%
		Total	61	100.0%
	Number of legs/approaches	Four legs	33	54.1%
		Three legs	28	45.9%
		Total	61	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Average total crash	61	0	15	4.92	3.470
Covariate	Ln_Q_sum	61	8.77	10.96	9.9505	.47979
	Ln_Q_proportion	61	-8.08	.67	-.8154	1.13843

Goodness of Fit[a]

	Value	df	Value/df
Deviance	67.458	52	1.297
Scaled Deviance	67.458	52	
Pearson Chi-Square	54.013	52	1.039
Scaled Pearson Chi-Square	54.013	52	
Log Likelihood[b]	-128.935		
Akaike's Information Criterion (AIC)	275.869		
Finite Sample Corrected AIC (AICC)	279.399		
Bayesian Information Criterion (BIC)	294.867		
Consistent AIC (CAIC)	303.867		

Dependent Variable: Average total crash

Model: (Intercept), Ln_Q_sum, Ln_Q_proportion, Left turn lane on the major approach , Presence of crosswalks on the minor approach, Number of legs/approaches

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
51.114	7	<.001

Dependent Variable: Average total crash
Model: (Intercept), Ln_Q_sum, Ln_Q_proportion, Left turn lane on the major approach , Presence of crosswalks on the minor approach, Number of legs/approaches
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	5.551	1	.018
Ln_Q_sum	11.126	1	<.001
Ln_Q_proportion	5.204	1	.023
Left turn lane on the major approach	10.847	2	.004
Presence of crosswalks on the minor approach	5.608	2	.061
Number of legs/approaches	11.235	1	<.001

Dependent Variable: Average total crash
Model: (Intercept), Ln_Q_sum, Ln_Q_proportion, Left turn lane on the major approach , Presence of crosswalks on the minor approach, Number of legs/approaches

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	Upper	Hypothesis Test Wald Chi-Square	df	Sig.
(Intercept)	-3.712	1.4530	-6.560	-.864	6.526	1	.011
Ln_Q_sum	.451	.1352	.186	.716	11.126	1	<.001
Ln_Q_proportion	-.150	.0656	-.278	-.021	5.204	1	.023
[Left turn lane on the major approach =2]	-.578	.2107	-.991	-.165	7.531	1	.006
[Left turn lane on the major approach =1]	.068	.1968	-.318	.454	.120	1	.730
[Left turn lane on the major approach =0]	0[a]	.	.	.	.	.	.
[Presence of crosswalks on the minor approach=2]	.466	.1967	.080	.851	5.604	1	.018
[Presence of crosswalks on the minor approach=1]	.446	.3602	-.260	1.152	1.531	1	.216
[Presence of crosswalks on the minor approach=0]	0[a]	.	.	.	.	.	.
[Number of legs/approaches =1]	.625	.1864	.259	.990	11.235	1	<.001
[Number of legs/approaches =0]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						
(Negative binomial)	3.613E-8[c]	.	.	.			

Dependent Variable: Average total crash

Model: (Intercept), Ln_Q_sum, Ln_Q_proportion, Left turn lane on the major approach , Presence of crosswalks on the minor approach, Number of legs/approaches

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

c. Hessian matrix singularity is caused by the scale or negative binomial parameter.

Appendix E: Simulation for the effect of the proportion of traffic volume on minor approach

The simulation was done for a random signalized intersection from the sample size. A total volume of 30,000 vehicles (AADT) was selected for simulation. Thus, by keeping other variables constant, only the percent of the proportion of the traffic volume on the minor road to the total traffic volume was changed to see the effect on the expected crash.

Model 4 (FF-4)

$$\ln \mu = (-3.712) + 0.451 \ln(Q_{MA} + Q_{MI}) - 0.15 \ln(\frac{Q_{MI}}{Q_{MA}}) - 0.578(LTL_{maj}) + 0.466(CW_{min}) + 0.625(N^{o}\ Approaches)$$

Total volume (Q_{maj}+Q_{min})=30,000

% (Q_{min}/Q_{tot})	Qmaj	Qmin	Ln(Qmaj+Qmin)	Ln(Qmin/Qmaj)	Predicted crash
17%	25000	5000	10.30895266	-1.609437912	4.849
33%	20000	10000	10.30895266	-0.693147181	4.228
40%	18000	12000	10.30895266	-0.405465108	4.049
42%	16000	14000	10.30895266	-0.133531393	3.888
50%	15000	15000	10.30895266	0	3.81