UHASSELT
KNOWLEDGE IN ACTION

School of Transportation Sciences

Master of Transportation Sciences

Master's thesis

Pedestrian crash prediction model for signalized intersections in Ethiopia

Bruk Belay Dejenie

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

SUPERVISOR :
Prof. dr. Elke HERMANS

CO-SUPERVISOR :
dr. Evelien POLDERS

2019
2020

UHASSELT
KNOWLEDGE IN ACTION

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

Pedestrian crash prediction model for signalized intersections in Ethiopia

Bruk Belay Dejenie

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

SUPERVISOR :

Prof. dr. Elke HERMANS

CO-SUPERVISOR :

dr. Evelien POLDERS

School of Transportation Sciences

Master of Transportation Sciences

Master Thesis

Pedestrian Crash Prediction Models for Signalized Intersections in Ethiopia.

Bruk Belay Dejenie

Id:1849913

Supervisor:	Co-supervisor:	Local Supervisor:
Prof. Dr. Elke Hermans	Dr. Evelien Polders	Dr. Bikila Teklu

August, 2020

TABLE OF CONTENTS

PREFACE

This master thesis was conducted in two parts. This first part mainly included the background, literature review, and methods of the thesis. While in part two, data collection, data analysis, and recommendations were made. These components are done under the supervision of Prof. Dr. Elke Hermans and co-supervisor Dr. Evelien Polders.

The increasing road fatality and injury numbers have become a major public health issue globally and in sub-Saharan Africa in particular. This main externality of the transportation sector is worse in developing countries like Ethiopia. Especially, in Addis Ababa, the impact is significantly high on vulnerable road users- they took a major portion both in fatality and severity numbers. High traffic conflict locations such as crossings (intersections) took the major portion of pedestrian road fatalities in Addis Ababa. Towards improving safety to pedestrians in the city it is vital to primarily identify the main risk factors associated with the occurrence of a crash for efficient and effective intervention. Besides, it is crucial to have a tool for the proper estimation of the number of crashes at intersections as it will help to identify crash-prone intersections. Based on these considerations, two pedestrian predictive models were developed. And these models will help in estimating pedestrian crashes at signalized intersections and for identification of major risk factors.

I would like to acknowledge the close supervision of Prof. Dr. Elke Hermans, co-supervisor Dr. Evelien Polders and Dr. Bikila Teklu. I am grateful for the insightful comments and guidance which were indeed helpful, featured with positive sense and quick responses. I also would like to express my gratitude for all concerned persons for the VLIR-UOS scholarship grant.

LIST OF FIGURES

LIST OF TABLES

SUMMARY

Pedestrians are the leading victims of a road crash in Addis Ababa. The main location for these crashes is road crossings. Studies in Addis Ababa showed several factors are associated with the existing high burden of crashes on pedestrians including poor road infrastructure, car-oriented road design, risky driving behaviors etc. Yet, the key factors associated with pedestrian crashes at an intersection are not well known. Hence, the current study aims to identify the key factors related to pedestrian unsafety at signalized intersections.

The city administration of Addis Ababa is working to improve pedestrian safety in collaboration with several non-governmental organizations. One of the undergoing projects is the 'Safe Intersection Program" which aims at improving the safety of all road-users at intersections by creating all road-user inclusive intersections. In developing countries like Ethiopia where road safety budget is a huge constraint, it is important to come up with solutions that ensure effective investments (e.g. identification of crash-prone locations and/or road facility prioritization). To easily identify crash-prone road facilities and to prioritize, predictive models play an important role by providing the predicted number of crashes at road facilities and enabling the use of the Empirical Bayes method to calculate the estimated number of crashes (when observed crash data is available). This study will work to develop a pedestrian crash prediction model for pedestrian crashes at 34 signalized intersections in Addis Ababa.

Primarily, to come up with possible risk factors related to pedestrian crashes, previous works of literature was reviewed and it was evidenced that traffic characteristics (traffic volume, pedestrian volume, posted speed limit), intersection characteristics(width of crossing, number of lanes to be crossed, presence of raised median refuge islands, pedestrian-related signings, average pavement condition, average sidewalk width, presence of sidewalk barrier), built-environment characteristics (transit density, school density, number of alcohol sales establishments, presence of parking, the presence of train or metro stations), land-use (Commercial, high-density mixed residential, Medium and low-density mixed residence, government and offices, and social services) and socio-demographic characteristics (neighborhood income) in the vicinity of intersections have a significant effect on pedestrian crash involvement. Further, the effect of those variables differs with the buffer width at which the variable is extracted (such as 150m for transit density, 300m for the presence of parking and number of alcohol sales establishment, 400m for school density and presence of light rail transit station). Regarding the modeling approach, a wide range of modeling approaches are used depending on the crash data characteristics and methodological issues (the assumed functional form). The widely used modeling approach for pedestrian crash prediction is negative binomial distribution, which mainly overcomes the problem of overdispersion in crash data.

The widely recommended and the standard model, according to the Highway Safety Manual (2010), for crash prediction modeling is the generalized linear model (GLM) with a negative binomial distribution and log-link function. GLM uses a power function (for exposure variables) and exponential function (for risk factors). The power function ensures that the predicted crash will only be zero when the exposure variable (traffic volume or pedestrian volume) is zero; otherwise, the predicted crash will be a positive value. While the exponential function ensures the crash number will not be zero or negative as a result of a zero or negative value from the linear predictor (regression of risk factors). A generalized linear model (GLM) with a negative binomial distribution and log-link function will be used for model development in this study. The cross-sectional study design will be used to model 3 years of crash data at 34 intersections. Primary data (onsite traffic count and data extracted using a buffer zone) and secondary data (From Addis Ababa Police Commission and Addis Ababa Road and Transport Bureau) was collected at each intersection.

Before model development, correlation and multi-collinearity among variables, was assessed by using the Pearson coefficient of correlation and the variance inflation factor (VIF). After model development, goodness-of-fit and model validation were conducted. Finally, important risk factors were identified, and remedial measures forwarded. Besides, the estimate of their effect on pedestrian crashes found in this study are compared with previous studies.

In model I, the effect of exposure variables and intersection characteristics variables on pedestrian crashes at a signalized intersection was modelled. Accordingly, a significantly high correlation was found between pedestrian volume and vehicle volume. This being one of the reasons why pedestrian volume was found to be insignificant in the study. There was no significant correlation nor multicollinearity found in all variables and within independent variables respectively. Accordingly, four variables – vehicle volume, pavement condition, number of lanes crossed by pedestrians, and pavement condition were found to be a significant predictor of the total number of crashes on pedestrians. Vehicle volume and sidewalk width have shown a positive association with the number of crashes. Whereas, pavement condition and width of walkway were negatively related. Except, the number of lanes crossed by pedestrians, which is negative and explainable within the scope of the study, the remaining results are consistent with previous studies.

In the second model the effect of incorporating built-environment and land-use variables together with traffic volume and intersection characteristics variables was assessed. As a result, a model with better goodness of fit was found. The model incorporated - vehicle volume, pavement condition, commercial land use, school density, and bus stop density as significant variables. For both Model I and Model II, vehicle volume was the main predictor, which indicates its direct relation with pedestrian crashes. Built-environment and land-use variables haves a lower value of the coefficient, which indicates the indirect relation between the variables and pedestrian crashes. Variables with a positive association with pedestrian crashes at the signalized intersections include poor pavement condition, school density in a 400-meter buffer zone of the intersection and commercial land use around the intersection. As opposed to the above fact, bus stop density in 100meter was found to have a negative relation with pedestrian crashes. This can be explained by considering the local situation - where there are a higher number of traffic regulators at transit locations, which might play a major role in reducing flow speed and ease of traffic flow.

1 BACKGROUND

1.1 INTRODUCTION

Road traffic injuries are posing a major problem in public health leading to physical impairment, social and economic problems worldwide, mainly in middle and low-income countries (WHO, 2007). According to the WHO report 2018, road traffic injuries are the 8th leading cause of death of people of all ages with fatalities increasing steadily to 1.35 million in 2016. The report also states that the crash risk is related to the income level of a country (WHO, 2018). Compared to all other continents (the minimum being 10.4 in Europe and the world average 18.2 per 100,000 inhabitants), the fatality rate is the highest in Africa (26.7 fatalities per 100,000 inhabitants) and the fatality rate in Ethiopia equals this average fatality rate (26.7 fatalities per inhabitants) (WHO, 2018).

Globally, vulnerable road users - pedestrians, cyclists, and motorcyclists - constitute more than half of all road traffic deaths. This indicates the urgency of actions and safety improvements concerning vulnerable road users worldwide. In middle and low-income countries, road crash related risks of vulnerable road-users is higher, with 40% pedestrian deaths on African roads. In Ethiopia, where 60% of the road users represent pedestrians (BIGRS, 2019), 37% of fatalities among road users refer to pedestrians with an increasing trend of total fatalities (WHO, 2018). In terms of collision types, pedestrian crashes take a big portion of the types of collisions and road crossings are a common location of crashes in Ethiopia (Tulu et al., 2013).

As crashes are a result of interaction between road users, several studies have been done in the past and continue to be done in the future too. The interaction between vehicles and pedestrians, and related crash prediction are addressed by researchers under two main categories; by considering non-crash-based measures of effectiveness and crash-based measures of analysis and estimation. Many studies have applied non-crash-based measures of effectiveness such as change in pedestrian behavior (e.g. crossing behavior (Liu & Tung, 2014; Oxley, Lenné, & Corben, 2006)), gap acceptance (Koh & Wong 2014; Pawar & Patil 2015), change in driver behavior and other surrogate safety measures (e.g. conflicts, avoidance maneuvers) (Harwood et al, 2008). Although, a crash is a rare event and it is needed to have crash data of several years in order to find correlations with contributing factors, there are several studies performed by using the number of crashes as a unit of analysis (Elvik & Goel, 2019; Kaygisiz et al., 2017; Lee, Abdel-Aty, & Cai, 2017; Guo et al., 2017; Guo et al., 2017; Daniels et al., 2011).

Most of the contributing factors for pedestrian crashes and activities are related to the road network, demographic characteristics, land use, and geometric and transit characteristics of the facilities (Haghighatpour et al., 2014). Several studies have considered variables, under the above-mentioned categories, to generate causal and predictive relationships of these factors with pedestrian-vehicle crashes (Cottrill & Thakuriah, 2010; Olszewski et al., 2015; Kaygisiz et al., 2017; Lee et al., 2017; Xie et al., 2018).

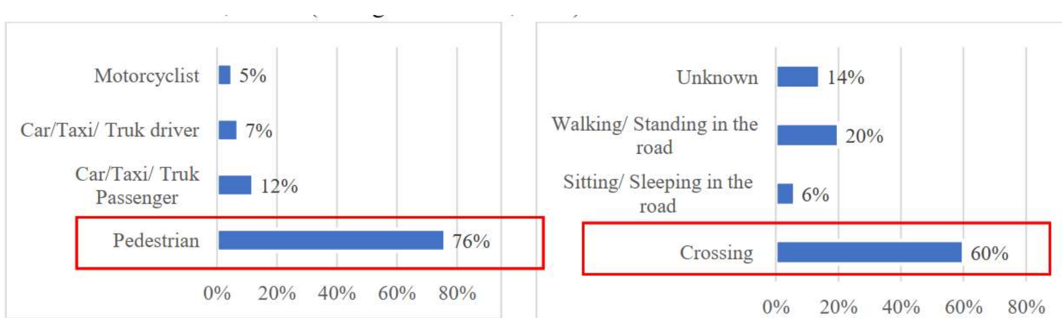
In the recent past, the effect of the built-environment, land use and sociodemographic characteristic of pedestrians was not much considered in the modeling (Harwood et al, 2008), while recent studies are giving much more emphasis on these factors and in some of these studies a significant relationship has been found (Miranda-Moreno et al., 2011; Pulugurtha & Sambhara, 2011).

Several research models have been used to relate the number of vehicle-pedestrian crashes at intersections with contributing factors. For the past two decades, as Xie et al., (2018) summarized, the following regression models have been used: multiple regression model with logarithmic function form, negative binomial model, Poisson model, fixed and random parameters negative binomial model.

In this thesis, the interaction between vehicles and pedestrians at intersections will be studied. The main objective of this study is to model the interaction between pedestrian crashes and contributing factors. This will assist in identifying major contributing factors and the extent of their influence. Unlike models solely based on traffic flow variables, this model considers additional contributing factors to pedestrian crashes by undertaking a literature review to filter out the possible and recommended contributing factors researches.

1.2 PROBLEM STATEMENT

Traffic collision and fatalities have emerged as a major concern in improving road safety in Addis Ababa (AARTB & ITDP, 2018). According to the annual road safety report of 2017-18, 478 road crashes occurred in Addis Ababa, in which pedestrians represented 76% of the fatalities and had a share of 69% of the over 3000 reported injuries (AACA & BIGRS, 2019). The report also presented that 60% of the pedestrians were killed while crossing the road both on zebra and non-zebra crossings. A study by Anteneh Kebede et al, (2019) on the police report of 2013/2014 indicated that pedestrian fatalities constitute 84% of the road fatalities in Addis Ababa. The authors also mentioned that pedestrian fatalities are more prevalent on the road median and road crossings. This makes safety on road crossings of Addis Ababa an alarming issue. In terms of collision types, pedestrian crashes take a big portion of the types of collisions and road crossings are the common location of crashes in Ethiopia (Tulu et al., 2013)



a) Traffic crash deaths by type b) pedestrian status at time of fatal crash

Figure 1 Road traffic crash deaths and pedestrian status at the time of the fatal crash (BIGRS, 2019)

There are several road safety improvement projects in Addis Ababa- for example the 'Safe Intersection Program'. As a road user, pedestrian interaction at road facilities needs to be studied. However, due to traditional black spot analyses, studies in the sector of road safety improvement programs did not consider pedestrian specific crash risk and exposure analysis (Tulu et al, 2015). This results in the underrepresentation of the pedestrian's share in the road safety improvement projects and worsens the unsafety of pedestrians.

The analysis and identification of crash-prone zones (crossings) are vital in reducing crashes (Connors et al., 2013). The identification of causes and contributing factors for such a crash are to be analyzed for remedial measures. There has been relatively little research focusing on identifying causes and contributing factors for pedestrian crashes done worldwide (Xie, 2018) and also in developing counties; Iran (Haghighatpour & Moayedfar, 2014), Turkey (Kaygisiz, 2017), Ethiopia (Tulu et al., 2015). As compared to the attention given and a wide range of studies available to other road user types, pedestrian safety is among the least studied. A pedestrian crash prediction model is prepared by some cities and for a specific facility type in developed countries. For example, combined pedestrian crash predictive models at intersections for Charlotte and Toronto (USA) (Harwood et al, 2008).

The relationship between demographic and land-use characteristics with pedestrian crash is becoming the new area of study, yet some studies indicate the possible difference in the effect of these characteristics between regions and countries. Establishing a significant relation between crashes involving pedestrians with certain demographic and/or land-use characteristics and of similar pedestrian crashes occurring in areas with those demographic characteristics might infer that demographic and land-use characteristics would correlate to crashes. Further, identification of areas with a pedestrian crash highly relies on roadway properties; even in some studies where land-use is included, there is the issue of high generalization of the variables (Harwood et al, 2008). In recent years studies by Miranda-Moreno et al., (2011); Pulugurtha et al., (2013); Cai et al., (2016); Wang, Huang, & Zeng, (2017) and Lee et al., (2017), the interaction of the macro-level environment has been studied. However, in the studies, it is indicated that limited frequency of crash data, limited data on variables and a small number of sample sites/facilities are mentioned as a limitation. In addition, there are enormous differences in walking culture, road environment characteristics, socioeconomic factors, and road user behavior among different regions of the world (Tulu et al., 2015). Thus, there is a need to do more studies in different regions of the world to indicate a variety of situations and strengthen knowledge about pedestrian safety. As this study will be done in one of the developing countries (Ethiopia), it will be a good predictor for countries in the developing region. And globally, it will provide significant input for the predicting capacity of models used in studies on pedestrian crashes on intersections worldwide.

It appears to be difficult to identify the reason behind the change in observed crashes and crash fluctuation over time. Statistically, it is likely that a period with a comparatively high crash frequency will be followed by a period with a comparatively low crash frequency (AASHTO, 2010). The inability to account for this effect will cause regression-to-the-mean bias. In Ethiopia, road safety management studies such as black spot identification and effectiveness of road facilities are done based on the observed number of crashes (Tulu et al., 2015). This traditional approach (also known as "reactive approach" (Fitzpatrick et al., 2018)) of blackspot identification introduces significant chance of a regression-to-the-mean (RTM) bias, which is also known as "selection bias". "RTM bias" happens, when a site is selected based on the short-term trend of observed crashes. This RTM can be accounted in two ways: site selection using long-term observed crash frequency and combing observed crashes with expected crash frequency by using the Empirical Bayes method. In developing countries like Ethiopia where obtaining long-term crash data is unfortunate, using a proactive approach (e.g. using predictive models) for road safety and urban design policies will insure reasonable estimation of crash frequency.

1.3 RESEARCH QUESTIONS

The main research question of the study is: "What are the key variables that influence pedestrian safety at an intersection?"

