ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

Evaluating the Major Causes of Road Traffic Accident Occurrence Related to Geometric Design Elements of Addis Ababa -Debre Brehan Trunk Road

A Thesis in **Road and Transportation Engineering**

By: MAHDER BEKELE

June 30, 2023

Addis Ababa

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science

The undersigned have examined the thesis entitled '**Evaluating the Major Causes of Road Traffic Accident Occurrence Related to Geometric Design Elements of Addis Ababa -Debre Brehan Trunk Road**

Dr. Getu Segni (Phd)	____________________	____________
Advisor	Signature	Date
Mr. Mesay Shemsu	____________________	____________
Internal Examiner	Signature	Date
Dr. Robeam Solomon (Phd)	____________________	____________
External Examiner	Signature	Date
Dr. Abreham Gebrie (Phd)	____________________	____________
Chair person	Signature	Date

UNDERTAKING

I certify that research work titled "**Evaluating the Major Causes of Road Traffic Accident Occurrence Related to Geometric Design Elements of Addis Ababa -Debre Brehan Trunk Road**" is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged / referred.

____

Mahder Bekele

ABSTRACT

Road safety issue becomes a worldwide concern specially for developing countries these days since it results in a serious social and economic problems. Several Geometric design and safety research have been undertaken by aiming to improve roadway designs and providing a forgiving roads to the road users. The impact of geometric design factors such as horizontal and vertical curves, lane width, shoulder width, super elevation, median width, curve radius, sight distance, and others on safety has been investigated. However, the relationship between geometric design components and accident rates is remained complicated and unclear since there is a scarcity of data on the association between geometric design features and the accident rates.

This study's primary objective is to investigate the geometric design related factors that influence the frequency and severity of traffic accidents on the Addis Ababa–Debrebrehan Trunk Road, by considering design elements such as the average degree of curvature, average gradient, number of lanes, lane widths, shoulder widths and types, median widths and types, total surface width, rural-urban classifications, speed limits, and the presence of structures.

As-built drawings, findings from an inventory assessment of the roads, information on traffic volumes, and traffic accidents reports from six consecutive years have been collected for this investigation. Empirical and descriptive statistical analysis models have then been used to analyses the data. In order to analyses the count/discrete outcome variable of traffic accident frequency, a generalized linear model family of Negative Binomial Regression model has been used and Ordinal Logistic Regression Model with Probit link function for severity analysis, by considering geometric design features as continuous and categorical independent variables. As a result, a total of 1192 traffic incidents were recorded and used for evaluation, of which 274 resulted in fatal, 293 in serious injury, 182 in minor injury, and 443 in property damage only crashes.

Thus, the model for accident frequency analysis has demonstrated a convergence value for the segment length, average daily traffic, median width, number of lanes, outside shoulder width, average degree of curvature, average gradient, and speed limit. In addition to the frequency model, the severity analysis model has also reached at convergence for the vertical gradient, degree of curvature, availability of structures, median width, surface width, right and left shoulder widths and availability of median barrier.

Key Words: Crash Frequency, Accident Severity, Geometric Design Elements, Negative Binomial

ACKNOWLEDGMENTS

I want to start by giving thanks to the Almighty God for his never-ending mercy, blessings, and kindness in my life.

Next, I would like to express my sincere gratitude to my advisor, Dr. Getu Segni Tulu, and my co-advisor, Mrs. Emnete Tadesse, for their understanding, steadfast support, and intellectual guidance throughout my study. Without their persistent support, push, and patience, this thesis would not have been finished.

I am incredibly grateful to everyone who helped me out by providing data, translating and interpreting the data, and lifting me up through their prayer.

Table of Contents

List of Abbreviations

DS Design Standard

ERA Ethiopia Roads Administration

VIF Variable Inflation Factor

IHSDM Interactive Highway Safety Design Manual

HSM Highway Safety Manual

HCM Highway Capacity Manual

GoF Goodness of Fit

GLM Generalized Linear Model

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

GLM Generalized Linear Models

NB Negative Binomial

1. INTRODUCTION

1.1. Background

93% of all road fatalities across the globe occur in low- and middle-income countries, according to a WHO report released on June 20, 2022. Each year, approximately 1.3 million people lose their lives as a result of traffic accidents, which cost most countries 3% of their Gross Domestic Product. Even in high-income countries, People from lower socioeconomic backgrounds are more likely to be involved in traffic accidents. In addition, the African region has the greatest rates of fatal traffic accidents. The report also mentions that even though the human factor takes the lion's share for causing the road traffic accident, among the major risk factors for accident, unsafe road infrastructure also contributes significant impact (Organization, 20 June 2022).

In Ethiopia, according to a report released by the Ethiopian Federal Police at the end of July 2021, the country had 15,034 road accidents in the fiscal year that concluded on July 7, 2021, with 4, 161 persons killed. In addition, 5,763 people had significant physical injuries, while 5,110 others suffered minor injuries because of the accidents. In addition, police reported 31,643 accidents that results a property damage of more than 2.28 million birr (Kussia, 2017).

As various studies have shown, it is challenging to confidently describe, demonstrate, and quantify where traffic accidents may happen or where black spots may form. Researchers have thus divided elements affecting road safety into behavioral (i.e., human-related) and non-behavioral (i.e., factors relating to the vehicle, characteristics of the road, and its environment) factors. Hence, it will be challenging to determine the root cause and contributing factors of the road traffic accident for countries like Ethiopia where there is no performance evaluation and/or assessment practice on the completed roads and planned to be constructed road projects over the duration of their design life. It will consequently result in a subjective assessment of the situation based on the person's experience of the subject matter.

Furthermore, as there is no structured and centralized system for collecting and maintaining data on traffic accidents, most reports of road traffic accidents attribute their causes to a single, human-related element. As a result, it underestimates the role played by other non-behavioral elements, which are often given a 1% contribution to the overall number of reported road traffic incidents.

Even though, human related factors contribute the lion's share to the road traffic accident, direct control and prediction of human factors are difficult. Thereby, it is believed that human factors can be indirectly controlled and predicted through investigation of roadway factors particularly by designing and constructing forgiving roads which are the outputs of consistent roadway geometric design elements.

Many evaluation studies of the safety implications of road design elements reveal the current inability to explain the occurrence of accidentality. The major causes of accidents are thought to be the driver's actions, which are largely influenced by his personality, skills, and experience. The driver's behavior is also influenced by external factors including the weather, the conditions of the roads, the date and time and lighting conditions.

Several investigations on different countries have looked into the empirical correlations between highway geometric elements and traffic accidents using various statistical models. Contrarily, in Ethiopia, despite the fact that the frequency and severity of traffic accidents are increasing rapidly throughout the country, only a little amount of work has been made to identify the major factors contributing to the road and its environment. Hence, in this research using descriptive and empirical approaches a little effort will be added to address the safety questions such as;

- on a section of highway with specified attributes, what vehicle accident involvement rate and accident probability can be reasonably expected?
- on a given set of highway geometric design parameters, which parameters are relatively more critical to the safety performance of the road?

1.2. Statement of the Problem

The prevalence of traffic accidents is increasing in the country as exposure to this risk increases as a result of fast increment of motorized vehicles (without adequate regulation), rapid population growth, expansion of the road network, and poor user behavior and safety culture.

According to data on road traffic accidents taken from the Ethiopian Federal Police Commission report for 2019–2020 and 2020–2021, the number of fatalities, severe injuries, and slight injuries has been increasing. The result of road accidents, however, Property damage only decreases slightly. The number of people killed in traffic accidents increased by 9.6% in 2020–21, while the number of accidents grew by 3.9%, a threefold increase.

In order for the appropriate authorities to take a number of measures to address the issue, the empirical research on the causes of these severe road traffic accidents in Ethiopia has not been properly studied. In reality, researches that focused on identifying root causes potential triggering factors that are contributing to the road traffic accidents frequency and severity can be undertaken based on two approaches (i.e. either at nationwide level using time series traffic accident data collected on all road links or case study by selecting specific road links on the bases of high traffic accidents and severity). However, we have adopted the second approach since it is preferable to identify factors related to defects in geometric design that are often different for different road section cases (Bedria Hussien, 2022).

Up to now, separate researches have been undertaken to understand the root causes and triggering factors for the road traffic accident on the selected road. Accordingly, the research undertaken by Abdulmalik Adem (Adem, 2017) has briefly explains the pattern and location of the accidents in depth even though the research has its own limitations since it is focusing on identifying the black spot locations and spatial and temporal distribution of accidents along the road.

Hence, to fill the gap, by using the aforementioned research as a baseline I have conducted a retrospective analysis of road accidents and their interactions with the roadway environment throughout the project to scrutinize the root causes of traffic accidents that are directly related to the geometric design elements and triggering factors that initiates the accidents so that Road Planners, Designers and Safety Engineers will find a proactive and preventive guidance to address safety-related issues on the given road or newly designed/upgrading road projects.

As a result, the effect of the horizontal and vertical road geometric design parameters on the cause of traffic accidents and severity of crash is assessed by linking traffic accident data with the actual location where the accident occurred and the geometric design parameters data from the as built documents of the road.

1.3. Research Questions

- What is the retrospective pattern of the road traffic accident frequency and severity and its interaction with the geometric design elements on Ababa – Debrebrehan Trunk road?
- What are the major causes and contributing factors that are related to the geometric design elements of the road for the traffic accident occurred on the Addis Ababa – Debrebrehan Trunk road?

1.4. Objectives of the Study

1.4.1. General Objectives

The general objective of this research is to explore the causes of the traffic accident occurrence and severity then identifying the major contributing factor for the traffic accident occurrence of Addis Ababa -Debre Brehan Trunk Road.

1.4.2. Specific Objectives

The specific objective of the research is:

- To analyze retrospective pattern of the road traffic accident and its interaction with the geometric design elements that exist on the road
- To identify the major causes and contributing factors of the Traffic accident that are related to geometric design elements of the road

1.5. Scope and limitation of the research

The road itself and its environment, which includes the geometric design specifications of the route, driver behavior, and vehicle variables, are one or a combination of the causes of traffic accidents on the road. In this study, non-behavioral characteristics that are major geometric design elements of the road were exclusively considered when evaluating the causes and severity of traffic crashes in the study area. As a result, both behavioral factors that are related to the driver, pedestrians and passengers and vehicle related factors have only been evaluated using only general descriptive terms. In addition to that, this study does not consider environmental conditions or characteristics of the road connected to the pavement surface.

1.6. Organization of the Thesis

This research paper has been organized in to five charters and appendixes as follows

- The first chapter presents an introduction to the study topic which contains background information, statement of the problem, objectives, research questions and scope and limitation of the study.
- The second chapter contains the literature review part that encompasses a review on related previous researches, guidelines and books that used for assessing the causes and severity of traffic accidents and evaluates influencing factors that contributes for the causes of traffic accidents and injuries severity.
- The third chapter presents the study methodology and materials used for the analysis.
- The fourth chapter presents the results and discussion part of the research. It includes the analysis of the traffic accident frequency and injury severity on the project by using descriptive analysis and empirical statistical models for analyzing the major causes and influencing factors of the traffic accident occurrence and severity.
- Finally, the conclusions, recommendations and for future study area are presented in chapter five

2. LITERATURE REVIEW

2.1. Introduction

As per HSM 2010, crash is defined as a set of events that results in fatality, injury or property only damage due to the collision of at least one motorized vehicle and may involve collision with another motorized vehicle, bicyclist, pedestrian or an object. Crashes are rare and random by their nature, when we say it's rare, we mean that they represent a very small proportion of event that occurs in the transportation system and while we are saying it's random, they occur as functions of a set of several deterministic and stochastic factors (Anon., 2010).

it is challenging to say that crashes are caused by a single factor rather it is a combination of a series of events that are influenced by a number of contributing factors and categorized under three broader groups;

- Human related factor-related to the driver's age, driving skill, fatigue, attention, sobriety etc.;
- Vehicle related factors-related to vehicle condition, brake efficiency etc.;
- Roadway and its environment related factors- geometric design parameters, cross sectional elements, traffic control devices, roadside conditions etc.

Different researches have been undertaken internationally to detect the root causes of the highway traffic accidents. Studies conducted on the late 1990's were used descriptive analysis to evaluate the safety performance of a given road which considers crash frequency, rate and equivalent property damage and summarized as the history of crash occurrence, type and severity. Whereas, the recent studies have used quantitative predictive models that considers expected number and severity of crashes at sites having similar geometric and operational characteristics for existing and future conditions or roadway design alternatives.

2.2. Traffic Accident Trend in Ethiopia

According to studies done between 2007 and 2018, there was an average annual increase in road traffic accidents in the country of 9.16%. Similar to this, annual average growth rates for the motorized vehicle and road network coverage were about 10.4% and 9.34%, respectively. The report demonstrates that the growth rates of motorized vehicles, road network coverage, and traffic accidents all increased at approximately the same rate. The study concluded that the development of the country's road network had no significant or substantial impact on the reduction of the frequency of accidents involving motor vehicles. This suggests that the frequency of traffic accidents over the study period was inversely related to the proportions of the road network. While the development of the road network was closely tied to the introduction of motor vehicles in Ethiopia (Deme, 2019).

According to research by Abegaz Gebremedhin, of the 75,271 household members counted, 123 experienced RTAs throughout the study period, with a rate of road traffic accident related injury of 163 per 100,000 people. 28 of the 123 causalities resulted in death, resulting in a fatality rate of 37 (per 100,000 people). According to estimates, there have been 21,681 and 4,922 road traffic accident related injuries and fatalities per 100,000 motor vehicles, respectively. (Teferi Abegaz, January 29, 2019).

After unintentional accidents, RTAs were the second most prevalent type of accidents and injuries, accounting for 22.8% of all incidents. 43.8% of all fatalities resulting from accidents and injuries were related to RTAs. RTA casualties included drivers (21.9%), passengers (35.0%), and vulnerable road users (21.0% for motorcycles, 12.1% for pedestrians, and 2.9%

for cyclists). 15.3% of the accidents were either children under the age of 15 or seniors above the age of 64, making up about half (47.1%) of the causalities' age range. Males made up over two-thirds (65.0%) of the victims (Teferi Abegaz, January 29, 2019).

2.3. Contributing Factors for Road Traffic Accident overview

According to the American Association of State Highway and Transportation's (AASHTO) (Anon., 2010), roadway factors account for 3% of all traffic accidents, while the remaining 97 % are caused by a combination of Human and vehicle related factors. Hence, it is tough to pinpoint a single cause of road accidents since the majority of the contributing factors are interrelated, and accidents frequently involve a number of different components at effect.

2.3.1. Non-Behavioral Factors

Non-behavioral factors are those that have no direct relationship with human being and have developed for the reason that of the difficulty in directly predicting or controlling behavioral factors. Therefore, human factors can be indirectly controlled and predicted by looking into non-behavioral elements associated with road variables, specifically roadway geometric design.

Horizontal curvature, vertical curvature, gradient, number of lanes, lane width, shoulder width and type, median width and type, number of access points, curve radius, Super-elevation, sight distance, and pavement conditions are some of the elements of roadway geometric design that have been investigated to influence road accidents. However, some of these elements may not have a significant impact on road accident rates(Harwood et al., 2000).

2.3.1.1. Road Traffic Accidents related to Geometric design elements

Geometric design elements are crucial in determining a roadway's traffic operational efficiency. As several studies have demonstrated, it is challenging to define, prove, and quantify where traffic accidents may occur or where black spots may form with confidence. However, everyone agreed that there is a strong link between traffic safety and the consistency of geometric design.

Studies that have shown the influence of various geometric design elements on the occurrence of accidents have concluded that not all variables have the same level of influence in all places Perkins et.al. (Perkins, 2014) and. This uncertainty in the impact of geometric variables on accidents has prompted researchers to develop mathematical models to better understand the relationship.

Mathematical models enable highway agencies to select design standards that are essential to highway safety and to allow comparison among alternative designs that will optimize the overall safety of the highway system under limited resources and other constraints. These models can also be used to test the sensitivity of accident rates to changes in a specific geometric variable.

By considering five independent variables—the radius of bend, the degree of curvature, the super-elevation, the widening of the curve, and the side freedom (at the inner side of the curve)—through three dependent variables—the EAN (Equivalent Accident Number), accident level, and accident number, Willy Kriswardhana et al. (W. Kriswardhana, 2020) used statistics to model the relationships between the traffic accident rate and road design parameters. As a result, they have concluded that; The horizontal alignment parameter with the closest relationship to EAN is a free area on the inner side of the curve with a coefficient of determination (R^2) of 0.5841, to mean that the free area on the inner side of the curve affects EAN by 58.41%. Super-elevation has the strongest relationship with the level of an accident, with a coefficient of determination (R^2) of 0.1955, implying that super-elevation affects the accident rate by 19.55%.

The horizontal alignment parameter with the closest relationship to the number of accidents has a coefficient of determination (R^2) of 0.6471, meaning the free area on the inner side of the curve affects the number of accidents by 64.71%. According to the data obtained at the reviewed curve, the type of accident is a head-on collision and a run-off-road collision. The main cause of the two types of accidents is poor visibility of the curve, which makes it difficult to maneuver.

Poul Greibe (2001) (Greibe, 2003) on his research Accident prediction models for urban roads (Based on information on land use, the number of minor side roads, and the posted speed limit, six types of urban road links have been established. Additionally, estimates of accident prediction models for each type of road analyzed). In two different investigations, a number of models were estimated as urban road connections and urban junctions using data from 1024 junctions and 142 km of road links. The primary objective of the study was to develop simple, practical accident models that can precisely forecast the anticipated number of collisions at urban intersections and road linkages. Techniques such as generalized linear modelling were employed to link accident frequencies to explanatory factors. By collecting road links and junctions' data then classified into homogenous sections, the regression analysis was performed by using GNMODE procedure by assuming that traffic counts are Poisson distributed. As a conclusion;

- AADT and its impact on safety: the findings revealed that the most crucial variable in the models is vehicle traffic flow and accident frequencies are related to the AADT with a power of 0.8-1.0.
- Speed limits: the modeling revealed that high-speed road links have a lower accident risk. This does not imply that high speeds are inherently safer; rather, the results for this variable highlight the correlation problems within the data set. High-speed roads typically have few vulnerable road users and are located in sparsely populated areas.
- Number of lanes: Compared to road links with two or more lanes, road links with one lane and no center line have more collisions involving cars travelling in the same direction.
- Road width: for the majority of accident types, it appears that road links with a road width (from curb to curb) of 8–8.5m have the lowest accident risks.
- A major problem was strong internal correlation within the data. Variables describing traffic flow tend to correlate strongly with other variables like road width, number of lanes, etc. As a result, it was challenging to assess the safety impacts from a single explanatory variable because it may be influenced by other model variables.
- Another problem in safety analysis is the relatively small number of observed accidents in the data, which may cause problems in the statistical studies

In their study, Travelling Speed and the Risk of Crash Involvement on Rural Roads, Kloeden CN et al. (Kloeden, 2001) used a case control study design to quantify the link between free travelling speed and the risk of being involved in a fatal crash in 80 km/h or higher speed limit zones in rural South Australia. Using the latest computer assisted crash reconstruction techniques, the Road Accident Research Unit studied and rebuilt the collisions involving the 83 case passenger automobiles that were present. The 830 control passenger vehicles' locations, routes, hours of operation, and days of the week were compared to those of the cases, and a laser speed gun was used to gauge their speeds.

Additionally, it was discovered that the likelihood of a passenger car going at a free speed causing a fatal collision grew more rapidly than exponentially compared to a car travelling at an average speed. When travelling 10 km/h over the average speed of non-crash involved vehicles, the risk of being involved in a fatal crash is more than twice as high, and when travelling 20 km/h over that average speed, the risk is over six times as high. A decrease in the absolute speed of traffic is much more crucial in reducing crash frequency than a decrease in traffic speed differences, according to the mechanisms investigated for this increase in risk (where higher speeds are associated with longer stopping distances, increased crash energy, and a greater likelihood of loss of control). In rural and outlying places, the danger of being involved in a fatal crash rises more or less exponentially as free travel speed increases. It was discovered that exceeding the average speed of other traffic by just 10 km/h doubled the probability of being involved in an accident. Additionally, it was discovered that lowering the speed limit on undivided roads to 80 km/h would likely have a noticeable impact on the frequency of casualty crashes. Small speed reductions in rural areas have the potential to significantly lower the number of accidents involving casualties in those areas.

