Developing road crash prediction models to investigate the combination of effects of roadway conditions on national rural road crashes

by

Robert Ambunda

Dissertation presented for the degree of Doctor of Philosophy in the Faculty of Engineering at Stellenbosch University

Promoter: Prof. Marion Sinclair

December 2021

Declaration

By submitting this thesis/ dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2021

Abstract

Namibia is faced with the reality of an increase in the frequency of fatal and serious injury (FSI) crashes on national rural roads, despite the roadway infrastructure considered to be in good condition. More so, an increase in roadway traffic volume has subsequently worsened the crash risk levels for road users. To address this issue, the study was aimed at exploring the combinatorial effects of road and traffic characteristics of national rural roads in Namibia on fatal and serious injury crashes and the crash risk factors preceding the crashes. The main crash dataset, for the period 2012 to 2016, and supplemented roadway design data were provided by the Namibian National Road Safety Council (NRSC) and Roads Authority (RA) respectively. The study applied novel robust multiple linear regression models and cluster analysis to the aggregated study dataset. The study objectives were five-fold. The first objective of the study was to examine the profiles and risk factors attributed to national rural road crashes. The goal of this objective was to create a new basis to assess the relationship between road characteristics and driver risk factors preceding road crashes. This will serve as a basis for crash risk factor comparisons for any future studies. The second objective was to identify high risk traffic crash locations on the different national rural road classifications. The third objective was to assess the distribution of fatal and serious injury crashes across the national rural road network by applying the KDE spatial analysis technique. The fourth objective was to investigate the compliance of the rural road design characteristics with road design guidelines. Recommendations on the suitability of the design standards were based on the results of the first three and fifth objectives of the study. The fifth objective of the study was to develop novel road crash predictive models; calibrated and within the context of the Namibian national rural road environment. This objective was underpinned by the other four objectives in examining the spatial distribution of the road crashes, the response of crash distribution to design compliance levels and the sensitivity of the novel CPMs to changes in design characteristics. The insights from the study will have a long-standing and significant impact on rural road safety in Sub- Saharan Africa (SSA) and beyond. The study has highlighted multiple areas in the rural road safety system that urgently need to be addressed to provide a safer environment for road users on the network. As Namibia prepares the new Decade of Action (DoA) Strategic Plan for the year 2021 to 2030, the insights from the study provide a backbone on which rural road safety can be addressed in the DoA, with an approach that is aimed at reducing and eliminating so-called latent gaps in the components of a safe road system.

Keywords: Road safety, Crash modelling, Risk factors, National rural roads, Namibia

Opsomming

Namibië word gekonfronteer met die realiteit van ’n toename in die frekwensie van noodlottige en ernstige beserings (Fatal and Serious Injury, FSI) op plattelandse nasionale paaie, ondanks die feit dat die ryvlak infrastruktuur in ’n goeie toestand is. ’n Toename in die verkeersvolume het gevolglik ook die ongeluks-risiko vir padverbruikers vererger. Om hierdie kwessie aan te spreek, was die studie gerig op die navorsing van kombinatoriese effekte van pad- en verkeers-neigings van plattelandse nasionale paaie in Namibië op noodlottige en ernstige beserings-ongelukke en die risiko faktore wat die ongelukke voorafgaan. Die hoofbotsing-datastelsel vir die tydperk 2012 tot 2016, en aangevulde data van die ryvlak, is onderskeidelik deur die Namibiese Nasionale Padveiligheidsraad (Namibian National Road Safety Council, NRSC) en die Padowerheid (Roads Authority, RA) verskaf. Die studie het nuwe robuuste meervoudige lineêre regressiemodelle en groepsanalise toegepas op die geaggregeerde datastelsel. Die studie doelstellings was vyf-voudig. Die eerste doelstelling van die studie was om die profiel en risikofaktore wat toegeskryf word aan plattelandse nasionale padongelukke, te ondersoek. Die oogmerk van hierdie doel was om ’n nuwe basis te skep om die verband tussen padkenmerke en bestuurder risikofaktore voor padongelukke te beoordeel. Dit sal dien as basis vir die vergelyking van botsing risikofaktore vir toekomstige studies. Die tweede doelwit was om hoë risiko verkeersongeluk areas op die verskillende plattelandse nasionale pad-klassifikasies te identifiseer. Die derde doelwit was om die verspreiding van noodlottige en ernstige beserings-ongelukke oor die plattelandse nasionale padnetwerk te beoordeel deur die KDE-ruimtelike ontledingstegniek toe te pas. Die vierde doel was om te ondersoek of die plattelandse padontwerp eienskappe aan die padontwerp riglyne voldoen. Aanbevelings oor die geskiktheid van die ontwerpstandaarde is gebaseer op die resultate van die eerste drie en vyfde doelstellings van die studie. Die vyfde doelstelling van die studie was om nuwe voorspellingsmodelle vir padongelukke te ontwikkel; gekalibreer en spesifiek binne die konteks van die Namibiese plattelandse padomgewing. Hierdie doelstelling was ondersteun deur die ander vier doelstellings om die ruimtelike verspreiding van padongelukke, die reaksie van verspreiding van botsings op die ontwerp-voldoeningsvlakke en die sensitiwiteit van die CPMe vir veranderinge in die ontwerpkenmerke, te ondersoek. Die insigte uit die studie sal ’n langdurende en belangrike invloed op padveiligheid in Afrika suid van die Sahara (SSA) en daarbuite stel. Die studie het verskeie areas in die plattelandse padveiligheidstelsels beklemtoon wat dringend aangespreek moet word om ’n veiliger omgewing vir padverbruikers op die netwerk te bied. Namate Namibië die nuwe ‘Decade of Action’ (DoA) strategiese plan vir die tydperk 2021 tot 2030 voorberei, bied die insigte uit die studie ’n grondslag waarop plattelandse padveiligheid in die DoA aangespreek kan word, met ’n benadering wat daarop gemik is om sogenaamde latente leemtes in die komponente van ’n veilige padstelsel te verminder en uit te skakel.

Sleutelwoorde: Padveilgheid, Botsingsmodellering, Risikofaktore, Plattelandse Nasionale paaie, Namibië

Submitted manuscripts included in this thesis

Title 1: Analysis of contributory factors to crash risk levels on national roads: Drivers perspective
Author(s): **R. Ambunda** & M Sinclair
Submitted:
Manuscript Number:

Title 2: Temporal and demographic analysis of fatal and serious injury crashes among drivers in Namibia
Author(s): **R. Ambunda** & M Sinclair
Submitted:
Manuscript Number:

Title 3: Combinational effects of roadway conditions on road safety on various national road classifications – a case study from Namibia
Author(s): **R Ambunda** & M Sinclair
Submitted:
Manuscript Number:

Title 4: Comparative study on the impact of existing and "ideal" road design characteristics on road safety
Author(s): **R Ambunda** & M Sinclair
Submitted:
Manuscript Number:

Publications during candidature

I. Peer reviewed journal papers:

N/A

II. Peer reviewed conference papers:

Title 1: Influence of road geometric and traffic features on road safety
Author(s): **R Ambunda** & M Sinclair
Submitted: Southern African Transport Conference 2021 (5-7 July 2021)
Manuscript Number:

Acknowledgement

I wish to acknowledge and thank the following people and institutions for their valuable support, advice and contribution throughout my PhD study at Stellenbosch University;

First and Foremost, I would like to thank the Lord for the health, wisdom, protection and strength to carry out and complete my PhD studies.

My deepest gratitude goes to my supervisor, Prof. Marion Sinclair, for the astute academic leadership, support and critical feedback provided throughout the PhD journey. Her continuous motivation and advice has been vital to the completion of the PhD study.

Special thanks goes to the Namibian National Road Safety Council, the Namibian Police Authority and the Motor Vehicle Accident Fund of Namibia for providing the traffic crash data used in the study. Your prompt response and willingness has been noticed and appreciated greatly.

Special thanks goes to the Roads Authority of Namibia for their guidance and help in acquiring road design and traffic data on the national roads.

Enormous gratitude goes to my parents and sister for the constant prayers, motivation and sacrifices throughout my studies. Your unwavering support has been a pillar of strength and motivation for me throughout the study period. Words may never be enough to express my appreciation.

Special thanks to my family and friends for your words of wisdom, advice and support.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway and Transportation Officials
AIC	Akaike Information Criterion
ARR	All Rural Roads
ANOVA	Analysis of Variance
AU	African Union
BIC	Bayesian Information Criterion
BMM	Base Mean Model
CCI	Cross-Cutting Issues
CF	Cluster Features
CPM	Crash Prediction Models
CR	Crash Rate
CSRA	Committee of State Road Authorities
DoA	Decade of Action
ESDA	Exploratory Spatial Data Analysis
FHWA	Federal Highway Administration
FSI	Fatal and Serious Injury
GIS	Geographical Information System
GP	Generalized Poisson
GPS	Global Positioning System
GSW	Ground Shoulder Width
HOR	Horizontal Curve Radii
HORR	High Order Rural Roads
HSM	Highway Safety Manual
HV	Heavy Vehicles
KS	Kolmogorov-Smirnov
KVR	Kaiser-Varimax Rotation
LORR	Low Order Rural Roads
LRDC	Law Reform and Development Commission of Namibia
LV	Light Vehicles
LW	Lane Width
MVA	Motor Vehicle Accident Fund of Namibia
MLR	Multiple Linear Regression
NB	Negative Binomial
NKDE	Network Kernel Density Estimation
NRSC	National Road Safety Council
NSA	Namibia Statistics Agency
OECD	Organisation of Economic Development and Cooperation

OPS	Operating Speed
QKF	Quartic Kernel Function
PC	Pavement Condition
PDF	Probability Density Function
PKDE	Planar Kernel Density Estimation
PSD	Passing Sight Distance
RSMB	Road Safety management Bill
SCM	Swiss Cheese Model
SEM	Structural Equation Modelling
SHOT	Shoulder Type
SSATPP	Sub-Saharan Africa Transport Policy Program
SSD	Stopping Sight Distance
SSE	Sum of Squares Error
SST	Sum of Squares Total
SSW	Surfaced Shoulder Width
ST	Surface Type
SU	Stellenbosch University
SW	Shapiro-Wilk
TKC	Trans-Kalahari Corridor
TRH	Technical Recommendations for Highways
TSC	Two-Step Cluster
TV	Terrain Vertical
UN	United Nations
WHO	World Health Organisation

Chapter 1: Introduction

Numerous research efforts on road safety analysis methods, including road safety statistical modelling (crash prediction models), descriptive road crash profiling and geospatial crash analysis have been conducted worldwide in recent years, in an attempt to investigate the link between the frequency and severity of road crashes, and road and traffic characteristics.

Reducing the frequency and severity of road traffic crashes has constantly been one of the most important tasks for transportation and traffic engineers. Traffic safety can be influenced by improving the geometric aspects of the roadway system and their influence on driver behaviour, coupled with developing, enforcing traffic rules and educating road drivers on the importance of road safety. Investigating the extent of the link between road crashes and the characteristics of the roadways underpins the efforts to improve the precarious road safety situation on the roads.

The ability to predict road crash rates is also important to transportation engineers, as it provides the capacity to identify potential high-risk road and traffic characteristics that influence the frequency and severity of road crashes, and potential hazardous road sections that warrant further road safety examinations. Moreover, an investigation into road crash profiles and crash causation factors reported in the historical crash data is vital in providing an insight into the behavioural aspects of the drivers on the roadways. Information on the factors influencing the occurrence of road crashes is key for road safety authorities to develop, identify and implement evidence-based proactive and remedial measures and treatments to provide a safer driving environment.

The goal of the study was to develop a method that quantitatively investigates the extent of the combined effect of various national rural road geometric, pavement and traffic characteristics on road safety. The developed method provides a straightforward and mathematically sound way of predicting road crash rates and identifying combinational crash risk factors that potentially precede road crashes and affect driver safety on the roadways.

1.1 Background

At a global level, fatalities and injuries resulting from road traffic crashes have been on an increasing trend. Road safety is one of the most significant issues in modern society, with the World Health Organisation (WHO) (2017) estimates showing that over 1.3 million road users die every year globally on the world's roads, and that another 20 to 50 million road users sustain non-fatal injuries of various severity. Traffic safety is a major concern for developing countries, due to a greater burden of higher injury severity crashes compared to other world regions. Developing countries are reported to account for 90 percent of road traffic crashes worldwide, while only having 48 percent of the world's vehicle population (Peden *et al.*, 2017).

Namibia is rated as one of the countries with the highest road traffic related fatalities on the African continent. In 2015, the Namibian Statistics Agency (NSA) (2015) reported that Namibia's road fatality rate (31.1) (see Figure 1.1) was higher than the African continental average (26.6), by more than 4 fatalities per 100 000 population. The Namibian National Road Safety Council (NRSC) (2012) states that road traffic crashes are one of the major and increasing causes of deaths in Namibia. This is despite considerable efforts by road safety stakeholders in Namibia, to reduce the frequency and severity of crashes. Moreover, with an increase in road traffic volumes, traffic safety has become and continues to be a serious concern for authorities in Namibia.

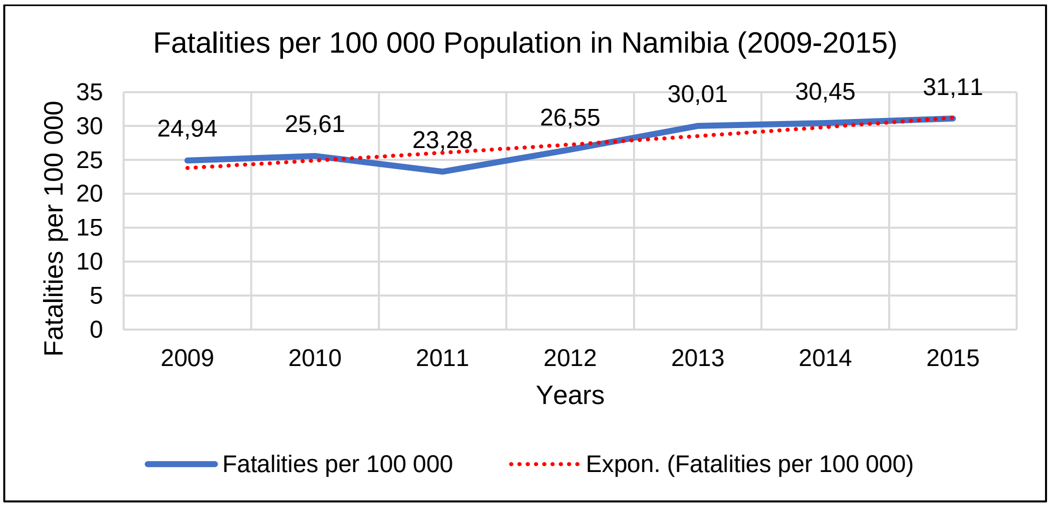


Figure 1.1 Fatalities per 100 000 Population in Namibia 2009-2015 (NSA, 2015)

In the same way to crash fatality rates, the Namibian Statistics Agency (2015) reported an increasing trend in the rate of seriously injured road users per 100 000 population on Namibian roads (see Figure 1.2). The period from 2011 to 2015 recorded a compounded increase of four (4) percent in the frequency of injury crashes.

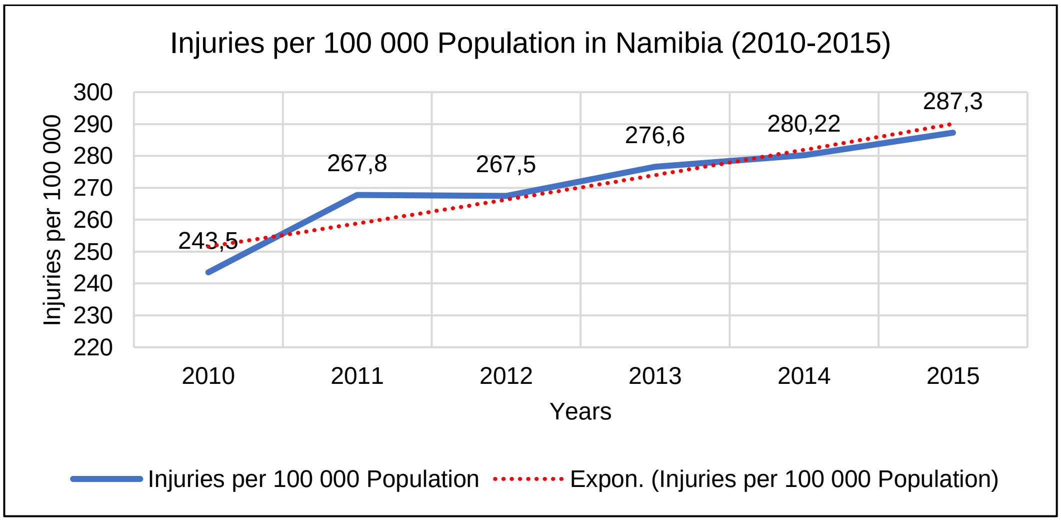


Figure 1.2 Injuries per 100 000 Population in Namibia 2010-2015 (NSA, 2015)

Road crashes are complex events and are influenced by multiple factors such as road geometric design, traffic volume and composition, speed differentials between vehicles of the same and different classes, weather and drivers physical and mental conditions (Vayalamkuzhi and Amirthalingam, 2016). Runji (2003) notes that a variety of factors influence the frequency and severity of road crashes, relating to driver behaviour and perceptions, the roadway environment and vehicle related factors. A graphical representation of the combination of main risk factors in crash occurrences is shown in Figure 1.3. It is important to note that the relative role of these three factors can differ significantly between countries.

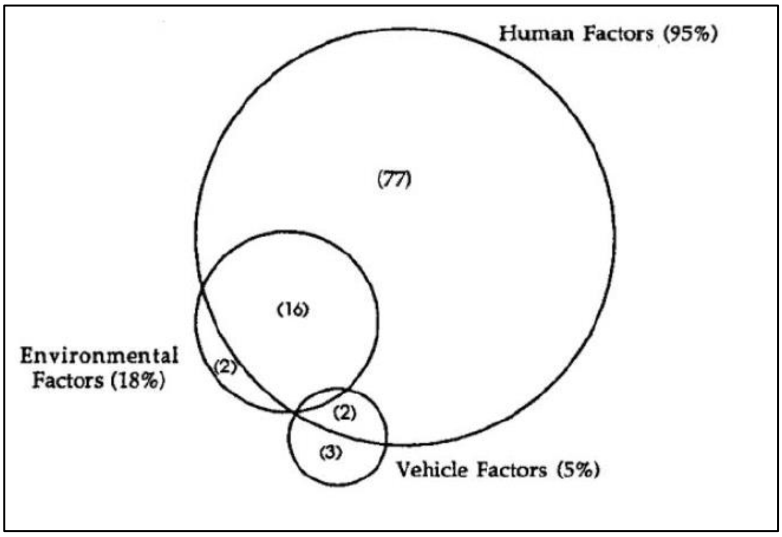


Figure 1.3 Factors influencing the occurrence of road crashes (Runji, 2003)

Yingxue (2009) reports that despite unsafe behaviour of drivers, such as excessive operating speeds, fatigued driving, driving under the influence of alcohol and overloading, contributing highly to road traffic crashes, many road crashes are simply the result of road design elements and the road environment, due to negative road designs that lead to hazardous driver perceptions.

The national rural roadway environment is often the location for higher severity road crashes due to undivided roads, high operating speeds and poor lighting conditions. Mohammed (2013) notes that aspects of the national rural road environment often included in road safety assessments include the geometrical characteristics of the road facilities and the traffic conditions on the roadway, relating to traffic composition and speeds. Ambunda & Sinclair (2019) mention that road safety analysis can be useful in identifying road sections prone to high road crash incidence and high injury severity, while determining the factors significantly contributing to the high road crash rates and influencing driver perceptions. Estimating the causes and factors influencing road crashes on a given national rural road is important in evaluating the different road design variables and alternatives (Glavić *et al.*, 2016). Road safety analysis plays an important role in ensuring a safe and efficient transportation system, with a variety of methods used to quantitatively assess and visually communicate the safety of transportation facilities (Stephan and Newstead, 2017). Various methods have been used to carry out road safety analysis, with several statistical and geospatial analysis methods commonly used by road traffic engineers (Hauer, 2014).

Hauer (2014) mentions that reducing road crashes on rural roads has always been one of the most important tasks for traffic engineers. Recently, the influence of rural road environments on road crash incidence has attracted considerable research interest, with road safety modelling taking the lead in statistical road safety analysis due to its wide variety of applications and practical implications (Karlaftis & Golias, 2002; Gaudry & Vernier, 2002). The study has used both geospatial and statistical road safety analysis methods to investigate and analyse the extent of the relationship between the rural roadway environment and crash risk levels on the Namibian national road network.