The research sub-questions include

1. How do the different variables contribute to the number of pedestrian-vehicle crashes at intersections?
2. Which method of regression is best to relate the number of crashes and contributing factors?

3. Are the results of the predictive model for Addis Ababa consistent with previous studies?
4. How can the safety of pedestrians at intersections be improved?

1.4 RESEARCH OBJECTIVE

The main objective of this study is to establish a model that can predict the number of pedestrian-vehicle crashes at signalized intersections in Addis Ababa.
The sub-objectives of the study will be to:

1. Identify potential risk factors contributing to pedestrian crashes at signalized intersections
2. Determine the influence of explanatory variables on the number of pedestrian crashes.
3. Formulate a recommendation to improve pedestrian safety at signalized intersections.

1.5 SCOPE OF THE STUDY

This study will be delimited in terms of road user and road facility type. Pedestrian crashes that occurred at a signalized intersection in the years 2017 to 2019 will be studied. Moreover, in terms of geographic location, it will be delimited to an urban region (Addis Ababa, Ethiopia).

1.6 LIMITATIONS

Due to time constraints and lack of available pedestrian volume data, conversion of daily count to annual traffic volume will be done. This might have a slight difference with the annual count data. The study will share some of the drawbacks of crash modeling studies, which are omitted variable bias and controlling for confounding factors.

1.7 CONCEPTUAL FRAMEWORK

The framework is proposed by considering the findings of studies which indicated the relation of risk factors with pedestrian crash risk and pedestrian crash frequency (Harwood et al, 2008; Miranda-Moreno, Morency, & El-Geneidy, 2011 and Kaygisiz et al., 2017). The conceptual framework shows that land-use, built environment, demographic characteristics, and intersection characteristics have a direct influence on pedestrian risk exposure. Whereas, vehicle traffic volume, pedestrian volume, operating speed, and geometric properties will have a direct effect on vehicle-pedestrian frequency and severity.

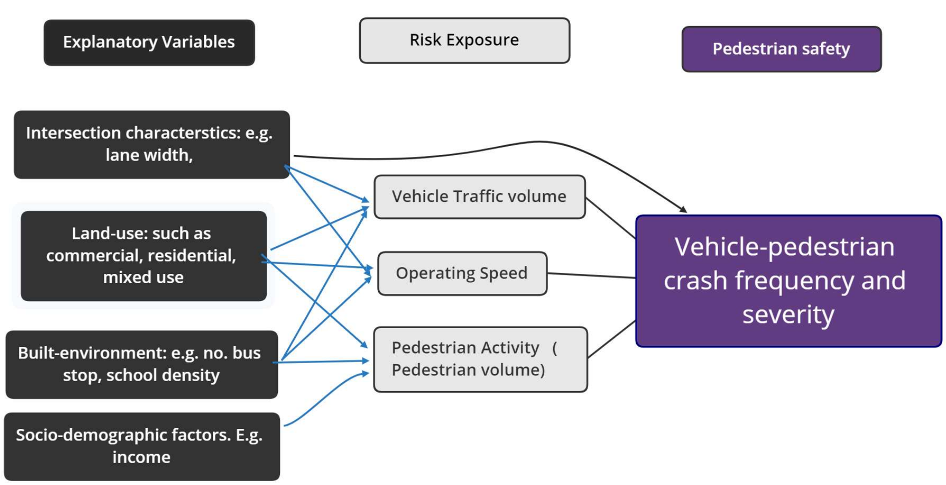


Figure 2 Conceptual framework: built environment, risk exposure, pedestrian safety (Miranda-Moreno, 2011)

2 Literature Review

The literature review aims at creating background knowledge for a pedestrian predictive model by exploring the pedestrian crash predictive variables and methods of regression used in previous studies. And the pedestrian safety condition and findings of studies in Addis Ababa are also reviewed.

2.1 PEDESTRIAN SAFETY IN ADDIS ABABA

Road traffic crashes in the capital city are rapidly increasing, based on the police report from 2009 G.C[1] (2002E.C) to 2014 G.C(2007 E.C), as reported by Hirpa (2016), the road fatality number in Addis Ababa increased by more than 30%. A recent report by BIGRS (2019), presented a fatality number of 463 in 2017 G.C of which 80% involved a pedestrian[2].

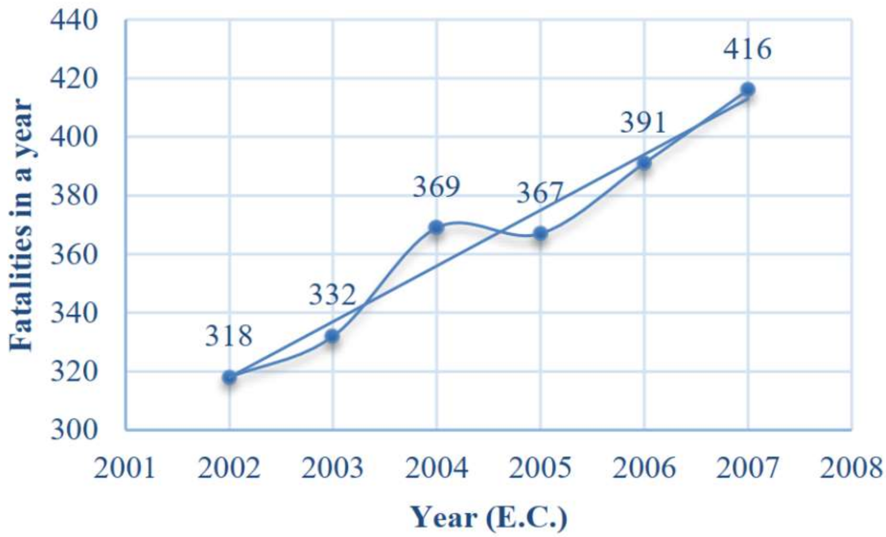


Figure 3 Fatality in Addis Ababa from 2002E.C to 2007E.C (Hirpa, 2016)

Even though there is a big share of walking in Addis Ababa, road infrastructure design and upgrading seems not to be inclusive of all road users. According to a non-motorized transport strategy, the existing city road design in Addis Ababa is characterized by a car-oriented design approach and prioritizing vehicle speed over pedestrian safety (AARTB & ITDP, 2018). The report further emphasized that this approach poses a major problem at intersections where these road users interact. Most roads in the city lack pedestrian facilities even in a place with pedestrian walkway problems such as blockage by roadside facilities (light-pole, tele-pole), uncovered drainage facilities, blockage by construction on the roadside and unmanaged street vendor and ring (Hirpa, 2016). According to iRap[3] rating, for pedestrians, only 14% of the roads are rated 3-stars and better. These poor road conditions plus the high prevalent risky driver behavior left pedestrians with the highest fatality rate in the city (AARTB & ITDP, 2018). Considering the above situation and crash data presented it is easy to recognize the high risk of pedestrians on Addis Ababa roads.

2.2 SAFETY AT INTERSECTIONS

At an intersection the chance of making an error, by a driver, is higher. This is because, at intersections,

[1] GC- Gregorian Calendar
[2] E.C- Ethiopia Calendar
[3] iRap rating- International Road assessment Program(www.irap.org)

drivers need to be more alert in terms of visual search, gap estimation, and decision-making (AASHTO, 2010).

A safety analysis study by Strauss et al., (2014), composed of 647 signalized and 435 non-signalized intersections, reported that at signalized intersections cyclists and pedestrians are at 14 to 12 times higher risk than motorists, respectively. The study further confirmed that all road users are at higher risk at a signalized intersection than non-signalized intersections. A recent study by Lee et al., (2018), has also found that the provision of signalized intersections is related to an increase in pedestrian activity. Demanding for safer crossing is explained as a reason for signalized intersections to attract more pedestrian than the unsignalized ones. This will raise the issue if an increase in pedestrian activity might indicate a rise in conflict between pedestrians and vehicles. These findings are a good indicator of the possible existence of additional risk factors at these intersections beyond the safety improvement gained through signalization of intersections.

2.3 VEHICLE-PEDESTRIAN CRASH PREDICTIVE VARIABLES

2.3.1 EFFECT OF TRAFFIC AND PEDESTRIAN VOLUMES

Several studies have indicated that traffic volume and pedestrian volume have a positive and statistically significant effect on the number of pedestrian crashes (Pulugurtha et al., 2011; Elvik et al., 2013; Wang et al., 2017; Olszewski et al., 2018; Xie et al., 2018).

Elvik, (2013) studied the effect of the marked pedestrian crossing on pedestrian safety. The analysis included 159 marked pedestrian crossings and 316 pedestrian crashes collected for five years in the city of Oslo, Sweden. In the study, it was found that pedestrian and vehicle volume have a positive and significant effect on pedestrian crash occurrence. However, the product of pedestrian and traffic volume was found to have a negative effect on pedestrian crash occurrence.

Alarifi et al,. (2017) studied the crash prediction model for intersections in the Orlando metropolitan area, Florida. The study analyzed 8347 intersections and three years of crash data. The modeling included the effect of macro-level data in addition to the intersection level data. In the study, a model was developed for crash type (vehicle, bicycle and pedestrian crash) and severity type (total and severe crashes). In the study, it was found that major and minor traffic volumes have a positive and significant relation with pedestrian crash occurrence. Further, it was identified that traffic volume on major roads has a higher potential for crash prediction.

In the study to develop a SPF[4] for pedestrian crashes at intersections in Seattle, United States (by Thomas et al., 2017), 12,266 intersections with seven-year crash data, it was reported that the natural logarithm of pedestrian volume estimates was related to pedestrian crashes at intersections. The findings of this research showed that the potential to predict the crash occurrence differs between traffic volume and pedestrian volume. Studies have proved that motor vehicle volume is the main determinant for pedestrian crash number prediction.

In a study done on 519 signalized intersections in the city of Montreal Quebec, Canada, where only pedestrian and traffic volume were used as a predictor, it was found that the coefficient for pedestrian volume and traffic volume was 0.45 and 1.15 (Miranda-Moreno et al., 2011). In the same study, a second model with built environment variables in addition to volume variables reported having lower traffic volume and pedestrian volume coefficients of 0.91 and 0.26 respectively. Thus, it is seen that the values of parameter estimates decreased in the model when built-environment characteristics are

[4] SPF- safety performance function

included. The author outlined that the low change in coefficients (pedestrian from 0.45 to 0.26 and traffic volume 1.15 to 0.91) when built environment variables are introduced in the model may suggest that a large portion of observed variability is explained by pedestrian and vehicle volume (Miranda et al., 2011). While this situation may not be the case in some other studies. A study in Warsaw, Poland on a sample of 52 unsignalized and 50 signalized intersections developed a model for pedestrian crashes (Olszewski, 2018). In this study, the author developed two models: one model with exposure variables and another model with exposure, geometric and traffic control variables. Firstly, the estimated coefficient for pedestrian volume was higher than for motor vehicle volume. Secondly, the author reported the estimated coefficients to increase when additional variables (land-use, the proportion of heavy vehicles, the presence of bus stops and traffic peak to off-peak ratio) are included.

In another recent study done on 279 intersections in Florida, USA Wang et al., (2017), reported a coefficient for traffic volume to be 1.19 (for a model with traffic volume and road variables) and 1.15 (for a model including macroscopic variables). In a study done in Hong Kong, based on the model fitted from 262 signalized intersections and 3 years of crash data, found estimates of the pedestrian and the traffic volume coefficients were 0.21 and 0.27, respectively, (Xie, 2018).

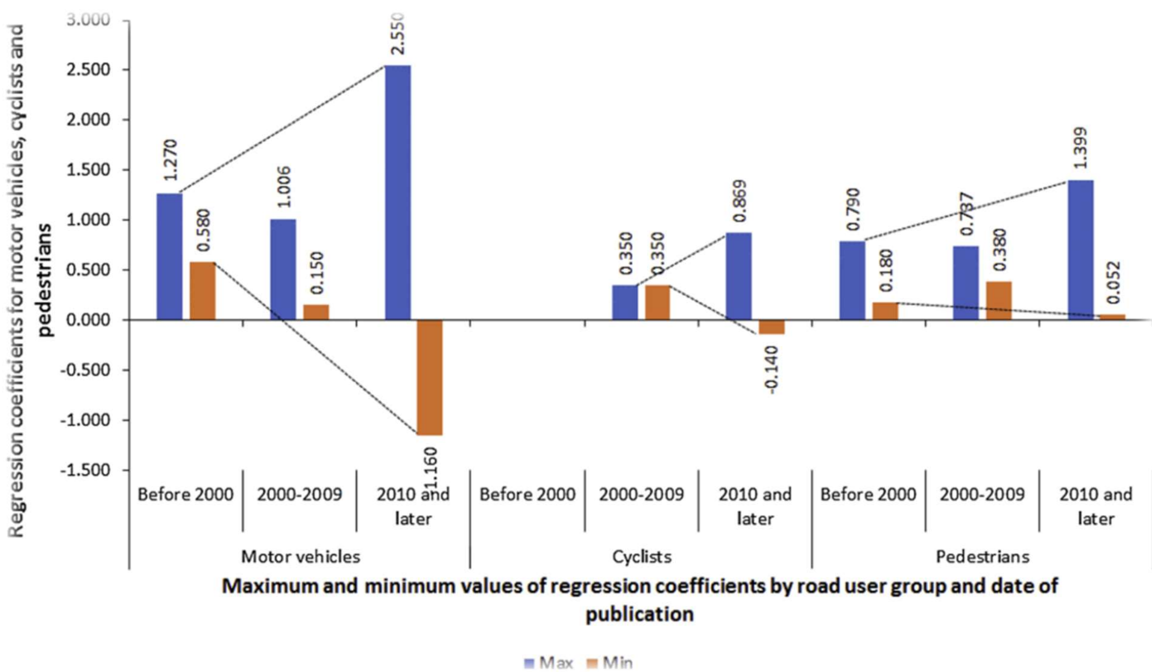


Figure 4 Regression coefficients for a motor vehicle, cyclist and pedestrian from previous researches (Elvik, 2019)

A good summary to view the trend of traffic volume and pedestrian volume coefficients in previous works of literature is a summary of the coefficients by a figure(4), prepared by Elvik et al., (2019). Thus, even if the traffic volume and pedestrian volume are positively related to the pedestrian crash occurrence and significant variables, the range of the estimated coefficient is higher (Elvik, 2019). Miranda, (2011), forwarded the possible reason for this to result from the difference in the number of intersections, intersection type (three-legged, four-legged), number of crash data involved for analysis, quality of traffic and pedestrian data, regression method and many more from one study to the other.

Finally, a study by Lyon and Persaud (cited in Torbic et al., 2010) has indicated that the left turn vehicle volume has vital importance in determining pedestrian crashes. For his study on three leg intersections with a stop-controlled minor road, the effect of the left-turn vehicle was significant.

2.3.2 EFFECT OF INTERSECTION CHARACTERISTICS

Several studies indicated that intersection characteristics (such as the number of crossing lanes, presence of exclusive pedestrian signals) have a significant effect on pedestrian crashes (Harwood et al, 2008; Torbic et al., 2010;Zegeer et al., 2017). Before the year 2008, there was no study that indicated a significant relationship between the width of the crossing and the presence of a median refuge island, and pedestrian crashes (Harwood et al, 2008). Torbic et al., (2010) have considered a wide range of intersection characteristics such as the width of crossing, raised pedestrian crosswalks, marking at crossings, median refuge islands, raised intersections, pedestrian-related signings, pedestrian signal types. In the study, it was revealed that the maximum number of lanes crossed by a pedestrian has a significant effect on a pedestrian crash. While a recent study by Zegeer et al., (2017), to develop crash modification factors, identified that roadway width and lane width are vital risk factors for pedestrian crashes.

Based on a study on 262 signalized intersection in Hong Kong, it was found that the presence of an exclusive pedestrian signal at all crossings of the intersection will reduce the risk of pedestrian crashes by 43% (Xie et'al., 2018). A study by Lalani, 2000 (as cited in Harwood et al, 2008), found that the presence of refuge islands related to fewer pedestrian crashes.

Lee (2018) has considered pavement condition, sidewalk width, and presence of a sidewalk barrier as an exposure variable for pedestrian activity and the presence of a sidewalk barrier as an explanatory variable for pedestrian crashes. The model results showed that intersections with sidewalk barriers, wider sidewalk width and good pavement condition are associated with higher pedestrian activity. At an intersection where sidewalk barrier is installed, there is a higher probability of having a large number of pedestrians (Lee, (2018). In the case of Addis Ababa, at the location with a huge number of pedestrians (e.g. Megenagna and Mexico intersections) and at intersections in the ring-road, sidewalk barriers are installed. In addition, volunteers are also involved to make sure pedestrians will only walk within the guide rails. Even though Lee, (2018) stated no causal relation between sidewalk width and pedestrian crashes a study by Siddiqui et al 2012 (cited in Chen, 2016) found a negative relation between them. This seems reasonable because wider footpaths will reduce mixed traffic, for example, if the footpath is wider the tendency of a pedestrian to walk on the vehicle roadway will be very low. The effect of pavement condition on pedestrian crashes was not widely studied. In the case of Addis Ababa road maintenance and poor road conditions are prevalent. In some cases, poor road conditions on pedestrian walkways may be the reason for pedestrians to use the vehicle roadway which then leads to a high risk of crashes. It is possible for poor road conditions to be one of the factors associated with crash occurrences. So, studying the association of pavement condition to pedestrian crash occurrence is vital.