By choosing 25 undivided and 29 divided road sections for which reliable and sufficient accident data are available, Girma Berhanu (2004) (Berhanu, 2004) conducted research on models relating traffic safety with road environment and traffic flows on divided and undivided arterial roads in Addis Abeba. The road sections were then grouped as homogenous sections in terms of adjacent land use and cross-sectional characteristics such as number of lanes, road width, median, and shoulder. To estimate the model parameters, Poisson and negative binomial regression approaches were applied. Both divided and undivided highways have a considerable influence from the number of lanes. Additionally, a strong correlation between lane width and the overall accident rate on undivided highways has been discovered. However, no such association has been discovered for split roadways. In addition, the frequency of accidents on split highways is strongly correlated with the density of minor junctions, in contrast to the accident risk on undivided roads.

Raised curbs and paved walkways were another crucial explanatory factor taken into account in the study and demonstrated a substantial association with accident frequency on the roadways under examination. According to research, installing raised curbs at road edges lowers the frequency of all pedestrian accidents on divided roads by 17% and on undivided roads by 46% (Berhanu, 2004).

Using a multivariate ratio of polynomials, Garber and Ehrhart (2000) (Ehrhart, 2000) discovered that the variables "standard deviation of the speed profile" and "flow per lane" were substantially connected to crash rate when they looked at just the road segments between two major junctions.

According to Zhang and Ivan (2005) (Ivan, 2019), there is a significant correlation between accidents and the speed limit, "the sum of the change rate of the horizontal curvature," "the sum of the change rate of the vertical curvature," and "the maximum horizontal curvature rate."

Negative binomial regression models were proposed by Pardillo and Llamas (2003) (Salvatore Cafisoa, 2003) for two definitions of segments: (1) segments of constant length of 1 km, and

(2) network linkages connecting nodes with variable lengths ranging from 3 to 25 km. Significant relationships between accidents and access density, average sight distance, average speed limit, and percentage of no passing zones were discovered. These relationships highlight the necessity of using these variables to define homogenous sections with adequate lengths of at least 400 m.

Abdel-Aty and Radwan (2000) (M A Abdel-Aty, 2000) divided a 227 km long road into 566 segments with uniform traffic flow and geometry (degree of horizontal curvature, shoulder and median widths, rural/urban classification, lane width and number of lanes), all of which were found to be strongly related to the accident occurrence. According to the research, sections should be 0.8 km or longer in order to produce a reliable accident prediction model. The calibration of models that served as the foundation for those used in the Interactive Highway Safety Design Model) (Paniati, 1996) and, more recently, the Highway Safety Manual scheduled for release in 2010 has perhaps been the most significant US research recently.

That study's findings (Bared and Vogt, 1998)) (Vogt, 1998) linked lane width, shoulder width, roadside hazard rating, horizontal curvature, vertical slope and curvature, and driveway density to accidents per unit of exposure (measured as the product of traffic volume and segment length). The average radius of horizontal curves for a roadway section shows promise as a measure of design consistency, but it does not seem to be as appropriate as the speed reduction for individual horizontal curves, according to another IHSDM-related effort (Fitzpatrick et al., 2000) (Fitzpatrick, 2000).

Emer T. Quezon et. Al (2017) (Teyba Wedajo, January-2017)on their research Analysis of Road Traffic accident related to geometric design elements in Alamata- mehoni- hewane road project, by conducting road safety inspection audits on the said project by focusing on the geometric design elements of the road (i.e. cross sectional elements; lane width, shoulder width, median type and width, sight distance, horizontal and vertical curves) and ancillary road features such as road markings, lighting conditions they have come to conclusion that, from the collected traffic accident data, around 56.4% was occurred at rural sections and the reason for that was the terrain condition (i.e. most of the section is under mountainous and escarpment terrain), narrow road right of way, deficient lane width, eroded shoulder, invisible pavement markings, over speeding and overloading.

From the overall accident recorded on the roadway for five years, 38.9 % of the record is considered as non-behavioral factor (i.e. roadway defect 33% and brake failure 5%) and 48.5% is regarded as accident caused by behavioral factors such as over speeding, over loading, failure to respect the right hand rule, failure to give way to pedestrian and lack of experience. the remaining 12.6% was regarded as accident cause unrecorded.

By taking into account logical and realistic geometrical and environmental aspects that affect accident frequencies, Venkataraman Shankar et al. (1994) (Venkataraman Shankar, June 1995) proposed a model by the type of accident that occurred to evaluate the effect of roadway geometrics and environmental factors on rural motorway accident frequencies. Through interaction variables, the model specifically gives insight on how weather and geometric elements interact. Designers can determine geometric thresholds, such as maximum grade, beyond which their interaction starts to significantly affect accident occurrences by using indicator-type interaction variables. Their findings have showed that,

- by taking horizontal curves with design speeds less than 96.5kph as an explanatory variable, tend to increase sideswipe and rear-end collisions but decrease crashes with stationary objects and parked cars. According to this research, drivers may tend to slow down on curves having design speeds less than 96.5 kph due to signage and visual perception, preventing serious crashes with stationary objects and other large vehicles. This potential speed reduction does not, however, seem sufficient to prevent sideswipe or rear-end collisions brought on by lane violations or speed differentials because of braking on the curve. It should also be mentioned that drivers are likely to avoid parking on the shoulder, which tends to decrease the number of parked vehicle collisions.
- The results reveal that under designed curves between 96.5 kph and 112.6 kph do not have a visual effect on drivers to slow down. This is true for the number of horizontal curves designed between 96.5 kph and 112.6 kph. The final effect is an increase in lane violations, which cause vehicle collisions, as well as vehicles driving off the road and striking stationary objects. Fixed-object collisions are more likely than vehicle crashes in the same direction to cause serious injuries from the viewpoint of severity. The slight underdesign curves (96.5 kph to 112.6 kph) should then be improved if fixed-object accidents exhibit rising trends in certain areas.
- Number of horizontal curves in a section, the finding suggests two distinct phenomena; As the frequency of horizontal curves increases, the likelihood of vehicle crashes in the same direction tends to increase since curve speeds do not decrease enough to avoid lane violations. The likelihood of fixed objects, such as guardrails, being present on sections with more curves is increased by the fact that fixed-object collisions tend to increase with the total number of curves in a section. A more severe accident, such a vehicle overturn, is avoided when such

objects are present. This finding leads to the conclusion that, in situations when the topography prevents the construction of straight sections, it is preferable to design longer but fewer horizontal curves.

- Average spacing of horizontal curves in a section, even though the research does not consider an effect of roadway geometry that was not noticeable in the overall accident frequency model, the spacing of horizontal curves in a segment significantly affects vehicle speeds, it makes sense that if curves are farther apart, drivers will be less cautious, which will increase vehicle speeds. As a result, if curves are spaced more apart in a segment, there is a higher chance of an overturn. The adoption of corrective actions should be given careful consideration in this regard. In order to reduce the incidence of overturn accidents, it works better to decrease the distance between the curves. However, it would seem paradoxical to physically locate curves closer as a protection. More advance warning signs should be placed in portions with longer curve spacing as an alternative action. The spacing of curves is slightly changed in the driver's thinking by strategically placing additional warning signs in advance.
- Sections having lower horizontal curve radius, Sections with a lower horizontal curve radius may have a different type of terrain than sections with a higher horizontal curve radius. Sections with higher minimum radii may induce lane violations and sideswipes in the driver. However, low radii curves are typically found on winding sections of the highway, which reduces the chance that enough speed will build up to cause an overturn accident.
- Sections having maximum gradient, the section with the steeper maximum upgrade between any two will see more rear-end collisions and other same-direction accidents. In addition, if the maximum gradient in that segment is higher than 2%, rear-end collisions will increase significantly. Speed differences brought on by the impacts of grades can explain both

effects. When there are downgrades, the effect of grades is reversed. There will be less rear-end and other same-direction incidents on the stretch with the steeper maximum downhill between any two sections, probably because to smaller speed differentials. The visual impact of brake lights warning drivers to the potential slowing of vehicles ahead appears to counterbalance much of the effect of the longer stopping distance on downgrades. On the other hand, drivers are less likely to apply the brakes when an upgrade eliminates a critical warning indication for speed reductions.

2.3.1.2. Road Traffic Accidents related to traffic volume and characteristics

It is difficult to point out the separate impact of traffic volume on road traffic accidents among various causes that are related to the roadway and its environment. However, many researches have been undertaken by considering the traffic volume as an independent explanatory variable especially for intersections.

Retallack et al.'s investigation on the correlation between traffic volume and accident frequency at intersections (Retallack AE, 2020) revealed that, for lower traffic volumes, the connection between traffic volume and accident frequency was roughly linear. As accident frequency rises at a faster rate in the highest traffic volumes, Poisson and negative binomial models revealed a strong quadratic explanatory component. This suggests that the most effective approach to lower accident frequency would be to concentrate management efforts on preventing these situations.

Using a power function, various studies have examined the link between a road segment's traffic volume and crashes. The form that has been employed most frequently for this purpose is

$$Y = L * e^{a} * (AADT)^{b} \;\ldots\ldots\ldots\ldots (Eq\ 1)$$

where: Y is the estimated number of accidents on a segment, L is the segment length, and a and b are regression coefficients to be estimated. This form is common because it is simple and fulfils the boundary condition that if there is no traffic (i.e., if AADT is zero), then the estimated number of accidents should be zero.

In the power model, it is generally accepted that b is positive since the number of crashes are expected to increase with increase in traffic volume. Depending on whether b is less than 1, equal to 1, or greater than 1.

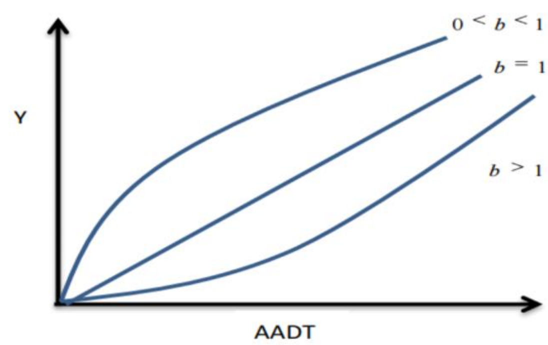


Figure 1:Shape of the Relationship between Accident Frequency and AADT as a Function of the Power b

In discussing the underlying relationship between accident frequency and traffic volume, Kononov et al., (2011) (Jake Kononov, 2011) state that the power model may not be the most appropriate functional form. As a starting point, he used neural networks (NN) to explore the underlying relationship between accident frequency and AADT using data from urban freeways from California and Colorado. The use of NN seemed to indicate that the relationship could be sigmoidal.

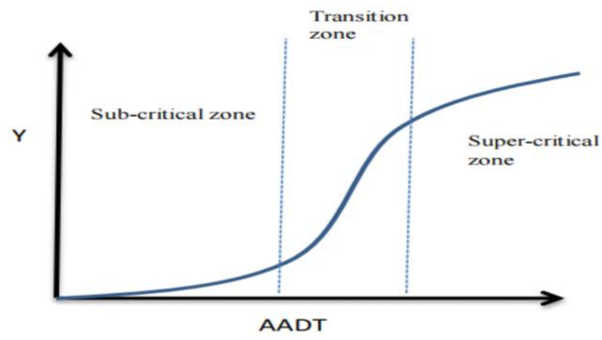


Figure 2:Sigmoidal Relationship between Accident Frequency and Traffic Volume (Kononov et al., 2011)

he then used the parameters from the Highway Capacity Manual (HCM) to explain the underlying relationship, by dividing the graph into three zones: sub-critical, transitional, and super critical zones. maintain that in the sub-critical zone (i.e., when freeways are not congested and traffic density is low), crashes increase gradually with AADT. However, in the transition zone, the crashes start increasing rapidly. Finally, in the super-critical zone, crashes still increase with AADT, but only gradually. Based on the HCM, he found that when the AADT increases from the sub-critical to the transition zone, operating speeds remain almost the same, but with a significant increase in traffic density, and this may be a possible reason for the rapid increase in crashes with AADT in the transition zone. In the super-critical zone, speeds start dropping, possibly leading to only a gradual increase in crashes with AADT.

Ezra Hauer (2004) (Ezra Hauer, 2004) highlighted that the oversimplified power function might not be adequate, particularly for collisions involving a single vehicle. Increasing traffic volume has been demonstrated to reduce the likelihood of a single-vehicle collision. Hence a functional from that allows a peak/valley and a point of inflection may be more appropriate for modeling the relationship between single-vehicle crashes and traffic volume – based on the above figure 2, the simple power model from the above equation does not allow a peak/valley or a point of inflection.

According to research by Angus Eugne et al. (2020) (Retallack, 2021), the link between traffic volume and accident frequency was essentially linear at lower traffic levels. Poisson and negative binomial models showed a substantial quadratic explanatory component, with accident frequency increasing at a quicker pace in areas with the largest traffic volumes. This implies that focusing management efforts on preventing these circumstances would be the most efficient strategy to reduce accident frequency.

2.4. Modeling Methods for Crash-Frequency Data Analysis

2.4.1. Evolution of Crash Prediction Models

formerly, when predictions of a roadway's current or future safety performance were required, they were created using one of the four methods:

2.4.1.1. estimates from historical accident data, from observed crash frequency and crash rate

Crash frequency and crash rates are used as a prediction tool to identify and prioritize sites that are in need of modification or for evaluation of the performance of effectiveness of treatments.

crash frequency and crash rate are usually used in conjunction with other analysis techniques, such as reviewing crash records by one or more of the following: year, collision type, crash severity, or environmental conditions to identify other apparent trends or patterns over time.

Using this method has an advantage of;

- Understandability—observed crash frequency and rates are easily understandable
- Acceptance—it is easy for the public to assume that observed trends will continue to occur;

- Limited alternatives—in the absence of any other available methodology, observed crash frequency is the only available method of estimation.

Crash estimation methods based only on historical crash data have several limitations. These include;

- The accuracy of estimation result will highly depend on
  - Accuracy and quality of the collected crash data
  - Crash reporting thresholds and the frequency-severity indeterminacy
  - Differences in data collection methods and definitions used by dominions
- Limitations due to randomness and change
  - Natural variability of crash frequency over time
  - Variations in roadway characteristics and its environment
  - Conflict between crash frequency variability and changing site conditions

2.4.1.2. Forecasts from statistical models based on regression analysis

Safety analysts have used statistical methods to create models that predict how and when accidents will occur on highways. Regression analysis is used to estimate the amounts of the coefficients or parameters in these models after a database of accident and roadway variables (such as traffic volumes, geometric design elements, and traffic control features) is compiled from the records of highway agencies. Historically, multiple regression analysis has been used to build the majority of these models since they are very accurate tools for estimating the expected total accident experience for a specific location or a group of locations, but they have not been successful in separating the effects of specific geometrical element or traffic control elements. However, regression model's coefficients can be easily misinterpreted as the actual impact of a small adjustment to the relevant roadway parameter. though regression models are centered on statistical correlations between road variables and accidents, it does not necessarily mean they should have cause-and-effect relationships which is a shortcoming for the regression models. Furthermore, it might be challenging to distinguish between the individual effects of independent variables in a model if they are strongly correlated to one another. Additionally, if a variable in the model has a high correlation with an important variable that simply happens to be missing from the data source, the variable's coefficient may then reflect the impact of the absent variable rather than its own. As a result, the coefficient of a particular geometric feature may be a reliable indicator of its actual impact on safety, or it may just be an artifact or a substitute for the feature's correlation to other variables.

2.4.1.3. Outcomes of before-after studies,

For many years, before-and-after studies have been used to assess how highway improvements have reduced accidents. But the majority of before-and-after studies that have been reported in the literature have design flaws that prevent them from taking into account the consequences of regression to the mean bias (i.e. because of the random nature of accidents, locations with high short-term accident experience are likely to experience fewer accidents in the future even if no improvement is made). As a result, it is inconceivable for the potential user of the before-and-after study results to know for sure whether they accurately reflect the improvement's potential effectiveness in reducing accidents or whether they are an overly optimistic forecast that is skewed by regression to mean.

2.4.1.4. Studies and professional conclusions drawn by skilled engineers

Making accurate safety predictions may need the use of expert judgment, which has been acquired through many years of experience in the field of highway safety. Experts may challenge to make quantitative estimations in the absence of a standard, but they typically succeed at making comparative assessments (e.g., A is likely to be less than B, or C is likely to be about 10 percent larger than D). Therefore, in order for specialists to reach on helpful conclusions, they require a frame of reference based on past accident data, statistical models, or before-and-after research findings.

However, employing each of these approaches separately has resulted in serious shortcomings to determine the safety performance of a given roadway so that it leads to seeing new methods by combining elements of each of abovementioned methods to predict the current and future safety performance of the road the is needed to be investigated. (D.W. Harwood, May 1997—September 2000)

Accordingly, many researches have been undertaken internationally by using statistical prediction models as a base but supplemented by the other three methods to predict safety performance of the road and the models are becoming more advanced through time.

Crashes are simulations of count data and are properly modeled using a specific family of statistical models called count data models. Many models have been developed to establish the relationships for different roadway classes, vehicle configurations, and accident severity types. However, most of these models were developed using the linear regression and have been reported to have several unsatisfactory statistical properties in describing vehicle accident events on the road (PAUL P. JOVANIS, 1986); (Buyco. C, 1987)). These unsatisfactory properties of the linear regression models have led to the investigation of the Poisson and negative binomial regression models in recent studies (Sarath C. Joshua, 2007); (Shaw-PinMiaou, August 1994).

These established models differed in (i) the geometric design and confounding variables (covariates or explanatory variables) that link the accident rate of a road section to its associated attributes, which can be additive or multiplicative; (ii) the choice of the functional form that delineates road sections, particularly in terms of their length; and (iii) the treatment of the distributional assumption of the occurrences of vehicle accidents and (vi) the consideration of data uncertainties, e.g. the uncertainties of accident location and vehicle exposure data, due to sampling and non-sampling errors.

2.4.2. Selected Modeling Methods for Crash-Frequency Data Analysis

2.4.2.1. Poisson Regression Model

The traditional least-squares regression, which implies a continuous dependent variable, cannot be used since crash-frequency data are non-negative integers. The Poisson regression model has been the starting point for much recent research in this area because the dependent variable is a non-negative integer. In a Poisson regression model, the probability that a highway entity (segment, intersection, etc.) i will experience yi crashes per some period of time (where yi is a non-negative integer) is given by:

$$p(y_i) = \frac{Exp(-\lambda_i)\lambda_i{}^{yi}}{y_i!} \text{..................(Eq2)}$$

where $P(y_i)$ is the probability of roadway entity i having y_i crashes per period and λ_i is the Poisson parameter for roadway entity i, which is equal to roadway entity i's expected number of crashes per year, E[yi]. Poisson regression models are estimated by specifying the Poisson parameter λ_i (the expected number of crashes per period) as a function of explanatory variables, the most common functional form being

$$\lambda_i = Exp(\beta x_i) \text{…………..(Eq3)}$$

where x_i is a vector of explanatory variables and β is a vector of estimable parameters.

Although the Poisson model has been used as a starting point for crash-frequency analysis for several decades, researchers have frequently discovered that crash data exhibit characteristics that make the use of the simple Poisson regression (as well as other models) unsuitable. Some Poisson model extensions are problematic. Poisson models, in particular, cannot handle over- and under-dispersion, and they can be affected by a small sample size in small samples, this can lead to biased results (Shaw-Pin Miaou, 2003).