1.2 Problem Statement

Namibia is faced with the reality of an increase in the frequency of fatal, serious and slight injury crashes, despite roadway infrastructure considered to be in good condition. Moreover, with an increase in roadway traffic volume, traffic safety has become a serious concern for traffic safety management authorities in Namibia.

Road traffic crashes occur as the result of a combination and interaction of several interrelated factors comprising driver related behaviour, the road environment and vehicle related factors (Turner *et al.*, 2015). Notwithstanding the general recognition that road user behaviour and perceptions on the roadways are the primary cause of road traffic crashes, the road environment and its geometric properties play a significant role on the crash risk level, due to its impact on road user perceptions and general safety on the roadway (Deller, 2013; Taylor *et al.*, 2000; Ambunda & Sinclair, 2019).

Most developing countries, including Namibia are faced with a lack of tools to predict and investigate the crash likelihood. Therefore, road safety authorities tend to be reactive instead of proactive to road safety issues. Moreover, little is known of the influence that the Namibian national rural road environment has on the occurrence of road crashes and the level of crash severity, as no literature was found in relation to examining the extent of the relationship between road design elements and traffic characteristics on the crash risk level.

Road elements are designed taking into consideration average driver behaviour, reactions and traffic conditions. Driver behaviour, however, is the direct result of how a driver ‘reads’ the road environment and determines what driving behaviour is appropriate given the physical environment. In this way driver behaviour is directly and immediately influenced by the combination of road design elements and traffic conditions. Road safety is abundant with studies investigating the influence of single road design and traffic elements on road safety – for example, the effect of the provision of a hard shoulder on driver perception and hence safety. Yet design elements work in tangent with each other – a hard shoulder is only one design detail among others which include lane width, horizontal and vertical curvature, pavement design, road marking and so forth. In addition, prevailing traffic flow offers another dimension to the information received and interpreted by the driver – the prevailing speed and traffic volumes, the proportion of heavy goods vehicles etc. All these factors together determine how the road environment is perceived and what behaviour then emerges. As such, it is vital that the road environment is considered as a whole in investigating its effect on road safety, without isolating single design and traffic variables.

Due to the lack of local studies on the relationship between road safety and combination of road elements, the authorities responsible for compiling road design standards have relied heavily on their own judgements or on standards imported from other countries, in the absence of appropriate local sources. The unavailability of local standards and the potential non-accordance of the adopted road design standards for the road network in Namibia increases the risks and contributes to the precarious road safety problem.

Road safety analyses require reliable and accurate historical crash data, with information on traffic characteristics, traffic exposure variables and the road environment vital for an appropriate geospatial and statistical analyses. The historical crash data collected by the Namibian road safety authorities is not geo-coded, with majority of the site-specific crash information missing. It was thus important for the study to address the deficiencies in the data by developing an approach to attempt to overcome the data shortcomings and by gathering additional site-specific information to carry out a comprehensive statistical and geospatial analysis focused on addressing road safety on national rural roads.

1.3 Study Aims and Objectives

The study develops road crash predictive models and investigates the relationship between road crashes and the Namibian national rural road environment, using historic crash data from the period 2012 to 2016. There is a need to inform and improve the road safety understanding on the implications of the rural road environment on the frequency and severity of road crashes in the Namibian context.

The main aim of the study is to investigate and develop road safety crash predictive models to explore the relationship between the combination of national rural road design, pavement and traffic conditions, and road crashes of numerous severity levels, using Namibia as a case study. Data analytics plays a significant role in the development of the road crash predictive models and their benchmarking against road safety conditions in countries with similar road conditions to Namibia. To this end, the study has the following specific objectives:

1. To identify high risk road traffic crash locations on the different national rural road classifications by using geospatial analysis methods
2. To assess how the geospatial analysis methods vary in performance at the different crash locations in identifying the high-risk road traffic crash locations.
3. To investigate the compliance of the rural road environment design variables with the road design standards used to design Namibian national rural roads, with the intent to:
   a) To develop a tool to compare road attribute data with current road design standards and identify sub- standard road elements considered to be deficits;
   b) To quantify the extent of the link between road design standard compliance of the high crash risk zones with road crashes
   c) To find ways of increasing the impact of the safety aspect in road design standards on road safety.
4. To examine and describe the road traffic crash profiles and crash risk factors attributed to the crashes from the historical crash database on the high-risk road traffic crash zones, with the intent to:
   a) To describe the road traffic crashes by injury severity
   b) To describe the road traffic crashes by the demographic characteristics
   c) To assess the road traffic crashes by the locations of the crashes
   d) To determine the extent to which temporal, demographic and roadway factors influence the combination of risk factors preceding road crashes and their overall impact on driver safety on national rural roads.
5. To develop road crash prediction model tools to investigate the relationship between the geometric design, pavement and traffic conditions, and road crashes on the Namibian national rural road network, with the intent to:
   a) To Identify the rural road design variables and traffic characteristics that influence road crash incidence on the identified study sections
   b) To quantitatively assess the extent of the relationship between the rural road environment design variables and the road crashes, and how this varies spatially for the different study sections.

1.4 Study Definition of terms

Road traffic safety: The methods and measures used to prevent road users from being killed or seriously injured (Ahmed, 2013).

Traffic crashes: Refers to a collision between vehicles or with an object. The term road crash reflects an element of causality, apportioning responsibility to road users and/ or traffic and road characteristics.

Road section: Uninterrupted flow facilities where traffic flow conditions result from the interaction among vehicles in the traffic flow, and between vehicles and the geometric and environmental characteristics of the roadway (Transportation Research Board, 2000).

Crash rate: The number of road crashes in a given period of time as compared to the traffic volume or other exposure variables.

Crash prediction model: Mathematical models that express the safety performance of road type/ network based on traffic and road characteristics (Duivenvoorden, 2010).

Road user: Refers to anyone that uses the road. Usually grouped into motorised and non-motorised road users.

Road fatality: A death resulting from a road traffic crash (usually within a 30 day period after the occurrence of a crash) (World Health Organisation, 2018).

Road injury: Damage done to a person’s body by a sudden transfer of energy exceeding physiological tolerance caused by a road crash.

1.5 Significance of the study

In addressing road safety issues on national rural roads, it is vital for both research and practical purposes to examine and understand the relationship between the rural road environment and road traffic crashes. This understanding enables road safety stakeholders to develop adequate and efficient strategies and tools, which serve as effective and efficient proactive and remedial road safety measures.

A limited body of research exists locally and internationally on studies that examine the influence of the combination of rural road environment variables on the frequency and severity of road crashes. This study serves as one of the few investigations into the development of road crash predictive models interrogating the relationship between the combination of numerous national rural road environment conditions (design characteristics, pavement and traffic conditions) and road crashes on a macro scale in the Namibian context.

The research findings are a crucial step in providing road safety stakeholders in Namibia with a basis to develop evidence-based proactive and remedial road safety measures, and examine their effectiveness in addressing road safety issues on rural roads. The findings on rural road safety from the study will also go a long way in addressing, supporting and building on the adopted "Wear. Believe. Act. A Decade for Road Safety 2011 to 2020" strategic road safety plan in Namibia, which is aimed at highlighting high risk road crash areas, to provide for public education on road safety, stricter traffic enforcement, safer vehicle practices, safer roads and improved road crash emergency responses (Namibia National Road Safety Council (NRSC), 2012).

The insights expected from the study include identifying and evaluating the high-risk rural road traffic crash zones; assessing and addressing the deficiencies of the road design variables and road design standards used on the identified high-risk rural crash zones; and addressing the shortcomings in crash data collection and management systems. The study findings could potentially reduce the frequency and severity of road crashes by establishing effective and efficient road safety measures; making roadway improvements and providing comprehensive road safety orientated educational programs aimed at increasing the relevance of road safety research locally and internationally.

The study on the development of crash predictive tools and the influence of the rural road environment conditions on the frequency and severity of crashes on the roadway can provide a better understanding of the factors that drive this association. These can guide the road safety stakeholders to develop evidence-based targeted measures to address the road safety issues on Namibian rural roads, in line with Sustainable Development Goals (SDG 3.6[1] and SDG 11.2[2]) and the African Union (AU) Cross-Cutting Issues (6-Rural and Urban Road Safety[3]).

1.6 Study statement

The hypothesis of the study states that "A quantifiable relationship exists between the rural road environment conditions and the frequency and severity of road crashes", based on the supposition that the rural road environment design variables, traffic operational characteristics and traffic exposure variables (traffic volumes, traffic speeds, traffic conflicts and road length) have a predictable influence on the frequency and severity of road crashes.

---

[1] SDG 3.6 – By 2020, halve the number of global deaths and injuries from road traffic crashes

[2] SDG 11.2 – By 2030, provide access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, notably by expanding public transport, with special attention to the needs of those in vulnerable situations, women, children, persons with disabilities and older persons

[3] AU Cross-cutting issues (Pillar 6) – The cross-cutting issues concern transport safety in rural areas. The objective is for states to undertake rural safety audits, ensure that this data is taken into account in the design and construction of roads in rural areas, improve transport safety through mixed transport measures and sensitise rural populations to road safety issues.

1.7 Study Assumptions

The study is centred on examining historical road traffic crash data, which includes fatal and serious injury traffic crashes. The road traffic crashes sourced from the Namibian National Road Safety Council (NRSC), the Namibian Police Authority and the Motor Vehicle Accident Fund of Namibia (MVA) were not geo-coded. Due to inaccurate crash data recording and capturing, the locations of several traffic crashes on the Namibian Police road crash forms were described using landmarks close to the roadway. During geo-coding, the closest kilometre marker on the roads to the landmark mentioned in the crash forms were assumed as the crash location on the study road in this study.

1.8 Limitations and Delineations

Improving road safety is one of the important objectives for transportation stakeholders. In order to improve road safety effectively, it is vital to understand what and how factors affect road safety. This study has offered a review of current literature on road safety theory and the effect of various road factors, with a focus on the factors related to traffic characteristics (speed, traffic flow) and road characteristics (road geometry), mainly for road crashes on major roads in Namibia.

Haddon (1972) notes that the safety of road users on roadways is affected by numerous factors; human factors, vehicle factors and environmental factors, which Krug & Sharma (2009) note form the basis of the Haddon matrix in relating the sequence of events in a road crash. The study was limited to investigating the pre-crash human and environmental factors of the Haddon matrix as illustrated in Table 1.1.

Table 1.1 Factors considered for the study in the Haddon Matrix (Krug and Sharma, 2009)

		FACTORS		
PHASE		HUMAN	VEHICLES & EQUIPMENT	ROAD & ENVIRONMENT
Pre-crash	Crash prevention	*Information* *Attitudes* *Impairment* *Police enforcement*	Roadworthiness Lightning Braking Handling Speed management	*Road design & layout* *Speed limits* *Pedestrian facilities*
Crash	Injury prevention	Use of restraints Impairment	Occupant restraints Other safety devices Crash-protective design	Crash-protective road side objects Forgiving infrastructure
Post-crash	Life sustaining	First-aid skills Access to medics	Ease of access Fire risk	Rescue facilities Congestion

With regard to factors influencing traffic safety, the study has limited the crash analysis models to aspects of the rural road environment (roadway design), all aspects associated with the crash risk level and aspects related to traffic safety (road crash frequency and road crash severity).

The study parameters considered in the study models were chosen according to the level of detailed information available in the Road management System (RMS) of the Namibian Roads Authority (RA) during the study period. The parameters include the following:

1. Average annual daily traffic (AADT) (Averaged across 8 years);
2. Segment lengths;
3. Design, posted and operating speeds;
4. Traffic composition;
5. Lane widths;
6. Shoulder widths and type;
7. Horizontal and vertical curve characteristics;
8. Access management;
9. Sight distances; and
10. Pavement condition.

The behavioural aspects of the drivers on the selected roads were limited to a descriptive analysis in the study, through road traffic crash profiling and identifying factors attributed to road crashes in the historical crash data. Moreover, the study limited the GIS-based spatial analyses (Kernel Density Estimation (KDE)), descriptive and statistical analyses and modelling of historical crash data to national rural roads classifications on the Namibian road network. The national rural road network is categorised into trunk, main and district roads, guided by the Roads Authority of Namibia practices. In addition to other national rural road classifications investigated, the major national rural trunk roads on the Namibian road network are illustrated in Figure1.4.

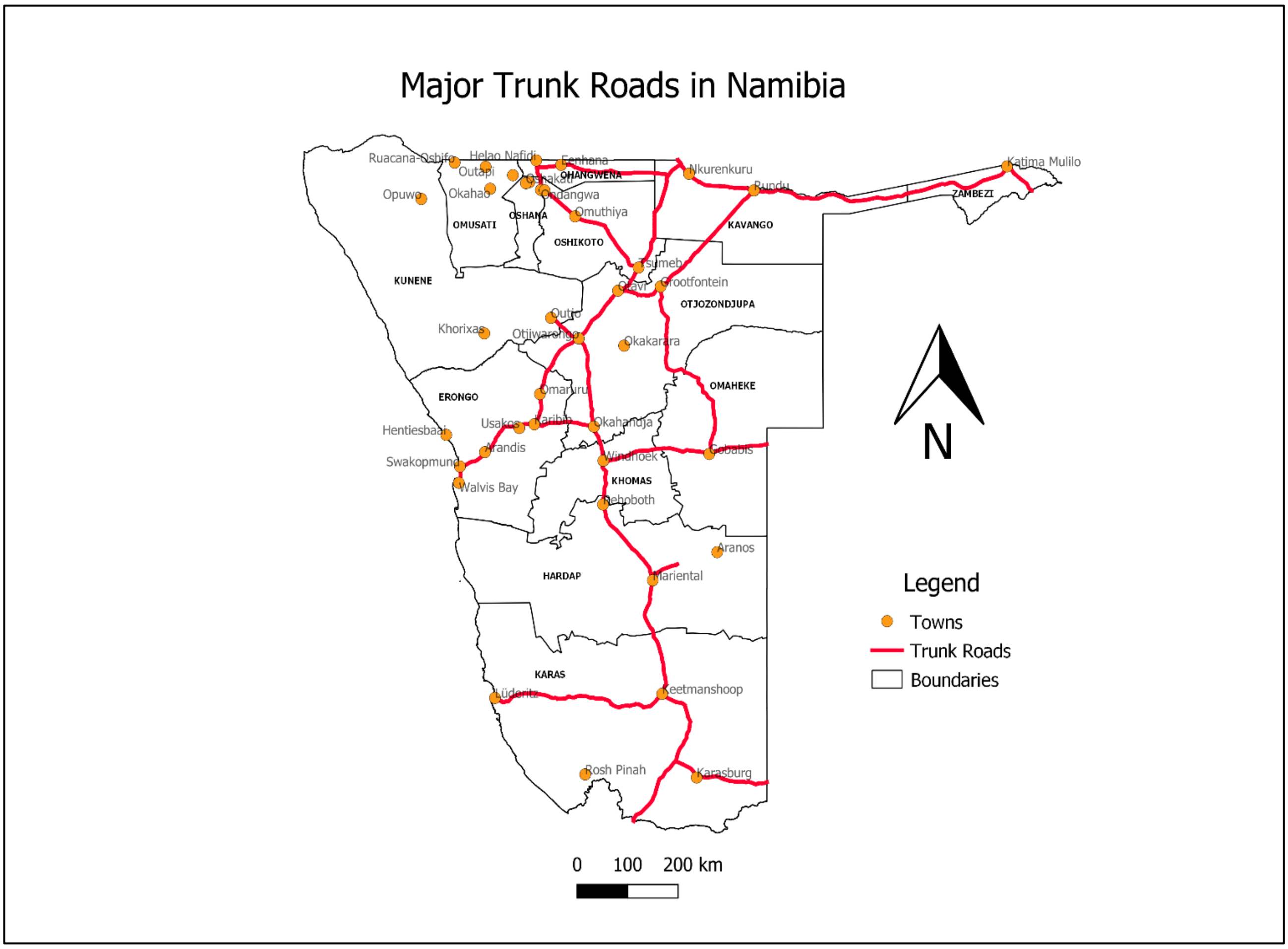


Figure 1.4 Major trunk roads in Namibia

1.9 Scope of the Study

Previous research that acknowledged and investigated the association between the rural road environment design variables the frequency and severity of road traffic crashes served as the basis to structure the study and supported the outline adopted in the study (Shalom Hakkert & Gitelman, 2014; Singh, 2017; Othman and Thomson, 2007; Dehuri, 2013). The study framework acknowledges that road crashes are the result of a complex combination of factors on the roadway, with the central factors in the framework considered in the study being the rural roadway environment (roadway design) and traffic safety on the roadway. Data on the mediators (shown in Figure 1.5) in the framework, that contribute to the crash risk level were available and considered in the study, as they were directly and indirectly linked with the design variables on the rural roadways. Figure 1.5 illustrates the conceptual framework of the study as discussed.

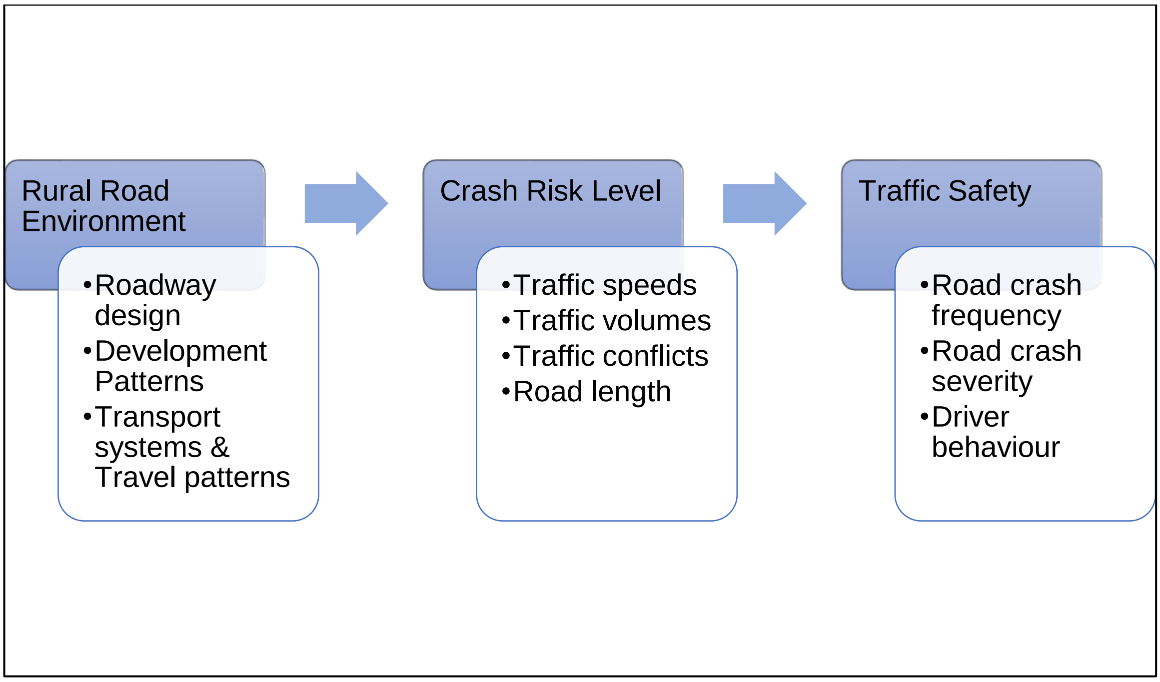


Figure 1.5 Study conceptual framework

1.10 Study Design

The nature of the study was empirical, with the safety effects of the rural road attributes and traffic conditions on road crashes examined using several methodical techniques; spatial black spot analysis and regression (GIS-based); descriptive; Analysis of variance (ANOVA); and statistical crash modelling (General Linear Regression Approach), to address the study questions and achieve the study aims and objectives. A review of existing literature was carried out to identify emerging themes in the study area, with a focus on the methods used by researchers to achieve the aims and objectives of their studies. Through the examination of existing literature, the research topic and problem statement were formulated, with research questions and study aims and objectives influenced by the problem statement.

Two methods of acquiring data necessary for the study were used, guided by research methods from relevant literature and road safety stakeholders in Namibia. Firstly, historical crash data was sourced from the NRSC, MVA and Namibian Police, whereas road geometric characteristic information was sourced from the Roads Authority of Namibia. Secondly, due to data deficiencies in the historical crash data and road geometric characteristic information sourced from the local institutions, site-specific information was collected from the selected study roads in an attempt to address the deficiencies in the data and improve data quality and reliability. The data was analysed using the numerous analysis techniques, furthermore, the results from the analysis techniques were discussed and compared to each other and to results from other studies in the same research area.

The study design is illustrated in Figure 1.6.