A meta-analysis study by Elvik et al., (2019) has summarized the estimates (crash modification factor) of explanatory variables (intersection characteristics) and presented that the presence of bus stops will result in a 114% increase in crashes while presence of a median was related to decrease crashes by 44% (Elvik, 2019). Even though it decreases the number of pedestrian crashes there is a chance of an increase in the total crash number as vehicles will crash to the raised refuge island (AASHTO, 2010).

The effect of a traffic signal on pedestrian safety was studied by Xie et al (2018); the study considered total cycle time, the number of signal stages, presence of right-turn pocket, and presence of exclusive pedestrian signals. The model results showed that an exclusive pedestrian signal has a significant relation with pedestrian crash frequency. It further stated that the provision of an exclusive pedestrian signal at intersections will bring a 43% reduction in pedestrian crash frequency, this finding can be related to the reason that an all-red-signal prohibits vehicles from entering the intersection and facilitates pedestrians crossing the road.

In the table below, explanatory variables that have been found to have a significant relation with pedestrian crash risk and frequency are summarized and proposed for use in model development for this study.

Table 1 Identified important exposure and explanatory variables for pedestrian crash risk and frequency estimation from previous researches

Category	Variables	Unit	Buffer zone	Reference
Traffic characteristics	Traffic volume	AADT	n/a	Torbic, (2010)
	Pedestrian volume	AAPT	n/a	Torbic, (2010)
	Posted speed limit	Km/hr	n/a	Torbic et al., 2010; Chen, 2016
Intersection characteristics	Width of crossing	meters	n/a	Torbic et al., 2010; Zegeer et al., 2017
	Number of lanes to be crossed	1 to 6	n/a	Elvik, 2013; Lee, 2018; Torbic, 2010
	Presence of raised median refuge islands	yes /no	n/a	Harwood et al., (2008)
	Pedestrian related signings	yes /no	100m	Xie et al., (2018)
	Average pavement condition	Good/fair/bad		Lee et al., (2018)
	Average sidewalk width	meters		Chen et al., 2016; Lee et al., 2018
	Presence of sidewalk barrier	yes /no		Lee et al., (2018)

2.3.3 EFFECT OF BUILT ENVIRONMENT AND LAND-USE

Previous studies have shown a significant relation between land-use characteristics and pedestrian crash risk and frequency (Miranda-Moreno et al., 2011; Pulugurtha & Sambhara, 2011; Chen & Zhou, 2016; Ding et al., 2018; Mansfield 2018). However, in terms of sign (positive or negative relation) of significance the findings are not consistent (Chen, 2016). In recent studies, the effect of built environment on pedestrian crashes has been studied from several perspectives including its effect on pedestrian activity (Miranda-Moreno et al., 2011; Mansfield et al, 2018), its zonal effect on predicting crash counts (Torbic et al., 2010; Pulugurtha & Sambhara, 2011; Chen & Zhou, 2016; Lee, Abdel-Aty, & Cai, 2017, Mansfield et al, 2018). As far as the same variable is considered, the effect of these variables on road segment and intersection, for spatial based studies is similar, which will allow to include studies on road segment for intersection studies (Cai et al., 2018).

Based on a model developed from 176 signalized intersections in the city of Charlotte, North Carolina, the increase in population, transit stops and the number of approaches at intersections were identified to result in an exponential increase in pedestrian crashes (Pulugurtha & Sambhara, 2011). While pedestrian crashes will tend to be less in areas such as single-family residential areas, urban residential areas, commercial center areas, and neighborhood service district areas. Contrary to the idea that an increase in pedestrian activity will lead to an increase in pedestrian crashes (Lee, 2017), here Pulugurtha

& Sambhara (2011) infer to the increase in alertness and attention by vehicle drivers, in areas where there is higher pedestrian activity, which indirectly increases pedestrian safety in those areas.

Several macro-level crash prediction model studies have indicated that demographic and socioeconomic zonal characteristics have an influence on traffic safety (Lee, 2017). Therefore, there is a need to incorporate their influence in crash prediction models. There are two methods to collect macro-level data (e.g., population density, proportions of specific age groups, commuters who walk, or commuters using a bicycle, etc.) and demographic characteristics: (1) using the existing geographic units zoning (e.g. Traffic Analysis Zone(TAZ), country) and (2) by creating a buffer zone around each intersection. The prior method for collecting macro-level demographic and land-use data was applied in various studies of pedestrian crash prediction (Pulugurtha et al., 2013; Cai et al., 2016; and Lee et al., 2017). A study by Lee et al., (2017), showed the influence of macro-level demographic factors (e.g. population density, median house income, proportion of age group) by developing three different crash prediction models: (1) with micro-level variables only, (2) with micro and macro-level variables and (3) micro and macro-level variables with random effect, over a variety of spatial units (census-based, traffic-based, or political boundaries) and different types of crashes (total crashes, bicycle crashes, and pedestrian crashes). Then it is revealed that the pedestrian crash model showed the best performance with census tract-based data ($\rho^2 = 0.170$) but also traffic analysis zone (TAZ)-based data can offer equally good performance ($\rho^2 = 0.1679$)[5].

The effect of built-environment and land-use characteristics on pedestrian activity is different from the buffer dimension used to collect the data on those variables (Miranda-Moreno, 2011). A study by Miranda (2011), considered the effect of land-use (commercial, park, industrial, residential, number of school, …), demography (population, worker, student, senior), transit characteristics (number of bus stops, presence of metro stations) and road network connectivity (length of road, road class) by using buffer zones of 50m, 150m, 400m and 600m. The study analyzed 519 signalized intersections with 5 years of crash data. The author reported that commercial land-use and number of bus stops are significantly related to pedestrian activity at 50m buffer zones, while the number of schools, number of employments, presence of metro stations, percent of major arterials and average street length are significantly related to pedestrian activity at 400-meter buffer zones. Based on the pedestrian crash prediction model they developed, commercial area, number of bus stops and schools are positively related to pedestrian crashes. Whereas, a study by Torbic et al., (2010), using a 300-meter buffer zone, identified the presence of bus stops, schools and parking as a significant predictor for a pedestrian crash. In the study by Miranda-Moreno et al., (2011), the important buffer zones with a significant correlation between pedestrian activity and the macro-level variables are indicated (Table 2).

Table 2 Significant variables for pedestrian activity with the respective buffer zone (log-linear model for pedestrian activity) (Miranda-Moreno, 2011)

Variables	Buffer	Parameter estimates	Elasticities
Intercept		4.08***	
Population (in 1000's)	400m	0.08**	0.34
Commercial (1000's)	50m	0.19***	0.20
No. of jobs (1000's)	400m	0.17***	0.28
No. of schools	400m	0.15***	0.20
Metro station	400m	0.33***	0.28

[5] ρ^2 = McFadden's ρ^2

No. of bus stops	150m	0.11***	0.37
% of major arterials	400m	-0.71**	-0.16
Average street length	400m	0.95**	0.46
Intersection indicator		0.36***	0.31
Goodness-of-fit		R^2=0.55	

[a] *4-legged intersection = 1 and 3-legged intersection = 0*

[**] *statistically significant at 5%*

[***] *statistically significant at 10%*

In order to extract data on built-environment and demographic characteristics, Pulugurtha et al., (2011) suggest to set the buffer width in accordance with the acceptable walking time of pedestrian in the study region and stated a range between 400 meters to 1.6km. Thus, in the study, they collect the data at 400m, 800m, and 1.6km, and developed a model for each buffer width. Even though the sign (positive or negative) of the relation between pedestrian crashes and the built environment variables stayed the same, the model for data collected at 800meter buffer width had the lowest QIC[6]. Thus, the author recommended to use an 800-meter buffer width for land use, and transit stops data collection. However, the study by Miranda (2011), contradicts the use of a uniform buffer width for all variables, as his study showed that the effect of some variables on pedestrian activity differs with the buffer width considered (Table 2).

Among the most recent studies on built environment effect on pedestrian safety, Mansfield (2018) has studied the effect of built-environment on fatality risk using census tract categorization for urban (50,027) and rural tracts (22,711) in United States. The study aimed to identify the effect of traffic characteristics and the built environment. A total of 25,251 traffic fatalities, collected for four years on roadway segments was used. Built environment variables used in the study include residential population density, employment density (office, retail, industrial, transportation, general services, and entertainment and food/accommodation services), activity mix index and work commute (%). According to the study, a higher density of retail employment indicates the likelihood of the tract to be categorized under the not-always-zero tract. Whereas, higher density of office, industrial and general services in the tract will infer the likelihood of the tract being categorized under always-zero tract. Always-zero tract means a tract with a high chance of not having pedestrian fatality. And these findings are consistent with a study done in Seattle by Chen & Zhou, (2016).

The results of the average marginal effects of the study, also, with an intention to show the sensitivity of fatality risk with change in traffic volume and land-use characteristics, revealed that retail employment density and entertainment (food/accommodation) services employment have a positive effect on pedestrian fatality risk, which is in line with the finding of Ding, Chen, & Jiao, (2018). Despite that, findings on the effect of residential land-use are not constant among studies. Mansfield et al, (2018) found a negative association with regard to office, general service employment, and residential population density. Similarly, Pulugurtha & Sambhara,(2011) have found a negative relation between the single-family residential areas and pedestrian crash frequency. However, Miranda-Moreno et al., (2011), pointed out that there is a lack of consistency in the effect of residential land-use as the buffer radius of signalized intersections changed.

[6]Quasi-likelihood under independence criterion (QIC)

Road user-related problems are one of the main causes of crashes. And under the road user-related problems, driver and/or pedestrian impairment are among them. Several studies have been conducted concerning driver impairment. Relative to driver impairment studies there are few studies on the impact of pedestrian impairment, especially alcohol-related, for crash occurrence (Ortiz et al., 2017). One type of pedestrian-related impairment is alcohol use or intoxication. Accessibility of alcohol establishment will imply the possibility of pedestrian impairment or intoxication, which will increase the risk of involvement in a crash. Some studies done on characteristics of alcohol-involved pedestrian crashes point out that these crashes mostly occur in high-speed zones, during the night, on straight roads and on-road segments (Hezaveh et al., 2018). Hezaveh & Cherry, (2018), further stated, the main reason for these crashes to occur at an intersection to be a driver turning maneuver (failure to give right of way to pedestrians) and/or aberrant pedestrian behavior resulting from alcohol use. However, more research is needed to understand the character of these crashes (alcohol-related pedestrian crashes) at intersections.

Lascala et al., (2000) has considered pedestrian impairment in the study, where the availability of alcohol in bars and restaurants is spatially related to pedestrian injury data. The author reported that alcohol establishment around the road facilities has a direct relation to pedestrian injury crashes. A similar implication was presented by Torbic et al (2010), even though the analysis was done at the intersection level. The study developed a pedestrian prediction model and CMF for some variables on 4-legged signalized intersections and found that the effect of alcohol sales establishment, at a 300m buffer zone, was statistically significant at 84% confidence interval. According to the crash modification factor developed in the study a 56% increase in crashes was estimated when the number of alcohol sale establishments (alcohol shop, restaurants, clubs) is more than nine (9).

Transit density in the vicinity of an intersection is associated with increased pedestrian crash frequency and risk (Miranda-Moreno et al., 2011; Pulugurtha & Sambhara, 2011; Chen & Zhou, 2016; Ding et al., 2018). Miranda-Moreno et al., (2011) consider pedestrian activity as a risk exposure for pedestrian crash occurrence. The study assessed the effect of bus stops and the presence of metro stations on pedestrian activity and pedestrian crash occurrence. Using the elasticities of two models – a pedestrian activity model and a crash frequency model (without built environment characteristics), it was found that 100% increase in bus stop number and presence of metro stations around an intersection will result in 37% and 28% rise in pedestrian activity, respectively (Miranda-Moreno, 2011). On the other hand, with respect to pedestrian crashes, doubling the population and employment density, and bus stops around the intersection will raise the crash frequency by 45% (Miranda-Moreno, 2011).

According to Pulugurtha et al., (2011) and Ding et al., (2018), the effect of transit density on pedestrian crashes is non-linear. Ding et al., (2018) has studied the contribution of contributing factors and their non-linear effect on pedestrian crashes (Figure-5). According to the study, the bus stop density, at TAZ[7] level, has 1.91% contribution to pedestrian crashes and has a nonlinear relation with pedestrian crashes - a rapid increase in pedestrian crashes is seen from 0.1 to 0.15 bus stops per hectare and the reverse till 0.3 bus stop per hectare, which then shows a gradual increase till 0.5 bus stop per hectare.

[7] Traffic Analysis Zone

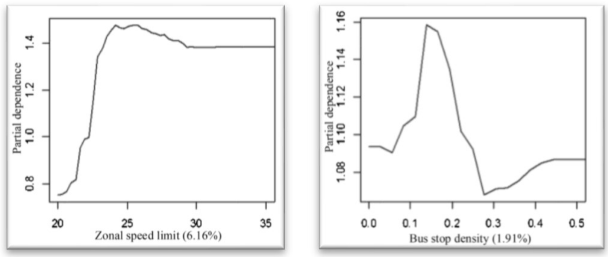


Figure 5 Non-linear effect of intersection characteristics on pedestrian crash frequency Ding et al., (2018).

Ding et al., (2018), applied the Multiple Additive Poisson Regression Trees (MAPRT) methodology to identify the relative importance of pedestrian crash contributing variables and to identify the non-linear effect of these variables on pedestrian crashes for 863 TAZ considered. The study ranked contributing factors in their relative contribution to pedestrian crashes and ranked them based on relative importance. According to the rank, the top 5 contributing factors are the number of trips, household density, commercial land-use, land-use mixture, and speed limit, with relative importance (%) of 23.78, 14.38,13.16,11.39 and 6.16 respectively. The graph below illustrates the relative contribution of variables with different levels of tree complexity.

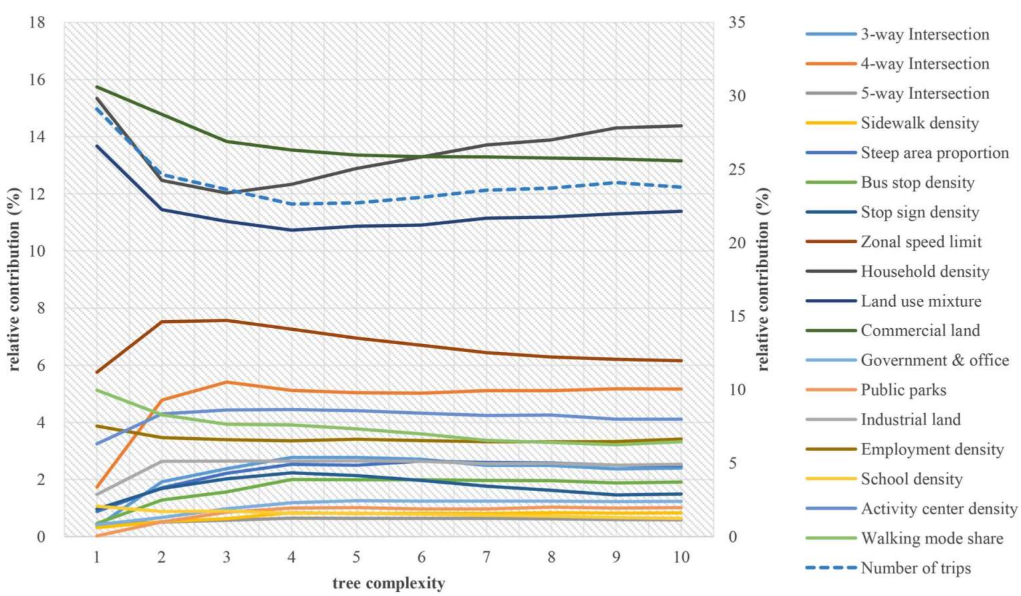


Figure 6 Relative contribution of variables with different level of tree complexity (Ding et al., 2018)

In conclusion, the following explanatory variables (table-3) for land-use, built-environment characteristics and socio-demographic characteristics of intersections are identified and proposed for the model development based on two criteria. First, based on the finding of previous studies. Second, based on the feasibility to collect those data at the study area of this research.

Table 3 Identified important land use, built-environment and socio-demographic factors for pedestrian crash risk and frequency estimation from previous researches

Category	Variables	Unit	Reference
Land-use			

	Commercial, Residential, Industrial, Mixed land-use, Gov.t, and office	%	Chen, 2016; Mansfield et al, 2018; Ding et al., 2018
Socio-demographic			
	Income (Neighborhood average per capita income)	Low, high	Torbic;2010; Mansfield et al, 2018
Built-environment	Transit density (# of bus/taxi stop)	Count	Chen, 2016; Torbic, 2010; Pulugurtha et al., 2011
	School density	Count	Wang et al., 2013; Ding et al., 2018
	Number of alcohol sales establishments	Count	Lascala et al., 2000; Torbic et al., 2010
	Presence of curb parking		Torbic et al., 2010; Xie et al., 2018
	The presence of LRT stations	Yes/No	Miranda et al., 2011

2.4 PEDESTRIAN CRASHES AND THE REGRESSION METHODS

2.4.1 DATA AND METHODOLOGICAL ISSUES

Issues related to data and methodology in crash-frequency-modeling researches have been studied and it has been found that these issues are the main source of error for incorrectly specifying statistical models. This leads to an erroneous crash frequency and wrong inference of explanatory variables considered (Lord et al., 2010). Based on the review and explanation done by Lord et al., (2010), these methodological and data-related issues are explained.