2.4.2.2. Negative Binomial (Poisson-gamma) Regression Model

The negative binomial (or Poisson-gamma) model is an extension of the Poisson model to overcome possible over dispersion in the data. The negative binomial/Poisson gamma model assumes that the Poisson parameter follows a gamma probability distribution. The model results in a closed-form equation and the mathematics to manipulate the relationship between the mean and the variance structures is relatively simple. (Shaw-Pin Miaou, 2003). The negative binomial model is derived by rewriting the Poisson parameter for each observation i as

$$\lambda_i = xp(\beta x_i + \varepsilon_i)\ ; \ldots\ldots\ldots\ldots\ldots\ldots.(Eq\ 4)$$

where EXP (ε_i) is a gamma-distributed error term with mean 1 and variance α. The addition of this term allows the variance to differ from the mean as

$$VAR\ [y_i] = E[y_i][1+ \alpha\ E[y_i]] = E[y_i]+ \alpha\ E[y_i].\ \ldots\ldots\ldots\ldots(Eq5)$$

The Poisson regression model is a limiting model of the negative binomial regression model as α approaches zero, which means that the selection between these two models is dependent upon the value of α. The parameter α is often referred to as the over dispersion parameter.

The Poisson-gamma/negative binomial model is probably the most frequently used model in crash-frequency modeling. However, the model does have its limitations, most notably its inability to handle under dispersed data, and dispersion-parameter estimation problems when the data are characterized by the low sample mean values and small sample sizes (Anon., 2018).

2.4.2.3. Poisson-Lognormal Model

Some researchers have proposed using the Poisson-lognormal model as an alternative to the negative binomial/Poisson-gamma model for modeling crash data (Miaou et al., 2003; Lord and Miranda-Moreno, 2008; Aquero-Valverde and Jovanis, 2008) (Lord, n.d.) (Aguero-

Valverde, (2008)) (Shaw-Pin Miaou, 2003). The Poisson-lognormal model is similar to the negative binomial/Poisson-gamma model, but the EXP (ε_i) term used to compute the Poisson parameter is lognormal rather than gamma distributed. Although the Poisson-lognormal potentially offers more flexibility than the negative binomial/Poisson-gamma, it does have its limitations. For example, model estimation is more complex because the Poisson-lognormal distribution does not have a closed form and the Poisson-lognormal can still be adversely affected by small sample sizes and low sample mean values (Shaw-Pin Miaou, 2003).

2.4.2.4. Zero-inflated Poisson and Negative Binomial

Zero-inflated models have been developed to handle data characterized by a significant number of zeros or more zeros than the one would expect in a traditional Poisson or negative binomial/Poisson-gamma model. Zero-inflated models work on the premise that a splitting regime that simulates a roadway segment's crash-free versus crash-prone propensity accounts for the excess zero density that cannot be accommodated by a conventional count structure. A binary logit or probit model can be used to determine whether a highway entity is more likely to be in the zero or non-zero states (Lambert, 1992); (Quintel Washington, 2014).

Others have criticised the use of this model in highway safety despite its broad applicability to a number of circumstances where the observed data are characterised by substantial zero densities. For instance, (Dominique Lord, 2007) claimed that the zero or safe state cannot accurately describe the process of generating crash data because its long-term mean is zero.

One well-known zero-inflated model is Diane Lambert's zero-inflated Poisson model, which concerns a random event containing excess zero-count data in unit time. (Lambert, 1992). The zero-inflated Poisson (ZIP) model mixes two zero generating processes. The first process generates zeros and the second process is governed by a Poisson distribution that generates counts, some of which may be zero. The mixture distribution is described as follows:

2.4.2.5. Conway-Maxwell-Poisson Model

Conway and Maxwell (1962) introduced the Conway-Maxwell-Poisson distribution as a generalization of the Poisson distribution for modeling queues and service rates. (Galit Shmueli, 2004) investigated the statistical properties of the Conway-Maxwell-Poisson distribution in greater depth, and (Joseph B. Kadane, 2006) created conjugate distributions for the parameters. The Conway-Maxwell-Poisson can deal with both under-dispersed and over-dispersed data, and several common probability density functions are Conway-Maxwell-Poisson special cases (for example, the geometric distribution, the Bernoulli distribution, and the Poisson distribution). The Conway-Maxwell-Poisson distribution's flexibility greatly expands the types of problems for which it can be used to model crash frequency data. This model was recently used in highway safety research and was found to be comparable to the Poisson-gamma model for data with overdispersion (Dominique Lord, 2008). However, its main advantage is related to data with low dispersion. On the downside, this model is susceptible to low sample mean and small-sample bias, and there have been no multivariate applications of the approach to date.

The Conway-Maxwell-Poisson distribution is defined as:

$$P(Yi = yi|Xi, Zi) = \frac{1}{Z(\lambda i, Vi)} * \lambda in/(Yi!)vi\ ,\ y_i = 0,1,2,\ldots\ldots\ldots.(Eq\ 6)$$

Where the normalization factor is $Z(\lambda i, Vi) = \sum_{n=0}^{\infty} \lambda \mathrm{in}/(\mathrm{n})!\,\mathrm{vi}$

And $\lambda i=\exp(X_i'\beta)$, $Vi=\exp(-g_i'\delta)$

The β vector is a (k+1)*1 parameter vector. (the intercept is β_0, and the coefficients for the K regresser are $\beta_{1,\ldots}$ $\beta_{k)}$. the δ vector is an (m+1)*1 parameter vector.

The intercept is represented by δ_0, and the coefficients for the m regressors are $\delta_{1,}$ δ_k)

The covariates are represented by x_i and g_i vectors

One of the restrictive properties of the Poisson model is that the conditional mean and variance must be equal. However, the Conway-Maxwell-Poisson distribution overcomes this restriction by defining an additional parameter, ν, which governs the rate of decay of successive ratios of probabilities such that,

$$P(Yi = yi - 1)/P(Yi = yi) = (yi)\ \nu i/\ \lambda i \quad \ldots\ldots(Eq\ 7)$$

The introduction of the additional parameter, ν, allows for flexibility in modeling the tail behavior of the distribution. If ν=1, the ratio is equal to the rate of the decay of the Poisson distribution. If ν>1, the rate of decay increases in a non-linear fashion, thus shortening the tail of the distribution (under dispersion). If ν<1, the rate of decay decreases, enabling to model processes that have longer tails than the Poisson distribution (over dispersed data)

2.4.2.6. Gamma Model

The gamma model has been proposed by (Jutaek Oh, 2006) to analyze crash data exhibiting under-dispersion. This model can handle over dispersion and under-dispersion and reduces to the Poisson model when the variance is roughly equal to the mean of the number of crashes. Although this model performs well statistically, it is still a dual-state model, with one of the states having a long-term mean equal to zero. The gamma model has seen limited use since it was first introduced by Oh et al. (2006).

The gamma distribution can be parameterized in terms of a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, called a rate parameter. A random variable X that is gamma-distributed with shape α and rate β is denoted:

$X \sim \Gamma(\alpha, \beta) = \text{Gamma}(\alpha, \beta)$

The corresponding probability density function in the shape-rate parameterization is

$$f(x;\ \alpha, \beta) = x\,\alpha - 1e - \beta x \beta \alpha / \Gamma(\alpha)\ for\ X > 0\ \alpha, \beta > 0 \ldots\ldots\ldots (Eq\ 8)$$

Where $\Gamma(\alpha)$ is the gamma function. For all positive integers, $\Gamma(\alpha) = (\alpha-1)!$

2.4.2.7. Generalized Estimating Equation Model

(Lord, 2000) used the generalized estimating equation model to model crash data with repeated measurements in highway safety analysis. it is common to have data from roadway entities (roadway segments or intersections) over multiple time periods, which creates a serial correlation problem. The generalized estimating equation is not a regression model, but rather a method for estimating models with serially correlated data. The generalized estimating equation model can handle serial correlation in a variety of ways, including independence, exchangeable, dependence, and auto-regressive type 1 correlation structures. When using count-data models with a complete dataset (few if any omitted variables), the correlation structure has a minimal influence on the modeling output; however, choosing the correlation type can be critical when the dataset contains omitted variables.

2.4.2.8. Generalized Additive Models

Traditional count-data models are less flexible than the generalized additive model (Hastie, 1990); (Wood, 2013). According to (Zhang, 2008), generalized additive models have a more flexible functional form that includes smoothing functions for the model's explanatory variables. The smoothing function represents a more flexible relationship in terms of how explanatory variables are considered, and it is not limited to linear or logarithmic relationships, as is common in traditional count models.

Although the generalized additive model is more flexible than traditional count models, it has limitations. First, because these models have more parameters, the estimation process can become quite complex, especially if the default values are not used. Second, generalized additive models are more difficult to interpret than traditional count models because they use spline functions. Third, if the explanatory variables are exogenous and the dependent variable has a linear or exponential relationship with them, the modeling results of the generalized additive model and traditional models are likely to be similar. So far, only a few papers have used generalized additive models for crash-frequency analysis.

2.4.2.9. Random-Effects Models

There may be reason to anticipate correlation between observations. This correlation could be caused by spatial considerations (data from the same geographic region may share unobserved effects), temporal considerations (as in panel data, where data collected from the same observational unit over successive time periods may share unobserved effects), or a combination of the two. To account for this correlation, random-effects models (in which common unobserved effects are assumed to be distributed over spatial/temporal units according to some distribution and shared unobserved effects are assumed to be uncorrelated with explanatory variables) and fixed-effects models (in which common unobserved effects are accounted for by indicator variables and shared unobserved effects are assumed to be correlated with independent variables) can be used. In the context of count models, (Jerry A. Hausman, 1984) first investigated random-effects and fixed-effects negative binomial models for data set (with temporal considerations) in their study of Research and development patents.

2.4.2.10. Negative Multinomial Models

The problem of correlation among observations can also be addressed with a negative multinomial approach (Guo, 1996). This model is similar to the negative binomial in that it uses

$$\lambda_i = Exp(\beta x_i + \varepsilon_i) \text{ ………..(Eq 9)}$$

except now EXP (ε_i) is associated with a specific entity (roadway segment, intersection) as opposed to a specific observation. This is a crucial distinction to make because, for instance, if one were to look at annual crash frequencies with five years of data, each roadway entity would produce five observations with their own EXP (ε_i), which could lead to a correlation issue. For the negative multinomial model, at the segment/intersection level, the EXP (ε_i) is again assumed a gamma-distributed error term with mean 1 and variance α. (Gudmundur F. Ulfarsson, 2003) presented an application of the negative multinomial model to crash frequency data and compare estimation results to standard negative binomial and random-effects negative binomial models. Negative multinomial models cannot handle Under-dispersion and are susceptible to problems in the presence of low sample means and small sample sizes.

2.4.2.11. Random-Parameters Models

An extension of random-effects models are random-parameter models. Random-parameter models, on the other hand, permit each estimated parameter of the model to vary across each distinct observation in the dataset, as opposed to merely having an impact on the model's intercept. These models make an effort to explain the unobserved heterogeneity between various roadway sites (Karim El-Basyouny, 2009). Panagiotis Ch. Anastasopoulos and Karim El-Basyouny (2009) (Panagiotis Ch. Anastasopoulos, 2009) both applied these models to crash-frequency data.

. Because each observation has its own parameters, the final model will often provide a statistical fit that is significantly better than a model with traditional fixed parameters. However, random-parameter models are very complex to estimate, they may not necessarily improve predictive capability, and model results may not be transferable to other data sets because the results are observation specific (Shugan, 2006); (Simon Washington, 2013).

The final model frequently offers a statistical fit that is noticeably better than a model with standard fixed parameters because each observation has its own parameters. Random-parameter models can be highly difficult to estimate, they may not always increase the capacity to forecast, and because the results are observation-specific, they may not be transferable to other data sets (Shugan, 2006); (Simon Washington, 2013).

2.4.2.12. Bivariate/Multivariate Models

Bivariate/Multivariate models become necessary in crash-frequency modeling when, instead of total crash counts, one wishes to model specific types of crash counts (for example, the number of crashes resulting in fatalities, injuries, etc.). Modeling the counts of specific types of crashes (as opposed to total crashes) cannot be done with independent count models because the counts of specific crash types are not independent (that is, the counts of crashes resulting in fatalities cannot increase or decrease without affecting the counts of crashes resulting in injuries and no injuries). To resolve this problem, Bivariate/multivariate models are used because they explicitly consider the correlation among the severity levels (for example) for each roadway entity (Shaw-Pin Miaou, 2005).

Multiple linear regression model is one of the multivariate models with the following form:

$$Y = B0 + B1(X1) + B2(X2) + B3(X3) \ldots + Bn(Xn) \ldots\ldots (Eq\ 10)$$

Where:

Y: the predicted number of accidents.

B_0: the constant coefficient of the regression line.

B_1, B_2, B_3, B_n: the regression coefficients

2.4.2.13. Finite Mixture/Markov Switching Models

Finite mixture/Markov switching models are an innovative type of model for studying heterogeneous populations. Although this type of model has been around for a long time, it has previously gained popularity due to advancements in computing power and technology (Frühwirth-Schnatter, 2006). The assumption for finite mixture models is that the overall data is generated from several distributions that are mixed together, implying that individual observations are generated from an unknown number of subgroups. Markov Switching models also assume that the data is generated by several underlying distributions and that individual observations can switch between these distributions over time.

Several studies have been conducted in recent years to investigate the application of finite mixture models (Byung-Jung Park, 2009) and Markov switching models (Malyshkina et al., 2009; (Nataliya V. Malyshkina Fred, 2009) to highway safety. Finite mixture and Markov switching models have the potential to provide significant new insights into crash data analysis, but they are also quite complex to estimate.

A random variable Y with density $p(y|\vartheta) = \eta 1 p(y|\theta 1) + \cdot\cdot\cdot + \eta K p(y|\theta K)$,

Where all component densities arise from the same parametric distribution family T (θ) with density p (y|θ), indexed by a parameter θ ∈ Θ, is said to arise from a (standard) finite mixture of T (θ) distributions, and abbreviated by (Fr¨uhwirth-Schnatter, December 2, 2008)

$$Y \sim \eta 1 T\ (\theta 1)\ + \cdot\cdot\cdot\ + \eta K T\ (\theta K)\ \ldots\ldots\ldots\ldots (Eq\ 11)$$

This could be seen as the marginal distribution of a model, where a hidden categorical indicator S is introduced which is assumed to follow a multinomial distribution:

$$Y|S\ \sim\ \eta 1 T\ (\theta S), S\ \sim\ MulNom\ (1, \eta 1, \ldots, \eta K) \ldots\ldots (Eq\ 12)$$

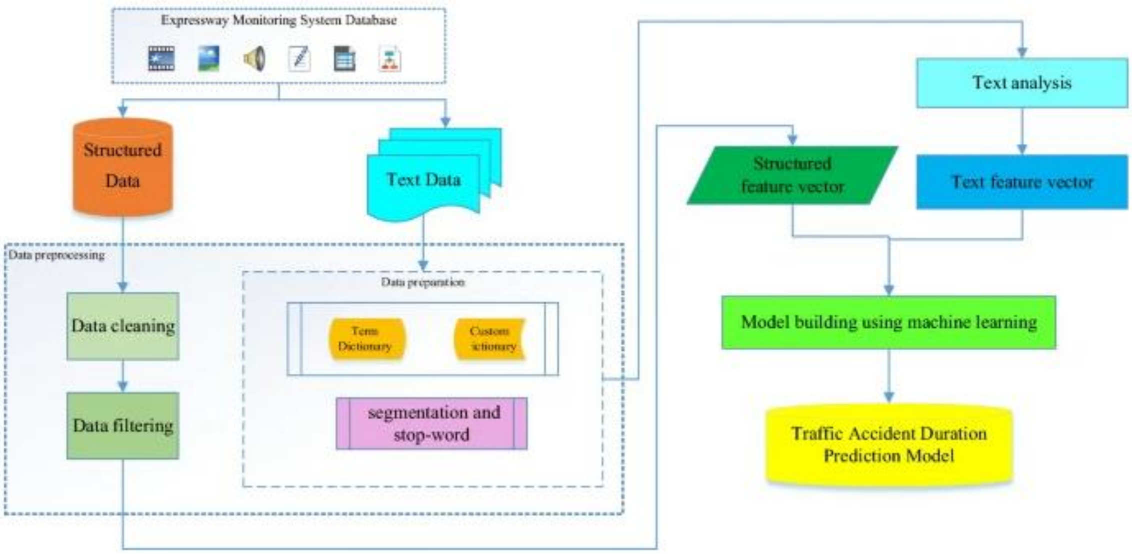
2.4.2.14. Duration Models

Another approach to the crash-frequency problem is to consider the time between crashes rather than the frequency of crashes over a given time period. The frequency of crashes and the time between crashes are undoubtedly related. Evidently, count-data models (such as the standard Poisson model) imply an underlying distribution of time between crashes (the underlying time distribution for the standard Poisson is exponentially distributed), and a model of the duration of time between crashes can be aggregated to produce an expected frequency in any given period of time.

The most common duration-model approach is a hazard-based model that takes into account the conditional probability of a crash occurring at some time $t + dt$ given that time t has passed since the last crash. Hazard-based models can be estimated using a wide range of distributional assumptions and non-parametric forms, allowing important inferences to be drawn about how the chances of a crash change over time (Ahmad Tavassoli Hojati, 2013).

Hazard-based duration models can be quite sophisticated in terms of their ability to deal with data and common problems associated with crash data (unobserved heterogeneity, for example), and can provide insights into duration effects. (Doohee Nam, 2000), for example, used a hazard-based duration model to study the factors influencing the time between their crashes rather than considering the number of crashes individual drivers had over their lifetimes. The duration effects found in this study were fascinating since they showed that the longer the guys went without a crash, the less probable it was that they would have one soon, but the duration of the females' crashes had no statistically significant effect on the likelihood of their occurring.

There is significant potential for the future application of duration models to crash frequency analysis, but the amount of data required in terms of crash timing and explanatory variable values and how they change over time can be prohibitive in many cases. Figure 3 shows the architecture of the duration prediction model for traffic accidents;



The architecture of traffic accident duration prediction model.

Figure 3: architecture of the duration prediction model for traffic accidents

2.4.2.15. Hierarchical/Multilevel Models

Hierarchical models are used to analyze data with correlated responses within hierarchical clusters. Crash data in highway safety can be thought of as having several levels of hierarchy. For example, the crashes themselves could be at the bottom of the hierarchy. The next level could be the vehicle type (passenger cars, trucks, etc.). The next one could be the location of the accident on the transportation network, and so on. The primary assumption in this type of model is that there may be a correlation between crashes that occur for the same type of vehicle and location because they may share unobserved characteristics related to the vehicle type or location. When traditional count-data modeling approaches are used, failing to account for the potential hierarchical structure of the data (the possibility of a complex correlation structure) may result in poorly estimated coefficients and associated standard errors ((A J Skinner, 1995); (Goldstein, 2011)).

The essential feature of hierarchical models is the fact that the model specifies the observations at the lower levels as being clustered into higher level units, and that these units themselves are considered to constitute a sample from a larger population (of accidents, of road sites, or segments, …). A multilevel/hierarchical model can thus be defined as a regression model (based on linear or generalized linear models) in which the regression coefficients are assigned a probability model. The higher levels of the model have parameters of their own – the "hyperparameters" of the model – which are also estimated from the data (Huang, 2010).

To further define the principles underlying ML models, a simplified two-level model will be used first. The response variable corresponds to the probability for driver *i* involved in accident *j* to die as a result of this accident.

The response y_{ij} is defined as being a function of the expected value defined for accident *j* (π_j) and of driver-specific variation (R_{ij}).

$$yij = \pi j + Rij \quad \text{………………(}Eq\ 13)$$

The expected average probability for each accident is in turn defined as being a nonlinear function of a linear combination of predictors.

$$\pi j = f(X\beta ij) \text{…………………}Eq\ 14$$

The expected probability for driver i to die in accident j is now defined as being a logit function of the linear combination of an average value holding for the accident population ($\gamma 0$), of the effect of level-1 (and/or) level-2 predictors $(\sum_{k=1}^{r} \gamma_k x_{hij})$ and of accident-related random deviation Uoj.

$$Logit\,(\pi_j) = \gamma_0 + u_{0j} \quad \ldots\ldots(Eq\ 15)$$

Where (γ_0) represents the average of logit (πj) across accidents and U_{0j} the accident-specific deviation from this population average value (the "accident level random variation). These deviations are assumed to be normally distributed, with mean 0 and variance σ^2_{u0} (Emmanuelle Dupont, n.d.).