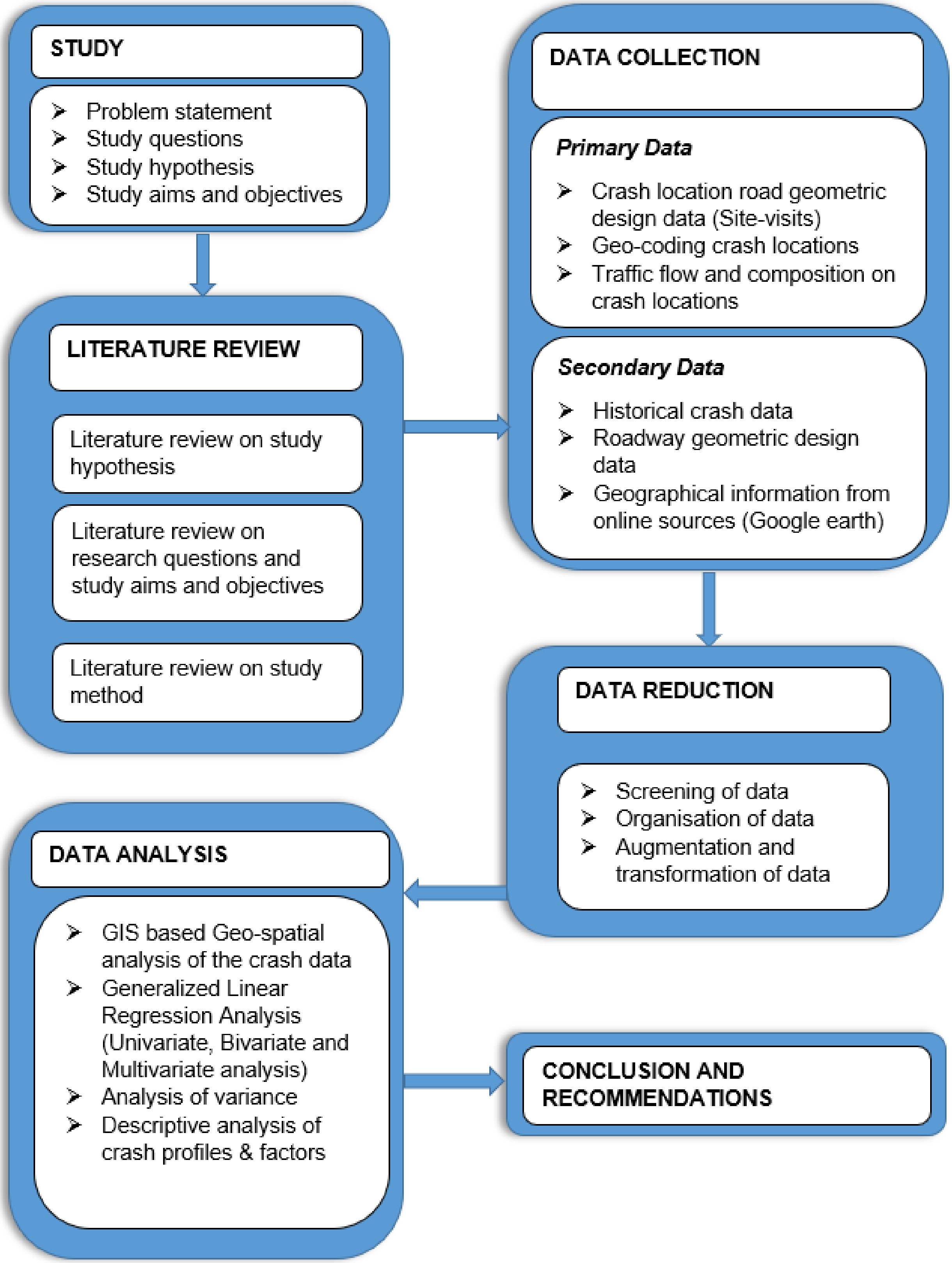


Figure 1.6 Study design

1.11 Chapter Overview

Chapter 1 introduces the study, by outlining the background, discussing the problem statement and the research questions addressed by the study, stating the delineations and limitations, the study assumption, the study procedure and detailing the aims and objectives of the study. Chapter 2 provides a comprehensive review of existing literature relevant to addressing the research questions, in order to provide a theoretical basis to achieve the aims and objectives of the study. Chapter 3 discusses and details the study procedure applied for the collection and analysis of the data, with the purpose of addressing the study questions. Chapter 4 presents and discusses in detail the findings of the study from the analyses carried out, furthermore, comparisons between the study findings and results from studies in the similar research area are discussed. The detailed discussion of the results presented in Chapter 4 is presented in Chapter 5. Chapter 6 presents the conclusions drawn from the study results, impact of the results on current and future research and practical techniques in road safety and recommendations for future research studies.

Chapter 2: Literature review

2.1 Introduction

Roadway infrastructure attributes and driver behaviour both play a significant role in road safety (Garber and Hoel, 2009). While a large proportion of the crashes are caused by driver behaviour, a significant number involve roadway factors in some way (Ahmed, 2013). The second pillar of the United Nations Global Plan for the Decade of Action for Road Safety 2011-2020 (United Nations, 2011) puts a lot of emphasis on raising the safety and protective characteristics of road networks for the benefit and safety of all road users. Knowledge of roadway parameters affecting and influencing road safety can help promote safety-conscious orientated planning, designing, building and maintaining of the road infrastructure to enable a safe road environment (World Health Organisation (WHO), 2017).

This section provides a comprehensive review of previous literature most relevant to the study and the issue of road safety on rural highways. The review of previous literature provides a background to the study questions formulated, with the aim of providing a basis to attempt to address and achieve the aims and objectives of the study. Furthermore, the literature review reviews the methodologies used in previous studies to investigate rural roadway safety, with the aim of identifying analysis techniques pertinent to realise the aims and objectives provided in Chapter 1. The literature review of the study is structured as follows:

1. Road classification
2. Road design standards
3. Road traffic safety
4. Rural-urban road crash divide
5. Road safety risk factors associated with traffic crashes
6. The impact of road design characteristics and traffic conditions on road safety
7. Road crash modelling and analyses techniques
8. Key conclusions from the literature

2.2 Road classification

In this study, it is important to distinguish between roads that can be regarded as part of the rural road network and those which are part of the urban road network for road crash analysis. The Technical Recommendations for Highways 26 on Road Classifications and Access Management –

TRH 26 (Committee of State Road Authorities (CSRA), 1988) notes that roads in rural[4] and urban[5] areas have the same six functional classes as shown in Table 2.1, however operating at different scales and standards (CSRA, 1988).

Table 2.1 Functional classes of rural and urban roads (CSRA, 1988)

Acronym	Rural Classes	Acronym	Urban Classes
R1	Rural principal arterial	U1	Urban principal arterial
R2	Rural major arterial	U2	Urban major arterial
R3	Rural minor arterial	U3	Urban minor arterial
R4	Rural collector road	U4	Urban collector street
R5	Rural local road	U5	Urban local street
R6	Rural walkway	U6	Urban walkway

A rural road is defined as a road leading through an area characterised by sparse development (CSRA, 1988). Roads that lead through urban areas, but do not have intersections and have restricted access for vehicles only are considered as rural (Through-way or a Bypass), provided their function remains that of a rural road (Archer and Vogel, 1999), as described in Section 2.2.1.

An urban road is defined as a road located within the boundaries of an urban area (CSRA, 1988). Urban roads are defined by the Swedish Institute for Transport and Communications Analysis (SIKA, 2000) according to the following requirements.

a) Roads which are often directly adjacent to large numbers of buildings where people live and work (urban areas);
b) Roads where there are numerous different types of road users (including pedestrians and cyclists) using the roadway;
c) Roads with a high density of intersections, roundabouts, pedestrian crossings, traffic control devices etc. to allow for a reasonable level of accessibility for all road users;
d) Roads where a maximum allowed speed is no greater than 60km/h, or where a higher speed limit is posted, however the density of the surrounding buildings and traffic conditions resemble those described above.

The TRH 26 notes that an urban road leaving an urban area automatically becomes a rural road, with a recommended class not lower than that of the urban area (CSRA, 1988). The urban-rural road classification changes at the boundary of the urban area, with the TRH 26 recommending the

[4] A rural area is defined as an area characterised by sparse development, mainly given over to nature or farming activities;

[5] An urban area is defined as an area that has been subdivided into erven, whether formal or informal. It includes informal settlements and areas on which townships have been formally declared. Rural settlements of one hectare or less are also included in the urban definition (CSRA, 1988).

adjusting of the road design in advance (500m) of the urban area to provide a transition area for drivers (CSRA, 1988).

In road crash modelling, Joanne (2013) notes that for a road section to be considered as rural, an average minimum threshold of 5kms on a single carriageways and 10kms on motorways from urban areas are recommended as minimum distances for rural road crash risk level assessments. Road sections shorter than 5kms from an urban area were found to show greater year on year variability in crash numbers and were likely to change risk ratings from one period to the other when compared over time (Laird *et al.*, 2010). The variances in crash numbers over time were found to be significantly high up to road section lengths of 10kms for motorways and dual carriageways (Joanne, 2013).

2.2.1 Road functionality

The road functionality principle aims at a clear distinction of roads into categories on the basis of their traffic function (SWOV, 2010). It is important to distinguish clearly the functions of different roads, and clear distinctions between roads with a through function or an access function need to be made (Karlaftis and Golias, 2009). The functionality of the roadway is vital in informing the road user the function of the road and ensuring that the road users uses the road for the purpose it was designed (Thomas *et al.*, 2013).

The Technical Recommendations for Highways 26 (TRH 26) manual on Road Classification and Access Management (Committee of State Road Authorities, 1988) classifies roads exclusively on the basis of their functions. The TRH 26 uses six –class rural and urban road classification system shown in Table 2.1, with the first three road classes[6] (Class1-3) consisting of mobility roads and the second three classes (Class 4-6) consisting of access/ activity roads. The distinctive functions of rural and urban roads are discussed in Section 2.2.1.1 and Section 2.2.1.2.

[6] Road classes' means that all public roads and paths in the country are allocated into one of six functional classes, numbered for ease of reference. Each class has a unique function to fulfil;

2.2.1.1 Rural road functionality

The main function of rural mobility roads is to connect areas that generate high volumes of traffic, typically cities, towns, airports and other mobility roads. In contrast, the main function of rural access/ activity roads is to provide access to individual properties, typically farms, mines, settlements and nature parks. In distinguishing between different road classes, the TRH 26 uses three primary criteria: Size and strategic importance of trip generator[7]; reach of connectivity[8]; and travel stage[9], in distinguishing between different road classes (CSRA, 1988). Table 2.2 shows the classification of rural roads into primary classes according to the primary distinguishing criteria.

Table 2.2 Road primary classes according to road classification criteria (CSRA, 1988)

Primary class	Trip generator	Reach of connectivity	Travel stage
Mobility roads	Large/ strategic generators	Longer travel	Through, destination not reached
Access roads	Individual properties	Shorter connections	Local, stop at destinations

The reach of connectivity criterion for rural roads is further described in Table 2.3. Note that the distances are provided on a logarithmic scale. Moreover, there is no exact cut-off between road classes as their functions can overlap.

Table 2.3 Rural road classes according to reach of connectivity criterion (CSRA, 1988)

Distance (km)		1	2	4	8	16	32	64	128	256	612	1024
Mobility	R1 (Principal arterial)							✓	✓	✓	✓	✓
	R2 (Major arterial						✓	✓	✓	✓		
	R3 (Minor arterial)				✓	✓	✓	✓				
Access/ Activity	R4 (Collector)					✓	✓	✓				
	R5 (Local road)			✓	✓	✓						
	R6 (walkway)	✓	✓	✓								

The TRH 26 (CSRA, 1988) states that it is not possible to provide an exact quantitative estimate (traffic volumes, trip length or vehicle-kilometres travel (veh-km)) to distinguish between rural road classes. However, a broad guidance on the percentage of the total of different road classes can be estimated (Federal Highway Administration, 1989), as provided in Table 2.4.

[7] A trip generator refers to a centre of development or zone from which trips originate or terminate;
[8] Reach of connectivity is an indication of the length of travel that can be accommodated on a particular road; and
[9] Travel stage describes that traveling is undertaken in three stages, local at the origin, through and local at the destination. Local in nature trips are served by access roads while through in nature trips are served by mobility roads (Committee of State Road Authorities, 1988).

Table 2.4 Traffic volume percentage of the different rural road classes (CSRA, 1988)

Rural road classes	FHWA description	% of veh-km	% of road length
R1, R2	Principal arterials	30-55	2-4
R1, R2, R3	Principal & minor arterials	45-75	6-12
R4	Collectors	20-35	20-25
R5	Local roads	5-20	65-75

2.2.1.2 Urban road functionality

The main function of urban mobility roads is to connect urban districts. Urban mobility roads should carry the traffic entering, leaving and traveling through urban areas. The efficiency of urban mobility roads is high when they serve the majority of urban travel with a minimum of road space and restricted access to individual properties (CSRA, 1988).

The primary function of urban access/activity streets is to provide access to individual properties and to accommodate traffic that is local in nature having an origin or destination along the street (CSRA, 1988). Urban access/activity streets are recommended not to serve traffic travelling through the urban area (Semar, 2003).

As with rural roads, the TRH 26 uses three primary criteria to distinguish between the primary urban road classes, namely; size of the trip generator, reach of connectivity and the travel stages, as shown in Table 2.2. An indication of the proportion of vehicle travel and linear length on urban roads is given by the Federal Highway Administration (FHA) (1989) in Table 2.5.

Table 2.5 Traffic volume percentage of the different urban road classes (FHA, 1989)

Urban road classes	FHWA description	% of veh-km	% of road length
R1, R2	Principal arterials	40-65	5-10
R1, R2, R3	Principal & minor arterials	65-80	15-25
R4	Collector streets	5-10	5-10
R5	Local streets	10-30	65-80

2.2.2 Road homogeneity

The road homogeneity principle aims to ensure relatively low variations in vehicle mass, speeds and the direction of road users, with the aim of reducing the occurrence and severity of road crashes (SWOV, 2010; Tolouei *et al.*, 2012). Homogeneity results in relatively uniform traffic flows and operating speeds. In practice, homogeneity involves the adaptation of the road environment to minimise speed variations between road users and taking measures to separate different types of road users, either physically or using traffic control devices (Ahmed, 2013; SWOV, 2010). The Dutch

Ministry of Transport (2005) notes that the following requirements for homogeneity are mainly a result of crash analyses studies.

1. Avoid conflicts with oncoming traffic;
2. Separate vehicle types;
3. Reduce speed at potential conflict points;
4. Reduce speed variations along the road segment; and
5. Avoid obstacles along the roadway.

Rural roadways are considered to be the safest roads globally, based on the number of crashes per kilometre travelled as a safety indicator (Choudhary *et al.*, 2018). Despite the higher operating speeds, rural roadways have been found to have relatively uniform speeds, with little variations in direction and vehicle mass (Dutch Ministry of Transport, 2005; Wegman & Elsenaar, 1997; Nusholtz, 2011). Urban area zones with posted speed limits between 30 and 50 km/h were found to have lower road crashes per kilometre travelled despite a considerable variation in the direction and vehicle mass (Dutch Ministry of Transport, 2005; Maqbool, 2019). Tolouei *et al.* (2012) notes that the increased safety is attributed to considerably low driving speeds and low variations in speeds between different road users.

Roads with a distributor function were found to be the most hazardous and to significantly impact the homogeneity of the roadway, due to greater vehicle mass and operating speed variations, and a considerably high amount of intersecting traffic (Eenink *et al.*, 2005; SWOV, 2010; Demasi *et al.*, 2018). Meng *et al.* (2006) found that separating motorised and non-motorised traffic, using pedestrian walkways and cycle paths, improved the homogeneity and safety of the distributor roads.

2.2.3 Self-explanatory roads (Road predictability)

Given the modest success of traditional road safety countermeasures including posted speed limits and road warning signs, Herrstedt (2015) notes that additional road safety solutions have been sought. The self-explaining road approach emerged as a road safety solution in the Netherlands in the 1990s (Theeuwes and Godthelp, 1995), centred on providing information to drivers through implicit cues (Lewis-evans and Charlton, 2006).

The self-explaining road concept is based on two cognitive psychology processes: categorisation and expectancy (Theeuwes & Godthelp, 1995 cited in Prestor *et al.*, 2014). The categories themselves must be internally consistent and mutually exclusive or at least clearly distinguishable (Mackie *et al.*, 2013). Theeuwes & Godthelp (1995) explain that road categories positively influencing driver behaviour can be achieved by assigning unique road category-defining properties, such as cues and affordances, to every road category. Weller *et al.* (2008) note that inadequate road categorisation is unsafe as it induces inadequate driver expectations.

Several studies have reported the importance of road categorisation and expectancy in a safe traffic system (AASHTO, 2010; Ambros, 2013; Edquist *et al.*, 2009; Shalom Hakkert & Gitelman, 2014). The self-explaining road concept involves designing a road system in which the driver's expectations created by the road environment are implicitly in line with the safe and appropriate driving behaviour (Ambros, 2013). Shalom Hakkert & Gitelman (2014) explain that self-explaining roads communicate to drivers the appropriate speeds to select on different road design elements and inform drivers on whether to expect traffic from access roads. Edquist *et al.* (2009) note that the speeding behaviour of the drivers may be influenced even without changing road geometry on internally consistent and clearly distinguishable roadway categories. The Highway Safety Manual (AASHTO, 2010) describes the following self-explaining roadway requirements.

1. Avoid unpredictable driver behaviour through clear road designs, signing and marking;
2. Make road categories clear and recognisable for appropriate driver speed selections and behaviour; and
3. Limit the number of design elements and provide uniformity in road design.

Similarly, Abele & Møller (2011) note that in a safe traffic system, road design should be consistent throughout the route, enabling drivers to correctly perceive the type of road and instinctively adopt their behaviour to the design and function of the road. To avoid uncertainty among road users, Hanno (2004) also states that roadways should be designed, constructed and marked to communicate the sort of behaviour expected from the drivers and for users to anticipate the behaviour of other drivers.

2.3 Road design standards

2.3.1 Road design standards: Global perspective

The Technical Recommendations for Highways 17 (TRH 17) of the Geometric Design of Rural Roads (Committee of State Road Authorities (CSRA), 1988) recognise road design standards to be vital principles to guide and control the design of the roadway. The Policy on Geometric Design of Highways and Streets (American Association of State Highway and Transportation Officials (AASHTO), 2011) reports that design standards are aimed at providing operational efficiency, safety, comfort and convenience to road users. The flexibility in the road design standards allow for localised solutions for numerous functional and operation requirements (Semar, 2003). However, Slop (1994) states that allowing space for interpretation may unintentionally lead to different road designs even in the same road area, which may cause safety issues.

Kopits & Cropper (2005) note that the unsuitability and inability of large parts of the road network to fulfil the combination of functions they are designed for plays a role in the hazardous road safety situation in various regions in the world. Pinard *et al.* (2003) argue that adopting road design standards from developed countries with the aim of addressing the precarious road safety situation in developing countries is considered a misjudgement. In developed countries, design standards are generally backed by road safety training and traffic law enforcement, which is often not the case in developing countries (Eggleston, Hansen and Carrera, 2016). Additionally, the traffic and road characteristics in developing countries differ greatly from those in developed countries (Agerholm *et al.*, 2017). Contingent on the required traffic capacity and the immediate road environment (rural or urban), whether in developing or developed countries, the design standards for individual road types are based on the following road safety principles (Dutch Ministry of Transport, 2005):

1. To prevent unintended use of the road;
2. To prevent significant speed and directional variances, thus reducing road user encounters with implicit risk; and
3. To prevent uncertainty amongst road users, through enhancing the predictability of the road design and the behaviour of other road users.

In most countries, geometric road design standards have been developed to aid transportation engineers to make sound decisions in developing efficient and safe roadways. Geometric design standards are largely underpinned on three main factors (Shalom Hakkert and Gitelman, 2014):

a) To ensure uniformity among the road design elements. This aids in making traffic conditions and road user behaviour more predictable, leading to safer road conditions.
b) To enable existing expertise in geometric design, often centred in major road authorities to be applied on a broad level; and

c) To ensure that road funds are spent satisfactorily through appropriate road designs.

Over the years, it has been assumed that that design standards and norms, as they evolved, were developed from a solid base of research, with road safety as a major consideration for the design standards and the road elements (Thomas *et al.*, 2013). However, during the past decades, the changing parameters of vehicle and the changing public attitudes have brought into question the solid foundations of the design norms (Padmanaban *et al.*, 2010).

Despite the acknowledgement of safety as a vital aspect of roadway design, empirical research necessary to establish the relationships between roadway geometry and safety are limited; sometimes contradictory, and otherwise insufficient to establish firm scientific and practically desirable relationships (Slop, 1994). Abele & Møller (2011) note that design standards that shape the road system are developed with safety in mind, but in some instances without quantitative knowledge of the link between the engineering decisions and their safety consequences.

2.3.2 Road design standards: Namibian perspective

The Southern African Development Community (SADC) Guidelines on Low-Volume Sealed Roads (Pinard *et al.*, 2003) note that the design of a road is linked to key factors, including the state of development of a road network, functional and performance requirements within the characteristics of the local road environment. Additionally, Wedajo *et al.* (2017) reports that the road geometric design philosophy varies between developing and developed countries.