Over-dispersion: In most cases, the variance of crash count data is larger than the mean. This will cause erroneous estimation in a situation when the modeling approach assumes equal mean and variance is used. For example, using a Poisson regression over-dispersed data set will lead to a biased estimate (Cai, 2016).

Under-dispersion: In most cases, crash data is not characterized by under dispersion, but sometimes the sample-mean value will be less than the value of variance, especially when there is a lot of zero observation which results in a low sample mean (Daniels, 2011).

Time-varying explanatory variables: As crash data are collected over a certain time period and data on explanatory variables are collected for a single time period, mainly due to lack of data for whole time periods, the model developed ignores the potential within-period variation of explanatory variables. Due to this unobserved heterogeneity, the model's estimation of the contribution of explanatory variables on crash frequency will be biased (Lord, 2010).

Omitted-variables bias: some researchers prefer to use few variables for simplifying models (for example developing a model with traffic and pedestrian volume variables only). However, several traditional predictive models excluding important explanatory variables will lead to a biased estimate of pedestrian crash count (Lord, 2010).

Functional form: The functional form of the model determines the relation between dependent and independent variables in the model. In the past researches, it was used to assume a linear relationship, while most recent studies are considering a nonlinear relation(Lord, 2010).

2.4.2 MODELING APPROACHES

In order to alter the methodological issues related to crash frequency data several methods have been used over the years, such as Poisson, Negative binomial, Poisson-lognormal, Zero-inflated Poisson, and Negative Binomial, Gamma, Generalized estimating, Random-effects, Random-parameters equation, Bivariate/multivariate models(Lord, 2010). The table below summarizes models used by recent researches related to crash prediction modeling.

Table 4 Summary of previous research analyzing crash-frequency data

Model type	Researches
Poisson	Geyer et al. (2006), Daniels, (2011), Wali et al., (2018)
Negative binomial	Torbic, (2010**)**, Daniels, (2011**)**, Pulugurtha, (2011**)**, Miranda-Moreno, (2011**)**,(Siddiqui et al., (2012**)**, Elvik, (2013**)**, Strauss, (2014**)**,Mooney et al., (2016**)**, Thomas et al., (2017**)**, Lee, (2018**);** Chimba et al., (2018)
Gamma	Daniels, (2011)
Random Parameter Negative Binomial	Wang, (2017)**;** Wali et al., (2018)
Random Parameter Poisson	Wali et al., (2018)
Poisson-lognormal	Siddiqui, (2012**);** Xie, (2018**)**
Zero-Inflated Negative Binomial	Cai, (2016)**;** Lee, (2018)
Mixed-effects negative binomial	(Lee, 2017)
Multilevel Poisson-lognormal (MPLN) joint	**(**Alarifi, 2017)

Poisson regression model

As crash data is a non-negative integer, least-squares regression cannot be used. In addition, least-square regression assumes a continuous dependent variable. Given that, at the beginning of crash modeling, most studies used the Poisson regression model (Lord, 2010). In Poisson regression, the probability of a crash ($\boldsymbol{y_i}$!) occurring at road facility (road segment/intersection) $\boldsymbol{i}$ per some time period is given by:

$$\boldsymbol{p(y_i)} = \frac{\boldsymbol{EXP(-\lambda_i)\lambda_j^{y_i}}}{\boldsymbol{y_i!}} e^{\Sigma y_i x_i} \qquad (1)$$

Where, $\boldsymbol{\lambda_i}$ is the Poisson parameter for the roadway entity $\boldsymbol{i}$**,** which is equal to the roadway entity expected number of crashes per year, the most common functional form of $\boldsymbol{\lambda_i}$ is $\boldsymbol{\lambda_i = EXP(\beta X_i)}$**,** where X$\boldsymbol{i}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of estimable parameters (Lord, 2010).

Even if the modeling approach served as a starting point for crash data analysis, it was found, in recent studies, problematic because it cannot handle over- and under-dispersion. Furthermore, it is significantly affected by low sample-mean and small sample size and can lead to biased results (Lord, 2010).

The negative binomial regression model

In order to overcome the problem of over-dispersion in crash-data, a new model called Negative Binomial, which assumes the Poisson parameter to follow a gamma distribution, was introduced. For the negative binomial model, parameter $\boldsymbol{\lambda_i}$ will be rewritten as

$$\boldsymbol{\lambda_i = EXP(\beta x_i + \varepsilon_i)},$$

where $\boldsymbol{EXP(\varepsilon_i)}$ is a gamma-distributed error term with mean 1 and variance α. The addition of α (over-dispersion parameter) allows the variance to differ from the mean, which in the case of Poisson regression variance was equal to mean,.

The Negative Binomial model is the most frequently used approach in crash-data modeling (Lord, 2010). However, it cannot handle under-dispersed data.

Poisson-lognormal model:

This modeling is recommended, as an option to the Negative Binomial model, by some researches. This approach enables more flexibility than NB even though model estimation will be more complex (Lord, 2010). Contrary to negative binomial modeling the error term ($\boldsymbol{EXP(\varepsilon_i)}$) in the Poisson-lognormal model follows a lognormal distribution. Similar to NB, it is affected by a low sample-mean and a small sample size and can lead to biased estimates (Lord, 2010).

Zero-inflated Poisson or negative binomial

This modeling approach aims to tackle issues related to excess zero in crash frequency data. Zero-inflated models enable splitting datasets into two parts, one crash-free (that accounts for the excess zero data which Poisson/ negative binomial models cannot handle to model), and the second one, crash-prone tendency of a roadway facility. The logistic regression model (Xie, 2018) or probit model can be used for determining the probability of the roadway element being zero or non-zero (Lord, 2010). For the issue described the model has been widely used by researchers (Cai, 2016**;** Lee, 2018). However, this approach is affected by a low sample-mean and a small sample size (Lord, 2010).

Gamma model:

The gamma model was introduced by Oh et al. (2006) (as cited in Lord, (2010)), in order to deal with crash data with under-dispersion characteristics. This model can address data with over-and-under dispersion characteristics. Lord, (2010) has indicated that the model has limited use by researchers. This might be related to the rareness of under-dispersed crash data, but it was observed by Daniels, (2011) during the roundabout crash modeling study.

2.5 CONCLUSION

In conclusion, as evidenced by previous studies, it was shown that traffic characteristics (traffic volume, pedestrian volume, posted speed limit), intersection characteristics (width of crossing, number of lanes to be crossed, presence of raised median refuge islands, pedestrian-related signings, average pavement condition, average sidewalk width, presence of sidewalk barrier), built-environment (transit density, school density, number of alcohol sales establishments, presence of parking, the presence of LRT stations), land-use and socio-demographic characteristics (neighborhood income) in the vicinity of intersections may result in possible pedestrian crash involvement. Further, the effect of those variables differs with the buffer width at which the variable is extracted (such as 150m for transit density, 300m for the presence of parking and number of alcohol sales establishment, 400m for school density and presence of LRT station). Regarding the modeling approach, a wide range of modeling approaches are used depending on the crash data characteristics and the assumed functional form. The widely used modeling approach is negative binomial distribution, which mainly overcomes the problem of overdispersion in crash data.

3 METHODS

3.1 STUDY AREA

This study is carried out on 34 signalized intersections located in Addis Ababa, Ethiopia (figure 7). Addis Ababa is one of the fast-growing cities in sub-Saharan Africa with rapid urbanization (AARTB & ITDP, 2018). Based on a report by UN-Habitat, the city is home to 17% of the country's urban inhabitants: 3.2 million inhabitants, which is estimated to reach 4.7 million by 2030 as cited in AARTB & ITDP (2018). Addis Ababa constitutes of 10 sub-cities and 99 district administrations under the 10 sub-cities. Addis Ababa is proposed for this case study because of the steadily increasing traffic fatality numbers and high number of pedestrian crashes on crossings.

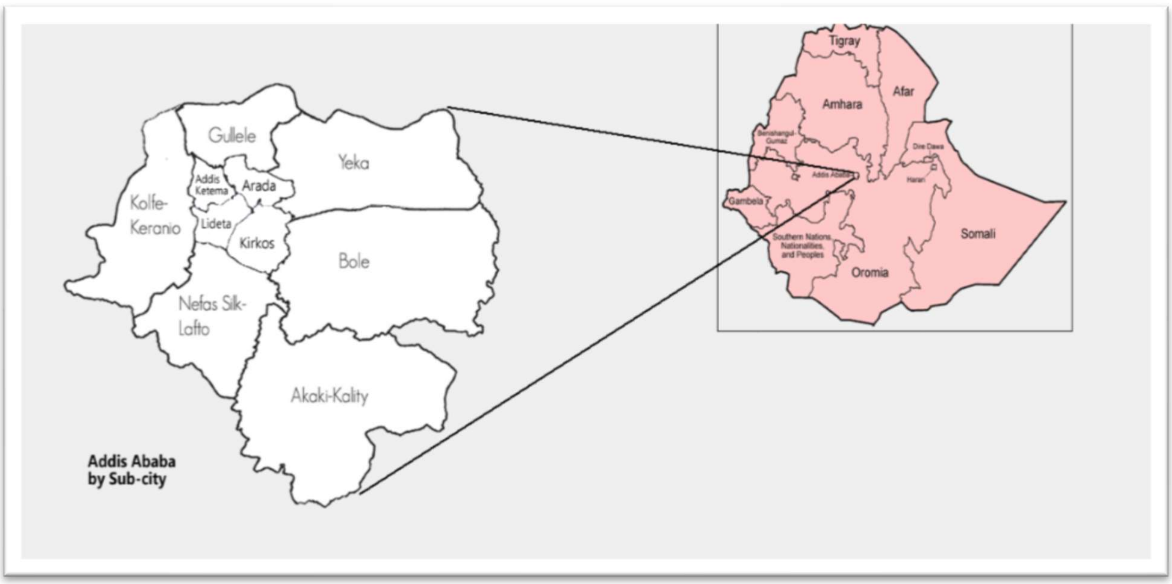


Figure 7 Map of Addis Ababa: from Addis Ababa city administration official website ("City Map - aaca," n.d.)

3.2 STUDY DESIGN

"Cross-sectional models are usually used to analyze pedestrian crash counts aggregated over multiple years and time series analysis is rarely used to account for temporal autocorrelations inside pedestrian crash counts" (Ding, 2018). A cross-sectional study design was used to conduct this study. The study was done on signalized intersections that are selected by purposive sampling; based on data obtained from AARTB there was a small number of signalized intersections if five years of crash data was used, so by reducing the years of crash data to 3 years (2017, 2018 and 2019) signalized intersections already installed in the considered crash year period were selected for this study.

The flow chart shown below illustrates the proposed methodology for the study.

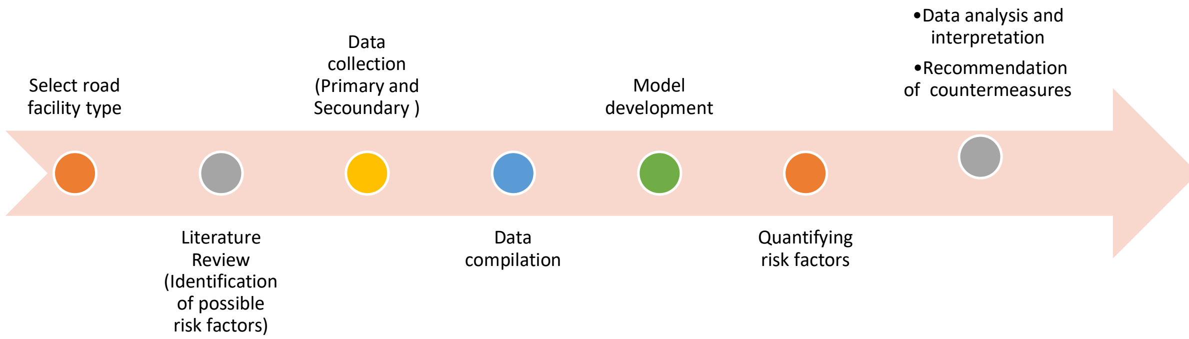


Figure 8 Study design flow chart

3.3 STUDY POPULATION

Step1: Select road facility type

As part of this phase, the study area, facility, location type, traffic control type, and target crash types will be defined. The study has been conducted in Addis Ababa, Ethiopia. In Addis Ababa, there are 183 junctions (including roundabouts (65), signalized intersections (47), non-signalized intersections (71)) connecting all kinds of road types. Due to limited resources and time, it was not possible to cover all the junction types in this study. Also, since the Addis Ababa City Administration was implementing the "Safe Intersection Program", several roundabouts in the city were being changed to a signalized intersection (AARTB & ITDP, 2018). Hence, a study focusing on the signalized intersection would have futuristic input in improving safety at intersections. Moreover, to collectively deploy the available resources to the study, focusing on a case that is in line with the government's policy and strategies would make the study valuable. Further, it will give quality time and resources for a deep dive into the specified area of study. As such, the study was done on crashes involving pedestrians at signalized intersections in Addis Ababa.

As there exists a probability that treated sites may exhibit changed vehicle-pedestrian interaction, it could cause a bias in the crash prediction model. Thus, an intersection that has undergone treatment in the years of crash considered was excluded from the study. Data on traffic signals installed were gathered from Addis Ababa Road and Transport Bureau (AARTB) and currently, there are 47 signalized intersections (figure 9). Initially, five years of crash data were aimed for modeling, yet only 19 intersections were found signalized as of 2015, which is a small sample size for modeling. Whereas, by reducing the years of observed crash data to three it was possible to study 36 signalized intersections. Due to the inconvenience caused during data collection on two intersections 34 intersections were surveyed. Therefore, to have a larger sample size with feasible years of observed crash data - three years of crash data at 34 signalized intersection was collected. A list of selected signalized intersections for the study is presented in appendix A.

Step 2: Identify possible risk factors/variables

At this stage of the study, possible risk factors were identified. Three main criteria that are indicated in the Highway Safety Manual (2010) were considered for the selection of the variables: (1) variables, from previous studies, that are found to have a major influence on the number of crashes, (2) the variables can be measured reliably, and (3) the variables exhibiting high correlation within explanatory variables should be avoided. Thus, variables qualifying the first two criteria are filtered out, whereas the variable selection using the third criterion will be done in step 4.

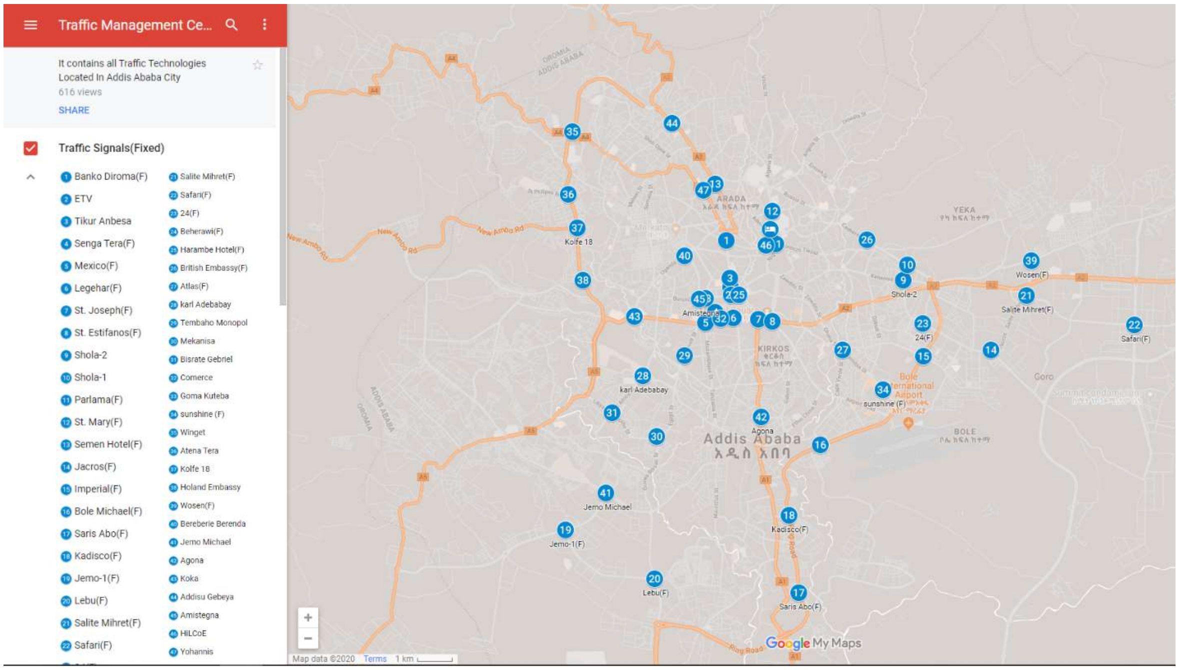


Figure 9 Signalized intersections in Addis Ababa (AARTB, 2019)

Based on the literature review and convenience of data collection 23 variables under four main categories have been identified: exposure variable, intersection characteristics variable, built environment variables, and land-use variables. The variables are listed in Tables 5 and 6.