2.4.2.16. Neural, Bayesian Neural Network, Support Vector Machine Models

Neural and Bayesian neural network models are multilevel network structures that define functions (Liang, 2005) the network structure is made up of nodes and weight factors that connect the various nodes in a hierarchical fashion: input layer, hidden layer, and output layer. Although Neural and Bayesian neural network models use similar modeling processes, they differ in how they predict outcome variables. The weights in neural networks are assumed to be fixed, whereas in Bayesian neural networks, the weights follow a probability distribution, and the prediction process must be integrated across all probability weights. These models have primarily been used in highway safety ((Abdelwahab, 2002); (Li-YenChang, 2005); (C. Riviere, 2006); (G. N. XieQ., 2007) as predictive models. Overall, these models outperform traditional count-model approaches in terms of linear/non-linear approximation properties. However, these models are frequently inapplicable to other data sets (G. N. XieQ., 2007).

The input variables for multi-level neural network structures are the number of vehicles per day and the road length in kilometers. The designed Multi-Layer Perceptron Neural Network (MLPNN) consists of the input layers, hidden layers and an output layer as shown on the figure.

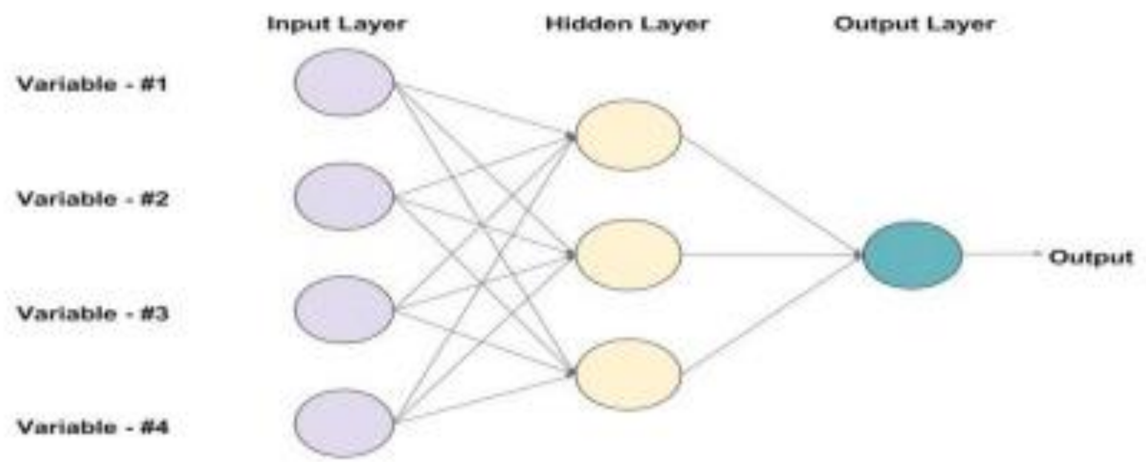


Figure 4: Multi-Layer Neural Network Structure

After clustering the entire datasets, (ANN) is used to get a model that is the best for predicting road traffic accident, as ANN always learns from past experience after it completes its first training, and so it becomes an appropriate methodology for prediction. The ANN technique shows some tolerance to a good extent over errors that may exist within the training set. It has the ability to show the veiled and dependencies that are not linear and still learn from its past experience after completing its first training, and this makes it appropriate method that is suitable for prediction (Francisca Ogwueleka, June 2014).

Based on statistical learning theory, support vector machine models are a new type of model that can be used to predict count-frequencies (Kecman, 2005). These models are a collection of related supervised learning methods for classification and regression that have the well-known ability to approximate any multivariate function to any desired degree of accuracy. The theoretical foundations for the learning algorithms of support vector machine models are statistical learning theory and structural risk minimization. It has been discovered that these models produce better or comparable results than neural networks and other statistical models (Kecman, 2005).

The mathematical thought of SVM is mapping low-dimensional nonlinear data to high-dimensional spatial data by nonlinear transformation and searching for the optimal hyperplane in high-dimensional space to make data linearly separable. The great performance of SVM is attributed to several outstanding features, including exceptional generalization function, effective avoidance of under fitting and overfitting and low requirements for input dimensions (Ye, et al., , 23–28 July 2017;). The optimal hyperplane can be described by the linear equation

$$\omega T x + b = 0, \ldots\ldots\ldots(Eq\ 16)$$

which can be produced by nonlinear mapping. Correct classification of data means the maximum classification margin is searched. For a linear-divisible sample set, the problem of searching the linear separating hyperplane is converted into the objective function and restraint conditions as shown in Equation;

$$\begin{cases} \min \ \frac{1}{2} \parallel \omega \parallel^2 \\ \hline y_i \left(\omega^T x_i + b\right) \geq 1 \end{cases} \quad \ldots\ldots\ldots\ldots(Eq\ 17)$$

Where

ω- weighted vector, It is orthogonal to the hyperplane; x_i—sample set. The premise of the dataset is that it can be linearly separable; yi—unique classification of x_i. It is classified into positive and negative categories (+1 and −1); b - offset vector It determines the deviation of the decision surface from the origin (Weifan Zhong, 2023)

Support vector machine models have recently been introduced for other transportation applications, including crash prediction (G. N. XieQ., 2007); (Li-YenChang, 2005). They are, however, difficult to estimate, and, like neural and Bayesian neural network models, they frequently cannot be generalized to other data sets. Another general criticism leveled at neural, Bayesian neural network, and support vector machine models is that they all tend to behave as black boxes, lacking the interpretable parameters found in traditional crash-frequency models.

Numerous research on various countries have looked into the empirical correlations between highway geometric design features and car accidents using various statistical methodologies and models. While the frequency and severity of traffic accidents are dramatically rising in Ethiopia, little effort has been made to pinpoint the primary cause of the highway traffic accident that is related to the roadway and its environment. Consequently, some further effort will be made in this study to address the safety concerns, such as;

1. on a section of highway with specified attributes, what vehicle accident involvement rate and accident probability can be reasonably expected?
2. on a given set of highway geometric design parameters, which parameters are relatively more critical to the safety performance of the road?
3. What percentage of reduction in vehicle accidents can be expected from various improvements in highway geometric design for the road selected for study?

To establish the correlations for various road classes, vehicle configurations, and accident severity categories, numerous models have been developed. However, the majority of these models were developed using traditional linear regression and have been found to have a number of statistical features that make them unsuitable for modeling car accident occurrences that occur on the road (Jovanis and Chang, 1986; Saccomanno and Buyco 1988). Recent studies have investigated the Poisson and negative binomial regression models as a result of the linear regression models' unsatisfactory characteristics (Joshua and Garber, 1990; Miaou et al. 1991)

The developed models were different in the assumption of the distributional occurrence of accidents, choice of geometric design parameters (covariates and explanatory variables), the choice of the functional form that links the accident rates of a road section with its associated attributes, the way that road cross sections are delineated, consideration of data uncertainties and other assumptions.

2.5. Goodness of Fit Evaluation Parameters

For linear regression models, the R^2 statistic, the proportion of the total variation in the dependent variable explained by the model, is the goodness-of- fit (GOF) measure typically reported. For GLMs, a direct analogous R^2 statistic is not available. However, there are other fit statistics that have been suggested to assess how well a GLM fits the data.

The Scaled Deviance and the Pearson chi-square are two traditional GOF statistics calculated in GLMs (G.R.Wood, 2002). In addition, many pseudo R^2 statistics have been used by statisticians and other safety researchers. Their approach involves the comparison of the empirical distribution of the observed counts to the negative binomial distribution with the mean estimated from the data. The probabilities from the two distributions are plotted. The extent of the overlap between the predicted and observed probabilities provides insight into the GOF of the model. The goal is to have nearly complete overlap between the predicted and the observed probabilities.

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are examples of measures that provide an assessment of the relative quality of specific models, for a given dataset. AIC and BIC penalize models based on the number of parameters (coefficient estimates). In other words, they deal with the trade-off between the GOF of the model and the complexity of the model. Therefore, AIC and BIC are statistics typically used as model selection criteria rather than for GOF assessment.

2.6. Summary of literatures

According to studies, there are many factors that contribute to traffic accidents and injuries, and these factors can be divided into three categories: factors related to the roadway and its environment, factors related to human, and factors related to vehicles. Numerous studies have been conducted on the impact of environmental and roadway elements on the causes of traffic accidents as well as the correlation between accident frequency and road geometric design parameters. In these studies, the relationship between accident rates and features of road geometry is examined. These elements include horizontal and vertical alignments, horizontal curve radius, super elevation, road gradient, lane number, sight distance, median width and type, and shoulder width. Numerous researchers discovered that certain geometric design features, such as a tiny curve radius and a steep gradient, can significantly raise the frequency and severity of traffic accidents. The majority of them concurred that crash rates start to become substantial at radius below 200 meters, and that the rate of accidents on roads with steep grade sections is higher than that of highways with flat sections.

The influence of the human factor on the factors that lead to traffic accidents is evident in the driver's behavior, which is greatly influenced by his personality, abilities, and experience. Younger drivers drive faster than older drivers, according to research on the impact of drivers' speeds at a tangent section. Male drivers also drive faster than female drivers, though the difference is not statistically significant. According to research on how drivers behave when driving, their operating speed is higher than the posted speed limit. This suggests that the driver, particularly the driver of a passenger car, was being careless or acting inappropriately. According to several studies carried out in many countries, the degree of adherence to the imposed speed limit varies with driver behaviors and enforcement methods.

On the relationship between traffic volume and accident frequency, numerous studies have been conducted. The majority of them concurred with the findings that accident frequencies rise as traffic volume rises.

The number of people injured in traffic accidents in Ethiopia is currently at an all-time high due to the country's dramatically rising vehicle traffic and unaltered driving habits. Although the frequency of traffic accidents has not decreased as much as Numerous research have been conducted in this field, examining the various features of traffic accidents while forecasting the crucial variables affecting their frequency and severity. The effects of road geometric design features on major roadways (trunk highways) with various demographic characteristics, roadway and environmental conditions have not, however, received much attention.

To date, several studies have been conducted to identify the root causes and aggravating circumstances of the chosen road's traffic accident. Accordingly, despite the research's own limitations because it is focused on locating the black spot locations and the spatial and temporal distribution of accidents along the road, Abdulmalik Adem's study has concisely explained the pattern and location of the accidents for sections identified as black spots only. Consequently, in order to fill the gap, this research makes a small attempt to assess the effects of the key geometric design components using the aforementioned research as a reference.

By linking traffic accident data with the precise location where the accident occurred and the geometric design parameter data from the as built documents of the road allows for the assessment of the impact of average daily traffic (ADT), number and widths of lanes, type and widths of medians, type and widths of inner and outer shoulders, and horizontal and vertical road geometric design parameters on the cause of traffic accidents and severity of crash.

3. RESEARCH METHODOLOGY

3.1. Introduction

This chapter contains information about the study area, the different categories of data used for the study, sampling procedures, the methods employed for gathering the different types of data, and the instruments used. The chapter also analyzes and explains the methods used to evaluate the data in order to find the answers to the research questions and achieve the study's objectives.

3.2. Study Design

The steps followed to come across the objectives of the study are identified and defining the problem from review of relevant literatures and formulation of appropriate methodology for data collection and analysis stages. Relevant information and key concepts were gathered from literatures which were used previously to model the correlation between road traffic accident and geometric design elements. Next to review, developing a research question and establishing an objective was followed by developing an applicable methodology to provide answers for the research questions. From the results found in the analysis stage, conclusions and recommendations will be drawn. (See Fig-5 below)

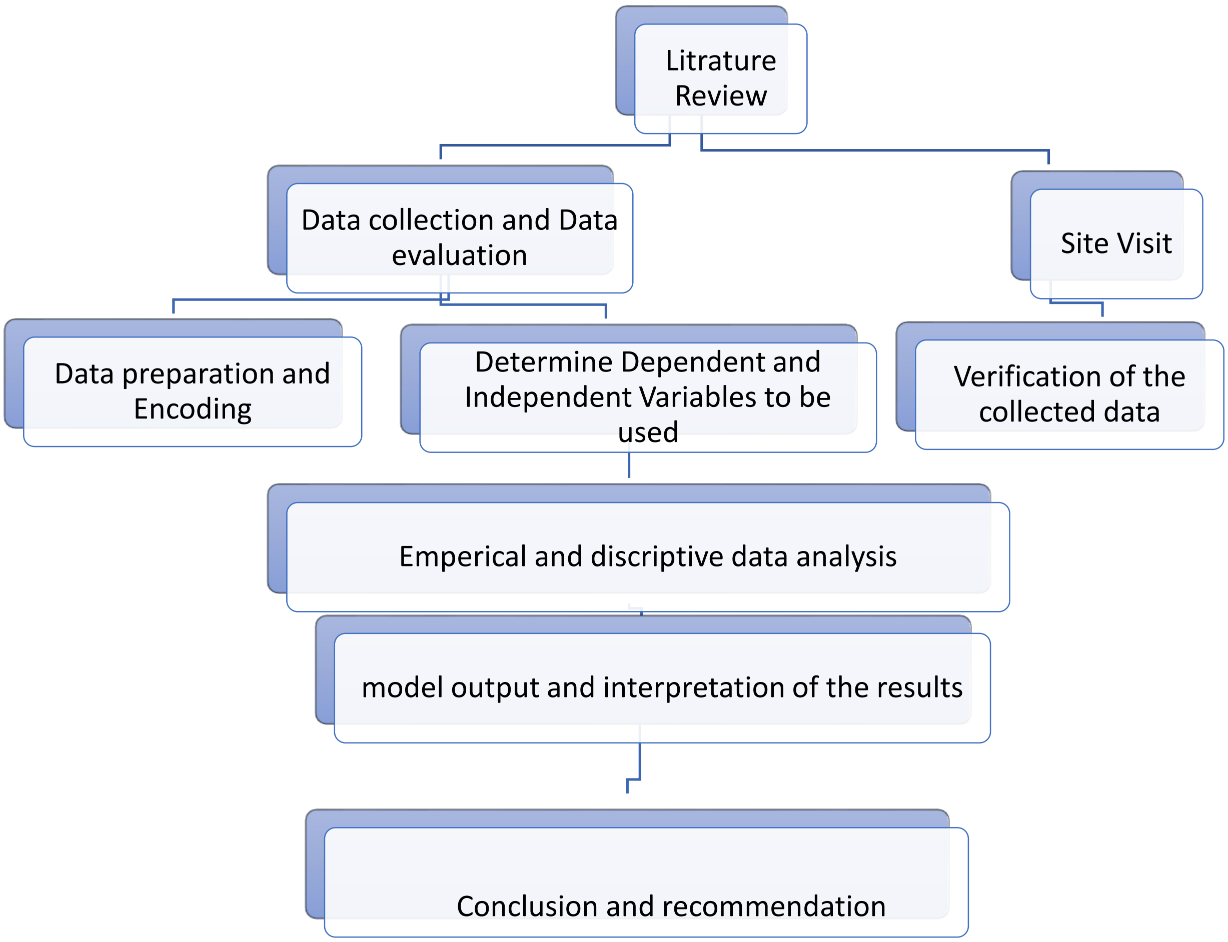


Figure 5:Organizational chart for Methodology

3.3. Study Area

The study site, Addis Ababa-Debrebrehan trunk road, having an estimated total length of 108.4 is located 19.7 km northeast of Addis Abeba. that connects the capital city Addis Ababa with Amhara and Oromia Regional states. The project's beginning and end are located at kilometer 17.9 (8.9738278N, 38.8368114E) and kilometer 108.4 (9.538779N, 39.4585387E) respectively.

The road traverses through Legetafo-Legedadi, Sendafa-Beke, Aleltu-Fiche Gelila, Hamus gebeya, Sheno, Sembo, cheki, chacha, and Tebase Towns. In addition, the dominant terrain classification of the road is rolling and the remaining section is flat terrain.

A total of 29.6 kilometers of the road pass through small to large towns and villages, with town lengths ranging from 0.8 to 3.5 kilometers. These towns have four lanes, an undivided, painted, raised median, and a total typical cross-sectional width that ranges from 25 to 42 meters. The remaining 78.4 km of the route is a two-lane rural road with typical cross sectional width ranges between 9.2 and 10.3 meters wide and paved shoulders between 1.1 and 1.5 meters wide.

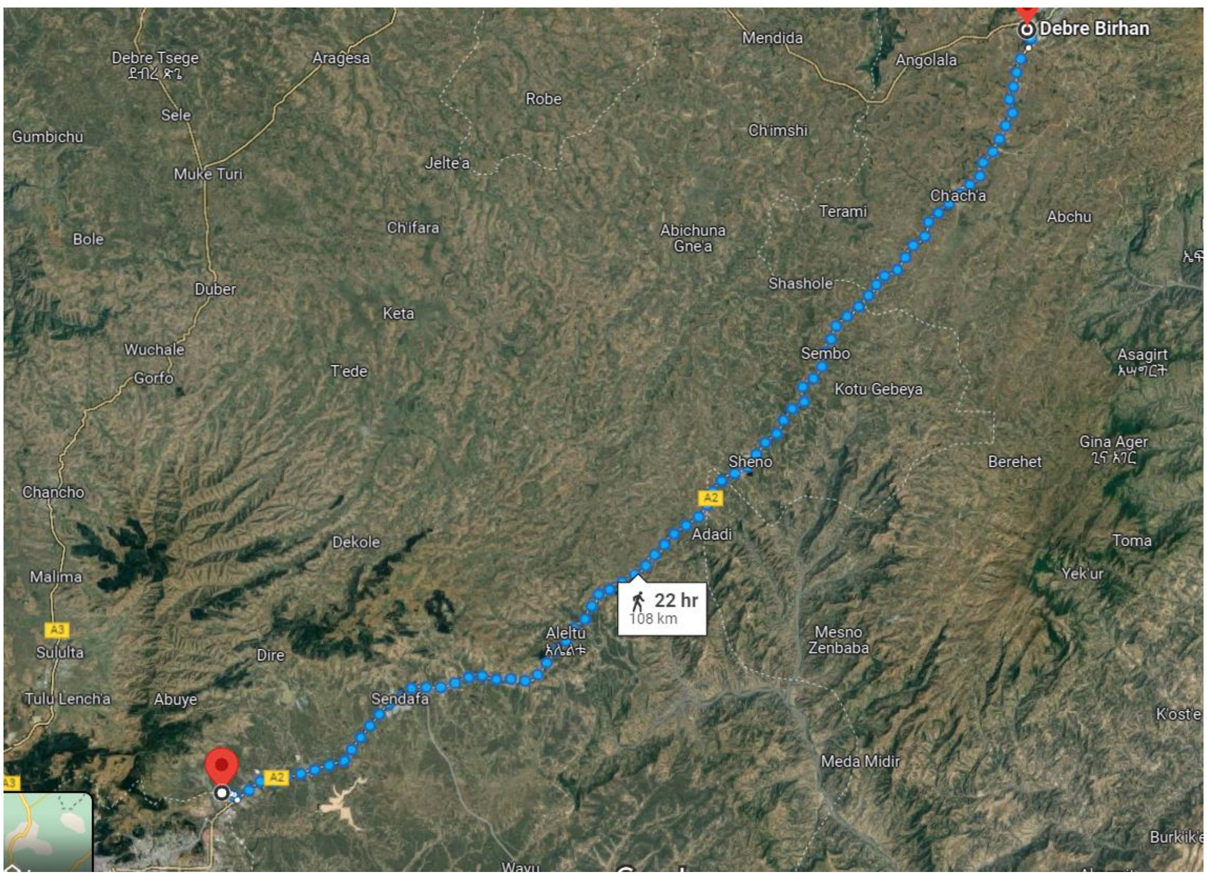


Figure 6: Study Area

3.4. Data types, sources, and collection procedures

The types of data collected are secondary data comprising of quantitative and qualitative types. The following data were collected and used as input and supporting features to explore the causes of traffic crash

3.4.1. Road geometric design data

Roadway data from as built drawings has been collected form Ethiopian Roads Administration those shows:

- Plan and profile of the roadway;
- Horizontal and vertical alignments data of the road;
- Curve radius;
- number and width of lanes;
- type and width of shoulders;
- type and width of medians;
- super elevation;
- Sight distance, to divide the highway as homogeneous sections with known length and homogeneous geometric design elements.