The South African Pavement Engineering Manual (Rose *et al.*, 2014) notes that the road network in the SADC region provides various complex characteristics and functionalities compared with road networks in developed countries. Design guidelines orientated towards developed countries are less suited to cater for the typically low traffic volumes, and complex network and operational efficiencies in the SADC region (CSRA, 1988).

In SADC countries, as in most developing communities, there are no existing design standards that are solely based on local studies regarding safety and economic factors (Pinard, Ellis and Eriksson, 2003). The design standards used in the design of SADC roads are rather a reflection of the standards in developed countries with which SADC countries have had ties. Most of the road design standards applied in SADC are a direct interpretation of global documents, with various modifications to address operational differences and deficiencies locally (CSRA, 1988).

Within Namibia, The TRH 17 on the Geometric Design of Rural Roads (CSRA, 1988) explains that the surrounding road environment has a major impact on the level of safety road safety provided to road users through the design of roadway facilities. Due to the absence of standards designed and focused on the SADC region road setting (South African National Road Agency Limited, 2003), standards designed for developed countries have been utilised and adopted to the local conditions in the SADC region, particularly the Policy of the Geometric Design of Highways and Streets (AASHTO, 2011). Ambunda (2018) found that alterations in international road design standards are often made without fully investigating the consequences that may arise from trying to incorporate local conditions in SADC. Similarly, in Namibia, practical measures to understand the impact of road design alteration to suit local conditions are not properly addressed, leading to potentially unsafe roadways for users (Nghishihange, 2018).

The Technical Recommendations for Highways (TRH) series of guidelines, largely derived from practices in South Africa are used in the design of Namibian roads. The TRH series of guidelines were accepted by the Committee of State Road Authorities (1988) for the design and maintenance of local roadways. The TRH series is orientated towards addressing the operational and functional requirements of the South African road environment through recommending the appropriate practices for highway engineering (CSRA, 1988).

For roads that traverse through the rural environment setting, the TRH 17 on Geometric Design of Rural Roads (CSRA, 1988) is used for the geometric design of the road elements. For urban roads in South Africa; the Urban Transport Guidelines (UTG) series; namely UTG 1, UTG 5 and UTG 7 are utilised for the geometric design of safe roads traversing through built up areas.

2.4 Road traffic safety

2.4.1 Road traffic crashes

A road traffic crash is defined as a collision or incident that may or may not result in an injury, occurring on a roadway and involving at least one moving vehicle (Peden *et al.*, 2017). Road traffic crash history is a key indicator of the safety performance of a road section (Hagenzieker *et al.*, 2014). Numerous techniques aid in determining the performance of road sections and serve as tools to analyse crash data, with the aim of identifying section with a greater need for safety improvements (Baguley *et al.*, 2006; Hyldekær & Giacomo, 2016). In this regard, different methods of using historical crash data to conduct road network screening and assess safety performance are discussed in the following sections, highlighting their benefits and drawbacks. The methods discussed are:

1. Crash frequency;
2. Crash rate; and
3. Critical rate.

2.4.2 Road safety performance indicators

2.4.2.1 Crash frequency

Cenek *et al.* (2012) defines crash frequency as a frequency-based method of identifying and evaluating the safety performance of a site, which has traditionally been used by transportation engineers, and is still used by most road safety stakeholders. The crash frequency determined from historical crash data over a certain period of time can be used for the purpose of comparing and ranking the safety performance of different locations (Chen *et al.*, 2016), at different injury severity levels (Cenek *et al.*, 2012; Mannering & Bhat, 2014; Sisiopiku, 2011).

Road crashes are relatively random events, such that a high crash frequency in any given year may simply be a random fluctuation around a much lower long-term crash average on a site, experiencing a phenomenon known as regression towards the mean (Choi *et al.*, 2019; Demissie, 2017). Therefore, relatively shorter periods of analysis are not recommended as the basis for a safety intervention (Thomas *et al.*, 2013). The Highway Safety Manual (AASHTO, 2010) recommends using data collected over a period of 3-5 years for a safety analysis, to minimise the effects of regression to the mean and unusual traffic activity (e.g. road reconstruction/ maintenance).

In addition to experiencing the phenomenon of regression to the mean, sites with higher traffic volumes typically have higher crash frequencies than sites with lower volumes (Kockelman, 2006). This does not always mean that a site is in need of safety improvements, as crash sites with higher traffic volumes more often than not have lower severity risks compared to low volume crash sites

(Vayalamkuzhi and Amirthalingam, 2016). Also, the availability of reference group crash frequencies can be problematic in analysing the safety of crash sites, as crash causation factors are ascribed to a certain grouping of road users only (Kassu & Anderson, 2018; Mannering & Bhat, 2014; Massie & Campbell, 1993).

2.4.2.2 Crash rate

The crash rate technique improves upon the average crash frequency in that it normalises the frequency of crashes against exposure (Ambunda, 2018; Cenek *et al.*, 2012). The road crash rates are determined by dividing the total crash frequency for a period of time by the estimated average annual daily traffic (AADT) during the time period investigated (Othman *et al.*, 2009; AASHTO, 2010). Crash rates provide an improved method to compare the safety of two different sites (Gaudry and Vernier, 2002).

Investigating sites with different traffic volumes requires the assumption that traffic volume and crash frequency have a linear relationship, which is often the case (Bamdad Mehrabani & Mirbaha, 2018; Taylor *et al.*, 2002). Earlier studies using crash rates to identify hazardous road crash sites did not consider crash severity (Jones, 1976; Ogden, 1994). Sites with higher crash rates may have fewer severe crashes (fatal and serious injuries). To identify location with higher crash severities, recent studies have used fatal and serious injury crash rates to rank sites with a greater need for safety interventions (Ambunda & Sinclair, 2019; Zimmerman *et al.*, 2012; Wang *et al.*, 2009).

The crash rates are determined from crash frequency, which fluctuates around a long-term average and experiences a regression towards the mean; For example, a site with an unusually recent period of high crash numbers might rank high compared to sites with an average higher number of crashes (AASHTO, 2010). To counter this phenomenon, similar to the crash frequency method, a study period of 3-5 years is recommended by the Highway Safety Manual (AASHTO, 2010) for a safety analysis.

2.4.2.3 Critical rate

The critical crash rate method is a widely used robust network screening technique, in which the calculated crash rate for a location is compared with a critical crash rate unique to each location (AASHTO, 2010). The critical crash rate is a function of the average crash rates of a reference group of locations with similar characteristics to the study being investigated (King, 2014). King (2014) states that the critical crash rate method provides a means of statistically testing how the crash rate at a particular location or section varies when compared to a reference group. The Highway Safety Manual (AASHTO, 2010) notes that locations with crash rates exceeding the critical crash rate warrant a further detailed investigation in the diagnosis[10] step of the road safety management process.

Similar to the disadvantages of crash frequency and crash rate methods, the critical rate works with the assumption that road crashes and traffic volumes experience a linear relationship (AASHTO, 2010). Also, the use of reference groups of critical rates can be problematic in examining the safety of roadways (Kassu & Anderson, 2018; Mannering & Bhat, 2014). Moreover, the regression to the mean is not addressed in this method (AASHTO, 2010; King, 2014).

2.4.3 Road traffic safety: Global perspective

Road safety remains one of the most significant issues globally, with current trends suggesting that it will continue to be the case in the foreseeable future, with estimates by the WHO (2017) indicating that road crashes kill over 1.2 million people annually and injure up to 50 million. Traffic safety has developed into a major concern in developing countries, with low- and middle-income countries (LMIC) reported to account for 90 percent of the road traffic crashes worldwide, while having only 48 percent of the world's vehicle population (Peden *et al.*, 2017). Figure 2.1 illustrates the stark differences in the road traffic mortality rate among the different regions of the world. Globally, the average fatalities per 100 000 population are less than 9 in HIC, while LMIC have an average road fatality rate of 20, with the African region demonstrating the highest road fatality rate of 26.6 fatalities per 100 000 population (World Health Organisation (WHO), 2018).

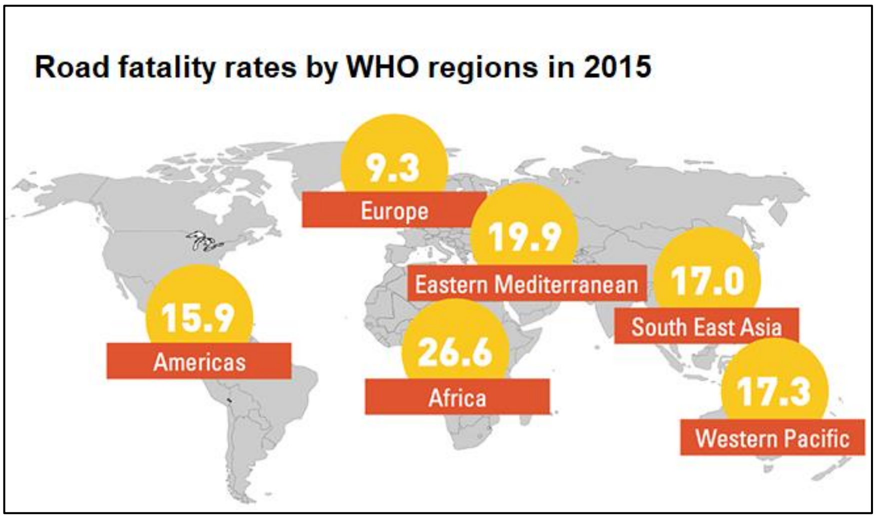


Figure 2.1 Road fatality rates in the various WHO regions in 2015 (WHO, 2018)

While traffic safety has been improving in high-income countries (HIC), road safety trends indicate that road fatalities are forecasted to rise to almost 2 million road fatalities annually by 2020 in LMIC only (Wegman, 2017). Projections of future traffic fatalities suggest that the global road death toll will

[10] The diagnosis step involves traffic engineers correctly diagnosing the type of safety problem on a road section/ location through reviewing the crash data, assessing field conditions and defining a problem statement (Rogers, 2003).

grow by approximately 66 percent between the years 2000 and 2020 (Kopits and Cropper, 2005). South Asia is estimated to have the highest change in the rate of fatalities over the 20-year period, with a 140 percent predicted. In contrast, countries in the high-income bracket are estimated to have a 24 percent decrease in the rate of fatalities over the same period. According to Kopits & Cropper (2005), Sub-Saharan Africa is estimated to have a growth of 70 percent in the rate of fatalities over the 20 year period. Figure 2.2 illustrates the average change of rates in the different regions of the world.

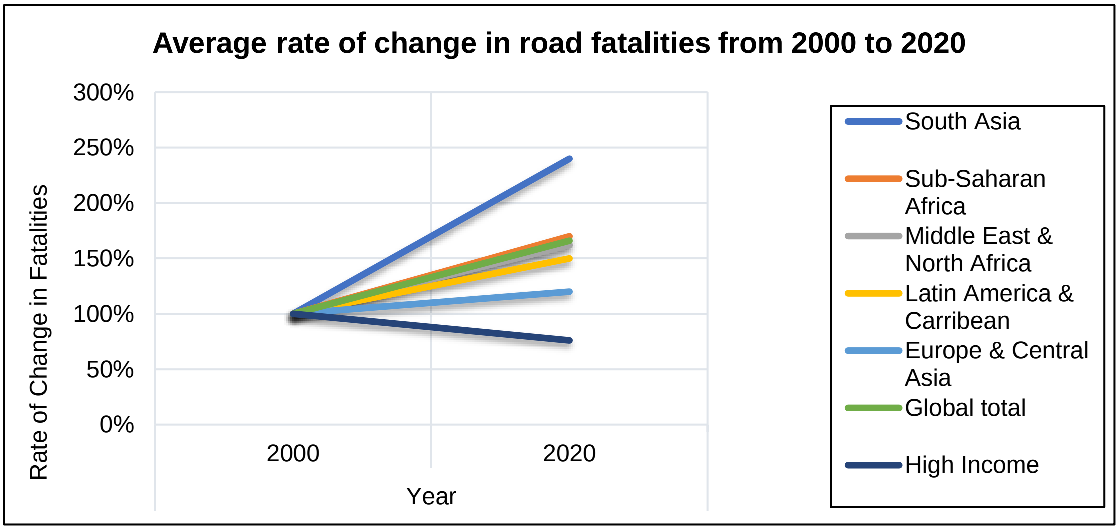


Figure 2.2 Predicted average rate of change in road fatalities from 2000 to 2020 (Kopits and Cropper, 2005)

Additionally, there is a stark difference in the road mortality rate in the types of road users in the different regions of the world (WHO, 2018). Figure 2.3 shows that the Africa has the highest risk for non-motorised users (NMU), with 43 percent of all road related deaths involving NMUs, compared to 19 percent globally. In contrast, the lowest proportion of NMUs deaths is reported in the Americas, with 16 percent of all road fatalities. The highest proportion of motorised users' (MU) deaths is reported in the European region, with 62 percent of all road related deaths, which is above the global average of 54 percent. In Africa, MU deaths account for half (50 percent) of all the road related deaths (WHO, 2018). The lowest proportion of MU deaths is reported in South-East Asia, with 48 percent of all road deaths in the region.

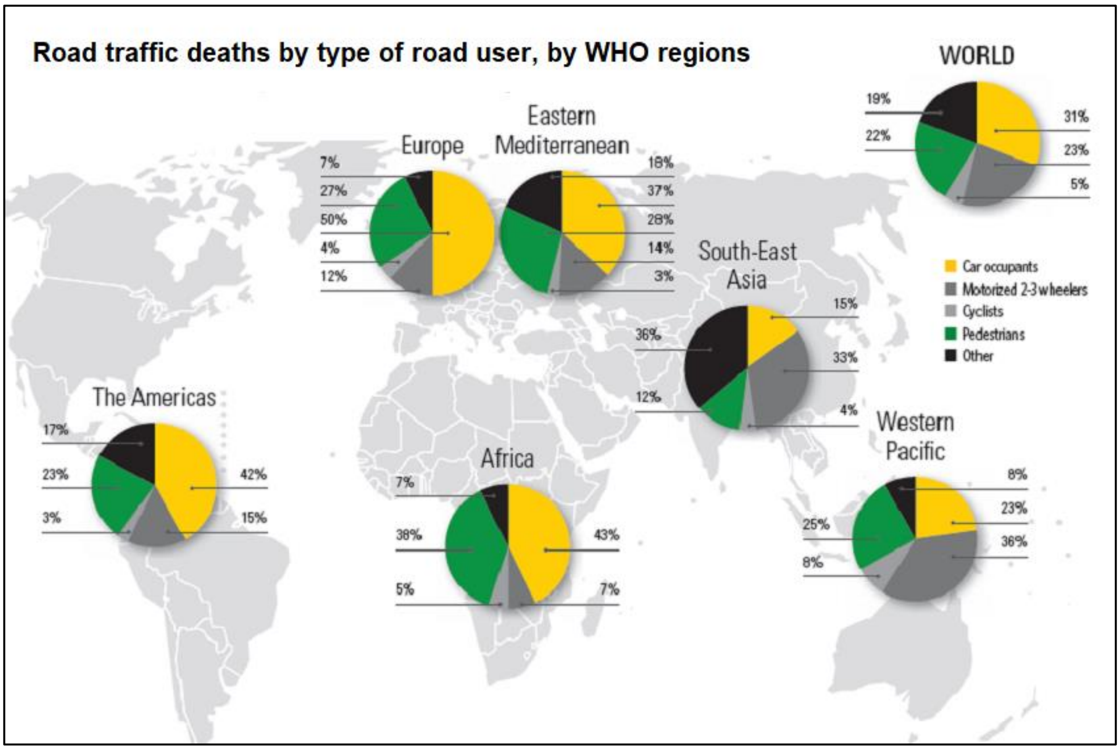


Figure 2.3 Road traffic deaths by type of road user in various WHO regions 2015 (WHO, 2018)

The future of road safety is uncertain and definitely not the same for all regions of the world (Singh, 2017). It is therefore important that LMIC work towards designing road safety strategies and implementing actions plans that align with local conditions, instead of adopting the road safety approaches taken by HIC which have different road environment and traffic conditions (Wegman, 2017).

2.4.4 Road traffic safety: Namibian Perspective

The road traffic safety situation has been a cause for major concern in Namibia in recent years (Ambunda and Sinclair, 2019). Road traffic related deaths are reported as one of the leading causes of death in Namibia, reaching approximately 4 percent of the total deaths in the country (Namibia National Road Safety Council (NRSC), 2012). A study by Eggleston *et al.* (2016) reported that the number of fatalities increased by 34% from 2011 to 2014. Statistics by the Namibian Statistics Agency (2015) reported that at the end of 2015, the road crash fatality rate on Namibian roads significantly rose above the African continental average of 26.6 fatalities per 100 000 population, with 31.11 road fatalities per 100 000 population. A study by the World Health Organisation (2015) identified Namibia as one of the countries with a precarious road safety situation, ranking 45th out of the 185 countries assessed. Amweelo (2016) reports that 70 percent of road crashes in Namibia are reported in built-up environments, with a high proportion of slight injury crashes. In contrast, a high proportion of fatal and serious injury crashes are reported on rural roadways, despite crashes on rural roadways comprising 30 percent of crashes on the road network.

A study conducted by the Namibia National Road Safety Council (2012) on the historical crash data, with the aim of monitoring the level of safety over several years (2002-2012), identified a number of primary indicators to measure the risk of exposure for road users on the Namibian road network, with the aim of making meaningful comparisons and establish road safety trends. Table 2.6 shows the absolute variations in numbers in the road safety condition from 2002 to 2012. The number of road traffic crashes steadily increased from year to year, with an average yearly growth of 5.33 percent. The number of fatalities vary from year to year, with the highest number of fatalities (549) recorded in 2010. The highest number of injury crashes (2 585) were recorded in 2011 according to the NRSC (2012).

Table 2.6 Road crash statistics from 2002 to 2012 in Namibia (NRSC, 2012)

Year	Road Safety Numbers								
	Crashes	Number of Vehicles Involved	Injury Crashes	Fatalities	Serious Injuries	Slight Injuries	Registered Vehicles [11]	Vehicle Kilometres Travelled (VKT) [12]	National Population
2002	10 915	17 708	2 125	508	1 396	3 053	180 342	4 722 048 700	1 860 145
2003	10 957	17 838	1 956	478	1 243	1 801	192 321	4 795 168 400	1 891 097
2004	10 262	17 074	1 763	491	972	2 480	204 460	5 089 239 800	1 923 347
2005	11 146	18 257	1 834	452	1 023	2 572	218 140	5 343 794 700	1 956 899
2006	13 396	19 870	1 248	530	795	1 991	232 348	5 747 261 300	1 991 746
2007	13 720	20 247	2 053	452	1 125	2 467	239 885	5 929 692 400	2 027 870
2008	13 825	21 710	2 279	459	1 822	2 991	213 939	6 409 643 700	2 065 224
2009	15 537	24 433	2 537	525	1 988	3 089	229 908	7 141 761 800	2 103 762
2010	17 387	24 817	2 570	549	2 088	3 131	249 421	7 969 687 101	2 143 411
2011	17 835	25 337	2 585	492	2 264	3 395	269 907	8 085 571 000	2 113 077
2012	17 892	25 189	2 461	572	2 596	3 172	280 583	8 271 980 501	2 155 440

Table 2.7 shows the results of the road safety risk indicators monitored to establish the level of safety on the Namibian road network during the period 2002 to 2012. According to the NRSC (2012), crashes per 100 000 population have increased steadily over the study period, with an average growth of 3.78 percent annually. In contrast, fatalities and injuries per 10 million VKT have steadily decreased, with a 3.89 percent and 1.23 percent average decrease annually. The fatalities per 100 000 population vary over the period under consideration, with the highest fatality rate (27.3)

[11] Registered vehicles information are obtained from the Namibian Roads Authority's Traffic Information System for 2012

[12] The national population figures are projections based on the calculations contained in the National Population Census Main Report 2011 (Namibia Statistics Agency, 2011).

recorded in 2002. Similarly, injuries per 100 000 population vary over the study period, with the highest injury rate (122.3) recorded in 2011.