3.4 DATA COLLECTION AND PREPARATION

<u>Step 3: Data collection and compilation</u>

The study deals with the development of a predictive model, and it relays a lot with several types of data. Both primary and secondary data were collected. The data has been collected for pedestrian crashes and 23 variables under the four categories listed above. The data collection format is shown in Appendix (B). Below the data collection is described in detail.

3.4.1 PEDESTRIAN CRASH DATA

The model is developed base on 3-years (2017, 2018, and 2019) of pedestrian crashes. The data was collected from the Addis Ababa Police Commission. According to the Highway Safety Manual, for a crash to be considered as an intersection crash it should be located within the 50 to 100-meter buffer zone of the intersection. Addis Ababa Police Commission has only some recent crashes with GPS location. Whereas for most crashes the location is designated by well-known places such as nearby buildings, schools, factories, hotels. Thus, the author locates the crash by considering the distance of the designated place of crash from the center of the intersection. A total of 244 crashes have been located within the 100-meter buffer zone of the 34 intersections.

3.4.2 EXPOSURE DATA

Exposure data were collected from both primary and secondary data sources. For 17 intersections, hourly (morning, mid-day, and afternoon) pedestrian volume data were obtained from AARTB. Whereas for vehicle volume data; 12-hour count data was obtained for 10 intersections and three-hour peak volume data was obtained for 15, both from AARTB.

The study at signalized intersections by Miranda-Moreno et al., (2011) collected three hours (peak morning, noon, and afternoon) vehicle and pedestrian volume data on normal weather conditions and weekday (Miranda-Moreno et al., 2011). Whereas, for this study, due to resource and time constraints two peak hour data, morning and afternoon peak hour data were collected for pedestrian and traffic volume. Based on the local knowledge, it was noticed that the peak-hour of an intersection was dependent of the location of the intersection - i.e. an intersection distant from the center of the city tends to have an earlier morning peak than those intersections at the center of the city. So, this situation has been considered when selecting the peak-hour for traffic data collection.

A traffic count was done on March and April, 2020. First, by video recording the traffic flow in the field, by installing two video cameras in two opposite directions, that ensured full coverage of the intersection area. Then, the traffic counts were performed in-office by trained traffic volume data collectors. Finally, the collected daily traffic volume was converted to ADT. Despite relentless effort exerted by the author, it was indeed a challenge to get data indicating a seasonal, weekly, and monthly variation of traffic volume in Addis Ababa. Thus, this study uses ADT instead of AADT. Indeed, ADT has been used for modeling in several works of literature and indicated as an alternative predictor in Highway Capacity Manual (2010).

The traffic counter found near intersections were used to estimate the Average Daily Traffic volume from the peak hour volume at the intersection. This was based on the approach indicated by the National Roads Authority (National Roads Authority, 2012). For this study, traffic counter refers to a location with 12 hours or more of traffic counts, located near the study intersection or location having a similar road type with the study intersection. And there was no major road intercepting the locations which might result in a significant difference in traffic flow between the traffic counter and the intersection. A total of 19 (9 traffic volume counts at road corridors and 10 traffic volume counts at intersections) traffic counters were used to convert the three-hour volume data to average daily traffic volume. Traffic volume data for 4 intersections (Kadisco, Imperial, Saris abo and Lebu) were counted in 2017, to address this, a growth factor of 0.35 was used to compute the traffic volume in 2019. The growth factor was suggested by a consultant company that performed the traffic count for the intersections.

$$ADT_i = \left(\frac{Q_i}{Q_{TC}}\right) \times (ADT_{TC})$$

Where: ADT_x=Average daily traffic at the location (intersection) i

ADT_{TC}= Average daily traffic at traffic counter

Q_i= Short period (morning or afternoon) peak traffic flow at intersection i

Q_{TC}= Short period (morning or afternoon) peak traffic flow at traffic counter

Whereas, for pedestrian volume, it was not possible to obtain traffic counts performed for 12hr for pedestrians to estimate the daily volume of pedestrians entering the intersection. Thus, the volume of pedestrians entering the intersection for one hour during peak period was used for modeling.

3.4.3 INTERSECTION CHARACTERISTICS DATA

Ten variables describing the characteristics of the intersection were collected, from March to April 2020, through on-site observation and measurement. The data collected includes the number of lanes crossed by the pedestrian on the major and minor road; the width of crossing for the major and minor road; average sidewalk width, presence of a sidewalk barrier, presence of a median refuge island, presence of pedestrian-related signings, presence of a dedicated right lane and pavement condition.

3.4.4 LAND-USE DATA

The land-use map of Addis Ababa was obtained from AARTB. For a buffer width of 800meters (Torbic et at.,2010), the proportion of land-use variables was calculated. ArcGIS overlay functions were used to calculate (quantify) the proportion of land-use types within the buffer zone of the intersection (Chen, 2016). These functions helped to measure the area covered by certain land-use types within the buffer zone of the respective intersection. This was done by overlaying the land use map over the road network map (with the buffer zone drawn on the selected intersections). The land-use data provided by AARTB were categorized into 6 land-use types: Commercial, high-density mixed residential, Medium and low-density mixed residence, government and offices, and social services. The Social services land-use was developed by summing the areas of schools, hospitals, and health centers areas together. The figure (10) below illustrates land-use overlay at intersections with a 800-meter buffer zone.

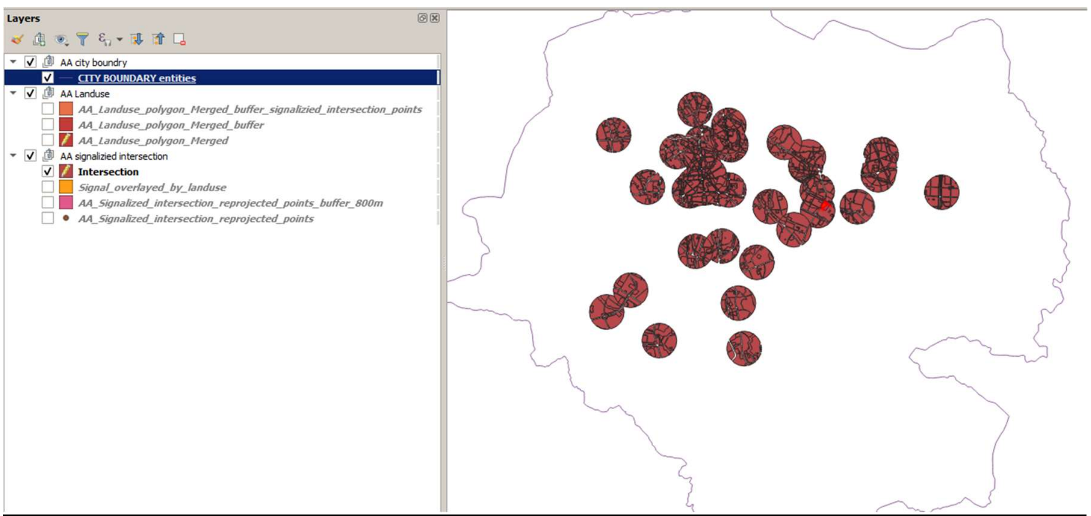


Figure 10 overlay of the land use map over the road network map with 800-meter buffer zone

3.4.5 BUILT-ENVIRONMENT CHARACTERISTICS DATA

Based on a review of literature, five built environment characteristics were identified to be studied. These are transit density (number of bus/taxis stops), school density, number of alcohol sales establishments, presence of parking, and presence of LRT stations. The data was collected through site observation done ,2020 and by locating the respective buffer zone of the variable using the google earth measuring tool. The buffer width, in which the explanatory variables were extracted, was chosen based on a review of previous work in literature. The buffer zone and the unit of measurement for variables are shown in Appendix C.

3.5 MODEL DEVELOPMENT

<u>Step 4: Model development</u>

<u>Correlation</u>

Checking for a high correlation among explanatory variables is crucial to identify the important contributing factors in the first step (Miranda et al., 2011). To avoid the problem of multicollinearity, the Pearson coefficient of correlation and variance inflation factor (VIF) were used. The Pearson coefficient of correlation identifies the correlation between two variables. If the Pearson correlation coefficient value is between -0.3 and +0.3, it implies a weak correlation between variables (Pulugurtha et al, 2011). If two variables are found to have a strong correlation, the one with weakest significance to pedestrian crash involvement was excluded from the model. Whereas, VIF will be used to identify collinearity among variables in the fitted model. A VIF value greater than 10 indicates major multicollinearity problems among variables (Chen et al., 2016). In case of higher VIF values, the variable with highest value will be discarded from the model. The analysis was made using SPSS software (Chen et al., 2016).

Develop a crash prediction model

At this step the predictive model will be developed. Accident prediction models relate the number of crashes with exposure variables (traffic volume) and explanatory variables (road geometry, land-use, demographic factors) (AASHTO, 2010). According to the Highway Safety Manual (2010), the generalized linear model (GLM) with a negative binomial distribution and log-link function is standard as modeling approach for pedestrian crash prediction. GLM uses a power function (for exposure variables) and exponential functions (for risk factors). The power function ensures that the predicted crash will only be zero when the exposure variable (traffic volume or pedestrian volume) is zero, otherwise the predicted crash will be a positive value. The exponential function ensures the crash number will not be zero or negative as a result of a zero or negative value from a linear predictor (regression of risk factors). In other words, even if the sum of the linear function becomes zero or negative, it does not result in a zero or negative crash value because the linear function is the exponent of the exponential function ($e^{\sum y_i x_i}$) – with the value of e^0 being one and e^{-x} being no negative value. Therefore, the generalized linear model (GLM) with a negative binomial distribution and log link function (equation 2) was used to develop a pedestrian crash prediction model in this study.

$$N_\rho = \alpha Q_1^{B_1} Q_2^{B_2} e^{\sum y_i x_i} \quad (2)$$

Where:

N_ρ= the predicted number of pedestrian crashes

Q_1 = vehicle traffic volume (AADT)

Q_2 =pedestrian traffic volume (ped/hr)

x_i= set of risk factors (explanatory variables)

α, B_1, B_1, y_i= model parameters

In the study by Miranda-Moreno et al., (2011), it is indicated that modeling demographic and land-use variables together with exposure variables might pose a multicollinearity problem. To check and evaluate the condition in this study two models were developed.

Model (1): N_ρ =f (traffic volume, pedestrian volume, variables indicating intersection characteristics). (3)

Model (2): N_ρ =f (traffic volume, pedestrian volume, a variable indicating intersection characteristics, land-use variable, built-environment variables). (4)

3.5.1 MODELING APPROACH

According to Olszewski et al., (2018) both Poisson and negative binomial distribution performed well based on goodness-of-fit measures, but he suggested using the Negative Binomial Model as it is flexible and versatile. One basic assumption of Poisson distribution is the mean of the data to be equal to the variance while in most practical cases the variance of pedestrian crashes is greater than the mean which will lead to overdispersion. Whereas, assuming a negative binomial distribution for the model fitting will have an advantage in avoiding overdispersion (AASHTO, 2010). Therefore, a generalized linear model (GLM) with a negative binomial distribution and log-link function will be used. SPSS software version 25 will be used to fit the model.

3.5.2 MODEL VALIDATION AND GOODNESS OF FIT

The main objective of the study is to develop a predictive model that will realistically estimate pedestrian crash frequency at signalized intersections. So, the model's goodness-of-fit and statistical adequacy need to be checked.

The leave-one-out cross-validation technique will be used to check the predictive performance of the model. This will be done by randomly selecting one sample which is used as a test set, and the remaining 33 are used as a training set. This process is repeated 34 times. Finally, the predicted crashes, using a model developed based on training sets, are compared with the observed number of crashes at the test sets by using the mean absolute deviation (MAD) and mean square predicted error (MSPE). “The MAD is the ratio of the sum of the absolute difference between an observed crash count and predicted mean value to the number of sites’’ (Mehta et al., 2013).

$$\text{MAD}= \frac{\sum_{i}^{n}|\hat{\mu}_i-y_i|}{n}$$

Where: $\hat{\mu}_i$= predicted number of crashes per year for site í.

y_i= observed number of crashes per year for site í.

n= number of sites

MAD helps to obtain the average variability of prediction; smaller values indicate a good predictive performance of a model (Mehta et al., 2013).

<u>Step 5: Quantifying risk factors</u>

At this stage, based on the estimates found from the model, the main risk factors are identified with their respective influence on the number of pedestrian crashes.

<u>Step 6: Data interpretation and recommendation</u>

Finally, the results are interpreted, and coefficients of explanatory variables compared with similar studies on pedestrian prediction models at signalized intersections. Once the main risk factors are identified, the appropriate countermeasures and recommendations to improve the safety of pedestrians are forwarded.

4 DATA ANALYSIS

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

4.1 DESCRIPTIVE STATISTICS

The mean, minimum, maximum, and standard error values of the identified variables are summarized in tables (5) and table (6). A total of 34 signalized intersections in Addis Ababa were studied.

Table 5 Descriptive statistics of the Response variable, Exposure variables and Intersection characteristics variables (n=34)

Variables	*Type*	*Minimum*	*Maximum*	*Mean*	*Std. Error*
A. Response variable					
Total number of pedestrian crashes	Continuous	1	21	7.18	0.752
B. Exposure variable					
Traffic Volume (ADT)	Continuous	16083	66216	42733	1924
Pedestrian Volume (peak hourly volume)	Continuous	977	5951	2990.3647	245.07233
C. Intersection characteristics					
Width of crossing (meter) Major road	Continuous	7.0	15.0	11.176	0.3680
Width of crossing (meter) Minor road	Continuous	7.0	12.0	8.632	0.3252
Average sidewalk width (meter)	Continuous	2	5	3.26	0.097
# of lanes to be crossed (one direction) Major road	Ordinal	Two lanes:3, Three lanes:21, Four lanes:8, Five lanes:2			
# of lanes to be crossed (one direction) minor road	Ordinal	Two lanes:19, Three lanes:13, Four lanes:2			
Presence of raised median refuge islands (YES/NO)	Dummy	Yes:19, No:15			
Presence of Dedicated Right turn lane road (YES/NO)	Dummy	Yes:8, No:26			
Pedestrian related signings (YES/NO)	Dummy	Yes:19, No:15			
Presence of sidewalk barrier (YES/NO)	Dummy	Yes: 16, No:18			
Pavement Condition	Categorical	Bad:12, Fair:13, Good:9			

Table 6 Descriptive statistics of variables in Built environment characteristics and Land-use (n=34)

Variable	Type	Minimum	Maximum	Mean	*Std. Error*
		Statistic	Statistic	Statistic	Std. Error
***D.* Built-environment characteristics**					
(# of bus/taxi stop)-150m	Continuous	1	8	4.65	0.329
School density (#)-400m	Continuous	1	9	3.53	0.354
Alcohol sales establishments (#)-300m	Continuous	0	18	9.03	0.851
Presence of curb parking-300m (YES/NO)	Dummy	Yes:29, No:5			
Presence of LRT stations-400m (YES/NO)	Dummy	Yes:12, No:22			
***E.* Land-use**					
Proportion of Commercial	Continuous	0.00%	47.14%	13.63%	2.80%
Proportion of HD Mixed Residential	Continuous	0.00%	74.90%	21.13%	3.28%
Proportion of Park/recreational	Continuous	0.00%	53.88%	15.09%	2.52%
Proportion of Med and low-density Mixed residence	Continuous	0.00%	91.52%	37.63%	5.83%
Proportion of Gov't and office	Continuous	0.00%	30.02%	4.21%	0.89%
Proportion of Social services (School, hospital, health center)	Continuous	0.00%	30.37%	8.29%	1.23%
Valid N (listwise)	34				

4.2 EXPLORATORY DATA ANALYSIS

4.2.1 CORRELATION

All variables in the study were analyzed in terms of their collinearity. Table (7) shows the result of the Pearson Correlation run in SPSS. A Pearson Correlation value of 0.6 was used as a threshold. Thus, for the variables with a correlation value above 0.6, one of the variables (the one with a better significance in terms of pedestrian crashes) was selected.

Based on the Pearson correlation analysis performed for all variables including the dependent variable, a high correlation (0.633**[8]) was found between pedestrian volume (.618**) and traffic volume (.793**)- Table (7). And pedestrian crash has correlation value of 0.618** and 0.793** significance value with pedestrian volume and traffic volume.

[8] **. Correlation is significant at the 0.01 level (1-tailed).