3.4.2. Traffic volume data

The AADT volume data for the past six years has been collected and also ERA Cyclic count for each of the ERA traffic counting season of each year has been collected.

3.4.3. Traffic accident data

From each of the town's police stations that are directly traversed by the road, a traffic accident data has been collected for fetching out the necessary data that allied with our study objective such as, the frequency of the accident and its pattern, location, accident type, the severity of individual accidents and the number of vehicles involved in the crash.

3.4.4. Sample size

The research employed the total traffic crash frequency and severity data that were gathered from each town's police stations throughout the six-year period beginning in September 2016 and running through September 2022. As a result, sampling for the traffic accident data was not carried out for the traffic accident and severity data. There were 1192 traffic incidents in total, 749 of which resulted in injuries, and the rest accidents caused property damage.

3.5. Data analysis

Once all the required data has been acquired and prepared to comply with the research objective and the software's requirements (in our case, SPSS, STATA, and EXCEL were employed), go on to

Step 1: Define the limit of the roadway and its type

At this stage of data analysis, the roadway type, length to be studied and the geometric design elements are fixed and limited as per the scope and objective of the study.

Step 2: Review the geometric design features and site characteristics

for all sections in the study area for which observed accident history data are available in order to determine the relevant data needs and avoid unnecessary data preparation. it is also necessary to understand the base conditions of the roadway and to classify the roadway and accident data in a manner that eases the data interpretation. Accordingly, the collected accident and roadway data for major junctions found along the roadway (100m before and after the location of the junctions) has been discarded since the model is implemented only on roadway segments.

The following geometric design and traffic control features are subjected to review that helps to determine the base conditions of the roadway:

- Average daily traffic volume (veh/day).
- Lane width.
- Shoulder width.
- Shoulder type.
- curve radius
- Vertical gradient
- Structure locations
- Beginning or end of a horizontal curve.
- Point of vertical intersection (PVI) for a crest vertical curve, a sag vertical curve, or an angle point at which two different roadway grades meet.

Step 3: divide Using the information from the above steps, the roadway is divided into individual sites, consisting of individual homogenous roadway segments

Step 4: evaluate the geometric design and traffic accident frequencies for each individual homogenous roadway segments

For each roadway segment, the following geometric and traffic control features on the existing road must be

- Length of segment (mi).
- ADT (veh/day).
- Lane width (ft).

- Shoulder width (ft).
- Shoulder type (paved/gravel).
- Presence or absence of horizontal curve (curve/tangent).
- Length of horizontal curve (m), if the segment is located on a curve. [This represents the total length of the horizontal curve, even if the curve extends beyond the limits of the roadway segment being analyzed].
- Radius of horizontal curve (m), if the segment is located on a curve.
- Presence or absence of spiral transition curve, if the segment is located on a curve.
- Superelevation of horizontal curve if the segment is located on a horizontal curve.
- Grade (percent), considering each grade as a straight grade from PVI to PVI (i.e., ignoring the presence of vertical curves).

Step 5: Assign observed crashes to the individual sites

Crashes that occur on the road segments and are not related to the presence of an intersection are assigned to the roadway segment on which they occur; such crashes are used in the analysis of the total crash frequency

Step 6: Model formulation for the selected site

a generalized linear model of Negative Binomial Model has been used for establishing the relationship between road traffic accidents frequency and roadway design explanatory variables and Multinomial, Binary and Ordinal Logistic Regressions for analyzing accident severity on the observed crash.

Step 7: Implementation of Goodness of Fit Evaluation Parameter

In order to decide which, set of independent variables should be included in accident predictive models, correlations between explanatory variables were studied to avoid multicollinearity problem, and the Akaike's information criterion (AIC) will be used. Starting with the full set of independent variables and following a stepwise procedure, the insignificant variables will be cancelled and models with smaller AIC values will be selected. Inclusion of individual variable in the relationships was made after checking that its estimated coefficient carries the expected sign and by examining whether the coefficient is significantly different from zero. The goodness of fit of the proposed models will then be assessed using the Pearson's $\chi 2$-statistic at the 0.05 significant levels.

Step 8: Discussion of the model results

Step 9: Conclusion

4. RESULTS AND DISCUSSION

4.1. Introduction

This section discusses with calculation of descriptive and frequency information displays for the total crash frequency and severity data. These techniques summarize various aspects about the data, giving details about the sample and providing information about the population from which the sample was drawn. total frequencies (raw counts), relative frequencies (proportions or percentages of the total number of observations), and cumulative frequencies for the chosen variables were the basic descriptive statistics methods utilized with discrete variables.

4.2. Analysis of factors contributing to traffic accident occurrence and severity

4.2.1. General Characteristics

4.2.1.1. Average daily traffic of the Project

Addis Ababa-Debrebrhan Trunk Road was built and made operational on 2005E.C./2013G.C with a DS3 road standard as per ERA Geometric Design Manual-2002, to accommodate a traffic volume ranges between 1000 and 5000 AADT in accordance with forecasted traffic during design by taking a project's design life of 20 years into account.

The project has an average daily traffic count volume ranged from 1826 to 3590 for each of the year's counting cycles, according to the ERA annual traffic count report. Accordingly, in order to get the variation of traffic volume, ERA cyclic ADT counts has been used during the analysis for each year of the analysis period considered.

Table 1:Average Daily Traffic for the study period

YEAR	**AADT**		
	Cycle 1*	**Cycle 2****	**Cycle 3****
2016/17	1826	2472	2683
2017/18	2753	2378	2004
2018/19	1870	1978	2552
2019/20	3590	3008	3338
2020/21	2685	2532	2696
2021/22	3311	2791	2936

*cycle 1 for the months of January up to April

**cycle 2 for the months of May up to August

***cycle 3 for the months of September up to December

4.2.1.2. Crash Data of the Project

The police stations located within the municipality that the project directly traverses have been used to gather the project's crash data. Legetafo Town, Sendafa Town, Aleltu Town, Sheno Town, Chacha Town, and Tebase Town are the major towns for which data has been acquired. The police stations in the aforementioned town were used to gather traffic accident data for additional small towns and villages that are directly accessed by the project, including 44, Legedadi, Lege Bery, Beke, Feche Gelila, Hamus Gebeya, Sembo, and Cheki. Nearly all of the accident reporting formats used by all municipalities include adequate descriptions of the accident or event; however, the majority of the reports lack fullness, either by failing to correctly narrate the incident and being ambiguous in their descriptions or by failing to fully document the event's key information. However, they were sufficient to satisfy our demands and achieve the objective of the research. A sample accident reporting format and summary of the collected crash data is attached in the appendix section.

4.2.1.3. Description of the Characteristics of the Roadway

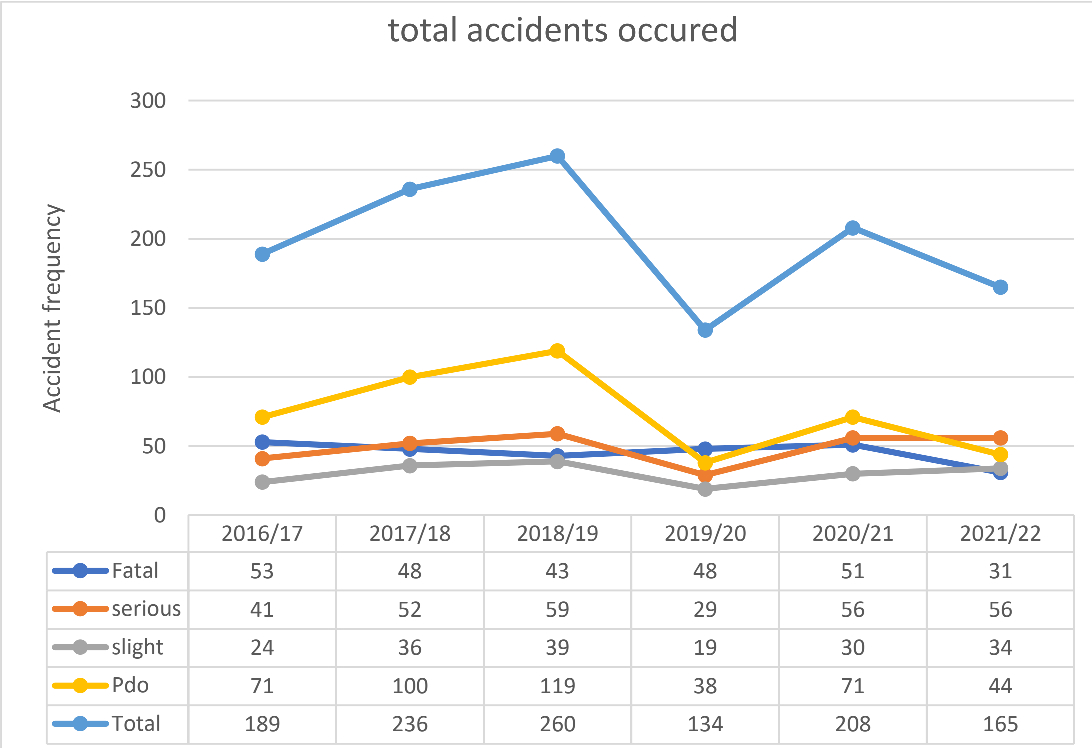
Functional classification (the grouping of highways by the character of service they provide) of roads was developed mainly for transportation planning purposes. ERA's Design Standard classifies roads in five (5) categories based on the character of service they provide. These are Trunk, Link, Main Access, Collector and Feeder roads. Accordingly, Addis Ababa – Debrebrhan Road is classified as Class I (trunk) road classification that links Addis Ababa to centres of international importance and to international boundaries.

The project's total length is 108.4 km (when Addis Abeba and Debrebrhan Sections are excluded), of which 78.4 km traverse through rural areas with two lane undivided carriageways and paved shoulders, and the remaining section traverses through urban settlements with four lane carriageways, including the parking lanes, with divided and undivided road sections. The project consists of a total of 109 vertical curves with a minimum gradient of 0.17% and a maximum gradient of 7.5%, as well as 144 horizontal curves with a radius range of 98 to 2000 metres. The summary of cross section and other necessary roadway elements data for the project are annexed in the appendix part of the report.

4.2.2. Descriptive statistical analysis of the traffic accident occurrence and severity

4.2.2.1. Total Crash Occurred on the Project

The vehicular crash data count on the project has shown that a total of 1192 causalities has been occurred during the study period of which 749 of the accidents are injury accidents and the remaining are property damage accidents. Annual distribution of the accidents is 189, 236, 260, 134, 208 and 165 for the years 2016/17 up to 2021/22 respectively. fig 5 shows the total traffic crash, fatality, serious injury, slight injury and property damage only (PDo) data for the last six years in Addis Ababa – Debrebrehan Road.



	2016/17	2017/18	2018/19	2019/20	2020/21	2021/22
Fatal	53	48	43	48	51	31
serious	41	52	59	29	56	56
slight	24	36	39	19	30	34
Pdo	71	100	119	38	71	44
Total	189	236	260	134	208	165

Figure 7: Total traffic accident occurred for the study period

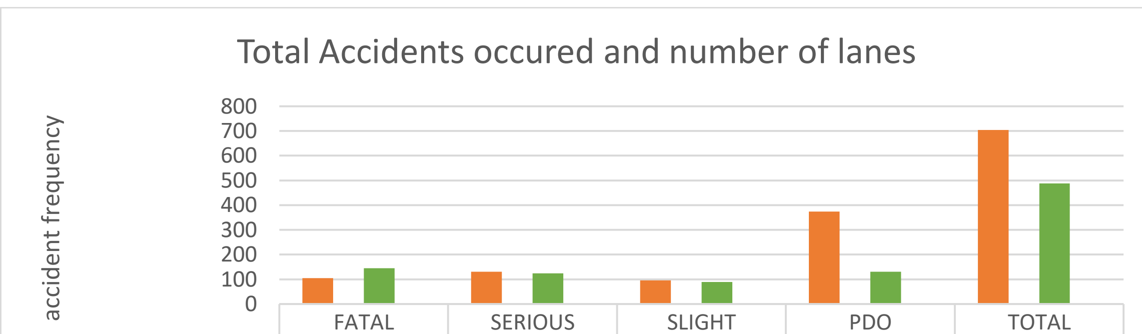
4.2.2.2. Accident Occurrence Location

Traffic crash frequency and severity for the whole section of the road (i.e. on conventional two lane rural sections and divided and undivided urban multilane sections) was assessed by the crash data collected from each counties police station as well as from the road geometry data and inventory survey data which is collected from ERA.

The project traverses through around thirteen (13) small to large town sections (i.e. Legetafo, Legedadi, 44, Legeberi, Sendafa, Beke, Aleltu, Fiche Gelila, Hamus Gebeya, Sheno town, Sembo, Cheki, Chacha and Tebase) which accounts for around 29km of the project passes through these towns. Almost all of the towns have a four-lane carriageway having a width range from 14.0 up to 14.m. besides, only two of the towns (i.e. Legetafo and Tebase) have raised and curbed median barriers with a width of 8.8m and 1.0m for respective town sections and the remaining towns carriageways are segregated by painted barriers even there are town sections having four lanes but does not have any median barrier that segregates the opposite direction traffic.

The remaining sections of the road passes through rural two lane roads having a carriageway width range from 7.0m up to 7.3m and with a paved shoulder (Surface Treatment) at both sides of the road with a width range of 1.0m up to 1.5m.

By taking the above-mentioned roadway inventory results into consideration, the traffic police accident report has been analyzed thoroughly. Accordingly, the results have showed that, out of 1192 total traffic accidents 704 of them has been occurred on four lane roads and the remaining 488 crashes were occurred on the conventional two lane rural roads. (See Fig:6 below)



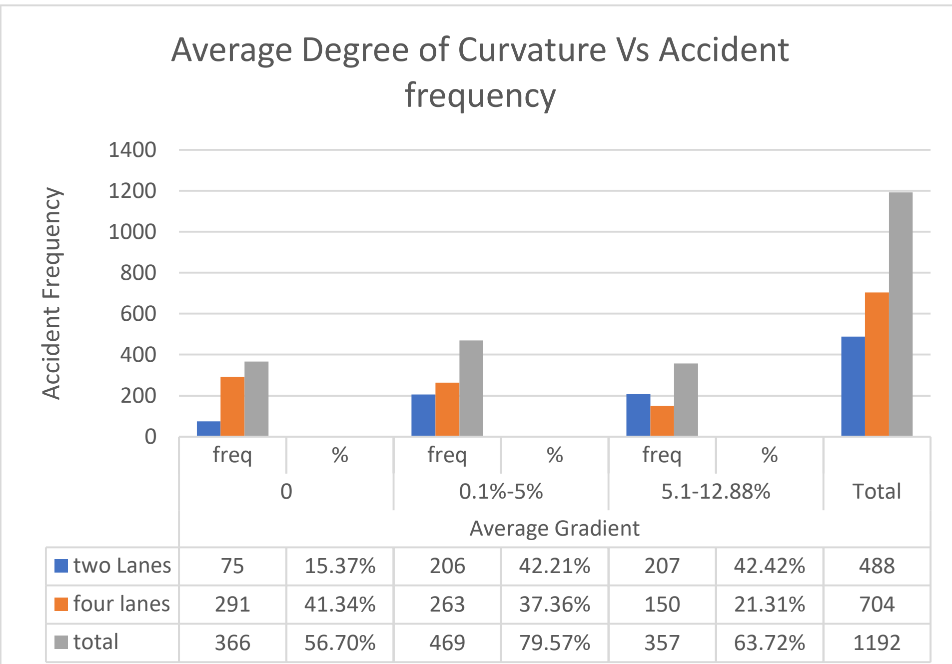
	FATAL	SERIOUS	SLIGHT	PDO	TOTAL
four lane roads freq	105	130	95	374	704
four lane roads %	14.91%	18.47%	13.49%	53.13%	100.00%
two lane roads freq	145	124	89	130	488
two lane roads %	29.71%	25.41%	18.24%	26.64%	100.00%

Figure 8: distribution of total accidents by number of lanes

As it has been shown from the result, 46.87% of the accident occurred on the four lane roads are injury accident and the remaining 53.13% percent of the accident is Property Damage only accidents. However, for two lane rural roads, 73.36% of the accident is injury accident and the remaining 26.64% is Property Damage only accident.

4.2.2.3. Degree of curvature

The impact of the availability of horizontal alignment to the crash occurrence has been analyzed by using average degree of curvature values since the accidents has been analyzed through segmented homogenous sections rather than specific accident spots due to the uncertainty of exactly locating the accident spots based on the traffic police report.



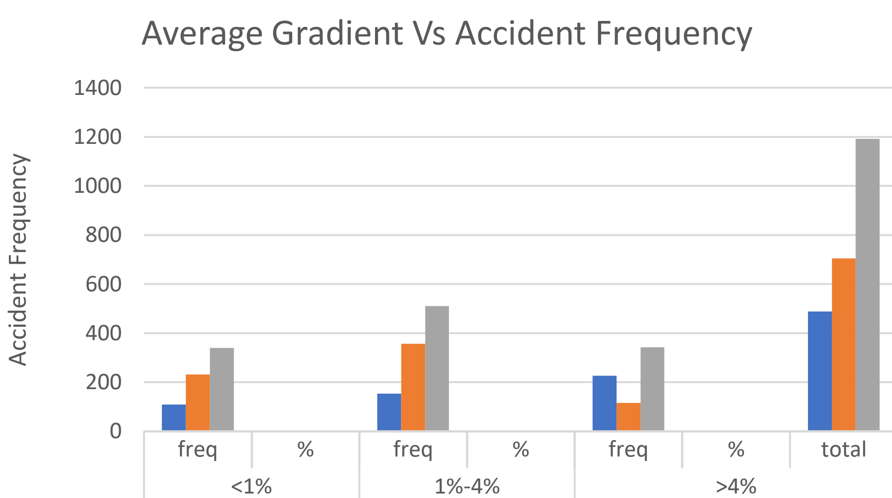
	0 freq	0 %	0.1%-5% freq	0.1%-5% %	5.1-12.88% freq	5.1-12.88% %	Total
two Lanes	75	15.37%	206	42.21%	207	42.42%	488
four lanes	291	41.34%	263	37.36%	150	21.31%	704
total	366	56.70%	469	79.57%	357	63.72%	1192

Figure 9:distribution of total accidents by average degree of curvature

As can be seen from the table, for two lane roads 85% of the accidents has been occurred on sections having horizontal curves and nearly similar number of accidents has been occurred on sections having degree of curvature between 0.1% and 5% and Sections having degree of curvature greater than 5%. However, on four lane road sections, higher number of accidents has been occurred on sections without horizontal curve which is 41.34% of the accidents. Next to those sections, sections with 0.1-5% average degree of curvature has higher number of accidents which is 37.36%.

4.2.2.4. Average gradient

The impact of the availability of vertical gradient to the crash occurrence has been analyzed by using average gradient per segment values since the accidents has been analyzed through segmented homogenous sections rather than specific accident spots due to the uncertainty of exactly locating the accident spots based on the traffic police report.



	<1% freq	<1% %	1%-4% freq	1%-4% %	>4% freq	>4% %	total
two lanes	109	22.34%	153	31.35%	226	46.31%	488
four lanes	231	32.81%	357	50.71%	116	16.48%	704
total	340	55%	510	42.79%	342	28.69%	1192

Figure 10:distribution of total accidents by average gradient

As can be seen from the graph, for two lane roads, higher number of accidents has been registered on sections with average gradient greater than 4%. Whereas, for four lane road sections higher number of accidents has been registered on sections with 1% - 4% average gradient.