Table 2.7 Road safety risk indicators from 2002 to 2012 in Namibia (NRSC, 2012)

Year	Road safety Risk Indicators (Rates of Comparison)						
	Crashes/ 1000 vehicles	Crashes/100 000 population	Injury crashes/ 100 000 population	Fatalities/ 100 000 population	Injuries/ 100 000 population	Fatalities / 10 million VKT	Injuries per 10 million VKT
2002	60,5	586,8	114,2	27,3	239,2	1,08	9,4
2003	57,0	579,4	103,4	25,3	161,0	1,00	6,3
2004	50,2	533,5	91,7	25,5	179,5	0,96	6,8
2005	51,1	569,6	93,7	23,1	183,7	0,85	6,7
2006	57,7	672,6	62,7	26,6	139,9	0,92	4,8
2007	57,2	676,6	101,2	22,3	177,1	0,76	6,1
2008	64,6	669,4	110,4	22,2	233,0	0,72	7,5
2009	67,6	738,5	120,6	25,0	241,3	0,74	7,1
2010	69,7	811,2	119,9	25,6	243,5	0,69	6,5
2011	66,1	844,0	122,3	23,3	267,8	0,61	7,0
2012	63,8	830,1	114,2	26,5	267,6	0,69	7,0

A study entitled “Enhancing the road safety situation in Namibia” by the Legal Assistance Centre (2016) notes that the issue of road safety on Namibian roadways is undoubtedly a cause for concern, and it is vital that the solutions recognise the potential road safety improvement measures available on the regional and international levels, which may be considered and revised to address the local road safety conditions.

2.5 Rural-urban road crash divide

Numerous spatial examinations of road crashes have found that road safety problems are hardly uniform over space (Loo *et al.*, 2011; Satria and Castro, 2016; Imprialou *et al.*, 2016). In fact, the rural-urban divide in road safety has been recognised worldwide. Several studies have found rural roads to be susceptible to higher severity crash rates compared to urban roads (Bayliss, 2009; Godavarthy and Russell, 2016; Loo *et al.*, 2011; Kassu & Anderson, 2018). The outcome of road crashes on rural roads are usually more severe as a direct result of higher operating speeds, low traffic volumes and less restrictions by the operating environment (Godavarthy and Russell, 2016), despite the low number of road crashes occurring on rural roads compared to urban roads (Bayliss, 2009). In contrast, the severity of crashes on urban roads is usually lower because of the greater limitations imposed on speed by the traffic conditions and traffic control measures (Kassu and Anderson, 2018), notwithstanding the higher occurrence of road crashes in urban areas (Bayliss, 2009). The high number of crashes on urban roads has prompted road safety researchers to focus on urban roads, at the expense of rural roads which tend to have greater severity risks (Vayalamkuzhi and Amirthalingam, 2016).

An early crash data analysis study by Robinson (1984) found that approximately two-thirds of fatal road crashes occur on rural roads while more than half to three-quarters of injury road crashes occur on urban roads. Also, a study by the Organisation for Economic Co-Operation and Development (OECD) (2003) on road safety in Austria, France, Germany, Italy and the United Kingdom found that 50 to 75 percent of road crashes causing low severity injuries happened on urban roads. The OECD (2003) also found that more than 60 percent of fatalities in road crashes happened on rural roads.

A study by Bayliss (2009) found that the proportion of serious and fatal crashes increased on rural roads compared to urban roads, despite an overall decrease in the occurrence of road crashes between 1972 and 2007 in Europe. Figure 2.4 shows that a higher proportion of fatal and serious road crashes occurred on rural roads compared to urban roads in the historical crash data analysed by Bayliss (2009). The fatal crashes on rural roads comprised of 57 percent of the analysed crash data compared to 43 percent on urban roads in 1972. Fatal crashes increased to 64 percent on rural roads while the proportion of fatal crashes on urban roads decreased to 34 percent of the crashes analysed in 2007. The rural-urban serious road crash divide also showed a similar trend. Serious rural road crashes comprised of 55 percent of all the crashes compared to 45 percent on urban roads in 1972. In 2007, the proportion of serious rural road crashes increased to 57 percent while the proportion of serious road crashes on urban roads slightly decreased to 43 percent. Singh (2017) notes that higher operating speeds on rural roads and improvements in the safety of urban roads are key factors contributing to the varied proportion of fatal and serious crashes on rural and urban roads.

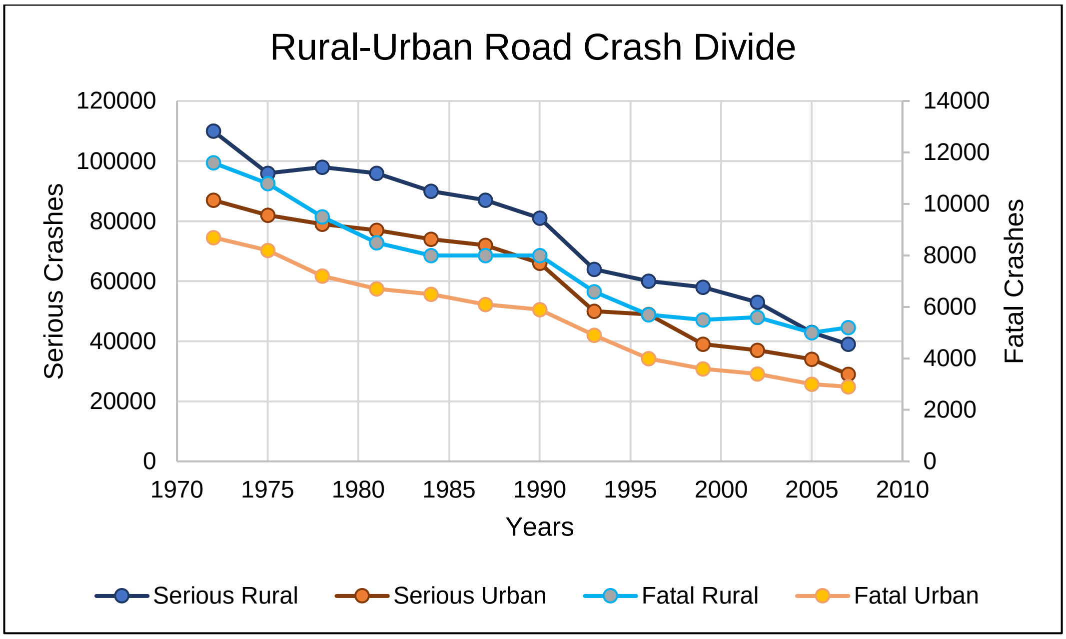


Figure 2.4 Fatal and serious road crashes reported between 1972 and 2007 (Bayliss, 2009)

Hakkert & Braimaister (2002) note that due to generally shorter travel distances and high traffic volumes on urban roads, a great number of people are encouraged to use bicycles or to walk to their destinations. Several road crash types occur on urban roads, with a great proportion of crashes occurring at intersections (Kassu and Anderson, 2018). Drivers on urban roads are also often involved in a great number of rear-end and turning crashes (Archer and Vogel, 1999). The urban roads are regarded as more complex compared to rural roads, due to vastly varied road user types and needs, and higher physical and mental demands placed on drivers, reflected in the high occurrence of crashes on urban roads (Bayliss, 2009).

Driver exposure, longer travel distances and higher operating speeds, are more common problems on rural roads than on urban roads (Shibani, 2016). Several driver behaviour and design factors have been identified as key factors influencing the occurrence of various crash types on rural roads, including distracted driving, unsafe passing behaviour, narrower road lanes, lack of physical traffic separation and poor lighting conditions (Shalom Hakkert & Gitelman, 2014; Amarasingha & Dissanayake, 2015; Yan *et al.*, 2012). Due to longer travel times on rural roads, drivers tend to get fatigued and become inattentive (Godavarthy and Russell, 2016), leading to a high occurrence of single-vehicle run-off road crashes on rural roads (Liu & Subramanian, 2009; Amarasingha & Dissanayake, 2015). Drivers on rural roads are also often involved in sideswipes and head-on road crashes, due to human errors (distracted driving and reckless overtaking), narrower road lanes and a lack of physical separation for opposing traffic (Yan *et al.*, 2012).

2.6 Road safety risk factors associated with traffic crashes

Road crashes are often caused by a combination of factors; human factors; roadway factors; and vehicle factors (Munteanu *et al.*, 2014).The promotion of road safety should be a priority for every road authority and safety stakeholder (Wretstrand *et al.*, 2014). Attention is generally focused on areas where a relatively high number of road crashes occur (Kundakci, 2014). Othman *et al.* (2009) state that measures designed to tackle the concentration of crashes should be based on a thorough and objective analysis of the causation factors. Understanding factors that influence the occurrence of crashes is vital in developing a proactive attitude towards avoiding situations that can create a hazardous road safety environment (Wegman, 2017; Dutch Ministry of Transport, 2005).

A concept of Sustainable Road Safety with a vision orientated towards safer road traffic systems was developed in the Netherlands (Wegman and Elsenaar, 1997). The sustainable safety concept aims to avoid burdening the future generation with the consequences of road traffic crashes that may arise from current and future mobility demands. Sustainable safety is based on a systematic approach where all road safety factors and the transport system are linked, and affect the performance of the whole safety system (Wegman, 2017). At the highest level of the safety system is the interaction between the driver, the roadway environment and the vehicle factors. At the next level is the relation between the function[13], form[14] and usage[15] of the roadway (Dutch Ministry of Transport, 2005). The systematic approach to Sustainable Road Safety is illustrated in Figure 2.5.

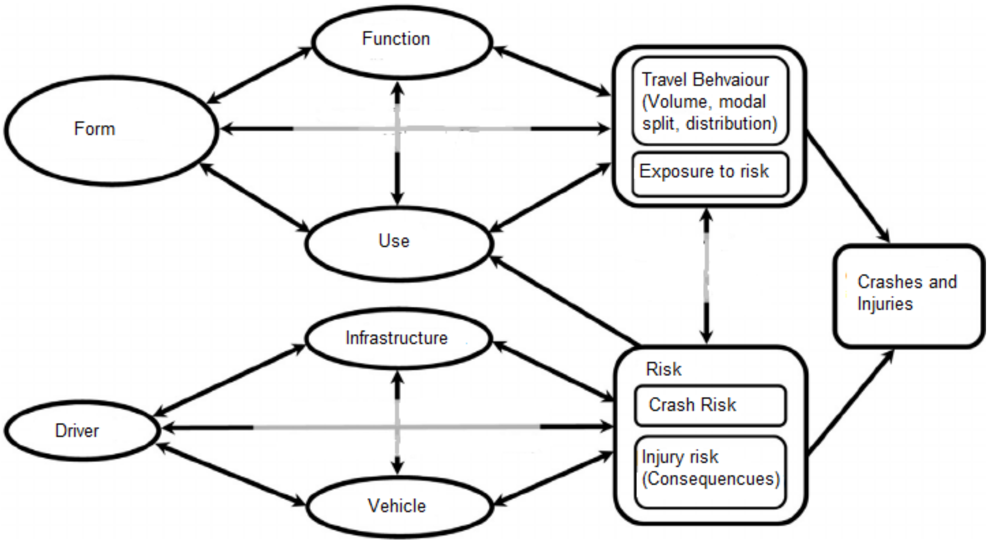


Figure 2.5 Systematic approach to Sustainable Road Safety (Dutch Ministry of Transport, 2005)

[13] Function related to the use of the roadway as intended by the road authority;

[14] Form related to the geometric design and layout characteristics of the roadway;

[15] Usage relates to the actual use of the roadway, the behaviour of the road users and the legislation relating to the requirements on the use of the roadway (Committee of State Road Authorities, 1988).

It is generally agreed that road safety risk factors should be investigated and understood towards delivering a safe and sustainable approach to road safety issues, with the aim of providing pro-active preventative measures addressing the functionality, homogeneity and predictability functions of road safety (Discussed in Section 2.2), as opposed to post-intervention measures. This section will examine the interactive relationship between risk factors that influence crash involvement; namely the behavioural aspects of the road user, the roadways environment; and the vehicle factors. The risk factors identified in the study and the Two-Step Cluster analysis technique applied to explore their correlation to crash occurrences are described in Chapter 3.

2.6.1 The Swiss cheese model of road crash causation

The Swiss Cheese Model (SCM) by James Reason in 1990, on the systems perspective theory of human error, focuses on the interaction between system wide inadequacies and errors and their influence on organisational failures (Reason, 1990). The SCM has served as one of the most central models in the explanation of road crash causation factors, as it considers a multilayer description of the complex factors and systems that contribute to road crashes (Hughes *et al.*, 2015; Grant *et al.*, 2018).

The SCM illustrated in Figure 2.6, describes the weaknesses that are created in the system's defences by the different stakeholders at different levels . These system inadequacies occur due to inaction and/ or inapt decisions by various stakeholders. The weaknesses predispose the users of the system to high crash risks, due to the accumulation and alignment of a multitude of risk factors – the holes in the cheese (Zhang *et al.*, 2018; Adanu *et al.*, 2019; Venter, 2019).

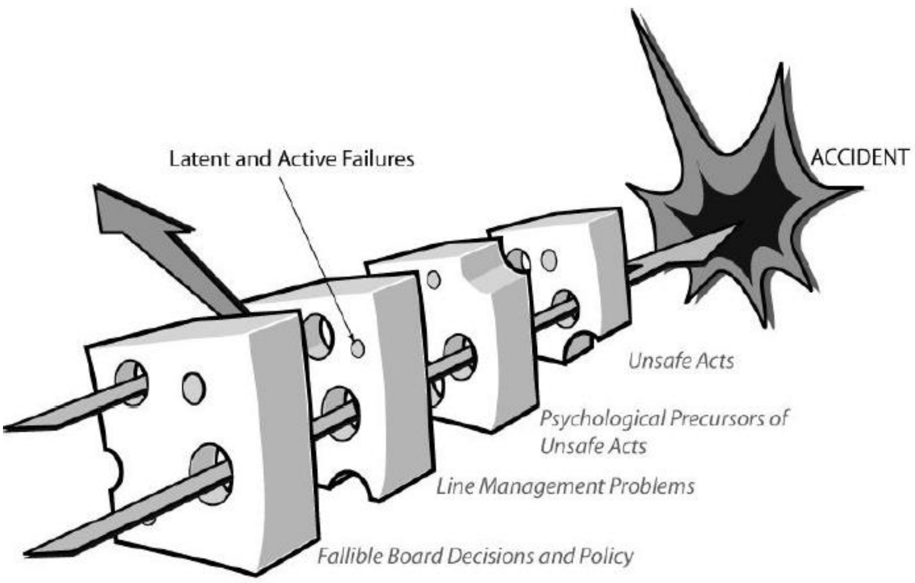


Figure 2.6 Swiss cheese model of road crash causation (Salmon and Johston, 2013)

The SCM presents a vivid and simpler description of the failures that can occur in a transport system. The failures in the interactive factors that constitute a road traffic system can impact how road users perceive the road environment, the behaviour and execution of tasks on the road and the planning and decision making of different stakeholder in road safety (Salmon and Johston, 2013;Afghari, 2018).

2.6.2 Driver behaviour

In both developed and developing countries, the behaviour of drivers has been recognised as an important risk factor associated with road traffic crashes (Peden *et al.*, 2017), with an estimated 90 percent of road crashes involving human error to a certain extent (Čičković, 2016). Peden et al. (2017) cited in Demissie (2017) states that among the risk factors considered to significantly impact the frequency and severity of road crashes are driver age, gender, public safety education, driver fatigue, socio-economic status and propensity for speeding. In an attempt to address and effectively improve the road safety situation, it is vital to understand the human factors associated with driving safety (Bax *et al.*, 2014).

2.6.2.1 Driver age

A report by the World Health Organisation (2018) indicated that road crash injury is the leading cause of death for young adult drivers between 15 and 44 years, accounting for 59 percent of global road traffic fatalities. Across the world, young drivers have a higher crash risk than older drivers. Literature from developed countries has shown that even with corrected exposure factors, young men have higher crash rates involvement than women (Parizel & Phillips, 2004; Butchart & Mikton, 2014). The elevated crash risk for young drivers was reported to be related to the following factors:

a) Mobility patterns and vehicle characteristics (For example, using a borrowed vehicle)
b) Psychological characteristics (over-confidence or thrill-seeking)
c) High blood alcohol concentration levels
d) Excessive or inappropriate speeds

2.6.2.2 Driver gender

The Global Status Report on Road Safety by the World Health Organisation (2018) indicated that approximately 77 percent of road traffic fatalities occurred among men. Road traffic fatality rates are reportedly higher in men than in women in all WHO regions globally across all age groups and income levels. The huge variation in fatality rates are significantly related to high exposure levels and thrill-seeking behaviour among men than in women (Peden *et al.*, 2017).

2.6.2.3 Driver speed

Speed is reported to be at the core of road safety, with both excessive[16] and inappropriate[17] speeds leading to unsuitable conditions on the roadway (World Health Organisation, 2018). Excessive driver speed has been found to have an exponentially detrimental effect on safety (Ahmed, 2013; Singh *et al.*, 2004; Deublein *et al.*, 2013). The NRSC (2012) found that driver excessive speeds and driving errors such as single-vehicle crashes and overtaking errors contributed to more than 63 percent of crashes on Namibian roads. A report by Peden *et al.* (2017) noted that the likelihood of road crashes and severity levels increased with higher average operating speed and speed variance. Foss & Goodwin (2003) identified numerous factors that significantly influence driver speed selections, as shown in Table 2.8.

Table 2.8 Example of factors influencing driver speed selection (Foss and Goodwin, 2003)

Road and vehicle related	Traffic and environmental related	Driver related
Road	***Traffic***	Age
Width	Density	Sex
Gradient	Traffic composition	Reaction time
Alignment	Prevailing speed	Attitudes
Surrounding	***Environment***	Thrill-seeking
Layout	Weather	Risk acceptance
Markings	Surface condition	Hazard perception
Surface quality	Natural light	Alcohol level
Vehicle	Road lighting	Ownership of vehicle
Type	Signs	Circumstances of trip
Power/weight ratio	Speed limit	Occupancy of vehicle
Maximum speed	Traffic enforcement	-
Comfort	-	-

2.6.2.4 Alcohol use

Several reports and studies report that drinking and driving increases the risk and likelihood of fatal and serious injuries resulting from a road crash (Peden *et al.*, 2017; Brookhuis, 2014; Schulze & Koßmann, 2010). Similarly, Shinar (2007) reports that alcohol impairment is directly related to the amount of alcohol consumed. The World Health Organisation (2018) reports that the risk of being involved in a crash increases significantly when the blood alcohol concentration (BAC) is above 0.04g/dl (WHO, 2018). The relative risks of involvement in fatal crash for BAC levels illustrated in

---

[16] Excessive speed is defined as vehicle speed exceeding the relevant speed limit;

[17] Inappropriate speed refers to vehicles travelling at a speed unsuitable for the prevailing road and traffic condition (Peden *et al.*, 2017).

Figure 2.7, in a study by Compton *et al.* (2002) cited in the Global Status report on Road Safety (WHO, 2018).

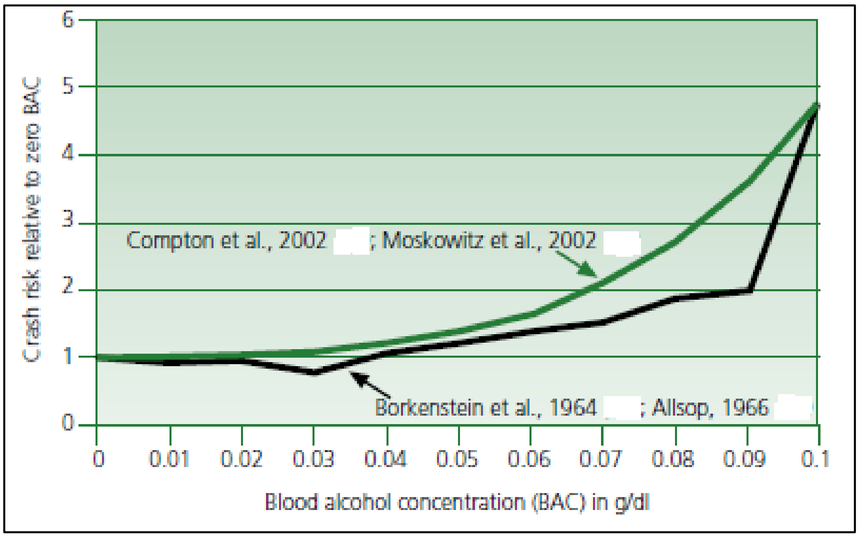


Figure 2.7 Relative risk of driver involvement in road crash in relation to blood alcohol concentration levels (Compton *et al.*, 2002)

2.6.2.5 Drug use

Driving under the influence of drugs has increasingly become a safety issue in many countries globally (Li *et al.*, 2013; Verstraete *et al.*, 2014; Peden *et al.*, 2017). Li *et al.* (2013) notes that driving performance can be impaired by a wide array of illicit and prescription drugs. Several studies have found marijuana as the most frequently detected drug substance in the general driver population and in drivers involved in road crashes (Ul *et al.*, 2014; Li *et al.*, 2013; Compton & Berning, 2015; Jones *et al.*, 2003). Marijuana has been found to double the risk of road crash occurrence (Asbridge *et al.*, 2012; Li & Baker, 2012),by impairing driver cognitive functions and driving performance, such as psychomotor skills, driver divided attention and lane tracking (Arria *et al.*, 2011; Kelly *et al.*, 2004; Hartman & Huestis, 2013).