Table 7 Pearson correlation

Correlations

	Total_No_casuality	Veh_Vol	LnVehicleVolume	Ped_Vol	LnpedestrianVolume	IC_Width_C_Maj_R	IC_Width_C_Min_R	IC_No_Lane_Maj_R	IC_No_Lane_Min_R	IC_P_Raised_Med_RI_Maj	IC_P_Ded_Right_Lane_Maj	IC_P_Ped_related_Signing	IC_Avg_SW_Width	IC_P_SW_barrier	IC_Pavement_Cond	BI_No_Bus_Stop	BI_School_density	BI_NO_Alcoholsales_est	BI_P_Curb_Parking	BI_P_LRT_Stn	LU_Commercial	LU_HD_Mixed_Residence	LU_Park_recreational	LU_MLD_Mixed_residence	LU_Govt_and_offices	LU_Social_services
Total_No_casuality	1																									
Veh_Vol	.843**	1																								
LnVehicleVolume	.793**	.983**	1																							
Ped_Vol	.692**	.629**	.602**	1																						
LnpedestrianVolume	.618**	.646**	.633**	.724**	1																					
IC_Width_C_Maj_R	0.212	0.106	0.088	-0.001	-0.234	1																				
IC_Width_C_Min_R	.416**	.348*	.363*	0.287	0.152	.368*	1																			
IC_No_Lane_Maj_R	.316*	0.150	0.128	0.039	-0.188	.963**	.435**	1																		
IC_No_Lane_Min_R	.505**	.380*	.381*	.324*	0.183	.425**	.980**	.520**	1																	
IC_P_Raised_Med_RI_Maj	-0.050	-0.154	-0.195	-0.098	-0.195	.318*	-0.126	.341*	-0.049	1																
IC_P_Ded_Right_Lane_Maj	-0.090	-0.029	-0.043	-0.160	-0.045	0.096	-0.221	0.012	-0.229	0.214	1															
IC_P_Ped_related_Signing	0.005	0.080	0.046	.333*	0.144	-0.186	-0.111	-0.252	-0.147	-0.074	0.074	1														
IC_Avg_SW_Width	0.066	0.018	0.048	-0.069	-0.154	.434**	.319*	.423**	.304*	0.003	0.015	0.003	1													
IC_P_SW_barrier	0.052	0.036	0.061	0.032	0.115	0.246	0.272	0.273	.292*	0.007	-0.106	0.007	.341*	1												
IC_Pavement_Cond	-.353*	-0.286	-0.230	-.480**	-.401**	0.250	-0.113	0.258	-0.093	0.025	0.026	-0.279	.390*	.422**	1											
BI_No_Bus_Stop	.385*	.464**	.426**	0.283	.297*	0.027	0.217	0.026	0.231	-0.116	-0.140	0.009	-0.190	-0.082	-0.220	1										
BI_School_density	.498**	0.180	0.115	.401**	0.277	0.036	0.241	0.150	.334*	0.089	-0.026	-0.057	-0.227	-0.073	-.471**	.324*	1									
BI_NO_Alcoholsales_est	0.143	0.167	0.197	0.173	0.054	-0.215	0.244	-0.149	0.243	-0.066	-0.266	-0.248	-0.089	-0.018	-0.192	0.185	0.235	1								
BI_P_Curb_Parking	.329*	0.141	0.124	0.077	0.025	0.221	0.237	.318*	.342*	-0.034	-0.161	-0.202	-0.048	-0.108	-0.059	0.034	.341*	0.031	1							
BI_P_LRT_Stn	-.369*	-.393*	-.361*	-0.204	-.370*	-0.186	-0.145	-0.249	-0.203	-0.087	-0.265	0.036	-0.201	-0.080	-0.005	-0.008	-0.020	0.231	-0.215	1						
LU_Commercial	0.157	-0.029	-0.042	-0.048	0.191	.302*	0.096	.364*	0.184	-0.050	-0.042	-0.037	.419**	.311*	0.238	-0.120	-0.047	-0.208	0.232	-.478**	1					
LU_HD_Mixed_Residence	-.404**	-.363*	-.326*	-.377*	-.319*	0.115	0.048	0.094	0.009	0.040	-0.124	-0.223	0.083	.337*	.303*	-0.280	-0.180	-0.055	-0.136	0.084	0.009	1				
LU_Park_recreational	0.110	-0.029	-0.018	0.023	0.066	0.265	.326*	.303*	.373*	-0.095	-0.184	-0.181	.507**	.394*	.375*	-0.114	-0.171	0.106	0.188	-0.248	.631**	-0.141	1			
LU_MLD_Mixed_residence	0.120	0.262	0.232	0.259	0.111	-.298*	-0.150	-.335*	-0.194	-0.037	0.184	0.230	-.495**	-.544**	-.485**	.368*	0.260	0.125	-0.089	.300*	-.728**	-.532**	-.648**	1		
LU_Govt_and_offices	-0.099	-0.122	-0.114	-.329*	-.346*	0.163	-0.216	0.146	-0.187	.321*	0.041	0.124	0.259	0.194	.296*	-.306*	-0.252	-0.216	-0.022	0.134	0.005	-0.035	-0.007	-0.153	1	
LU_Social_services	0.000	-0.061	-0.018	0.075	0.005	-0.242	-0.152	-0.223	-0.155	0.146	-0.098	-0.129	-0.058	0.023	-0.034	-0.270	-0.110	-0.032	-0.116	-0.144	-0.152	0.152	-0.039	-0.222	0.097	1

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

Besides, the total number of causalities is also highly correlated with pedestrian volume and traffic volume. Whereas, the total number of causalities has no high correlation (>0.6**) with any of the explanatory variables. This indicates that crashes are influenced by several explanatory variables, suggesting the use of multivariate analysis. Checking correlation within the explanatory variables, the width of crossing on major roads and the number of lanes crossed on major roads were found to have a high correlation (0.963**). Considering the relative higher importance of the number of lanes for estimation of the number of casualties, it is chosen for model prediction. While the width of crossings both on major and minor road parameters are discarded.

Correlation between the proportion of medium and low-density mixed residences was highly correlated with the proportion of commercial land-use and proportion of park/recreation land-use. At the same time, commercial land-use was also highly correlated with park/recreation land-use. Considering the higher importance of the proportion of commercial land-use for crash prediction, as it has a higher correlation with the number of casualties, commercial land-use will be prioritized and will not be fitted together with these two variables.

Proceeding to check multicollinearity within explanatory variables, the variance inflation factor (VIF) was used. The first analysis was done with all explanatory variables excluding the width of crossing on the major and minor road. The collinearity analysis showed a maximum value variance inflation factor (VIF) of 5.943. Further analysis by excluding the proportion of medium and low-density land-use, and the proportion of park/recreation land-use resulted in a lower VIF (5.242), which indicates no major multicollinearity problem among independent variables (Chen et al., 2016). A summary of collinearity for each variable can be found in appendix D. Finally, based on the correlation and collinearity analysis it was crucial to discard one of the variables which had a high correlation to erroneous prediction. The discarded variables are the proportion of Medium and low-density Mixed residence, the proportion of park/recreation land-use.

4.3 NEGATIVE BINOMIAL MODEL DEVELOPMENT

This section summarizes the results of the Negative Binomial GLM models applied to the 3 years vehicle-pedestrian crash data. The distribution of the crash data revealed a higher value of variance (19.2) than the mean (5.8). In such a situation, it is highly recommended to assume negative binomial distribution.

Two types of models were developed for the total number of crashes. Two separate models were needed to identify the effect of incorporating built environment and land-use variables on the most common type of model, which is a model with exposure variables and intersection characteristics.

Model type (1): pedestrian crash prediction model with exposure variables (Pedestrian and traffic volumes) and intersection characteristics.

Model type (2): pedestrian crash prediction model with exposure variables (Pedestrian and traffic volumes) and intersection characteristics, built environment characteristics, and land-use characteristics.

4.3.1 MODEL-I: EXPOSURE VARIABLE AND INTERSECTION CHARACTERISTICS VARIABLE

The forward selection process was used to fit the GLM model. The modeling started with testing the significance of the exposure variables. As indicated in the table (7), vehicle volume was found to be significant. Because pedestrian volume was not significant at 5%, it was removed from the model. Even though pedestrian volume was discussed by several works of literature to have a significant effect on pedestrian crashes, it was not found significant in this study. The one-hour pedestrian volume taken to fit the model in this study could be the reason for the non-significant result associated with pedestrian volume, as most researchers used more than three-hour counts (i.e. Miranda (2011) used 3-hour counts pedestrian volumes; Pulugurtha (2011) used 12-hour pedestrian volumes). Furthermore, this result may be an indicator of the necessity of higher hours of pedestrian counts and a larger sample size requirement for the pedestrian volume to fit in the pedestrian crash prediction model.

The process continued by adding intersection characteristic variables. The number of lanes crossed by pedestrians on the major road was found to be significant. Sidewalk width and pavement conditions were added consecutively, and the model fit improved. However, the remaining intersection characteristics variables were not found to be significant at 5%. As summarized in the table (8), the best fitted model incorporates vehicle volume (exposure variable) and three intersection characteristics variables - the sidewalk width, pavement condition, and the number of lanes crossed on the major road.

According to the fitted model (GLM), the vehicle volume has a positive relation with pedestrian crashes with a significantly high coefficient (1.990) at 99% significance level. Even though this result is consistent with several other studies (i.e. Lee, 2018; Miranda et al., 2011), it is in contrast with the study done on 173 signalized intersections by Pulugurtha (2011) where vehicle volume was not observed to have a significant role. Traffic volume has a direct and causal association with the occurrence of crashes (Lee, 2018; Miranda, 2011). Thus, it is expected that traffic volume is the main predictor for pedestrian crashes.

Regarding the intersection characteristics, the number of lanes crossed on the major road has a negative relation with pedestrian crashes. As the number of lanes increases the number of crashes will decrease. 45.1%, 54.1%, and 66% for two, three, and four lanes respectively. Two separate studies by Torbic et al., (2010) and Lee et al., (2008) argued a positive relation between the number of lanes and the number of crashes. Whereas, a study on marked pedestrian crossings by Elvik et al., (2013) found a negative relationship with a coefficient (-0.063) far from statically significant. One can assume that as the number of lanes increases the crash risk will also increase, but the existing studies do not show a significant and concluding statement regarding the relation. Considering the local context, the author argues that signalized intersections with a higher number of lanes tend to be ringroads and highways where pedestrian activity tends to be lower. As a study by Pulugurtha et al., (2011) showed, lower pedestrian activity is related to a lower risk of crash. Thus, to understand the effect of the number of lanes on pedestrian crash occurrence, further study needs to be conducted by considering the pedestrian activity at intersection locations with a higher number of lanes. This plays a major role to improve policy measures regarding pedestrian crossing management.

The average sidewalk width for all legs has a significantly positive relation. For one unit increase in average sidewalk width, the predicted number of crashes will increase by 17.7%. Intersections with higher pedestrian activities are expected to have a wider pedestrian walkway. This indicates higher pedestrian exposure to crashes (Lee et al., 2018). Thus, this study linked the effect of a wider walkway with pedestrian crashes.

According to Lee et al., (2018), the pavement condition around intersections will tend to improve the accessibility for pedestrians walking around the intersection. In this study, it was witnessed through observation that poor pavement conditions force pedestrians to walk on the roadway which contributes to an increase in the risk of a crash. As such, ‘Poor pavement condition’ has a significantly positive relation with pedestrian crash occurrence. It was found that poor pavement conditions led to a 25.5% increase in pedestrian crashes at a signalized intersection. Whereas, the coefficient for 'fair pavement conditions' is not significant and has a negative relation with a pedestrian crash. Despite that, the negative relation can be explained, as better pavement conditions reduce the chance of crash occurrence related to pavement irregularities and encourage pedestrians to stay on their walkways.

Table 8 Coefficient estimates for Model-I (best model obtained)

Parameters	*β*	*Std. Error*	*95% Wald CI*		*Hypothesis Test*		
			Lower	Upper	Wald χ^2	df	Sig.
(Intercept)	-19.463	1.6721	-22.741	-16.186	135.492	1	<0.001
Ln (Vehicle Volume)	1.990	0.1527	1.690	2.289	169.804	1	<0.001
[# of lanes to be crossed (one direction) Major road=2]	-0.451	0.1924	-0.828	-0.074	5.494	1	0.019
[# of lanes to be crossed (one direction) Major road=3]	-0.541	0.1673	-0.869	-0.213	10.470	1	0.001
[# of lanes to be crossed (one direction) Major road=4]	-0.667	0.1853	-1.030	-0.304	12.958	1	0.000
[# of lanes to be crossed (one direction) Major road=5]	0a						
Average sidewalk width (meter)	0.177	0.0783	0.024	0.331	5.124	1	0.024
[[Pavement Condition =1]	0.255	0.1145	0.031	0.479	4.957	1	0.026
[Pavement Condition- *Fair*=2]	-0.077	0.1154	-0.304	0.149	0.451	1	0.502
[Pavement Condition- *Good*=3]	REF [a]						
(Scale)	.038[b]						
(Negative binomial)	1[c]						

a. Set to zero because this parameter is redundant.

b. Computed based on the Pearson chi-square.

c. Fixed at the displayed value.

The model outperformed the simple model of (Ln (Vehicle Volume) +Average sidewalk width (meter) and Pavement Condition). Furthermore, the model is significantly better than the intercept only model. The summary of the goodness-of-fit indicators is indicated in the table (9). The transformed form of the model is as follows. The SPSS output of Model-I can be found in appendix E.

Best fitted Model-II

$$ln\big(\text{Total number of crashes}\big)_i = intercept + \beta 1 ln(\text{ADT}) + \beta 2 * \text{Average sidewalk width (meter)} + \beta 3 * \text{number of lanes to be crossed} + \beta 4 * \text{Pavement Condition}$$

Table 9 Goodness of fit indicators for the Model type 1

	Value[a]	df	Value/df
Deviance	1.008	26	0.039
Scaled Deviance	26.785	26	
Pearson Chi-Square	0.979	26	0.038
Scaled Pearson Chi-Square	26.000	26	
Log Likelihood[b,c]	-97.767		
Adjusted Log Likelihood[d]	-2596.823		
Akaike's Information Criterion (AIC)	211.534		
Finite Sample Corrected AIC (AICC)	217.294		
Bayesian Information Criterion (BIC)	223.745		
Consistent AIC (CAIC)	231.745		

"Dependent Variable: Total Number of casualties
Dependent Variable: Total Number of causalities
Model: (Intercept), Ln (Vehicle Volume), # of lanes to be crossed (one direction) Major road, Average sidewalk width (meter), Pavement Condition
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria
c. The log likelihood is based on a scale parameter fixed at 1.
d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

4.3.2 MODEL-II: ALL CATEGORIES INCLUDED

The negative binomial model is fitted by incorporating built-environment and land-use variables in the previously discussed Model-I. Through the forward selection process, three models were fitted, and the best model was chosen by comparing the AIC and BIC values of each model. The fitted models are listed in the table (10) below. Comparing the three models, model A and B have a better fit than Model C as the AIC and BIC values are lower. Model A and B indeed are very close to each other; however, Model B is chosen as the best fit as all variables in the model are significant at 5%. Whereas, in model A, the proportion of social services is significant at 10% level.

Table 10 Comparison of developed crash models

	MODEL	(AIC)	(BIC)
A	(Intercept), Ln (Vehicle Volume), Pavement Condition, Proportion of Commercial, School density (#)-400m, Proportion of Social services (School, hospital, health center)	209.463	220.147
B	Model: (Intercept), Ln (Vehicle Volume), Pavement Condition, Proportion of Commercial, School density (#)-400m, (# of bus/taxi stop)-150m	209.420	220.105
C	Model: (Intercept), Ln (Vehicle Volume), (# of bus/taxi stop)-150m, # of lanes to be crossed (one direction) Major road, Pavement Condition	211.609	223.820

The best model obtained for Model-II

$$ln\left(\text{Total number of crashes}\right)_i = intercept + \beta 1 ln(\text{ADT}) + \beta 2 * \text{Pavement Condition} + * \beta 3 * \text{Proportion of Commercial} + \beta 4 * \text{number of school} + \beta 5 * \text{number of bus/taxi stop}$$

The model incorporating all categories (Model-II) outperforms Model-I, as the AIC and BIC value for this model is lower than that of the Model-I. This is a crucial indicator that pedestrian crashes can be better explained by incorporating built-environment and land-use parameters. This result agrees with previous works in Canada (Miranda, 2011) and the US (Mansfield et al., 2018)

As expected, vehicle volume has a significant and positive relation with pedestrian crashes (Table-11). The coefficient estimate is significantly higher than 1, indicating its direct relations with pedestrian crashes. However, a higher coefficient is found in Model-II (2.27) than in Model-I (1.99). As indicated by Lee (2018), built environment characteristics have no direct relation with pedestrian crashes, rather they have a direct relation with pedestrian activity. This means variables with an indirect relation with crash occurrence tend to have lower prediction capacity resulting in an increase in the coefficient for variables with a direct relation (i.e. Vehicle volume). This result is consistent with the study of Miranda (2011) and Pulugurtha et al., (2011), where the major observed variability of a pedestrian crash is explained by traffic and pedestrian volume and where the effect of built environment characteristics occurs through its direct relation with pedestrian activity. The interaction between variables is illustrated in the conceptual framework (Fig. 2).

Out of the built environment variables used in the study, school density and transit density were found significant at the % level. These variables were also found significant in the study by Miranda et al., (2011) but the association with the pedestrian crashes is opposite to what is found in this study. The number of schools found within the 400meter buffer zone of a signalized intersection has a positive and significant effect on pedestrian crashes. Leaving other variables in the model constant, a unit increase in schools in the buffer zone results in a 7% increase in the predicted number of crashes. However, in the study by Miranda-Moreno et al., (2011), the school density was negatively associated with pedestrian crashes, where the author stated that this might result from some speed calming measures in the area.