4.2.2.5. Driver's characteristics

Since the behavioral factors of the general road users and driver take the lion's share on traffic accidents the same cannot be neglected on this research. Even though it is not the objective of this research to analyze the behavioral impacts of the accidents, it will provide us an insight to the type and behaviors of Drivers, Pedestrian and Passengers involved in the incident.

Table 2:distribution of accident frequency by Driver's Characteristics

Driver's Characterstics					**Casuality class**					
	Fatal		**Serious injury**		**slight injury**		**PDO**		**total**	
	freq	%	freq	%	freq	%	freq	%	freq	%
Age Group (yrs)										
<18	2	0.77%	0	0.00%	1	0.67%	1	0.20%	4	0.34%
18-30	155	59.62%	167	60.73%	97	64.67%	239	47.14%	658	55.20%
31-50	71	27.31%	81	29.45%	36	24.00%	214	42.21%	402	33.72%
>50	22	8.46%	21	7.64%	8	5.33%	37	7.30%	88	7.38%
not coded	11	4.23%	10	3.64%	10	6.67%	17	3.35%	48	4.03%
total	260	100.00%	275	100.00%	150	100.00%	507	100.00%	1192	100.00%
Gender										
Male	259	99.62%	275	100.00%	146	97.33%	505	99.61%	1185	99.41%
fimale	0	0.00%	0	0.00%	1	0.67%	0	0.00%	1	0.08%
not coded	2	0.77%	4	1.45%	5	3.33%	3	0.59%	14	1.17%
total	260	100.00%	275	100.00%	150	100%	507	100.00%	1192	100.00%
Educational BackGd										
Illitrate	15	5.77%	8	2.91%	9	6.00%	11	2.17%	43	3.61%
1_8	73	28.08%	73	26.55%	47	31.33%	141	27.81%	334	28.02%
9_10	69	26.54%	81	29.45%	44	29.33%	157	30.97%	351	29.45%
11_12	23	8.85%	25	9.09%	6	4.00%	79	15.58%	133	11.16%
>12	29	11.15%	30	10.91%	9	6.00%	74	14.60%	142	11.91%
not coded	52	20.00%	62	22.55%	37	24.67%	46	9.07%	197	16.53%
total	260	100.00%	275	100.00%	150	100.00%	507	100.00%	1192	100.00%
Vehicle Driver R/n S/p										
Owner	73	28.08%	64	23.27%	50	32.89%	153	30.18%	340	28.48%
employee	144	55.38%	189	68.73%	87	57.24%	301	59.37%	721	60.39%
Other	2	0.77%	3	1.09%	0	0.00%	13	2.56%	18	1.51%
not coded	42	16.15%	23	8.36%	15	9.87%	41	8.09%	121	10.13%
total	260	100.00%	275	100.00%	152	100.00%	507	100.00%	1194	100.00%
Driving Experience										
<1	45	17.31%	34	12.36%	21	14.00%	52	10.26%	152	12.75%
1_2	44	16.92%	56	20.36%	36	24.00%	83	16.37%	219	18.37%
2_5	76	29.23%	78	28.36%	35	23.33%	98	19.33%	287	24.08%
5_10	27	10.38%	34	12.36%	12	8.00%	53	10.45%	126	10.57%
>10	46	17.69%	42	15.27%	25	16.67%	201	39.64%	314	26.34%
not coded	23	8.85%	35	12.73%	23	15.33%	21	4.14%	102	8.56%
total	260	100.00%	275	100.00%	150	100.00%	507	100.00%	1192	100.00%

4.2.2.6. Pedestrian Characteristics

Pedestrians' behavior and observance to road traffic rules is vital in the hindrance and control of road traffic. There was a total of 402 pedestrians participating in the crash out of which 56 were females. Out of the total crash, 115 resulted in fatal crashes, 116 in serious injury crashes and 74 in slight injury crashes (See Table-3 below).

Table 3:distribution of accident frequency by Pedestrian's Characteristics

Ped's	**Casuality class**							
	Fatal		**Serious injury**		**slight injury**		**total**	
	freq	%	freq	%	freq	%	freq	%
Age Group (yrs)								
<18	15	9.38%	12	7.14%	10	13.51%	37	9.20%
18-30	23	14.38%	37	22.02%	23	31.08%	83	20.65%
31-50	50	31.25%	24	14.29%	16	21.62%	90	22.39%
>50	35	21.88%	17	10.12%	3	4.05%	55	13.68%
not coded	37	23.13%	78	46.43%	22	29.73%	137	34.08%
total	160	100.00%	168	100.00%	74	100.00%	402	100.00%
Gender								
Male	63	39.38%	60	35.71%	56	75.68%	179	44.53%
fimale	30	18.75%	35	20.83%	13	17.57%	78	19.40%
not coded	67	41.88%	73	43.45%	5	6.76%	145	36.07%
total	160	100.00%	168	100.00%	74	100%	402	100.00%
Occupation								
Driver	0	0.00%	0	0.00%	6	8.11%	6	1.49%
Farmer	3	1.88%	0	0.00%	0	0.00%	3	0.75%
Housewif	15	9.38%	0	0.00%	0	0.00%	15	3.73%
Private w	1	0.63%	0	0.00%	1	1.35%	2	0.50%
not coded	141	88.13%	168	100.00%	67	90.54%	376	93.53%
total	160	100.00%	168	100.00%	74	100.00%	402	100.00%
MOVEMENT OF PEDESTERIAN								
Crossing	4	2.50%	0	0.00%	6	8.11%	10	2.49%
Crossing	2	1.25%	0	0.00%	0	0.00%	2	0.50%
Moving o	0	0.00%	0	0.00%	3	4.05%	3	0.75%
Moving o	1	0.63%	0	0.00%	0	0.00%	1	0.25%
Moving o	0	0.00%	3	1.79%	0	0.00%	3	0.75%
not coded	153	95.63%	165	98.21%	65	87.84%	383	95.27%
total	160	100.00%	168	100.00%	74	100.00%	402	100.00%

4.2.2.7. Passenger's characteristics

Even though, most of the crash involves passengers in addition to the driver's and pedestrians' involvement, the traffic police report does not properly depict their characteristics for most of the crash occurs on the road under study.

4.2.2.8. Vehicular characteristics

There is evidence in linking certain vehicle characteristics to crash involvement that influence drivers' risk-taking behavior (See Table-4 below)

Table 4:distribution of accident frequency by Driver's Characteristics

Vehicular Characterstics	Casuality class									
	Fatal		Serious injury		slight injury		PDO		total	
	freq	%	freq	%	freq	%	freq	%	freq	%
Vehicle ownership										
Company	24	9.23%	26	9.45%	10	6.67%	30	5.92%	90	7.55%
Government	8	3.08%	12	4.36%	6	4.00%	22	4.34%	48	4.03%
Private	195	75.00%	194	70.55%	105	70.00%	420	82.84%	914	76.68%
not coded	33	12.69%	43	15.64%	29	19.33%	35	6.90%	140	11.74%
total	260	100.00%	275	100.00%	150	100.00%	507	100.00%	1192	100.00%
Vehicle year of service										
<1	1	0.38%	0	0.00%	3	2.00%	2	0.39%	6	0.50%
1_2	22	8.46%	9	3.27%	9	6.00%	32	6.31%	72	6.04%
2_5	10	3.85%	9	3.27%	1	0.67%	15	2.96%	35	2.94%
5_10	2	0.77%	6	2.18%	2	1.33%	5	0.99%	15	1.26%
>10	112	43.08%	142	51.64%	57	38.00%	352	69.43%	663	55.62%
not coded	113	43.46%	109	39.64%	78	52.00%	101	19.92%	401	33.64%
total	260	100.00%	275	100.00%	150	100%	507	100.00%	1192	100.00%
Vehicular defect										
Brake fault	0	0.00%	1	0.36%	0	0.00%	0	0.00%	1	0.08%
Light problem	1	0.38%	0	0.00%	1	0.67%	5	0.99%	7	0.59%
mechanical problem	0	0.00%	0	0.00%	0	0.00%	1	0.20%	1	0.08%
tyre problem	0	0.00%	1	0.36%	0	0.00%	0	0.00%	1	0.08%
No problem	254	97.69%	267	97.09%	148	98.67%	488	96.25%	1157	97.06%
not coded	5	1.92%	6	2.18%	1	0.67%	13	2.56%	25	2.10%
total	260	100.00%	275	100.00%	150	100.00%	507	100.00%	1192	100.00%
Crash causing vehicle										
Light Vehicles (Automob	129	49.62%	151	54.91%	95	63.33%	236	46.55%	611	51.26%
Medium Vehicle (Isuzu, I	80	30.77%	81	29.45%	35	23.33%	133	26.23%	329	27.60%
Large Vehicle (Trucks, L	46	17.69%	35	12.73%	19	12.67%	132	26.04%	232	19.46%
not coded	5	1.92%	8	2.91%	1	0.67%	6	1.18%	20	1.68%
total	260	100.00%	275	100.00%	150	100.00%	507	100.00%	1192	100.00%
Number of vehicles involved										
Single Vehicle	77	29.62%	95	34.55%	65	43.33%	250	49.31%	487	40.86%
Two vehicles	178	68.46%	177	64.36%	85	56.67%	240	47.34%	680	57.05%
multiple vehicles	5	1.92%	3	1.09%	0	0.00%	17	3.35%	25	2.10%
not coded	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
total	260	100.00%	275	100.00%	150	100.00%	507	100.00%	1192	100.00%

4.2.2.9. Crashes by Temporal Variation (Hour of the Day)

As it has been on the police reported data, a road traffic accident varies upon the nature of the hour of the day, the day of the week and the month of the year mostly manifested when the pedestrian and passenger movements stretched maximum levels. In this regard, the study revealed that highest crashes were reported during daylight and between (6:00am to 5:59pm) with 910 (76.34%) crashes (See Table-5)

Table 5: distribution of accident frequency by Temporal Variation

	Causality class							
Time of the day	Fatal		Severe		Slight		PDO	
Day (6:00am to 5:59pm)	192	73.85%	212	77.09%	122	81.33%	384	75.74%
Night (6:00pm to 5:59am)	68	26.15%	60	21.82%	27	18.00%	122	24.06%
not coded	0	0.00%	3	1.09%	1	0.67%	1	0.20%
TOTAL	260	100.00%	275	100.00%	150	100%	507	100.00%

4.2.2.10. Crashes by Different Types of Causes

The majority of road crashes were caused by drivers' errors like exceeding authorized speed limit followed by improper overtaking and not giving priority to the pedestrian as it was recorded in the police report (Figure-11).

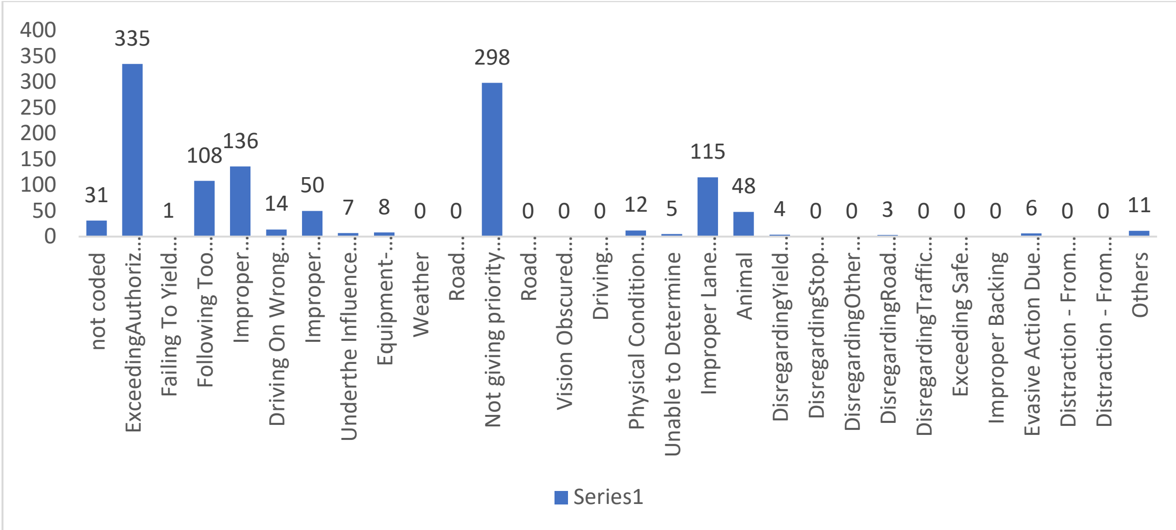


Figure 11:distribution of total accidents by types of causes

4.2.2.11. Types of Crash occurred on the road way

As it has been shown on the graph, Pedestrian accident, followed by turning of roadway then rear end collision is the major types of accidents caused on the roadway.

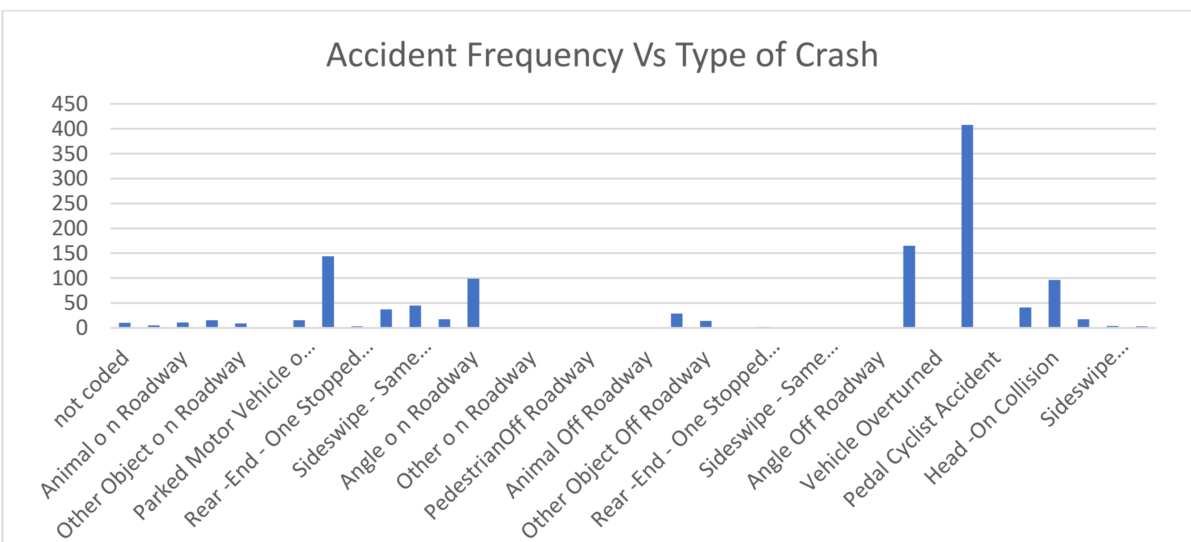


Figure 12: distribution of total accidents by types of crashes

4.2.3. Statistical analysis of accident occurrence and severity by regression model

Beside to the descriptive statistics analysis undertaken in an effort to come up with the most appropriate approach that relates accident frequencies and degree of severities with their explanatory variables, a number of empirical statistical regression analysis models has been tested in SPSS and STATA including Poisson model, Negative Binomial model, Zero-inflated Poisson and Zero-inflated Negative Binomial for frequency analysis of total number of crashes and Binary Logistic model, Multinomial Logistic model and Ordinal Logistic model for severity analysis of the dependent variable (i.e Fatal, Severe injury, Slight injury and Property Damage only crashes).

4.2.3.1. Data processing

The roadway has been segmented to homogenous sections in terms of key roadway design characteristics including number of lanes, lane width, horizontal and vertical alignments as an alternative to fixed length segment, which only splits the roadway into small segments of suitable equal length, which is less popular alternative segment division technique that has been employed in past accident modelling studies. Comparatively speaking, even though using homogenous road segments greatly increase the difficulty of data collection and data processing, it is more beneficial to accounting the effects of road related factors such as traffic volumes and geometric characteristics on accident frequency studies. (Pan Chengye, 2013).

A minimum highway section length of 500m were set to exclude short sections that might be influenced by adjacent section characteristics. A section boundary was formed when one or more of the geometric characteristics changed.

Following to setting out the homogenous sections for the roadway, observed crash frequency and severity data, traffic volume data and geometric design elements are assigned to the individual homogenous sections.

In this regard, the above referred segmentation of homogenous sections has only been used on the frequency analysis. Whereas, on the severity analysis individual accident locations were used for analysis.

4.2.3.2. Model Development

4.2.3.2.1. Poisson and Negative Binomial Regression Models for Accident Frequency Analysis

The number of accidents along a highway segment is modeled as a random variable of Poisson type. Given a mean value for the number of accidents per unit time at such a location, the Poisson distribution describes the distribution of the number of accidents, assigning probabilities to each possible number of accidents. The distribution forms a unimodal curve with maximum at or near the mean, and with variance equal to the mean. The modeling approach is to assume that this mean can be represented by a product of highway variables raised to various powers. This is equivalent to assuming that the mean is an exponential applied to a linear function of other highway variables, including a constant or intercept term. The goal is to find a suitable choice of highway variables and the coefficients of each in a model of the form

$$\hat{Y} = EXP(a0 + a1x1 + a2x2 + a3x3 \ldots anxn \ldots\ldots\ldots \text{(Equ 18)}$$

where $\hat{Y}$ is the proposed mean, x1, x2, . . . , xn are the highway variables, and a_0, a_1, a_2, . . . , a_n are the intercept and the desired coefficients.

An elaboration on this approach is the negative binomial model. This model of Poisson type allows the variance to be a quadratic function of the mean of the form

$$\hat{Y} + K(\hat{Y})^2 \ldots\ldots\ldots \text{(Equ 19)}$$

where K is a nonnegative constant called the overdispersion parameter. If K is zero, an ordinary Poisson model results. A positive value of K is a way of including additional nonrandom multiplicative variation in the model due to variables not explicitly present.

4.2.3.2.2. Ordinal logistic regression models for Accident Severity Analysis

Let Y be an ordinal outcome with J categories. Then $p(Y \leq j)$ is the cumulative probability of Y less than or equal to a specific category J=1,…,J-1. Can be defined as

$$\frac{p(Y \leq j)}{p(Y > j)}$$

For j=1,…, J-1 since $p(Y > j) = 0$ and dividing by zero is undefined. Alternatively, it can be written as $(Y > j) = 1 - P(Y \leq j)$. The log odds is also known as the logit, so that

$$log\left(\frac{P(Y \leq j)}{P(Y > j)}\right) = logit(p(Y \leq j)) \ldots\ldots\ldots \text{(Equ 20)}$$

The ordinal logistic regression model can be defined as

$$logit(p(Y \leq j)) = \beta jo + \beta 1x1 + \cdots + \beta pxp \ldots\ldots\ldots \text{(Equ 21)}$$

4.2.3.3. Preliminary examination of the data

Preliminary examination of the processed data has been undertaken for selection of regressor/independent variables. The selection of variables to retain, and the form in which to use them, are to some extent arbitrary since not all possibilities can be examined and some are more or less equivalent. The decisions are guided by criteria of simplicity (use of variables that are easily understood), comprehensiveness (inclusion of as many types of variables as possible), completeness of data (proportion of missing data) and significance (coefficients that are significantly different from zero according to statistical tests in one or more models).