Benzodiazepines have also been frequently detected in drivers (Carfora *et al.*, 2018; Kelly *et al.*, 2004) and have been consistently found to significantly increase the risk of road crash involvement and crash culpability (Carfora *et al.*, 2018; Li *et al.*, 2013). Several studies have found that nonmedical stimulants pose a threat to driving safety, when used in high doses, in combination with alcohol or other drugs, or with lack of sleep (Kelly *et al.*, 2004; Ramaekers *et al.*, 2012).

2.6.2.6 Distracted driving

Numerous factors have been identified to influence impaired driving, with a recently marked increase in the use of mobile phones becoming a growing concern among road safety stakeholder. A Global Report on Road Safety by the WHO (2018) concluded that using mobile phones while driving results in cognitive distraction, consequently reducing drivers alertness and perceptual skills . Drivers using mobile phones while driving are approximately four times more likely to be in a road crash (Karlaftis and Golias, 2009). Another study described that mobile phones impair the drivers cognitive whether they are used in a hand-held or hand-free manner (Thomas *et al.*, 2013).

2.6.2.7 Fatigued driving

Fatigue is defined as a gradual and cumulative process associated with a loss of efficiency and a disinclination for any kind of effort (Grandjean, 1979 cited in Dagli, 2004). Zhang *et al.* (2016) note that fatigue increases as time-on-task progresses. Thus, Dagli (2004) defines driver fatigue as drivers' loss of efficiency to drive a vehicle due to prolonged driving, sleep deprivation and exhaustion. Sleep deficiency and prolonged physical and mental activities have been found to significantly impact of the cognitive functions of drivers, including alertness, perceptual skills, risk proclivity and decision making (Swart & Sinclair, 2015; Hartley, 1998). Several human, temporal, environmental and sleep related factors shown in Table 2.9 were found to predispose a driver to fatigue. (Hartley & Arnold, 1996 cited in Dagli, 2004; Peden *et al.*, 2017).

Table 2.9 Factors that predispose a driver to fatigue (Peden *et al.*, 2017)

Driver at risk of fatigue	Temporal factors	Environmental factors	Sleep-related factors
Young drivers (up to 25 years)	Driving between 2am and 5am	Driving in remote areas with featureless terrain	Driving with sleep debt
Drivers over 50 years	More than 16 hours of wakefulness before trip	Monotonous roads	Driving with sleep-related conditions
Males	Long work period before trip	Long-haul driving	Driving after poor-quality sleep
Shift workers and those working extended hours	Long time since start of trip	Main arterial roads	Drivers disposed to nodding off
Those with medical conditions (such as narcolepsy)	Irregular shift work before trip	Extreme climatic conditions	
Driving after consuming alcohol	Driving after successive nights of shift work	Driving on unfamiliar routes	
Driving after inadequate rest and sleep	Driving under time pressure		
	Driving between 2pm and 6pm (Especially after eating or taking even one alcoholic drink)		

Estimates of the proportion of road crashes attributed to driver fatigue vary in different part of the world. A population based case-control study in New Zealand by Connor (2002) found that factors that significantly increased the risk of a fatal/ serious injury road crash were:

a) Driving while feeling sleep;
b) Driving after less than five hours of sleep in the preceding 24 hours; and
c) Driving between 2am and 5am

Results from surveys have indicated that more than half of drivers have at some time fallen asleep while driving or are vulnerable to driver fatigue (ETSC, 2001; Swart & Sinclair, 2015). A study by Davidović *et al.* (2018) on the working hours and habits of professional drivers indicated the risk of driver fatigue related crashes increased when drivers were driving at night, the length of their working day had increased or when they were working irregular hours. Similarly, a temporal analysis of road crashes by Noce *et al.* (2008) indicated that peak levels of road crashes related to fatigue at night are often 10 times higher than daytime road crash levels as shown in Figure 2.8.

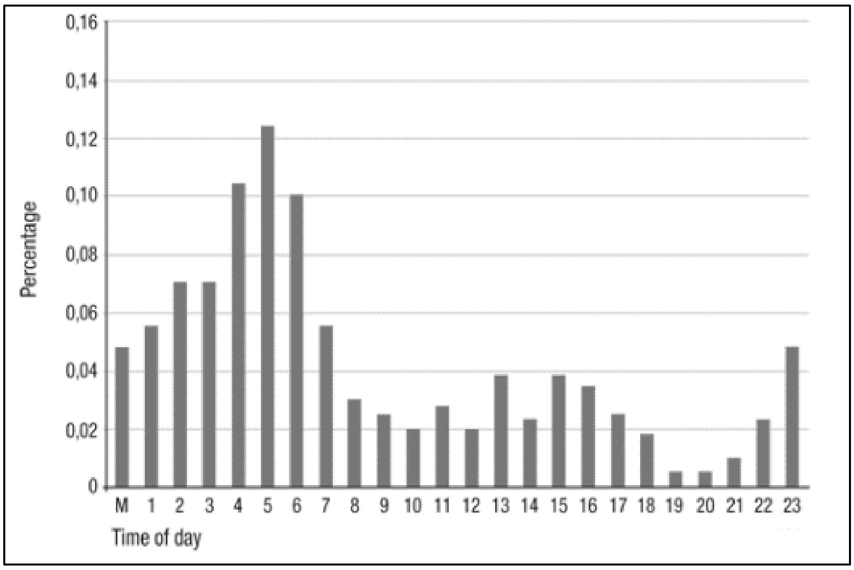


Figure 2.8 Heavy vehicles involved in fatigue related road crashes according to time of day (Noce *et al.*, 2008)

Connor (2002) concluded that fatal/serious road crash incidence could reduce by up to 19 percent, with a reduction in all three of the driver behaviour factors. In the United States, studies by the National Transportation Board (NTSB) found that 52 percent of single-vehicle crashes involving heavy vehicles were related to driver fatigue and that nearly 18 percent of the road crashes were attributed to drivers falling asleep (NSTB, 1995; NSTB, 1999 cited in Peden *et al.*, 2017). The European Transport Safety Council (ETSC) Identified driver fatigue as a significant factor in 20 percent of commercial road crashes in European countries (ETSC, 2001).

2.6.3 Roadway environment

Road crashes tend to be distributed throughout a road network, occurring in clusters on a single site or along a road section (Songpatanasilp *et al.*, 2015). While good road design and regularly maintained roads can greatly aid in reducing the frequency and severity of road traffic crashes, negative road designs can significantly contribute to a precarious road environment (Batrakova and Gredasova, 2016). The road environment can trigger human factor failures as it influences the information and instructions communicated to the road users (Peden *et al.*, 2017; Krug & Sharma, 2009).

Negative road engineering factors include those where a road defect directly causes a traffic crash, where a road design element communicates ambiguous information to drivers thereby causing a driver error, or where a feasible alteration to the road would have reduced the likelihood to a road crash has not been done (World Health Organisation, 2018). Krug & Sharma (2009) note that road environments that promote and allow risky driver behaviour (e.g. through encouraging high traffic speed) or that have not considered safety in all conditions (e.g. at night or in poor weather conditions) indirectly increase the likelihood of a road crash occurring. In the planning, design and maintenance of the road network, the following four key elements affecting road safety have been identified by Barrel *et al.* (2014).

1. Safety awareness in the planning of new roads;
2. The inclusion of safety features in the design of roads;
3. Pro-active safety improvements to existing roads; and
4. Remedial measures on high-risk crash locations.

The contribution of road factors to the occurrence of road traffic crashes varies significantly between developing and developed countries. In Europe, a review of the road risk factors showed that road environment factors were highly influential in 28 percent of road traffic crashes (Hyder *et al.*, 2017). A road safety analysis carried out in the Philippines found that poor road conditions only contributed to 5 percent of the road traffic crashes (Tamayo, 2009). Similarly, a study carried out by Demissie (2017) on the impact of the road environment on road traffic crashes revealed a low contribution of 2.9 percent to traffic crashes in Kenya. This significant variance can likely be attributed to inter-observer variations (Demissie, 2017). Despite the differences in the magnitude of the contribution of the road environment to road crashes, it is notable that a road designed according to operational and functional requirements, and that is maintained regularly, is vital in influencing the perception of drivers, leading to a safer road environment for all road users (Munteanu *et al.*, 2014; Wedajo *et al.*, 2017).

It has been reported that the road environment factor is worse in developing countries due to the poor road design and maintenance (Wegman, 2017). In addition, a variety of traffic mixes requiring with different infrastructural needs, often not provided, are commonly observed on the roads such as high speed vehicles, heavy commercial traffic, pedestrians, cyclists and motorcycle users (Mitra *et al.*, 2017). The rapidly increasing motorisation rates in developing countries are outpacing the current transportation infrastructure capacity, leading to an increase in crash rates and severity levels (Wang *et al.*, 2013).

2.6.4 Vehicle-related factors

Vehicle defects are considered as a key factor in influencing road traffic crashes globally (Demissie, 2017). Defective vehicle parts such as tyres, brakes and vehicle driving lights affect driver's ability to maintain control of a vehicle and can lead to road crashes (Al-Matawah, 2009; Hakkert *et al.*, 2007). Defective safety tools including warning lights and vehicle indicator lights may also inhibit drivers from communication their intentions to other road users, leading to a dangerous interactions between vehicles and other road users (Demissie, 2017). The maintenance and inspection of the vehicle safety systems is crucial to ensure the safety of drivers and all road users (G Botha, 2005).

In Namibia, the Namibian National Road Safety Council (NRSC) states that a significant number of vehicles on the rural roads are poorly maintained and this lack of maintenance affects the likelihood of crashes occurring (NRSC, 2012). Vehicle defects are only seldom reported to contribute to road crash occurrence. They are reported to contribute to 3 percent of crashes in developed countries, with examples of approximately 5 percent in Kenya and 3 percent in South Africa (World Health Organisation, 2018). A study by Kilawa & Nyongole (2015) reported that 15 percent of all road crashes in Tanzania were due to defective breaks. These regional differences could well reflect differences in crash reporting rather than differences in the actual extent of the problem. However, in all cases, vehicle defects appear to be a significantly less common cause of crashes than human errors. A study by Moodley & Allopi (2008), for example, found human error to be the most significant factor in affecting road crashes, with vehicle defects contributing less frequently.

Most developing countries lack the effective regulations and, in some instance, poor implementation of regulations to ensure that vehicles are inspected and maintained with the aim of keeping defective vehicles off the roadway. A study carried out in South Africa found that transportation authorities are critical stakeholders in reducing road crashes through properly identifying vehicles with defects and ensuring that vehicles are roadworthy (van Scoor *et al.*, 2001).

2.7 The Impact of road design characteristics and traffic conditions on road safety

A large body of research exists that investigates the contributing factors to road crashes from a wide range of aspects and approaches, with relationships between road design elements, traffic conditions and road crashes explored on several occasions (Mohammed, 2013; Gaudry and Vernier, 2002; Dwikat, 2014). This section provides a look at several relationships, assumed and proven, between road crashes, road design and traffic related characteristics, and the extent to which they impact road safety on rural roads worldwide.

Road crashes are characterised by multiple causes (Dwikat, 2014). The alignment of the road influenced by the surrounding road environment is an important factor in road safety: dimension of radii, ratio of consecutive curves, dimension of vertical curves and sight distances conditions. In various evaluations of road safety effects, driver behaviour, influenced by personality, skills and experience plays a considerable role in the cause of road crashes (Mohammed, 2013).

Deller (2013) affirms that geometric design elements play an important role in defining the traffic operational efficiency of any roadway. Key geometric design elements that influence traffic operations and impact the safety of the roadway include the number and width of road lanes, the presence and widths of shoulders and the horizontal and vertical alignment of the highway (Mohammed, 2013).

Ahmed (2013) mentions that the road network has an effect on crash risk because it determines how road users perceive their environment. Roadway factors, including roadway and roadside design elements, play an important role on determining the risk of road crashes (Stephan and Newstead, 2017). Negative road engineering factors include those where road defects directly triggers a crash or where some element of the road environment misleads a road user and thereby creates human errors (Parizel and Phillips, 2004).

The geometry of the roadway plays a significant role in road crash frequencies as well as the crash severity levels (Dwikat, 2014). Different elements of the road design are important. However, a few parameters are considered to be more important in influencing road safety than others. This section provides an extensive review of literature on rural roads design and traffic characteristics on road safety.

2.7.1 Speed

Traffic speed is probably the most important factor impacting crash frequency and severity on the roads (Elvik *et al.*, 2004). Empirical data shown by several studies led to the assumption that increased speeds result in more severe road crashes, should other factors (environment and vehicles factors) remain the constant (Kockelman, 2006; Thomas *et al.*, 2013). However, as the association between speed and the road crash frequency can be influenced by a multitude of other roadway factors, the extent of the relationship between speed and the likelihood of road crash incidence can vary depending on traffic and roadway conditions (Edquist *et al.*, 2009; Vadeby *et al.*, 2018).

Speed can affect the likelihood of a crash occurring in several ways. A distinction between excessive speeds (driving faster than the speed limit) and inappropriate speed (driving too fast for the road conditions, although speed may be under the posted speed limit) is made by the literature (Edquist *et al.*, 2009; Organisation for Economic Co-operation and Development, 2006). Driving too fast makes lateral control more difficult and reduces the available time and distance to recognise and respond to hazards in the roadway (Edquist *et al.*, 2009; Sjogren *et al.*, 2012; Turner *et al.*, 2015). In addition, the severity levels of a crash are highly affected by the impact speed (Organisation for Economic Co-operation and Development, 2006).

The extent of the relationship between speed and road crashes has been investigated by several studies, with most study findings concluding that higher speed selections have a direct relationship with higher road crash rates (Godavarthy and Russell, 2016; Nilsson, 2004; Feuillet *et al.*, 2015; Taylor *et al.,* 2002). Taylor *et al.* (2002) found positive associations between speed changes and crash frequency by employing a cross-sectional analysis on 174 road segment in England. Taylor *et al.* (2002) created dummy variables to represent the different categories in the Poisson regression models and included the set of road characteristics used to classify the segments in the models.

A before-after study by Nilsson (2004) extensively investigated the impact of change in speed on road safety using the Power Model. The study found positive correlations between change in speed and the severity of the road crashes, with the extent of the relationship influenced by the severity of the crash according to the power function as illustrated in Figure 2.9. Similarly, an extensive evaluation on the effects of speed change on road crashes by Elvik *et al.* (2004) concluded a linear causal relationship between speed changes and changes in road crashes. An area wide-level investigation by Nilsson (2005) on road speed and casualties concluded that an increase in the average operating speeds positively associated with an increase in road fatalities and injuries

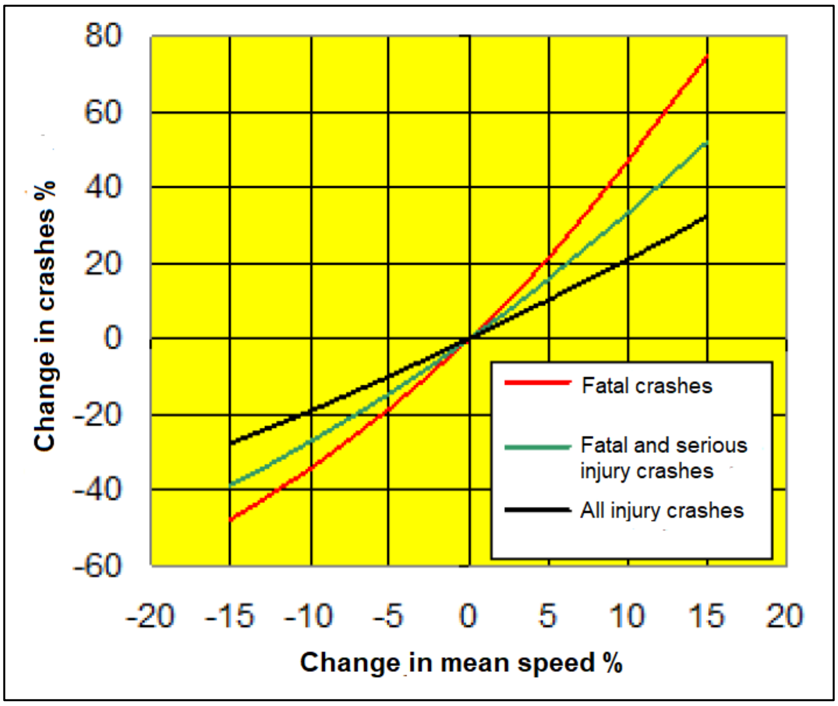


Figure 2.9 Relationship between percentage change in speed and percentage change in road crashes in power model (Nilsson, 2004)

The speed limits on the roads captures the characteristics of both the speed and speed variance (Wang *et al.*, 2013). Several studies explored and found extensive relationships between road crashes and speed limits (Deller, 2013; Woolley *et al.*, 2002; Richter *et al.*, 2016). Studies examining speed limit impact on road safety are often based on either a highly aggregate region level speed or a disaggregate road level speed. An extensive investigation by Kockelman (2006) on highway speed limit change in Washington State using time series found that increasing the speed limit had a direct impact on higher crashes rates. Deller (2013) found that crash frequency would reduce if the speed limits were to be reduced on Australian highways. Richter *et al.* (2016) investigated the influence of speed limits on overtaking on two-lane rural highways and found that reduced speed limits led to a decrease in the number of road crashes.

Speed was found to have significant effects on the safety of the roadway in the literature. Several studies concluded that increasing speed and greater speed variations create a hazardous road environment. Also, studies investigating speed limits concluded that changes in speed limits often resulted in changes in travel speed independent of the design conditions of the road, which resulted in a linear relationship between changes in posted speed limits and road crash rates.

2.7.2 Traffic volume

Traffic volume is defined as the number of vehicles crossing a particular point on the road study segment per hour, often expressed in terms of average daily traffic (ADT) and measured in vehicles per day (May, 1990). The traffic volume is instrumental in determining the annual average daily traffic (AADT), which is vital in developing road crash prediction models (El-basyouny & Sayed, 2009; Eenink *et al.*, 2005; Glavić *et al.*, 2016).

In understanding how road safety is affected by the volume of traffic and the interaction between vehicles, it is important to note that the operational conditions of the traffic flow on a roadway are characterised by the flow[18]-speed[19] and the density[20]-flow diagrams, which serve as the basic theoretical traffic flow correlations (Marchesini and Weijermars, 2010). The actual field conditions need to be described while distinguishing more sophisticated correlations. The Highway Capacity Manual produced by the Transortation Research Board (2000) describes the correlations between flow characteristics illustrated in Figure 2.10, that when there are hardly any vehicles and therefore density approaches zero, speed will approach free-flow speed (u_f), meaning that a driver's speed is not influenced by that of other drivers. Simultaneously, flow will approach zero as well. Speed will decrease to an optimum speed (u_o) when density increases to the optimum value (k_o). As there are more vehicles on the roadway, there is more interaction of vehicles. At the same time, traffic flow will increase to the maximum flow called capacity (q_m). A further increase of density to the maximum value or jam density (k_j) will result in a further reduction of speed until speed approaches zero. Flow will also decrease and approach zero.

[18] Flow (q): The number of vehicles passing a specific point or short section in a given period of time in a single lane;
[19] Speed: The average rate of motion;
[20] Density: The number of vehicles occupying a section of roadway in a single lane (May, 1990).