On the other hand, transit density (bus or taxi stops) within a 150-meter buffer zone of a signalized intersection is negatively associated with pedestrian crash occurrence (Miranda-Moreno et al., 2011; Pulugurtha & Sambhara, 2011; Chen & Zhou, 2016; Ding et al., 2018). Based on the local context rationale, there are a higher number of traffic regulators at transit locations who might play a major role in reducing flow speed and ease of traffic flow. Furthermore, reduced speed due to increased congestion at transit locations might involve improving visibility for both driver and pedestrian before crossing the intersection. The pavement condition exhibits similar characteristics as in Model I. Poor pavement condition is related to an increase in crash occurrence.

Table 11 Coefficient estimates for model-II (best model obtained)

Parameters	*β*	*Std. Error*	*95% Wald CI*		*Hypothesis Test*		
			Lower	Upper	Wald χ^2	df	Sig.
(Intercept)	-22.424	1.5024	-25.369	-19.479	222.765	1	<0.001
Ln (Vehicle Volume)	2.270	0.1461	1.983	2.556	241.267	1	<0.001
Proportion of Commercial	0.006	0.0023	0.002	0.011	7.384	1	0.007
School density (#)-400m	0.071	0.0196	0.033	0.109	13.111	1	<0.001
Number of bus/taxi stop-150m	-0.046	0.0214	-0.088	-0.004	4.706	1	0.030
[Pavement Condition- *Poor*=1]	0.087	0.1073	-0.123	0.297	0.658	1	0.417
[Pavement Condition- *Fair*=2]	-0.152	0.0989	-0.346	0.042	2.354	1	0.125
[Pavement Condition- *Good*=3]	REF[a]						
(Scale)	.033[b]						
(Negative binomial)	1[c]						

Dependent Variable: Total Number of casualties
Model: (Intercept), Ln (Vehicle Volume), Pavement Condition, Proportion of Commercial, School density (#)-400m, Number of bus/taxi stop)-150m
a. Set to zero because this parameter is redundant.
b. Computed based on the Pearson chi-square.
c. Fixed at the displayed value.

Table 12 Goodness of fit indicators for the Model type 2

	Value[a]	df	Value/df
Deviance	0.895	27	0.033
Scaled Deviance	27.234	27	
Pearson Chi-Square	0.887	27	0.033
Scaled Pearson Chi-Square	27.000	27	
Log Likelihood[a,b]	-97.710		
Adjusted Log Likelihood[c]	-2974.001		
Akaike's Information Criterion (AIC)	209.420		
Finite Sample Corrected AIC (AICC)	213.728		
Bayesian Information Criterion (BIC)	220.105		
Consistent AIC (CAIC)	227.105		

a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria
b. The log likelihood is based on a scale parameter fixed at 1.
c. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

The commercial activity within a 800-meter buffer zone of a signalized intersection has a positive and significant association with pedestrian crashes, which is consistent with a study by Miranda (2011); and Chen (2016).

Referring to model A, the other land-use variable which has a positive relation with pedestrian crashes is the proportion of social activity (health-centers, schools). As can be argued social activities tend to increase pedestrian activity which will increase pedestrian exposure at intersections. However, both variables have a coefficient closer to zero, which indicates the indirect relation of land-use to crash occurrence and its higher inclination to indicate pedestrian activity at an intersection. All in all, incorporating built-environment and land-use characteristics has resulted in a better model fit. The SPSS output of Model-II can be found in Appendix F.

5 CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

In recent years, researchers focus on identifying risk factors to develop knowledge and guidelines for pedestrian safety. This study adopted risk factor identification to understand the pedestrian unsafety at signalized intersections.

The study applied negative binomial models. According to the model traffic volume was found as the main predictor for number of pedestrian crashes. Several studies have identified intersection characteristics to correlate with pedestrian crash. In this study, the width of sidewalk and pavement condition were found to associate positively with pedestrian crash occurrence. On the contrary, the characteristics of intersection was found to have a negative correlation with the number pedestrian crashes on the study area. Good pavement condition will reduce the chance of pedestrian crash occurrence. Wider sidewalk width was related to the increase in pedestrian crash occurrence. Thus, signalized intersections with poor pavement condition and wider walkway will have higher pedestrian crash occurrence. The number of lanes crossed by pedestrian, which is negative and explainable within the scope of the study, the remaining results are consistent with previous studies.

For the past years, most researchers do not consider built-environment and land-use characteristics of signalized intersection. However, the current study has proved the paramount importance of incorporating built -environment and land use characteristics in explaining pedestrian crash at intersections. Among the many land use and built-environment factors, this study examined the role vehicle volume, pavement condition, commercial land use, school density, and bus stop density in relation pedestrian crash occurrence. Hence, the study affirmed that signalized intersection with higher school density in the vicinity of the intersection will have higher pedestrian crash. Similarly, higher proportion of commercial land-use at signalized intersection was also found to increase the chance of pedestrian crashes. On the contrary, bus stop density in 100meter around signalized intersections appeared to have a negative relation with the occurrence of pedestrian crash. This can be explained by considering the local situation - where there are a higher number of traffic regulators at transit locations, which might play a major role in reducing flow speed and ease of traffic flow. All in all, this study has identified pavement condition, sidewalk width, number of lanes, school density, bus-stop density and commercial land-use to have impact on pedestrian crash occurrence.

5.2 RECOMMENDATION AND DIRECTION FOR FUTURE STUDIES

The improving pedestrian safety does not relay on single stakeholder rather it demands a cooperated effort by several stakeholder. Accordingly, this study will indicate the role of stakeholder to reduce pedestrian crash at signalized intersection by working on the identified risk factors.

- Urban planners: as indicated by this study, built-environment and land-use characteristics has impact on pedestrian crash occurrence. Special intersection treatment is required at location with higher school and commercial land-use density.
- Traffic regulators: at signalized intersections with higher school and commercial land-use, there need to deploy additional traffic regulators and traffic signs indicating higher pedestrian activity in the area.

- Addis Ababa Road Authority: Aside with the existing annual inspection and maintained of roads in the city, the authority shall need to give special emphasize to signalized intersections under its program; Project for Development of Road Maintenance Capacity of Addis Ababa City.
- Sub-city or Woreda administration: at intersections with higher school and commercial land use proportion, the administration can play its role by providing awareness campaigns to improve road safety awareness of the pedestrians (Students) and drivers.
- Addis Ababa Road and Transport Bureau: “Safe intersection Program”: this study provides vital input for the improvement and efficiency of the program. the program needs to give special emphasis on pedestrian walkway condition and to provide special safety improvement measure at signalized intersections with high school density and commercial land-use.

The current study serves as a first impetus to understand the occurrence of pedestrian crashes at signalized intersections and their surrounding land use and built environment factors in a manner that has never been done in the past. However, it is also a broad avenue for future studies to consider the role of built environment and land use characteristics on pedestrian activity and consequent pedestrian crash risk. In addition to that, future studies shall also assess and compare the findings of the current study using un signalized intersections.

REFERENCES

AARTB. (2019). Traffic Management Center(operation). Retrieved January 10, 2020, from https://www.google.com/maps/d/u/0/viewer?ll=9.010073952139127%2C38.795639454930324&z=13&mid=1O2TrNbRfylYvaYzh6FWILbp4AgVOdKoz

AARTB & ITDP. (2018). *ADDIS ABABA NON-MOTORISED TRANSPORT STRATEGY*. Retrieved from http://www.tpmo.gov.et/sites/default/files/NMTStrategyEN.pdf

AASHTO. (2010). *Highway Safety Manual* (1st ed.). Washington D.C, USA: American Association of State Highway Transportation Professionals.

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. https://doi.org/10.1016/j.amar.2017.08.002

Anteneh Kebede et al. (2019). Road Traffic Accident Related Fatalities in Addis Ababa City, Addis Ababa, Ethiopia: An Analysis of Police Report 2013/14. *Ecronicon*, *3*, 128–138.

BIGRS, A. &. (2019). *Addis Ababa Annual Road Safety Report 2017-2018 (No. 3). Addis Ababa, Ethiopia.*

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. https://doi.org/10.1016/j.amar.2018.05.001

Cai, Q., Lee, J., Eluru, N., & Abdel-Aty, M. (2016). Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. *Accident Analysis and Prevention*, *93*, 14–22. https://doi.org/10.1016/j.aap.2016.04.018

Chen, P., Wu, C., & Zhu, S. (2016). Interaction between vehicles and pedestrians at uncontrolled mid-block crosswalks. *Safety Science*, *82*, 68–76. https://doi.org/10.1016/j.ssci.2015.09.016

Chen, P., & Zhou, J. (2016). Effects of the built environment on automobile-involved pedestrian crash frequency and risk. *Journal of Transport and Health*, *3*(4), 448–456. https://doi.org/10.1016/j.jth.2016.06.008

Chimba, D., Musinguzi, A., & Kidando, E. (2018). Associating pedestrian crashes with demographic and socioeconomic factors. *Case Studies on Transport Policy*, *6*(1), 11–16. https://doi.org/10.1016/j.cstp.2018.01.006

City Map - aaca. (n.d.). Retrieved December 9, 2018, from http://www.addisababa.gov.et/zh/web/guest/city-map

Connors, R. D., Maher, M., Wood, A., Mountain, L., & Ropkins, K. (2013). Methodology for fitting and updating predictive accident models with trend. *Accident Analysis and Prevention*, *56*, 82–94. https://doi.org/10.1016/j.aap.2013.03.009

Cottrill, C. D., & Thakuriah, P. (2010). Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis and Prevention*, *42*(6), 1718–1728.

https://doi.org/10.1016/j.aap.2010.04.012

Daniels, S., Brijs, T., Nuyts, E., & Wets, G. (2011). Extended prediction models for crashes at roundabouts. *Safety Science*, *49*(2), 198–207. https://doi.org/10.1016/j.ssci.2010.07.016

Ding, C., Chen, P., & Jiao, J. (2018). *Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach*. https://doi.org/10.1016/j.aap.2017.12.026

Elvik, R., & Goel, R. (2019). Safety-in-numbers: An updated meta-analysis of estimates. *Accident Analysis and Prevention*, *129*, 136–147. https://doi.org/10.1016/j.aap.2019.05.019

Elvik, R., Sørensen, M. W. J., & Nævestad, T. O. (2013). Factors influencing safety in a sample of marked pedestrian crossings selected for safety inspections in the city of Oslo. *Accident Analysis and Prevention*, *59*, 64–70. https://doi.org/10.1016/j.aap.2013.05.011

Fitzpatrick, & Kay. (2018). *Guidebook on Identification of High Pedestrian Crash Locations*. Retrieved from http://www.ntis.gov

Guo, Q., Xu, P., Pei, X., Wong, S. C., & Yao, D. (2017). The effect of road network patterns on pedestrian safety: A zone-based Bayesian spatial modeling approach. *Accident Analysis and Prevention*, *99*, 114–124. https://doi.org/10.1016/j.aap.2016.11.002

Haghighatpour, P. J., & Moayedfar, R. (2014). Pedestrian Crash Prediction Models and Validation of Effective Factors on Their Safety (Case Study: Tehran Signalized Intersections). *Open Journal of Civil Engineering*, *04*(03), 240–254. https://doi.org/10.4236/ojce.2014.43021

Harwood et' al, D. (2008). Pedestrian Safety Prediction Methodology. In *Pedestrian Safety Prediction Methodology*. https://doi.org/10.17226/23083

Hezaveh, A. M., & Cherry, C. R. (2018). Walking under the influence of the alcohol: A case study of pedestrian crashes in Tennessee. *Accident Analysis and Prevention*, *121*, 64–70. https://doi.org/10.1016/j.aap.2018.09.002

Hirpa, J. (2016). *Fatal Injury Crash Data Analysis and Directions to Reverse the Growing Rate of Road Traffic Related Fatality in Addis Ababa*. Retrieved from https://www.scribd.com/document/398230768/Fatal-Injury-Crash-Data-Analysis-and-Directions-to-Reverse-the-Growing-Rate

Kaygisiz, Ö., Senbil, M., & Yildiz, A. (2017, June 1). Influence of urban built environment on traffic accidents: The case of Eskisehir (Turkey). *Case Studies on Transport Policy*, Vol. 5, pp. 306–313. https://doi.org/10.1016/j.cstp.2017.02.002

Koh, P. P., & Wong, Y. D. (2014). Gap acceptance of violators at signalised pedestrian crossings. *Accident Analysis and Prevention*, *62*, 178–185. https://doi.org/10.1016/j.aap.2013.09.020

Lascala et'al, E. (2000). Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. In *Accident Analysis and Prevention* (Vol. 32). Retrieved from www.elsevier.com/locate/aap

Lee, J., Abdel-Aty, M., & Cai, Q. (2017). Intersection crash prediction modeling with macro-level data from various geographic units. *Accident Analysis and Prevention*, *102*, 213–226. https://doi.org/10.1016/j.aap.2017.03.009

Lee, J., Abdel-Aty, M., & Shah, I. (2018). Evaluation of surrogate measures for pedestrian trips at intersections and crash modeling. *Accident Analysis and Prevention*. https://doi.org/10.1016/j.aap.2018.05.015

Liu, Y. C., & Tung, Y. C. (2014). Risk analysis of pedestrians' road-crossing decisions: Effects of age, time gap, time of day, and vehicle speed. *Safety Science*, *63*, 77–82. https://doi.org/10.1016/j.ssci.2013.11.002

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A*, *44*, 291–305. https://doi.org/10.1016/j.tra.2010.02.001

Mehta, G., & Lou, Y. (2013). Calibration and Development of Safety Performance Functions for Alabama Two-Lane, Two-Way Rural Roads and Four-Lane Divided Highways. *Transportation Research Record: Journal of the Transportation Research*, *2398*, 75–82. https://doi.org/10.3141/2398-09

Mansfield et'al, T. J. (2018). The effects of roadway and built environment characteristics on pedestrian fatality risk: A national assessment at the neighborhood scale. *Accident Analysis and Prevention*, *121*, 166–176. https://doi.org/10.1016/j.aap.2018.06.018

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 and Prevention*, *43*(5), 1624–1634. https://doi.org/10.1016/j.aap.2011.02.005

Mooney, S. J., DiMaggio, C. J., Lovasi, G. S., Neckerman, K. M., Bader, M. D. M., Teitler, J. O., … Rundle, A. G. (2016). Use of google street view to assess environmental contributions to pedestrian injury. *American Journal of Public Health*, *106*(3), 462–469. https://doi.org/10.2105/AJPH.2015.302978

National Academies of Sciences, Engineering, and M. 2018. (2018). *Systemic Pedestrian Safety Analysis | The National Academies Press*. https://doi.org/10.17226/25255

National Roads Authority, D. (2012). Project Appraisal Guidelines. In *Project Appraisal Guidelines for National Roads Unit 16.1 - Expansion Factors for Short Period Traffic Counts*. Retrieved from www.nra.ie/publications/projectappraisal

Olszewski, P., Osińska, B., Szagała, P., & Włodarek, P. (2018). Development of accident prediction models for pedestrian crossings. *MATEC Web of Conferences*, *231*, 1–8. https://doi.org/10.1051/matecconf/201823103002

Olszewski, P., Szagała, P., Wolański, M., & Zielińska, A. (2015). Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland. *Accident Analysis and Prevention*, *84*, 83–91. https://doi.org/10.1016/j.aap.2015.08.008

Ortiz, N., & Ramnarayan, M. (2017). Analysis of Alcohol-Involved Pedestrian Fatalities in the United States, 2003–2015. *Journal of Transport & Health*, *5*, S93. https://doi.org/10.1016/j.jth.2017.05.250

Oxley, J., Lenné, M., & Corben, B. (2006). The effect of alcohol impairment on road-crossing behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, *9*(4), 258–268. https://doi.org/10.1016/j.trf.2006.01.004

Pawar, D. S., & Patil, G. R. (2015). Pedestrian temporal and spatial gap acceptance at mid-block street crossing in developing world. *Journal of Safety Research*, *52*, 39–46. https://doi.org/10.1016/j.jsr.2014.12.006

Pulugurtha, S. S., Duddu, V. R., & Kotagiri, Y. (2013). Traffic analysis zone level crash estimation models based on land use characteristics. *Accident Analysis and Prevention*, *50*, 678–687. https://doi.org/10.1016/j.aap.2012.06.016

Pulugurtha, S. S., & Sambhara, V. R. (2011). Pedestrian crash estimation models for signalized intersections. *Accident Analysis and Prevention*, *43*(1), 439–446. https://doi.org/10.1016/j.aap.2010.09.014

Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis and Prevention*, *45*, 382–391. https://doi.org/10.1016/j.aap.2011.08.003

Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2014). Multimodal injury risk analysis of road users at signalized and non-signalized intersections. *Accident Analysis and Prevention*, *71*, 201–209. https://doi.org/10.1016/j.aap.2014.05.015

Thomas, L., Lan, B., Sanders, R. L., Frackelton, A., Gardner, S., & Hintze, M. (2017). Changing the Future? Development and Application of Pedestrian Safety Performance Functions to Prioritize Locations in Seattle, Washington. *Transportation Research Record: Journal of the Transportation Research Board*, *2659*(1), 212–223. https://doi.org/10.3141/2659-23

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), 65–74. https://doi.org/10.3141/2198-08

Tulu, G. S. (2015). *Pedestrian crash in Ethiopia: Identification of Contributing Factors through Modelling of Exposure and Road Environment Variables*. Queensland University of Technology.