Several trials were made in order to explore the properties of the basic data. then, several exploratory tabulations were prepared in order to obtain an initial appraisal of the behavior of the data on several factors related to highway geometry for setting thresholds above and below which the Independent Variables have significant increment or reduction of the accident frequency and severity. The factors considered were:

- Amount of gradient
- Degree of curvature
- Number of structures
- Availability of Median

Table 6: a statistical summary of continuous and categorical Independent Variables used for frequency model

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Segment length	1962	500	1400	1004.59	145.542	21182.415
average degree of curvature	1962	1.0	12.9	4.033	2.3857	5.692
average gradient	1962	-4.3	7.5	.842	2.7156	7.374
STRucture	1962	0	2	.29	.494	.244
average daily traffic	1962	1826	3590	2633.50	489.984	240084.395
median width	1962	0	9	.27	1.445	2.087
surface width	1962	9.2	42.0	16.034	10.5242	110.760
number of lanes	1962	2	4	2.57	.902	.814
lane width	1962	7.0	14.4	9.221	3.0800	9.487
outside shoulder width	1962	0	1	.72	.451	.204
inside shoulder width	1962	0	1	.72	.451	.204
speed limit	1962	30	50	47.52	5.275	27.823
Valid N (listwise)	1962					

outside shoulder type

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no shoulder	558	28.4	28.4	28.4
	st	1404	71.6	71.6	100.0
	Total	1962	100.0	100.0	

rural urban

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	rural	1404	71.6	71.6	71.6
	urban	558	28.4	28.4	100.0
	Total	1962	100.0	100.0	

inside shoulder type

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	558	28.4	28.4	28.4
	3	1404	71.6	71.6	100.0
	Total	1962	100.0	100.0	

roadway classes

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 lane rural roads	1404	71.6	71.6	71.6
	urban multilane divided roads	486	24.8	24.8	96.3
	urban multilane undivided roads	72	3.7	3.7	100.0
	Total	1962	100.0	100.0	

Please note that the same procedure has been followed for both accident frequency and severity models in this regard.

4.2.3.4. Selection of Regressor Variables

In order to decide which subset of independent variables should be included in a crash estimation model, the Akaike's information criterion (AIC), was used. AIC was defined as follows:

$$AIC = -2 \cdot ML + 2 \cdot K \ldots\ldots\ldots\ldots(Equ\ 22)$$

where K is the number of free parameters in the model and ML is the maximum log-likelihood. The smaller the value of AIC, the better the mode. Starting with the full set of independent variables listed in the previous section, a stepwise procedure was used to select the best model based on minimizing the AIC value.

then, correlation analysis has been run for each of the independent variables to check whether they are multicollinear with each other or not. Accordingly, lane width, surface width, median width and median type have shown Variance Inflation Factor (VIF) value greater than 10 which is an indicator of severe dependency among each other. Hence, the model removes lane width and median type by considering them as multicollinear variables then all the remaining independent variables shows a VIF value of less than 10 (please note that the multicollinearity test result is attached to the Appendix Part of the report).

After that, separate regression model has been done for each of independent variables and pair of them for evaluating the best model that shows significance and evaluate the AIC and BIC values of each independent variables. Accordingly, for frequency analysis models Outside shoulder type, inside shoulder type, inside shoulder width and rural urban classification has been removed since they do not show any additional merit to the model. Hence, the same has been shown on the full model output result that has been run by considering Segment length, Average degree of curvature, average gradient, average daily traffic, median width, surface width, number of lanes, outside shoulder width and speed limit as continuous independent variables and outside shoulder type, structures and median type as categorical variables. For severity analysis models, inside shoulder width has shown significance contrary to the frequency models in addition to the above-mentioned Independent Variables.

4.2.3.5. Model Estimation and Testing

For frequency analysis, after omitting the above referred insignificant continuous independent variables, the model has been re-run for frequency analysis of total crash using Poisson, Negative Binomial, Zero-Inflated and Truncated Poisson and Negative Binomial regression models.

After evaluating the predictive performance of all of the above-mentioned models, it is found that Negative Binomial Regression model is the most desirable approach for our case. Accordingly, even though most of the models show significance to both the data and also to the variables considered, Negative binomial model is the one having lower AIC and BIC values.

The variables selected on the basis of these factors for possible inclusion as an independent variable in the derived models were number of lanes, section length, ADT, lane width, outside shoulder width and type, inside shoulder width and type, median width and type, speed limit, surface width, roadway class and rural urban variations. Each influencing variables has two or more sub variables and consequently the effect of each of the variables and sub variables has been considered in the analysis.

Accordingly, for severity analysis of crashes; Binary Regression model, Multinomial Logistic Regression model and Ordinal Logistic Regression model (with logit and probit link functions) have been used. After evaluating the predictive performance of all of the above-mentioned models, it is found that Ordinal Logistic model with probit link function becomes the most desirable approach for our case. As a result, gradient, median width, surface width, lane width, outside shoulder width, inside shoulder width, degree of curvature, speed limit, structures and median type has been considered in the analysis.

The descriptive summary of the independent variables considered in the full model for both accident frequency and severity models are as shown below:

Categorical Variable Information

			N	Percent
Factor	STRucture	1	1422	72.5%
		2	504	25.7%
		3	36	1.8%
		Total	1962	100.0%
	number of lanes	2 lanes	1404	71.6%
		4 lanes	558	28.4%
		Total	1962	100.0%
	roadway classes	2 lane rural roads	1404	71.6%
		urban multilane divided roads	486	24.8%
		urban multilane undivided roads	72	3.7%
		Total	1962	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Acc. freq	1962	0	15	.60	1.469
Covariate	seg_lng	1962	500	1400	1004.59	145.542
	avg.deg.curv	1962	.0	12.9	3.850	2.6373
	Avg.G	1962	-4.3	7.5	.842	2.7156
	ADT	1962	1826	3590	2633.50	489.984
	medwid	1962	.0	8.8	.270	1.4447
	surf_wid	1962	9.2	42.0	16.034	10.5242
	outshwd1	1962	.0	1.5	.835	.5460
	speedlim	1962	30	50	47.52	5.275

Figure 13:. Categorical and Continuous Variables Information for crash frequency models

Case Processing Summary

		N	Marginal Percentage
severity	Fatal	269	22.9%
	slight injury	172	14.6%
	Severe injury	287	24.4%
	PDO	447	38.0%
STR.	0	779	66.3%
	1	375	31.9%
	2	21	1.8%
no_lanes	2	478	40.7%
	4	697	59.3%
med_type	no median	536	45.6%
	painted	415	35.3%
	curbed and raised	224	19.1%
rururb	1	478	40.7%
	2	697	59.3%
rodwycls	1	478	40.7%
	4	639	54.4%
	5	58	4.9%
Valid		1175	100.0%
Missing		0	
Total		1175	

Figure 14:Categorical Variables Information for crash severity models

4.2.3.6. Model Results for both Crash Frequency Analysis and Crash Severity Analysis

a) **Models Derived to Estimate Total Crash Frequency using Negative binomial regression analysis:**

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	1.305	.6953	0.443	3.636	6.040	1	.014	7.685	.000	.707
[STR.=1]	-.514	.2957	-.452	.065	3.026	1	.000	.598	.335	1.067
[STR.=2]	.070	.2946	.114	.647	.057	1	<.001	1.073	.602	1.911
[STR.=3]	0[a]	.	.	.	.	.	.	1	.	.
[no_lanes=2]	-.716	.6079	-4.246	-.822	1.385	1	.013	.489	.149	1.610
[no_lanes=4]	0[a]	.	.	.	.	.	.	1	.	.
[med_type=1]	.468	.3529	.276	.224	1.758	1	.018	.626	.314	1.251
[med_type=2]	-.098	.3508	-1.480	-.105	5.108	1	.024	.453	.228	.901
[med_type=3]	0[a]	.	.	.	.	.	.	1	.	.
seg_lng	.457	.1739	.116	.798	6.900	1	.009	1.579	1.123	2.221
avg.deg.curv	.435	.7870	.281	.589	30.600	1	<.001	1.545	1.324	1.803
Avg.G	.002	.0166	-.030	.035	.021	1	.884	1.002	.970	1.036
ADT	.654	.1779	.306	1.924	21.924	1	<.001	1.924	1.358	2.726
medwid	-.646	1.3300	-.906	-.385	23.584	1	<.001	.524	.404	1.318
surf_wid	.056	.0144	.027	.084	14.868	1	<.001	1.057	1.028	1.088
outshwd1	-.455	.6500	-.582	-.327	48.946	1	.006	.634	.559	4.344
speedlim	.230	.0136	.134	.632	2.787	1	.003	1.023	.996	1.051
rodwycls	0[a]	.	.	.	.	.	.	1	.	.
(Scale)	1[b]									
(Negative binomial)	1[b]									

Dependent Variable: Acc. freq
Model: (Intercept), STR., no_lanes, med_type, seg_lng, avg.deg.curv, Avg.G, ADT, medwid, surf_wid, outshwd1, speedlim, rodwycls

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Figure 15:Full model output for parameter estimates of crash frequency model

The result of the statistical analysis using Negative Binomial Regression model shows that, out of the eleven independent variables, nine of them are statistically significant. As a result, it can be deduced that these independent variables have significant influence on accident frequency. To summarize the independent variables which have significant contribution to the accident frequency;

- Segment length (Seg_lng) - Accidents are more likely to happen on roadway segments that are longer in length. However, the shorter segment length has a positive coefficient, which shows a tendency of increasing accident frequency, based on the results shown after creating a scenario of the simultaneous presence of multiple geometric design elements (Presence of gradient, curvature) and with the presence of structure for both conventional two-lane rural roads and multilane urban roads. This is due to the psychological congestion brought on to the driver as a result of frequently changing geometry over shorter distances as well as the narrowing of the cross-section at the locations of the structures. Additionally, since most non-motorized traffic, including pedestrians, vehicles, and other non-motorized traffic, uses the same road, this results in a higher accident rate.

- Average Daily Traffic (ADT) - a similar conclusion was reached for ADT, an increase in ADT has a positive impact on the likelihood of accidents for both conventional two lanes, multilane road sections and on other formed scenarios.
- Average Degree of Curvature (avg.deg.curv.) - Since the coefficient is positive when the variable is seen as a continuous variable, a larger horizontal degree curvature has a positive effect on the likelihood of accidents. However, a separate model was employed for both rural two-lane roads and urban multilane roads in order to determine the threshold above/below which the degree of curvature has a positive or negative impact. as a result, the model coefficient for the model with zero degree of curvature is zero, and when the value increases from zero to 4^0, there is a noticeable increase in accident frequency. However, even if the coefficient value is positive for values between 4^0 and 8^0, the value is lower for those values, and similarly for values of the degree of curvature greater than 8^0 for the entire model, as well as for conventional two-lane and four-lane roads, it can be shown that the majority of the horizontal curves in that area

are milder curves. Furthermore, the presence of structure embodies the worst-case scenario on the model for simultaneous presence of geometric element on the segment, such as gradient and higher degree of curvature.

- Average Gradient (avg.G) - it does not show any significance of either increasing or decreasing the traffic accident rate/frequency. However, based on the descriptive analysis results there is a likelihood of increasing accident rate with increase in degree of curvature. However, to check the threshold above and below which the increment or reduction will occur, separate model has been prepared for the full model and also for both conventional two lane and multilane urban roads. Accordingly, the gradient value has been considered as categorical variable having value of 0 for no grade, 1 for gradient between 0.1- 4% and 2 for gradient values greater than 4%. As a result, the model with zero gradient has a coefficient of negative value which shows a reduction of accident rate as compared to others. Besides, for the model having gradient between 0.1-4% a higher positive value has been shown which shows a significant increase of accident rates. Contrary to that, for gradients having value greater than 6% the coefficient becomes negative which shows reduction of accident.
- Outside Shoulder Width (outshwd1) - As the model output shows, the increase in external shoulder width reduces the frequency of accidents as general and the same has been shown for separate model of conventional two-lane roads.
- Median Width (MedWid) - As the median width increases the frequency of accidents has been reduced in the general model and it is the same for multilane urban roads. Whether the roadway is divided or not, is accounted for implicitly in the ‘median width’ variable. If the roadway is undivided, then the median width would be equal to zero

and when the median width becomes zero there is a significant increase in accident rate and the rate becomes lower with increase in median width.

- Availability of structures (Str) - it also shows a significance for increasing the accident frequency as compared to the sections without having structure since the total surface width reduces at the structure locations and the same has been supported by the coefficient of the total surface width which shows a significant increase in accident frequency when the width is narrower.
- Number of lanes (No_lanes) - they have shown a tendency of increasing accident with increase in the number of lanes (i.e. conventional two lane rural roads experienced lower number of accidents that urban multiple lanes) since the total roadway width increases and exposure increases accordingly
- Speed limit (speedlim) – has showed a tendency of increasing an accident frequency with the availability of speed limits on the road
- Finally, even though the over dispersion parameter α does not have a value different from zero, with the reduced values of AIC and BIC values of the Negative binomial regression model is preferred to the more restrictive Poisson formulation.

b) Models Derived to Estimate crash severity using Ordinal Logistic Regression model with Probit link function:

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[severity = 1]	-15.062	6.087	6.124	1	.013	-26.991	-3.132
	[severity = 2]	-14.603	6.086	5.758	1	.016	-26.532	-2.675
	[severity = 3]	-13.912	6.085	5.227	1	.022	-25.839	-1.985
Location	Grade	.698	.259	4.896	1	.040	.156	1.249
	medwid	.089	.032	7.704	1	.006	.026	.153
	surf_wid	-.073	.013	32.273	1	<.001	-.098	-.048
	lanewid	-.920	.461	3.984	1	.046	-1.823	-.017
	outshwd1	-.853	.396	2.875	1	.015	-.076	-.324
	inshwd1	-9.937	3.695	7.231	1	.007	-17.179	-2.694
	speedlim	.011	.016	.491	1	.483	-.020	.043
	D_curv	.756	.299	6.396	1	.011	.170	1.342
	[STR=0]	-.005	.013	.134	1	.012	-.067	-.027
	[STR=1]	.330	.254	1.686	1	.050	.120	1.023
	[STR=2]	0[a]	.	.	0	.	.	.
	[no_lanes=2]	0[a]	.	.	0	.	.	.
	[no_lanes=4]	0[a]	.	.	0	.	.	.
	[med_type=1]	.946	.372	6.475	1	.011	.217	1.674
	[med_type=2]	.381	.225	2.854	1	.091	-.061	.822
	[med_type=3]	0[a]	.	.	0	.	.	.
	[rururb=1]	0[a]	.	.	0	.	.	.
	[rururb=2]	0[a]	.	.	0	.	.	.
	[rodwycls=1]	0[a]	.	.	0	.	.	.
	[rodwycls=4]	0[a]	.	.	0	.	.	.
	[rodwycls=5]	0[a]	.	.	0	.	.	.

Link function: Probit.

a. This parameter is set to zero because it is redundant.

Figure 16:Full model output for parameter estimates of crash severity model

The result of the statistical analysis using ordinal logistic model (1=fatal accidents, 2= Severe accidents, 3= Slight accidents and 4= property damage only accidents) shows that, out of the ten independent variables, nine of them are statistically significant. As a result, it can be deduced that these independent variables have significant influence on accident severity. To summarize the independent variables which have significant contribution to the accident severity;

- Gradient has shown significance on the accident severity in a way that with increase in vertical gradient, the severity of the accident will also increase. In addition to this, the same also confirmed with the descriptive analysis results (please note that most of the gradients of the project were not significantly deviate from the desirable gradient requirement of the project based on its design class and terrain classifications so that excessive speed reduction of vehicles is not expected while negotiating the vertical curves).
- Median width also shows significance in a way that increasing the median width will results in the reduction of accident severity.
- As surface width increases, the severity will decrease with a coefficient of 0.073 and significance value of less than 0.001.
- Outside and inside shoulder widths have a significance value of 0.015 and 0.007 and coefficients of -0.853 and -9.937 respectively. Which shows there is a tendency of reduction of accident severity with increase in shoulder widths.
- As it has been shown on the output result, speed limit does not show any significance to the severity of the accident.
- Degree of curvature has also shown a significance having a value of 0.011 with a coefficient of 0.756.
- Likewise, structure has also shown significance of increasing the severity with the availability of structures.
- There is a significant reduction of severity of accidents with the availability of physical median barrier as compared to no having median barriers.

Note that the full model output for both crash frequency and severity models will be attached in the appendix part of the document.

5. Conclusion and Recommendation

5.1. Conclusions

Based on the abovementioned analysis results it is concluded that;

- **From the Statistical Analysis Model Results of**

1. Accident Frequency analysis using Negative Binomial Regression Model:

- ✓ it is managed to reach at convergence and remarkable result of the empirical analysis for crash frequency analysis using the Generalized Linear Models for count/discrete outcomes or dependent variables that fits to both the data and to the independent variables considered and the same was supported by literatures and scholastic articles as it has been shown on the analysis part.
- ✓ As a result, by considering Crash Frequency as a dependent variable and segment length, average daily traffic, number of lanes, surface width, average degree of curvature, median width and outside shoulder width as continuous independent variables and availability of structure, availability of median, speed limit as a categorical independent variable's convergence has been reached.

2. Accident Severity Analysis using Ordinal Logistic Regression Model:

- ✓ Similarly, for crash severity analysis, it is also reached at convergence for ordered outcomes /dependent variables (i.e. 1=fatal, 2=severe, 3=slight and 4=PDO accidents) that fits to both the data and to the independent variables considered and the same was supported by literatures and scholastic articles as it has been shown on the analysis part.
- ✓ In line with this, by considering Crash Severity as an ordinal dependent variable and Gradient, Median width, Degree of curvature, Surface width, Lane width, Outside and

inside shoulder widths and speed limit as continuous independent variables and availability of structure, availability of median and number of lanes as a categorical independent variable's convergence has been reached. Accordingly, we can conclude that availability of vertical curves, availability of horizontal curves and availability of structures tend to increase the severity of accidents even though the thresholds above and below which the severity of the crashes will increase or decrease cannot be determined. In addition, increasing surface width, median width, inside and outside shoulder widths and availability of median barrier has a tendency of reducing severity of an accident.

❖ **Interpretation of the results from Engineering Perspective:**

- ✓ The availability of structures has an impact of increasing accident (sections having structures have higher rate of accident as compared to sections without structure) this works for both conventional two-lane roads and urban multi-lane roads but the risk is higher for two lane rural roads as it has been shown on the model output. This is because, the total roadway width will reduce since at the structure locations there is no sufficient walkway width to accommodate pedestrian and other road users.

  Beside to that, on the Addis-Debrebrehan Road, there are bridges that are found fully on the sharp curve and with a sight distance problem. At those locations several accidents have been witnessed.
- ✓ Number of lanes, they have shown a tendency of reducing accident with increasing number of lanes (i.e. two lane roads are expected to have higher number of accidents that urban multiple lanes) since the total roadway width increases accordingly. However, it should be reviewed in conjunction with the availability of median barrier and it's width

since as many literatures proved urban multilane undivided roads have higher accident rates. (MOHAMMED A. HADI, n.d.).

- ✓ Median width, as it has been shown above the availability of median has a greater impact on increasing traffic accident and likewise to the median width. In our case, sections having painted median has higher number of accidents than sections having curbed and raised median.
- ✓ The longer the length of the roadway section, the more likely accidents would occur on these sections. However, at urban sections longer section lengths and higher speed limits were associated with lower crash rates. This could be because longer section lengths indicate more uniform cross-section design.
- ✓ Average degree of curvature has also shown a positive relation to increasing the traffic accident (i.e. sections having horizontal curve have more likelihood of traffic accident than sections without horizontal curves).
- ✓ As it has been shown on the model result, average gradient doesn't show any significance. However, the descriptive analysis results show that it has a positive influence in increasing the traffic accident.
- ✓ Average daily traffic has higher degree of increasing the likelihood of accident occurrence for all types of roadways.
- ✓ Outside shoulder width has a tendency of reducing traffic accident as it increases the total roadway way width that increases a psychological safety to the driver and also provide a space for pedestrian and other road users.
- ✓ Speed limit has also an impact of increasing accident with increasing its limit which shows traffic violation and driving with a speed higher that the posted speed is inevitable.