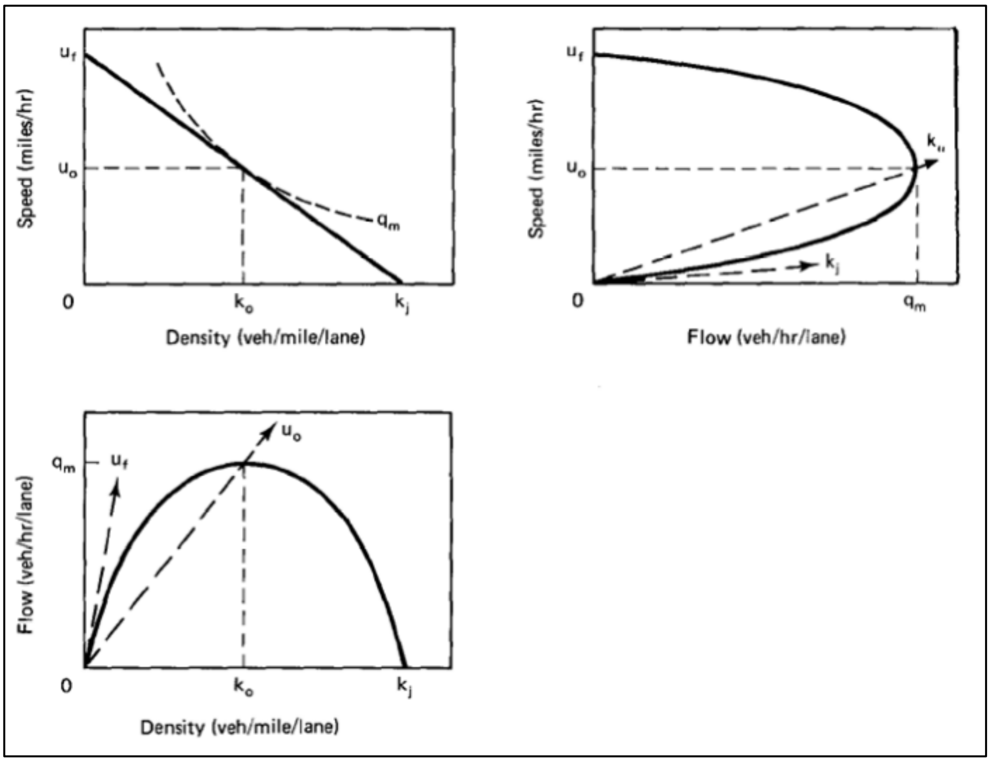


Figure 2.10 Traffic flow diagrams (Transportation Research Board, 2000)

A study by Golob *et al.* (2004) demonstrated that a strong statistical relationship exists between traffic flow and road crashes. The study results found that an increase in the traffic flow resulted in an increase in the road crashes on the roads. Similarly, an investigation by Golob and Recker (2003) in Southern California found a linear relationship between traffic flow and related road crashes while controlling weather and light conditions in the multivariate statistical analyses used. The combination of high traffic volumes and narrower lane widths increases the likelihood of a road crash occurring (Dehuri, 2013). The low traffic volumes on rural roadways, compared to urban roadways, result in high speed impact road crashes, with a high risk for head-on or run-off crashes and higher injury severity levels (Alsubeai, 2017; Karlaftis & Golias, 2002; Nambahu, 2018).

A study by Eenink *et al.* (2005) reported that in investigating the impact of traffic volumes on road safety, crash prediction models were developed with the aim of providing an insight into the safety levels on the roadways. Equation [2.1] was used in developing crash prediction models assess the extent of the impact of traffic volume on crash occurrence.

$$E(\lambda) = \alpha Q^{\beta} e^{\Sigma \gamma_i x_i} \quad [2.1]$$

Where; $E(\lambda)$ = estimated number of crashes

Q = represents traffic volume

x_i = represents the risk factor (*i* = 1, 2, 3,…, *n*)

γ = represents corresponding coefficient

β = represents the effect of traffic volume on road crashes

As illustrated in Equation [2.1], the traffic volume is a variable in univariate crash prediction models, with AADT used to represent the traffic volumes. To study the effect of traffic volume on road safety, a considerable amount of data is required, in particular data related to the length of the road segment and the AADT on the roads to be investigated. A study by Eenink *et al.* (2005) illustrated the relationship between traffic volumes and road crashes in Figure 2.11.

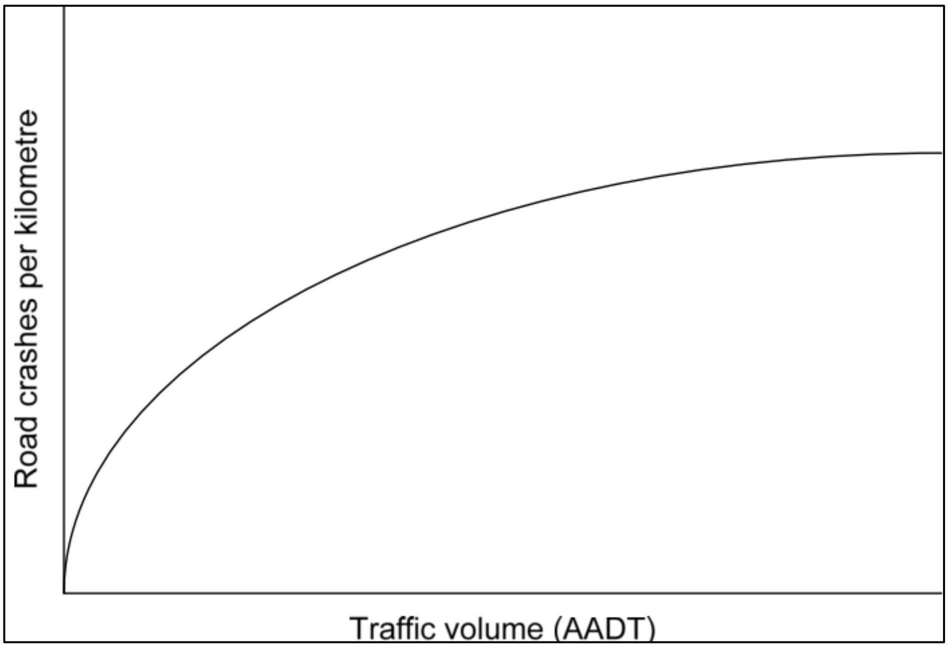


Figure 2.11 Relationship between AADT and road crashes (Eenink *et al.*, 2005)

Duivenvoorden (2010) states that these models can be used to monitor the safety performance of a road network as traffic volumes changes, this gives authorities the opportunity to improve the level of safety offered to all road users. Overall, the examined literature indicated that the number of road crashes increased as the traffic volumes increased.

2.7.3 Lane width and number of lanes

The road lane is defined by Housley (2015) as the portion of the roadway used for a single line of vehicles. The TRH 17 on the Geometric Design of Rural Roads (CSRA, 1988) notes that the selection of lane width is based on traffic volume, vehicle type and speed. The widest lane width of 3.7 m is recommended for roads with higher volumes and speeds, while the roads with expected lowest volumes are recommended to have the narrowest lane width of 3.1 m. Roads expected to have intermediate traffic conditions are recommended to have a lane width of 3.4 m (CSRA, 1988). Figure 2.12 illustrates that on paved roads, the lane width excludes the edge line markings as they are considered part of the road shoulder (CSRA, 1988).

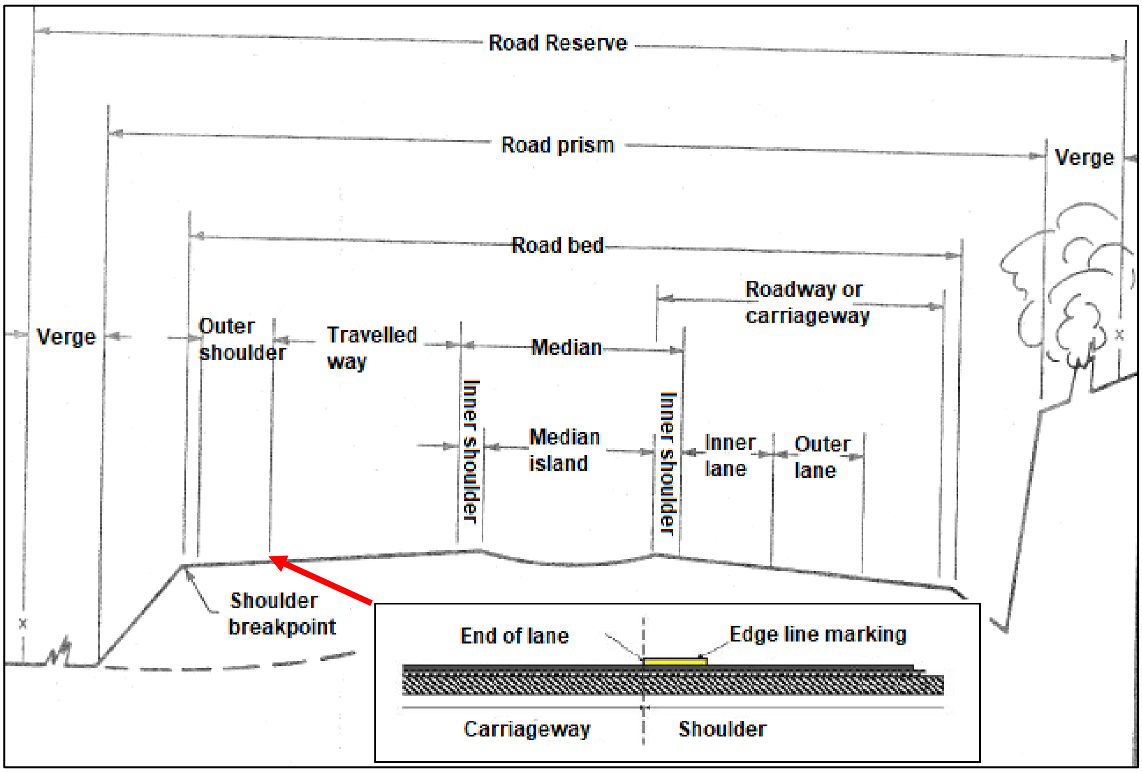


Figure 2.12 Lane width in the road cross sectional design (CSRA, 1988)

Lane width is a vital parameter affecting the road crash rates (Meng *et al.*, 2006; Papadimitriou *et al.*, 2018). A linear relationship exists between the travel lane width and road crash rates (Wedajo *et al.*, 2017; Park *et al.*, 2010). In addition, the comfort of driving and operational characteristics of a roadway improve significantly with increasing travel lane width (Mohammed, 2013). Investigations into the impact of lane widths on roadway safety have found that for any functional classification of roadway, a reduction in the lane width resulted in a drastic increase in the likelihood of crashes occurring (Ambunda & Sinclair, 2019; Wang *et al.*, 2013; Othman *et al.*, 2009).

An investigation by Ahmed (2013) found that increasing the lane width from 2.75m to 3.65m reduced the likelihood of head-on and other related crashes by approximately 50 percent. An earlier study by Iyinam *et al.* (1997) reported that road crash rates decreased from 1.5 to 1.1 crashes per million vehicle kilometres travelled with an increase in lane width from 2.7 m to 3.1 m. The crash rates further decreased on lanes widths between 3.4 m to 3.7 m to 0.9 crashes per million vehicle kilometres travelled (Iyinam *et al.*, 1997).

The Highway Safety Manual (HSM) (AASHTO, 2010) developed crash modification factors[21] (CMF) to investigate the relationship between lane width and road crashes. The HSM found that an increase in lane width does not always result in an increase in road safety, particularly on roads with wider

[21] Crash modification factor is a multiplicative factor used to compute the expected number of road crashes after implementing a given countermeasure at a specific site (Garber and Hoel, 2009).

shoulder widths (AASHTO, 2010). Figure 2.13 illustrates the relationship between lane width and rod safety on rural two-lane roadways.

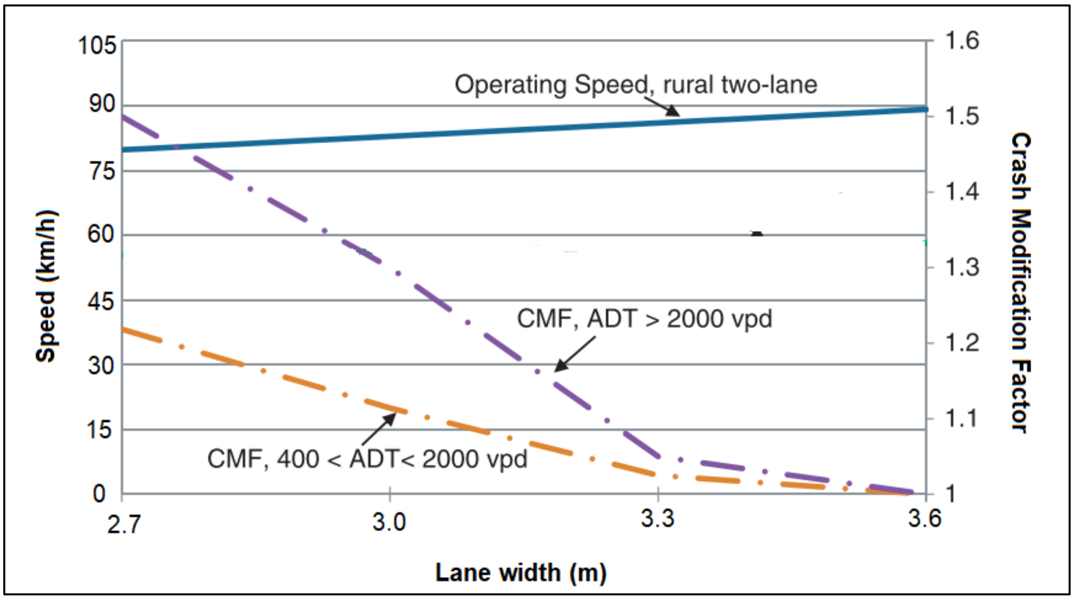


Figure 2.13 Relationship between lane widths, operating speed and safety on two-lane rural roadways (AASHTO, 2010)

A study by Kononov *et al.* (2008) found that increasing the number of lanes available to the road users increased the crash frequency, arguing that the increased potential lane-change related conflict opportunities contribute to an unfavourable safety situation. Similarly, studies by Noland and Oh (2004) and Haynes *et al.* (2008) investigated the impact of number of lanes on road safety at an aggregate area level. The research results revealed that an increase in the number of lanes was associated with an increase in road fatalities on roadways with lower traffic volumes and higher traffic speed conditions. The effect of lane width and number of lanes on driver speed choice depend on the amount of roadway width (lane width and number of available lanes) the driver perceives as usable. Drivers perceived a wider road width as safer (perceived space to correct driver errors), resulting in higher speed choices (Edquist *et al.*, 2009; Park *et al.*, 2010; Ben-Bassat & Shinar, 2011).

2.7.4 Shoulder width and type

Roadway paved shoulders have several functions, including recovery area for driver errors and the emergency stop and pull off function as detailed by the Policy on the Geometric Design of Highways and Streets (AASHTO, 2001). Several studies have investigated the correlations between the shoulder width and the likelihood of a road crash, with considerable variations in findings cited (Liu *et al.*, 2016; Zegeer *et al.*, 1981; Ben-Bassat and Shinar, 2011).

Ben-Bassat and Shinar (2011) state that in addition to the shoulder width, the shoulder type also impacts road crash frequencies. The presence of a paved shoulder is the best type of shoulder in terms of road safety, in contrast to a gravel shoulder (Karlaftis and Golias, 2009). Othman *et al.*

(2009) explained that while it is desirable that a shoulder be wide enough for a vehicle to be driven completely off the travelled way, narrow shoulders are better than no shoulder at all.

An earlier study by Zegeer *et al.* (1981) in Oregon, found that an increase in the shoulder width led to an increase in road crashes, except on roads with an AADT of 3 600 to 5 500 vehicles. An area-wide level investigation by Ambunda and Sinclair (2019) in Namibia, found a lack of correlation between shoulder widths and road crashes on two-lane rural roads with AADT less than 2000 vehicles. The findings of a study by Huanghui (2012) in Kansas, illustrated in Figure 2.14 found that composite shoulders and wider shoulders had a more positive impact on road safety compared to narrow shoulders.

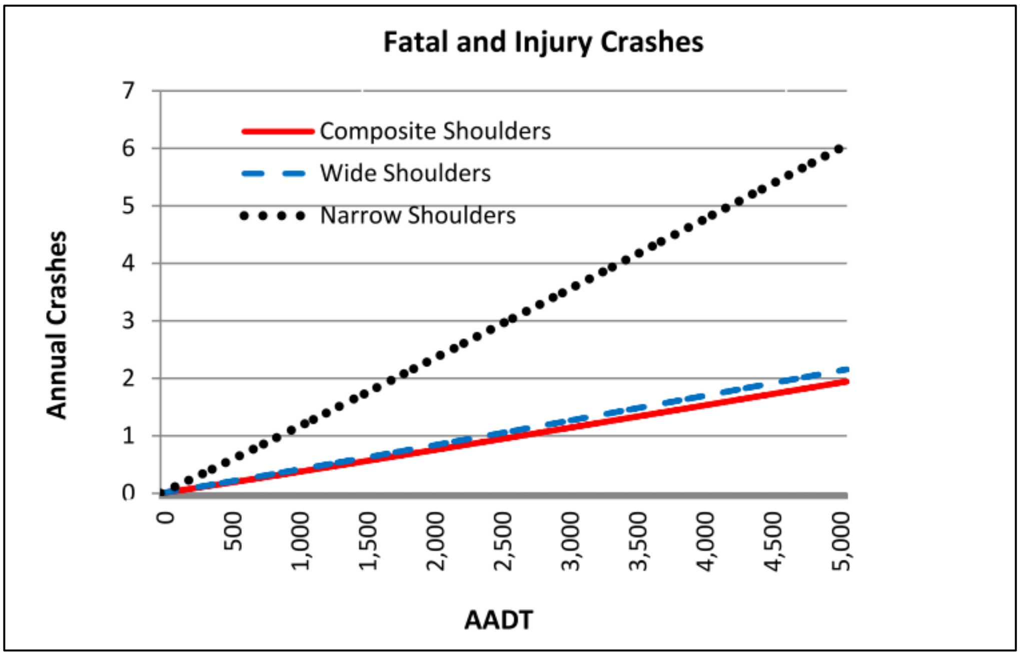


Figure 2.14 Effect of shoulders on road crashes (Huanghui, 2012)

An investigation by Karlaftis and Golias (2002) in Greece, found that as the shoulder width increases on two-lane rural roads up to 7.5 m, the road crash rates decreased significantly. In contrast, a study by Bamdad Mehrabani and Mirbaha (2018) found that paved shoulders with a width of 1.2 to 2.1 m had a significant positive effect on road crash rates. Park *et al.* (2010) identified the paved shoulder widths of 3.1 to 3.7 m as having a positive impact on road safety on roads with AADT between 3000 and 5000 vehicles in Texas. The crash modification factors (CMF) estimated by The Highway Safety Manual (AASHTO, 2010) concluded that shoulder width has a larger positive effect on road safety when road lanes are narrow, but the effect of shoulder widths decreases as lane widths are increased. The effects of shoulder widening on road crashes shown in Table 2.10 were determined for paved and unpaved shoulders in an early before and after study conducted by Zegeer V *et al.* (1987) in North America.

Table 2.10 Effects of shoulder widening for related crash types on rural two-lane roadways (CSRA, 1988)

Shoulder widening (m) per side	Percent (%) reduction in related crash types	
	Paved	Unpaved
0.6	16	13
1.2	29	25
1.8	40	35
2.4	49	43

Several studies have shown that shoulder widths can also have a conflicting effect on driver behaviour, leading to hazardous road safety conditions on the roadway (Abele and Møller, 2011; Čičković, 2016; Ben-Bassat and Shinar, 2011). Drivers tend to select lower speeds on narrow roads with narrow paved shoulder widths due to the perception of lower safety for the road user. This creates a safer driving behaviour compared to higher speed selections on roads wider roads with a wider paved shoulder (Godley *et al.*, 2004). Huanghui (2012) explained that speed selections are higher on roads with wider shoulders as they give drivers a sense of security and perceived space for correcting errors. In contrast, narrower roads and shoulders are perceived as less tolerant and therefore more dangerous, leading drivers to be more cautious to avoid risky situations (Liu *et al.*, 2016). In contrast, a study by Ben-Bassat and Shinar (2011) found that narrow shoulders led to drivers to steer away from the left shoulder and drive closer to the centre of the road, thus increasing the likelihood of a head-on road crash.

A study by Abele and Møller (2011) reports that driver speed selections are lower on roads with gravel shoulders due to visual cues (colour difference between the paved roadway surface and the gravel surfaced shoulder) that give a perception of a narrower driving lane, compared to conditions where a paved hard shoulder is present.

The TRH 17 on the Geometric Design of Rural Roads (CSRA, 1988) recommends the widest shoulder width of 3m for roads with the highest operating speeds and heavy traffic volumes. Roads with intermediate traffic volumes and higher operating speeds are recommended to have shoulder widths ranging from 1 to 2.5m. The South African National Roads Agency Ltd (SANRAL) Geometric Design Manual (South African National Road Agency Limited (SANRAL), 2003) provides Table 2.11 showing recommended shoulder widths for use on undivided two-lane rural roads.

Table 2.11 Shoulder widths recommended for undivided rural roads (CSRA, 1988)

Design Speed (km/h)	**Design hour volume (veh/h)**		
	<250	250-450	>450
	Shoulder width (m)		
50	1.0	-	-
60	1.5	1.5	-
70	1.5	2.5	-
80	2.5	2.5	2.5
90	2.5	2.5	3.0
100	2.5	2.5	3.0
110	-	3.0	3.0
120	-	3.0	3.0
130	-	-	3.0

2.7.5 Horizontal and vertical alignment (Road alignment)

Hanno (2004) defines the road alignment as the combination of vertical and horizontal geometric elements providing the location of the road through a terrain. The TRH 17 on the Geometric Design of Rural Roads (CSRA, 1988) states that the ease, comfort and safety of operations of a vehicle on rural roadways are determined by the consistency of design, among other factors. This consistency is achieved partly by relating the magnitude of successive elements of horizontal and vertical alignment to speed.