Tulu, G.S, Washington, S., & King, M. (2013). Characteristics of police-reported road traffic crashes in Ethiopia over a six-year period. *Proceedings of the 2013 …*, 1–13. Retrieved from http://eprints.qut.edu.au/63884/

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. https://doi.org/10.1177/0361198118767409

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

WHO. (2007). Drinking and Driving: a road safety manual for decision-makers and practitioners. In *Global Road Safety Partnership*. Geneva.

WHO. (2018). *Global status report on road safety 2018*. Geneva: World Health Organization.

Xie, S. Q., Dong, N., Wong, S. C., Huang, H., & Xu, P. (2018). Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. *Accident Analysis and Prevention*, *121*, 285–294. https://doi.org/10.1016/j.aap.2018.09.030

Zegeer, C., Lyon, C., Srinivasan, R., Persaud, B., Lan, B., Smith, S., … Sundstrom, C. (2017). Development of crash modification factors for uncontrolled pedestrian crossing treatments. *Transportation Research Record*, *2636*, 1–8. https://doi.org/10.3141/2636-01

APPENDIX A LIST OF SIGNALIZED INTERSECTIONS

List of signalized intersections selected for the study

No.	Name	Code	No. of Legs	Traffic volume data available	pedestrian volume data available
1	Legehar	6	4	Yes (AARTB)	Yes (AARTB)
2	Shola-1	10	4	Yes (AARTB)	Yes (AARTB)
3	Jacros(F)	14	3	Yes (AARTB)	on field count
4	Imperial(F)	15	4	Yes (AARTB)	Yes (AARTB)
5	Bole Michael -1	16	4	Yes (AARTB)	Yes (AARTB)
6	Saris Abo(F)	17	4	Yes (AARTB)	Yes (AARTB)
7	Kadisco -1	18	4	Yes (AARTB)	on field count
8	Jemo-1	19	4	Yes (AARTB)	Yes (AARTB)
9	Lebu	20	4	Yes (AARTB)	Yes (AARTB)
10	24	23	4	Yes (AARTB)	on field count
11	British Embassy	26	3	Yes (AARTB)	Yes (AARTB)
12	Comerce	32	4	Yes (AARTB)	on field count
13	sunshine	34	4	Yes (AARTB)	on field count
14	Bereberie Berenda	40	4	Yes (AARTB)	Yes (AARTB)
15	Jemo Michael	41	4	Yes (AARTB)	Yes (AARTB)
16	Banko Diroma1	1	4	on field count	
17	ETV	2	4	on field count	
18	Tikur Ambessa	3	4	on field count	
19	Senga Tera	4	4	on field count	
20	Mexico	5	4	on field count	
21	St. Joseph	7	4	on field count	
22	St. Estifanos	8	4	on field count	
23	Shola-2	9	4	on field count	
24	Parlama -1	11	4	on field count	
25	St. Mary	12	4	on field count	
26	Semen Hotel	13	4	on field count	
27	Salite Mihret	21	4	on field count	
28	Safari	22	4	on field count	
29	Beherawi	24	5	on field count	
30	Harambe Hotel	25	4	on field count	
31	Atlas	27	4	on field count	
32	Kolfe 18	37	4	on field count	
33	Ethio-China	39	3	on field count	
34	Stadium	42	3	Yes (AARTB)	Yes (AARTB)
35	Estifanos	43	4	Yes (AARTB)	Yes (AARTB)
36	Brass	45	4	Yes (AARTB)	Yes (AARTB)

APPENDIX B FIELD DATA COLLECTION FORMAT

Annex B-1 Field data collection format for intersection characteristics data

ID	Intersec-tion	Width of crossing	Number of lanes to be crossed	Presence of raised median refuge islands	Pedestrian related signings	Average sidewalk width	Presence of sidewalk barrier	Pavement Condition
1								
2								

Annex B-2 Field data collection format for built-environment characteristics and land-use

ID	Intersec-tion	(# of bus/taxi stop)s	School density (#)	alcohol sales establish ments (#)	Presence of curb parking	Presence of LRT stations	Neighbor-hood Income (average per capita income)	Land-use_ •Commercial, high-density mixed residential, Medium and low-density mixed residence, government and offices, and social services (%)
1								
2								

APPENDIX C LIST OF PROPOSED EXPOSURE AND EXPLANATORY VARIABLES FOR MODEL DEVELOPMENT

List of proposed exposure and explanatory variables for model development

Category	Variables	Unit	Buffer Zone	Data source	Reference
Dependent variable	Pedestrian crash data	count	30meter	Addis Ababa Traffic Police	
Traffic characteristics	Traffic volume	AADT	n/a	AARTB	Torbic et al (2010)
	Pedestrian Volume	AAPT	n/a	Field data and AARTB	Torbic et al (2010)
	Posted speed limit	Km/hr	n/a	Field data	Torbic et al. (2010); Chen (2016)
Intersection characteristics	Width of crossing	Meters	n/a	Field data	Torbic et al., 2010; Zegeer et al., 2017
	Number of lanes crossed by pedestrian	1 to 6	n/a	Field data	Elvik et al., 2013;Lee et al., 2018; Torbic et al., 2010
	Presence of raised median refuge islands	Yes /No	n/a	Field data	Harwood et al., (2008)
	Pedestrian related signings	Yes /No	100m	Field data	Xie et al., (2018)
	Average pavement condition	Poor/Fair/Bad		Field data	Lee et al., (2018)
	Average sidewalk width	Meters		Field data	Lee et al., 2018; Chen et al., 2016
	Presence of sidewalk barrier	Yes /No			Lee et al., (2018)
Socio-demographic	Income (Neighborhood average per capita income)	Low, high	800m	Sub-city & Wereda	Torbic;2010; Mansfield et' al, 2018
Land-use	Commercial	Proportion (%)	800m	AARTB	Chen, 2016; Mansfield et al, 2018; Ding et al., 2018
	Residential	%	800m	AARTB	
	social services	%	800m	AARTB	

	Mixed land-use	%	800m	AARTB	
	Gov't and office	%	800m	AARTB	

Category	Variables	Unit	Buffer Zone	Data source	Reference
Built-environment	Transit density (# of bus/taxi stops)	Count	150m	To be Collected on-site	Miranda et al., (2011)
	School density (#)	Count	400m	To be Collected on site	Miranda et al., (2011)
	Number of alcohol sales establishments	Count	300m	To be Collected on site	Torbic et al., (2010)
	Presence of curb parking	Yes/No	300m	To be Collected on site	Torbic et al., (2010)
	Presence of LRT stations	Yes/No	400m	To be Collected on site	Miranda et al., (2011)

APPENDIX D SUMMARY OF COLLINEARITY ANALYSIS

Summary of collinearity analysis

Independent variables	unstandardized coefficients		t	collinearity statistics	
	b	std. error		tolerance	VIF
(constant)	-111.258	24.184	-4.601		
Log (vehicle volume)	10.201	3.144	3.245	0.168	5.943
Log (pedestrian volume)	0.487	1.556	0.313	0.180	5.557
# of lanes to be crossed (one direction) major road	0.913	1.181	0.773	0.223	4.494
# of lanes to be crossed (one direction) minor road	0.142	1.119	0.127	0.330	3.032
presence of raised median refuge islands (yes/no)	0.017	1.145	0.014	0.470	2.128
presence of dedicated right turn lane major road (yes/no)	-0.232	1.196	-0.194	0.590	1.696
pedestrian related signings (yes/no)	0.052	1.053	0.050	0.555	1.801
average sidewalk width (meter)	-0.040	1.037	-0.039	0.452	2.211
presence of sidewalk barrier (yes/no)	-0.607	1.295	-0.468	0.363	2.754
pavement condition	-0.409	0.852	-0.480	0.343	2.915
(# of bus/taxi stop)-150m	-0.041	0.284	-0.145	0.524	1.907
school density (#)-400m	0.855	0.313	2.737	0.376	2.659
alcohol sales establishments (#)-300m	-0.073	0.122	-0.598	0.426	2.347
presence of curb parking-300m (yes/no)	0.067	1.499	0.045	0.538	1.858
presence of LRT stations-400m (yes/no)	0.665	1.321	0.503	0.381	2.626
proportion of commercial	0.034	0.045	0.744	0.283	3.529
proportion of HD mixed residential	-0.014	0.034	-0.407	0.369	2.707
proportion of park/recreational	0.039	0.058	0.670	0.213	4.699
proportion of gov't and office	0.066	0.115	0.575	0.437	2.287
proportion of social services (school, hospital, health center)	0.073	0.073	0.995	0.566	1.767

APPENDIX E MODEL I SPSS-OUTPUT

MODEL I SPSS-OUTPUT

```
* Generalized Linear Models.
GENLIN Total_No_casuality BY IC_Pavement_Cond (ORDER=ASCENDING) WITH
LnVehicleVolume
    IC_Width_C_Maj_R IC_Avg_SW_Width
  /MODEL LnVehicleVolume IC_Width_C_Maj_R IC_Avg_SW_Width
IC_Pavement_Cond INTERCEPT=YES
 DISTRIBUTION=NEGBIN(1) LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=DEVIANCE 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 CORB
  /SAVE MEANPRED CIMEANPREDL CIMEANPREDU XBPRED XBSTDERROR RESID
PEARSONRESID DEVIANCERESID
    STDDEVIANCERESID LIKELIHOODRESID.
```

Model Information

Dependent Variable	Total Number of casuality
Probability Distribution	Negative binomial (1)
Link Function	Log

Case Processing Summary

	N	Percent
Included	34	100.0%
Excluded	0	0.0%
Total	34	100.0%

Categorical Variable Information

			N	Percent
Factor	Pavement Condition	POOR	12	35.3%
		FAIR	13	38.2%
		GOOD	9	26.5%
		Total	34	100.0%

Continuous Variable Information

		N	Minimum	Maximum
Dependent Variable	Total Number of casuality	34	1	21
Covariate	Ln (Vehicle Volume)	34	9.685518092495647	11.100677404565232
	Width of crossing (meter) Major road	34	7.0	15.0
	Average sidewalk width (meter)	34	2	5

Continuous Variable Information

		Mean	Std. Deviation
Dependent Variable	Total Number of casuality	7.18	4.386
Covariate	Ln (Vehicle Volume)	10.622068197527193	.306668381203939
	Width of crossing (meter) Major road	11.176	2.1458
	Average sidewalk width (meter)	3.26	.567

Goodness of Fit[a]

	Value	df	Value/df
Deviance	1.524	28	.054
Scaled Deviance	28.000	28	
Pearson Chi-Square	1.487	28	.053
Scaled Pearson Chi-Square	27.311	28	
Log Likelihood[b,c]	-98.025		
Adjusted Log Likelihood[d]	-1800.445		
Akaike's Information Criterion (AIC)	208.050		
Finite Sample Corrected AIC (AICC)	211.161		
Bayesian Information Criterion (BIC)	217.208		
Consistent AIC (CAIC)	223.208		

Dependent Variable: Total Number of casuality

Model: (Intercept), Ln (Vehicle Volume), Width of crossing (meter) Major road, Average sidewalk width (meter), Pavement Condition[a]

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

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

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus

Likelihood Ratio Chi-Square	df	Sig.
192.793	5	.000

Dependent Variable: Total Number of casuality

Model: (Intercept), Ln (Vehicle Volume), Width of crossing (meter) Major road, Average sidewalk width (meter), Pavement Condition

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	136.984	1	.000
Ln (Vehicle Volume)	146.800	1	.000
Width of crossing (meter) Major road	.379	1	.538
Average sidewalk width (meter)	2.419	1	.120
Pavement Condition	5.657	2	.059

Dependent Variable: Total Number of casuality

Model: (Intercept), Ln (Vehicle Volume), Width of crossing (meter) Major road, Average sidewalk width (meter), Pavement Condition

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test
			Lower	Upper	Wald Chi-Square
(Intercept)	-21.447	1.7978	-24.971	-17.923	142.311
Ln (Vehicle Volume)	2.115	.1746	1.773	2.457	146.800
Width of crossing (meter) Major road	.014	.0226	-.030	.058	.379
Average sidewalk width (meter)	.144	.0924	-.037	.325	2.419
[Pavement Condition =1]	.292	.1343	.029	.555	4.728
[Pavement Condition =2]	.094	.1245	-.150	.338	.574
[Pavement Condition =3]	0[a]	.	.	.	.
(Scale)	.054[b]				
(Negative binomial)	1[c]				

APPENDIX F MODEL-II: SPSS OUTPUT

Model-II: SPSS output

```
GET
  FILE='C:\Users\Test\Desktop\Master thesis part 2\Phase 17\FINAL Phase 17 with 34
intersections .sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
* Generalized Linear Models.
GENLIN Total_No_casuality BY IC_Pavement_Cond (ORDER=ASCENDING) WITH LnVehicleVolume
BI_No_Bus_Stop
    BI_School_density LU_Commercial
  /MODEL IC_Pavement_Cond LnVehicleVolume BI_No_Bus_Stop BI_School_density
LU_Commercial
    INTERCEPT=YES
 DISTRIBUTION=NEGBIN(1) LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=DEVIANCE 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 CORB
  /SAVE MEANPRED CIMEANPREDL CIMEANPREDU XBPRED XBSTDERROR RESID PEARSONRESID
DEVIANCERESID
    STDDEVIANCERESID LIKELIHOODRESID.

[DataSet1] C:\Users\Test\Desktop\Master thesis part 2\Phase 17\FINAL Phase 17 with 34
intersections .sav
```

Model Information

Dependent Variable	Total Number of casuality
Probability Distribution	Negative binomial (1)
Link Function	Log

Case Processing Summary

	N	Percent
Included	34	100.0%
Excluded	0	0.0%
Total	34	100.0%

Categorical Variable Information

			N	Percent
Factor	Pavement Condition	POOR	12	35.3%
		FAIR	13	38.2%
		GOOD	9	26.5%
		Total	34	100.0%

Continuous Variable Information

		N	Minimum	Maximum
Dependent Variable	Total Number of casuality	34	1	21
Covariate	Ln (Vehicle Volume)	34	9.685518092495647	11.100677404565232
	(# of bus/taxi stop)-150m	34	1	8
	School density (#)-400m	34	1	9
	Proportion of Commercial	34	0.00%	47.14%

Continuous Variable Information

		Mean	Std. Deviation
Dependent Variable	Total Number of casuality	7.18	4.386
Covariate	Ln (Vehicle Volume)	10.622068197527193	.306668381203939
	(# of bus/taxi stop)-150m	4.65	1.921
	School density (#)-400m	3.53	2.063
	Proportion of Commercial	13.6251%	16.34140%

Goodness of Fit[a]

	Value	df	Value/df
Deviance	.895	27	.033
Scaled Deviance	27.000	27	
Pearson Chi-Square	.887	27	.033
Scaled Pearson Chi-Square	26.768	27	
Log Likelihood[b,c]	-97.710		
Adjusted Log Likelihood[d]	-2948.494		
Akaike's Information Criterion (AIC)	209.420		
Finite Sample Corrected AIC (AICC)	213.728		
Bayesian Information Criterion (BIC)	220.105		
Consistent AIC (CAIC)	227.105		

Dependent Variable: Total Number of casuality

Model: (Intercept), Pavement Condition , Ln (Vehicle Volume), (# of bus/taxi stop)-150m, School density (#)-400m, Proportion of Commercial[a]

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

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

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
335.746	6	.000

Dependent Variable: Total Number of causalities

Model: (Intercept), Pavement Condition, Ln (Vehicle Volume), (# of bus/taxi stop)-150m, School density (#)-400m, Proportion of Commercial

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	216.212	1	.000
Pavement Condition	7.808	2	.020
Ln (Vehicle Volume)	239.198	1	.000
(# of bus/taxi stop)-150m	4.666	1	.031
School density (#)-400m	12.998	1	.000
Proportion of Commercial	7.321	1	.007

Dependent Variable: Total Number of casuality

Model: (Intercept), Pavement Condition, Ln (Vehicle Volume), (# of bus/taxi stop)-150m, School density (#)-400m, Proportion of Commercial

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	95% Wald Confidence Interval Upper	Hypothesis Test Wald Chi-Square
(Intercept)	-22.424	1.5089	-25.381	-19.467	220.855
[Pavement Condition =1]	.087	.1078	-.124	.298	.653
[Pavement Condition =2]	-.152	.0993	-.346	.043	2.334
[Pavement Condition =3]	0[a]	.	.	.	.
Ln (Vehicle Volume)	2.270	.1467	1.982	2.557	239.198
(# of bus/taxi stop)-150m	-.046	.0214	-.088	-.004	4.666
School density (#)-400m	.071	.0197	.032	.109	12.998
Proportion of Commercial	.006	.0023	.002	.011	7.321
(Scale)	.033[b]				
(Negative binomial)	1[c]				

Dependent Variable: Total Number of casuality

Model: (Intercept), Pavement Condition, Ln (Vehicle Volume), (# of bus/taxi stop)-150m, School density (#)-400m, Proportion of Commercial

a. Set to zero because this parameter is redundant.

b. Computed based on the deviance.

c. Fixed at the displayed value.

APPENDIX G Model VALIDATION OUTPUT

The model validation out was computed, and the MAD value comes out to be 1.75.