- ✓ Since most of the geometric design parameters shows significance to both the accident frequency and severity of the traffic accident on the project under study, the road safety auditors and designers should give due consideration and ponder to these design parameters while performing safety performance analysis (safety audit) on the road under study or on other roads having similar roadway characteristics for prioritizing sites based on their crash severance or to upgrade existing roads in. As a remark charts for accident trend analysis have been plotted for major design elements of the road as a reference,

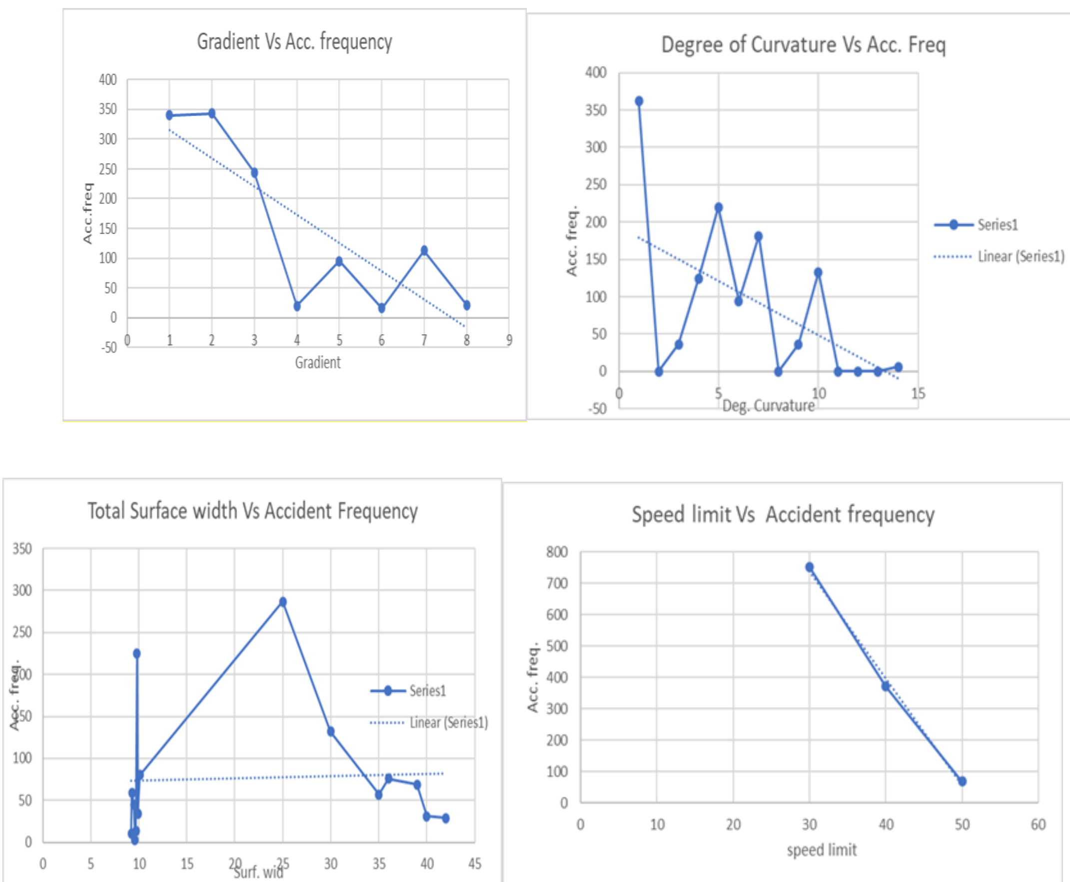


Figure 17: trend analysis chart for geometric design elements

5.2. Recommendations

Based the abovementioned Empirical and Descriptive statistical analysis results I have proposed the following recommendations.

- Since the major cause of the accident in our case is exceeding authorized speed limit and as can be seen from descriptive statistics results of the driver's behavior, most of the drivers are youth (at the age range between 18 to 30) and driving experience between 1-5 years more novice drivers are expected on the roadway. Hence, low enforcements related to the speed limits and also checking their licenses should be strict. In addition to that, low-cost engineering measures should be implemented for speed reduction.
- The other problem is not giving priority to pedestrian, this can be mitigated through creating awareness to the pedestrian and road user to obey roadway rules in addition to improving roadway sign and markings to the site.
- The third problem is improper overtaking followed by improper lane usage which can be mitigated by strict low enforcements and
- As most of the traffic accidents has been occurred on urban multilane road sections and most of these section does not have median barrier that segregates the directional traffic. Hence, construction of median barriers is recommended.
- Related to the accident data collected, even though they do have sufficient formats for registering the accident, the officers don't properly record the event. This may be due to lack of awareness. Hence, it is recommended if there is a centralized system for traffic accident data collection with sufficient details that is managed by highway official that can understand the sensitivity of the data. In addition to creating awareness for traffic police who recorded the event at the first place.

5.3. Future Study

Since the safety impact of the interaction between these factors has not been examined, it is suggested that future work expand accident prediction models by taking into account the combined effect of various explanatory variables that combine human related factors, roadway and its environment factors, and vehicular factors on the roadway.

6. References

A J Skinner, J. Q. B., 1995. Modelling and Simulation in Materials Science and Engineering. *Modelling and Simulation in Materials Science and Engineering.*

Abdelwahab, H. T. A.-A. M. A., 2002. Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. *Transportation Research Record,* 1784(1), p. 115–125.

Abdulhafedh, A., April 30, 2017. Road Crash Prediction Models: Different Statistical Modeling Approaches. *Scientific Research Publishing,* Volume 7.

Adem, A., 2017. Statistical Analysis for Identification of Motor Vehicle Crash Black Spots and Low-Cost Improvements (Case of Addis Ababa To Debre Birhan Road). *AAU Institutional Repository.*

Ahmad Tavassoli Hojati, L. F. S. W. P. C., 2013. Hazard based models for freeway traffic incident duration. *Accident analysis and Prevention,* 52(28), pp. 171-181.

Anon., 2010. Highway Safety Manual 1st edition 2010. In: *Highway Safety Manual 1st edition 2010.* s.l.:American Association of State Highway and Transportation Officials.

Anon., 2018. Developing a Random Parameters Negative Binomial-Lindley Model to analyze highly over-dispersed crash count data. *Dominique Lord, Mohammad Razaur Rahman Shaon, Xiao Qin, Mohammadali Shirazi, Srinivas Reddy Geedipally,* Volume 18, pp. 33-34.

Anon., n.d.

Bedria Hussien, G. S., 2022. ABABA INSTITUTE OF TECHNOLOGY SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING Exploring the causes of Traffic Crash and Injury Severity on Addis Ababa-Adama Expressway. *AAU Institutional Repository.*

Berhanu, G., 2004. Models relating traffic safety with road environment and traffic flows on arterial roads in Addis Ababa. *Accident analysis and prevention.*

Bowman, B. L. R. L. V. a. J. M., January 1994. Estimating Vehicle and Pedestrian Accidents Resulting from Different Median Types. Presented at the 73rd Annual Meeting of the TRB. *Transport Reasearch Board.*

Brewer RD, M. P. C. T. W. S. P. M. P. C., August 25, 1994. The Risk of Dying in Alcohol-Related Automobile Crashes among Habitual Drunk Drivers. *The new England Journal of Medicine.*

Buyco. C, S. F., 1987. *LONGLINEAR ANALYSIS OF TRUCK ACCIDENT RATES USING GLIM --ROAD USER BEHAVIOR. THEORY AND RESEARCH.* s.l., Transport and Road Research Laboratory (TRRL).

Byung-Jung Park, D. L., 2009. Application of finite mixture models for vehicle crash data analysis. *Accident analysis and Prevention,* 41(4), pp. 683-691.

C. Riviere, P. L. J. F. M. R. Y., 2006. A Bayesian Neural Network approach to estimating the Energy Equivalent Speed. *Accident Analysis and Prevention.*

D.W. Harwood, F. C. E. H. W. H. a. A. V., May 1997—September 2000. *PREDICTION OF THE EXPECTED SAFETY PERFORMANCE OF RURAL TWO-LANE HIGHWAYS,* McLean, Virginia 22101-2296: Office of Safety Research and Development.

Daniel Brookoff, C. S. C. C. W. a. C. S. M., August 25, 1994. Testing Reckless Drivers for Cocaine and Marijuana. *the new England Journal of Medicine.*

Deme, D., 2019. Road Traffic Accident in Ethiopia from 2007/08-2017/18. *American International Journal of Sciences and Engineering Research,* 2(2).

DimitriosTrichopoulos, X. Z. K. K. A. T. P. D.-V., 26 February 1983. PSYCHOLOGICAL STRESS AND FATAL HEART ATTACK: THE ATHENS (1981) EARTHQUAKE NATURAL EXPERIMENT. *The Lancet,* Volume 321(Issue 8322), pp. 441-444.

Dominique Lord, S. D. S. R. G., 2008. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes. *Accident analysis and Prediction,* 40(3), pp. 1123-1134.

Dominique Lord, S. W. J. N., 2007. Further notes on the application of zero-inflated models in highway safety. *Accident Analysis and Prevention,* 39(1), pp. 53-57.

Donald A. Redelmeier, M. a. R. J. T. P., 13 february, 1997. Association between Cellular-Telephone Calls and Motor Vehicle Collisions. *the new England Journal of Medicine.*

Doohee Nam, F. M., 2000. An exploratory hazard-based analysis of highway incident duration. *Transportation Research Part A: Policy and Practice,* 34(2), pp. 85-102.

Ehrhart, N. J. G. a. A. A., 2000. Effect of Speed, Flow, and Geometric Characteristics on Crash Frequency for Two-Lane Highways. *Transportation Research Record: Journal of the Transportation Research Board,* 1717(1).

Eleni Petridou, M. M., September 2000. Human factors in the causation of road traffic crashes. *Europan Journal of Epidemology.*

Emmanuelle Dupont, E. P., n.d. Multilevel analysis in road safety research;.

Ezra Hauer, Y. M., 2004. Safety Models for Urban Four-Lane Undivided Road Segments. *Transportation Research Record: Journal of the Transportation Research Board,* 1897(1).

Fitzpatrick, K. J. M., 2000. Speed-Profile Model for Two-Lane Rural Highways.. *Transportation Research Record,* 1737(1), p. 42–49.

Fr¨uhwirth-Schnatter, S., December 2, 2008. *Finite Mixture and Markov Switching Models, Implementation in MATLAB using the package bayesf Version 2.0,.* s.l.:s.n.

Francisca Ogwueleka, S. M. T. C. O. L. F.-S., June 2014. An Artificial Neural Network Model for Road Accident Prediction: A Case Study of a Developing Country. *Acta Polytechnica Hungarica 11(5):177-197,* pp. 177-197.

Frühwirth-Schnatter, S., 2006. *Finite Mixture and Markov Switching Models.* s.l.:s.n.

G. N. XieQ., W. W. M. Z. L. Q. L., 2007. Heat transfer analysis for shell-and-tube heat exchangers with experimental data by artificial neural networks approach. *Applied Termal Engineering.*

G.R.Wood, 2002. Generalised linear accident models and goodness of fit testing. *Accident Analysis and Prevention,* 34(4), pp. 417-427.

Galit Shmueli, T. P. M. J. B. K. S. B. P. B., 2004. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Journal of Royal Statstical Society.*

Goldstein, H., 2011. *Multilevel statistical models..* s.l.:s.n.

Greibe, P., 2003. @article{Greibe2003AccidentPM,. *Accident analysis and prevention,* 35(2), pp. 273-85.

Gudmundur F. Ulfarsson, V. N. S., 2003. Accident Count Model Based on Multiyear Cross-Sectional Roadway Data with Serial Correlation. *Transportation Research Record: Journal of the Transportation Research Board,* 1840(1).

Guo, G., 1996. Negative Multinomial Regression Models for Clustered Event Counts. *Sociological Methodology,* Volume 26, pp. 113-132 .

Hastie, T. J. T. R., 1990. Generalized additive models. 43, 335.. *Monographs on statistics and applied probability,,* 43(8), p. 335.

Howat, P. D. S. a. I. S., 1991. "Alcohol and driving: is the 0.05% blood alcohol concentration limit justified?.". *Drug and alcohol review; Wiley Online Library,* pp. 151-166.

Huang, H. A.-A. M., 2010. Multilevel data and Bayesian analysis in traffic safety.. *Accident Analysis and Prevention 42 ,* pp. 1556-1565.

Ivan, C. Z. a. J. N., 2019. Effects of Geometric Characteristics on Head-On Crash Incidence on Two-Lane Roads in Connecticut. *Transportation Research Record: Journal of the Transportation Research Board,* 1908(1).

Jake Kononov, C. L. B. K. A., 2011. Relation of Flow, Speed, and Density of Urban Freeways to Functional Form of a Safety Performance Function. *Transportation Research Record: Journal of the Transportation Research Board,* 2236(1).

Jerry A. Hausman, B. H. H. ,. Z. G., 1984. Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *National Bureau of Economic Research.*

Joseph B. Kadane, T. P. M. G. S. P. B. S. B., 2006. Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Analysis,* 1(2), pp. 363-374.

JULIE R, A. G. A. G., 24 Oct 2007. Speeding in relation to perceptions of risk, utility and driving style by British company car drivers. *Ergonomics,* pp. 2525-2535.

Jutaek Oh, S. P. D., 2006. Accident prediction model for railway-highway interfaces. *Accident analysis and Prevention,* 38(2), pp. 346-356.

Karim El-Basyouny, T. S., 2009. Urban Arterial Accident Prediction Models with Spatial Effects. *Transportation Research Record: Journal of the Transportation Research Board,* 2102(1).

Kecman, V., 2005. Support Vector Machines– An Introduction. *Support Vector Machines: Theory and Applications. Studies in Fuzziness and Soft Computing.*

Kloeden, C. N. &. M. A. J. &. P. G. &. A. T. S. B., 2001. Travelling speed and the risk of crash involvement on rural roads. [Canberra] : ,. *Australian Transport Safety Bureau.*

Knuiman, M. W. F. M. C. a. D. W. R., 1993. The Effect of Median Width on Highway Accident Rates. *Transportation Research Record 1401, TRB, National Research Council, Washington, D.C., ,* pp. 70-80.

Kussia, A., 2017. Trends, Causes, and Costs of Road Traffic Accidents in Ethiopia. *GRIN,* p. 44.

Lambert, D., 1992. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing". *Technometrics,* p. 34 (1): 1–14.

Lambert, D., 1992. Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics,* 34(1).

Liang, F., 2005. Bayesian neural networks for nonlinear time series forecasting. *Statstics and computing.*

Li-YenChang, 2005. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. *Safety Science,* 43(8), pp. 541-557.

Lord, D. P. B. N., 2000. Accident Prediction Models with and without Trend: Application of the Generalized Estimating Equation Procedure. *Transportation Research Record 1717. Transportation Research Board of the National Academies, Washington, .*

M A Abdel-Aty, A. E. R., 2000. Modeling traffic accident occurrence and involvement. *Accident analysis and Prevetion,* 32(5), pp. 633-42.

McMurray, L., 1970. Emotional stress and driving performance: The effect of divorce. *American Psycological Association ,* p. 100–114.

Miaou, S. P. H. S. P. T. W. A. K. R. a. S. C. D., 1992. Relationship between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach.. *Transportation Research Record 1376, TRB, National Research Council,,* pp. 10-18.

MOHAMMED A. HADI, J. A. L.-F. C., n.d. Estimating Safety Effects of Cross-Section design for various highway types using negative binomial regression. *Transportation reserch recored.*

Myers, R. H., 1992.. Classical and Modern Regression with Applications.. *PWS-KENT Publishing Company, Boston, Mass., .*

Nataliya V. Malyshkina Fred, L. M. A. P., 2009. Markov switching negative binomial models: An application to vehicle accident frequencies. *Accident Analysis and Prevention,* 41(2), pp. 217-226.

Organization, W. H., 20 June 2022. *road traffic injuries,* s.l.: World Health Organization.

Pan Chengye, p. R., 2013. Modeling motorway accidents using negative binomial regression. *Joutrnal of Eastern the Asia Society for transsportation studies,* Volume 10.

Panagiotis Ch. Anastasopoulos, F. L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident analysis and Prevention,* 41(1), pp. 153-159.

Paniati, J. t. J., 1996. INTERACTIVE HIGHWAY SAFETY DESIGN MODEL (IHSDM): DESIGNING HIGHWAYS WITH SAFETY IN MIND. *Transportation research circular.*

PAUL P. JOVANIS, H.-L. C., 1986. Modeling the Relationship of Accidents to Miles Traveled. *Transportation Research Board,* Volume 1060.

PetterGregersen, N., March 1996. Young drivers' overestimation of their own skill—an experiment on the relation between training strategy and skill. *Elsevier,* Volume 28(Issue 2), pp. 243-250.

Quintel Washington, A. S. N. P. S. B. J. J. T. N. S. P., 2014. Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities. *A Journal of Mathmatics,* 7(6), p. 751–767.

Rajalin, S., October 1994. The connection between risky driving and involvement in fatal accidents. *Elsevier,* Volume 26(Issue 5), pp. 555-562.

Retallack AE, O. B., 2020. Relationship Between Traffic Volume and Accident Frequency at Intersections. *International Journal of Environmental Research and Public Health,* 17(4).

Salvatore Cafisoa, A. D. G. G. D. S. G. C. B. P., 2003. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. *Accident Analysis & Prevention,* 42(4), pp. 1072-1079.

Sarath C. Joshua, N. J. G., 2007. estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation Planning and Technology.*

Shaw-Pin Miaou, D. L., 2003. Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods. *Transportation Research Record: Journal of the Transportation Research Board,* 180(1).

Shaw-Pin Miaou, J. J. S., 2005. Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accident analysis and Prevention,* 37(4), pp. 699-720.

Shaw-PinMiaou, August 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident analysis and prevention,* 26(4), pp. 471-482.

Shugan, S. M., 2006. Errors in the Variables, Unobserved Heterogeneity, and Other Ways of Hiding Statistical Error. *Marketing Science,* 25(3).

Simon Washington, P. C. A. T. H., 2013. Hazard based models for freeway traffic incident duration. *Accident analysis and prevention,* 52(28), pp. 171-181.

Simon, F. C., 1996. Road traffic offending, stress, age, and accident history among male and female drivers. *Ergonomics,* pp. 757-778.

Teferi Abegaz, S. G., January 29, 2019. Magnitude of road traffic accident related injuries and fatalities in Ethiopia.

Teyba Wedajo, E. T. Q. M. M., January-2017. Analysis of Road Traffic Accident Related of Geometric Design Parameters in Alamata-Mehoni- Hewane Section. *International Journal of Scientific & Engineering Research,* 8(1).

Trichopoulos, D. e. a., 1983. Psychological stress and fatal heart attack. *The Lancet 321.8322 ,* pp. 441-444.

Venkataraman Shankar, F. M. W. B., June 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident analysis and prevention,* 27(3), pp. 371-389.

Vivienne M. Moore, J. D. A. J. W., May, 1995. Vehicle Speed and Risk of a Severe Crash. *Epidemiology,* Volume Volume 6, No. 3 , pp. pp. 258-262.

Vogel-Sprott, M. S.-J. K., July 1989. Learning alcohol tolerance: the contribution of response expectancies. *Psychopharmacology.*

Vogt, A. B. J., 1998. Accident Models for Two-Lane Rural Segments and Intersections. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 1635, pp. 18-29..

W. Kriswardhana, A. H. I. M. P., 2020. Modelling road traffic accident rate and road geometric parameters relationship. *Semantic Scholar.*

Weifan Zhong, O. a. D., 2023. Predicting Traffic Casualties Using Support Vector Machines with Heuristic Algorithms: A Study Based on Collision Data of Urban Roads;. *Sustainability ,* 15 (4), p. 2944.

Wood, S. N., 2013. A simple test for random effects in regression models. *Biometrica,* 100(4), pp. 1005-1010.

Ye, X., Li, Y., Tong, L. & He, L., , 23–28 July 2017;. *Remote Sensing Retrieval of Suspended Solids in Longquan Lake Based on Ga-Svm Model.* Fort Worth, TX, USA, s.n.

Zhang, J. D. C. S. Y. e. a., 2008. Polymorphisms of tumor necrosis factor-alpha are associated with increased susceptibility to gastric cancer: a meta-analysis. *Journal of human Jenesis,* Volume 53, p. 479–489.

7. APPENDIXES

Appendix 1: Sample Traffic Data Collecting Format

Appendix 2: Summary of the collected crash data

Appendix 3: Summary of the collected roadway data

Appendix 4: Output result for collinearity tests

Appendix 5: Output result for full Negative Binomial Regression Models

Appendix 6: Output result of separate models for Two lane rural roads and Urban multilane roads

Appendix 7: Output result for full Ordinal Logistic Model with probit link function