A study by Hanno (2004) reports that most design practices are based on design guidelines considering the road alignment in two dimensions only. These guidelines were developed without considering the three-dimensional (3D) effect of the combined road alignment illustrated in Figure 2.15, consequently resulting in a design process that does not ensure road user safety.

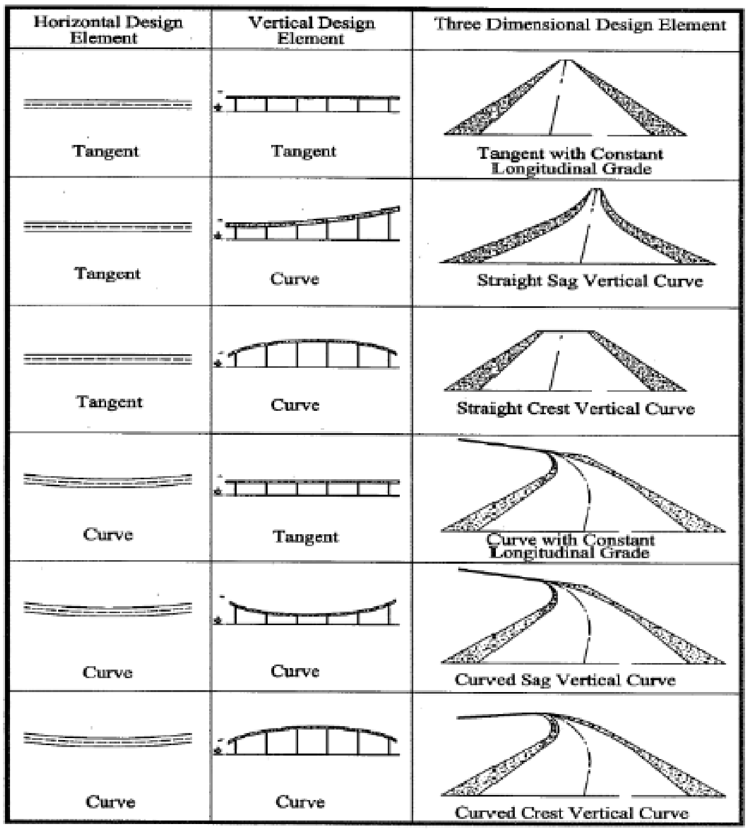


Figure 2.15 Three-dimensional combination of horizontal and vertical alignments (Hanno, 2004)

Hanno (2004) states that a poor coordination between the horizontal and vertical alignments leads to poor perceptions and driving errors, which consequently compromise the safety of the road. Similarly, Krug and Sharma (2009) report that inconsistent roadway design increases drivers' workload and results in a road alignment that does not meet driver expectations. The properties of the horizontal and vertical curves and their influence on road safety are described in Section 2.7.5.1 and Section 2.7.5.2.

2.7.5.1 Horizontal curves

Hassan and Easa (2003) report that the road crash rates on horizontal curves are significantly higher than the road crash rates on road tangents. Drivers' speed selections are significantly affected by the presence of a horizontal curve on a road section (Easa, 2003). An investigation by the National Cooperative Highway Research Program (NCHRP) Guide for Addressing Run-Off-Road (ROR) Collisions (Transportation Research Board, 2003) reports that 42 percent of ROR fatal crashes are reported on horizontal curves, with road fatalities increasing to 50 percent on two-lane rural roads.

Easa *et al.* (2007) observed a reduction in operating speeds by drivers traversing horizontal curves, with a desire to maintain satisfactory side road friction, expressed in Equation [2.2] by the TRH 17 (CSRA, 1988). Similarly, the study by Ambunda and Sinclair (2019) found that drivers speed selection on horizontal curves was lower than speed selection on straight sections.

$$f_D = \frac{v^2}{gR} - \frac{e}{100}$$ [2.2]

Where; f_D = side friction demand factor

e = vehicle speed (m/s)

g = gravitational acceleration (9.807 m/s^2)

R = radius of curve (m)

e = superelevation (%)

The safety prediction tool developed in the Highway Safety Manual (AASHTO, 2010) investigated the relationship between the design speeds, horizontal curve radii and operating speeds on rural two-lane roadways. Figure 2.16 illustrates the influence of horizontal curve radii on vehicle operating speeds, considering temporal conditions on the roadway. The effect of horizontal curve radii on operating speeds is marginal until the radius falls below approximately 350 m (AASHTO, 2010; Bauer & Harwood, 2014). Similarly, the impact of the curve radius on the expected road crash frequency changes nominally until the curve radius falls below 350m, as indicated by the change in the crash modification factor (CMF) (AASHTO, 2010; Glavić *et al.*, 2016).

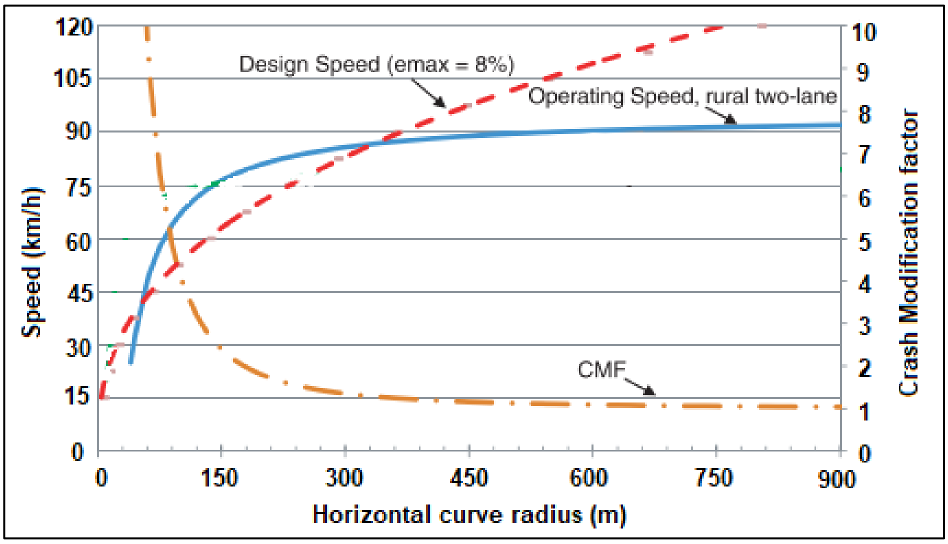


Figure 2.16 Relationship between horizontal curves radii, operating speed and safety on rural two-lane roadways (AASHTO, 2010)

A study by Garcia and Abreu (2016) found that a reduction in crash rates occurred with an increase in the horizontal curve radius. Congruently to the Highway Safety Manual (AASHTO, 2010) findings, Turner (2005) found that the risk off a road crash on horizontal curves increases with a reduction in the radii of the curves, with horizontal curve radii considered critical at radii less than 350 m as illustrated by Figure 2.17.

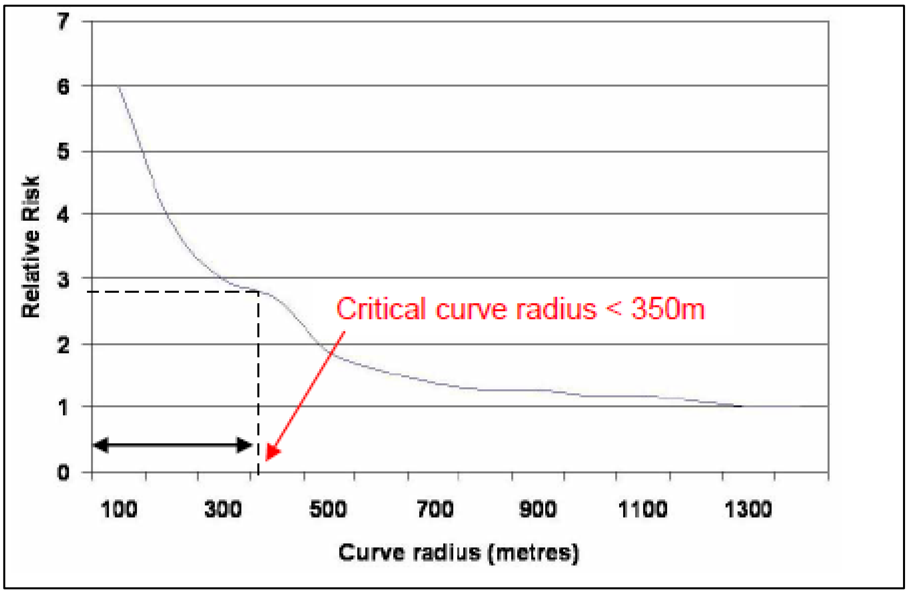


Figure 2.17 Relative crash risk on various horizontal curve radii (Ambunda, 2018)

The TRH 17 on the Design of Rural Roads (CSRA, 1988) provides the following guidelines to enable transportation engineers to design safe horizontal curves for road users:

1. Minimum radii

The minimum radius is a limiting value for a given design speed determined from the maximum rate of superelevation and the maximum allowable side friction factor. Minimum radii shown in Table 2.12 are recommended by the TRH 17 (CSRA, 1988) only under critical road environment conditions, as the deviation angle of each curve should be as small as the physical conditions permit, so that the road can be as straight as possible. It should also be considered that excessively curves may cause operational problems leading to safety issues.

Table 2.12 Minimum radii of horizontal curvature (CSRA, 1988)

Design Speed (km/h)	Radius (m)
50	80
60	110
70	160
80	210
90	270
100	350
110	430
120	530
130	640
140	760

2. Horizontal curve length

The TRH 17 (CSRA, 1988) states that for small deflection angles, curves should be long enough to avoid the appearance of a kink[22]. A minimum length of 300 m is recommended, which can be reduced to 150m if operational space is limited. For deflection angles less than 5 degrees, it is recommended that the minimum length of the curve to be increased from 300m, by 150m by 30m for each 1 degree decrease in the deflection angle.

For long curves, particularly near-minimum radius, may cause tracking problems. These are suffered principally by vehicles travelling at speeds significantly different from the roadways design speed. Gooch *et al.* (2016) reports that long curves may limit and negative affect the safety of overtaking manoeuvres on two-lane roads left-hand curves, as overtaking manoeuvres would have to start at a considerable distance behind the leading vehicle, due to the greater distance to be traversed on a left-hand curve compared to a right-hand curve during an overtaking manoeuvre. Consequently, the TRH 17 (CSRA, 1988) recommends that the length of the horizontal curve does not to exceed 1000m.

2.7.5.2 Vertical Curves

Gichaga (2017) notes that the vertical curves have properties of length and gradient, representing the height gained or lost in metres, divided by a horizontal distance of 100 m, expressed as a percentage. Vertical curves provide a gradual change from one tangent grade to the next, to enable drivers to safely and smoothly traverse vertical road sections (Garber and Hoel, 2009).

Easa (2003) found that vertical grades equal to or less than 5 percent have an insignificant influence on the occurrence of road crashes, while a steep increase in operational speeds and the occurrence of road crashes on vertical grades greater than 6 percent. A study by Hamzeie *et al.* (2017) found that drivers change their speeds as soon as a vertical curve in combination with a horizontal curve becomes visible. When approaching a vertical crest curve, drivers perceived the horizontal curve as sharper. Subsequently, drivers reduced their speeds. In contrast, drivers perceived the horizontal curve as less sharp and increased their speeds as they approached sag vertical curves. Hassan and Easa (2003) explained that the misperception of the combination of vertical and horizontal curves relates to the fact that drivers react to how the perceive the road alignment, independently of speed, warning or other regulatory safety signs.

A study by Bella (2005) noted that on roadway sections where the vertical and horizontal curve are combined, the value of the horizontal curve radius influenced by the vertical alignment may appear different to the driver than the actual value, which can be detrimental to the safety of drivers on the

[22] Kink: a sharp twist or curve on an otherwise straight road section (Karlaftis and Golias, 2009).

roadway. A study by Bidulka *et al.* (2002) discussed and developed a model shown by Equation [2.3] on how the type of vertical curve and the radius of the horizontal curve influences the perceived horizontal radius.

$$R_p = -51.28 + 0.953R_a + 132.11V + 0.125R_aV \qquad [2.3]$$

Where V is equal to 0 for crest vertical curves and equal to 1 for sag vertical curves. The model (units in metres) developed by Bidulka *et al.* (2002) states that the horizontal curve radius (R_p) would be perceived as sharper than the actual radius (R_a) when it overlaps with a crest vertical curve. Conversely, the radius would be perceived as flatter when it overlaps with a sag vertical curve, leading to higher operational speeds and an increase in the likelihood of a road crash.

The TRH 17 on the Geometric design of Rural Roads (CSRA, 1988) provides the following guidelines for the design of vertical curves, with aim of providing a safe roadway system for road users.

1. Minimum curve length

The TRH 17 (CSRA, 1988) explains that where the algebraic differences between successive grades are small, the intervening minimum vertical curve becomes very short. This can create the impression of a kink in the grade line particularly where the tangents are long. For differences in grade greater than 0.5 percent, a certain minimum length is proposed depending on the design speeds illustrated in Table 2.13, with a minimum curve length of 240 m recommended for highways.

Table 2.13 Minimum length of vertical curves (CSRA, 1988)

Design speed (km/h)	Length of curve (m)
40	60
60	100
80	140
100	180
120	220
140	260

2. Gradients

TRH 17 (CSRA, 1988) states passenger car speeds are relatively unaffected by the vertical curve gradient, as the horizontal alignment tends to influence driver speed selections. In contrast, truck speeds are significantly influenced by gradient. Therefore, the design of vertical curves targets grades that will not reduce the speed of heavy vehicles enough to cause hazardous condition for following drivers. Several studies globally have indicated that when truck speeds are reduced by more than 15km/h on vertical curves, the frequency of road crashes increased sharply (Dong *et al.*, 2015; Pais *et al.*, 2013; Choudhary *et al.*, 2018). For South African conditions, a 20km/h speed reduction in heavy vehicle speeds is accepted as representing hazardous conditions for road users. Notably, it may be necessary to provide auxiliary lanes for slower moving vehicles if the appropriate grade cannot be provided economically. Table 2.14 shows the TRH 17 (CSRA, 1988) recommended maximum grades for various topographies.

Table 2.14 Maximum vertical curve gradients (CSRA, 1988)

Design speed (km/h)	Topography		
	Flat (%)	Rolling (%)	Mountainous (%)
60	6	7	8
80	5	6	7
100	4	5	6
120	3	4	5

The critical length of any given grade is defined as the length that causes the speed of the design heavy vehicle to be reduced by 20km/h (CSRA, 1988). The starting point of the grade is approximated as a point halfway between the preceding vertical point of intersection and the end of the vertical curve (CSRA, 1988). The critical lengths shown in Table 2.15 therefore indicate where the provision of an auxiliary lane may have to be considered.

Table 2.15 Critical length of grade (CSRA, 1988)

Gradient (%)	Length of grade (m)
3	400
4	300
5	240
6	200
7	170
8	150

Notably, the road horizontal curve was found to have an inverse relationship with road crashes. Studies in the literature indicated that decreasing the radius of the horizontal curves resulted in the increase of the likelihood of a crash occurring on the curve segments. In contrast, studies in the literature found that increasing the grade of the road vertical curve resulted in a precarious road safety situation on the roads.

2.7.6 Sight Distance

Sight distance plays a vital role in determining the operational safety of a road (Housley, 2015; Yannis *et al.*, 2016). It is critically important that sufficient sight distance is provided to ensure that drivers are able to safely control the operations of the vehicles while on the road (Mollel *et al.*, 2011; Khan *et al.*, 2014; Rogers, 2003). The alignment of the roadway has a great impact on road safety because a drivers ability to see ahead is necessary for the safe operation of the vehicle and thus for the overall safety of the road system (Wang *et al.*, 2009). The Technical Recommendations for Highways 17 on the Geometric Design of Rural Roads by the Committee of State Road Authorities (CSIR) (1988) affirms that the best visual cue to the driver is the roadway ahead.

2.7.6.1 Stopping Sight Distance

The stopping sight distance (SSD) refers to the ability of the driver to bring their vehicle to a standstill, and thus is based on speed, driver reaction time and skid resistance on the road surface. The SSD is expressed in Equation [2.4]:

$$s = 0.694v + v^2/254f \qquad [2.4]$$

where; s = total distance travelled (m)

v = speed (km/h)

f = brake force coefficient

Table 2.16 provides the SSD distances based on traffic operating speeds and the appropriate brake coefficients adopted for design in Namibia.

Table 2.16 Stopping sight distance on level roads (CSRA, 1988)

Design Speed (km/h)	Stopping Sight Distance (m)
40	50
50	65
60	80
70	95
80	115
90	135
100	155
110	180
120	210
130	230
140	255

Figure 2.18 provides the SSD requirements on roads passing through hilly terrain with various road grades. The SSD is based on the traffic operating speed, braking coefficients and the road gradient. Figure 2.19 indicates the horizontal curve radius requirements on a roadway to provide a safe and appropriate SSD on the road segment.

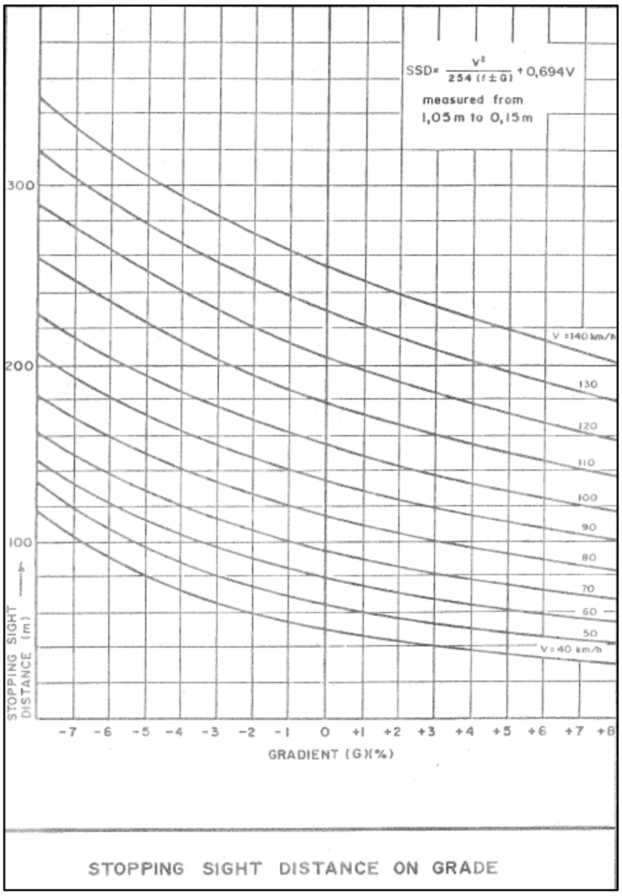


Figure 2.18 Stopping sight distance on roadway grades (CSRA, 1988)

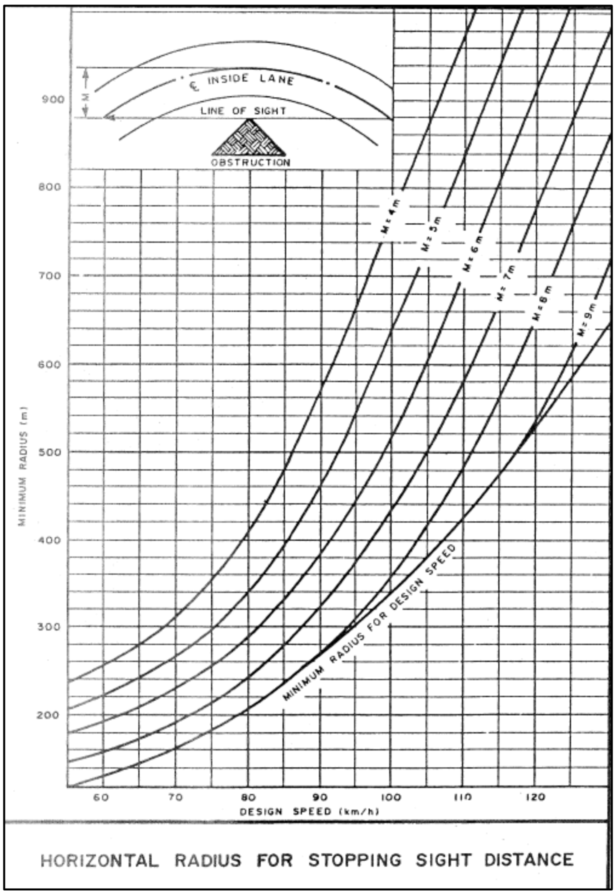


Figure 2.19 Stopping sight distance on roadway horizontal radius (CSRA, 1988)

2.7.6.2 Passing Sight Distance

The Passing Sight Distance (PSD), is critically important on two-lane roads to enable drivers to use the opposing traffic lane for overtaking other vehicles without interfering with oncoming vehicles (Karlaftis and Golias, 2009).The TRH 17 on the Geometric Design of Rural Roads confirms that the PSD is an important criterion indicative of the quality of service provided by the roadway (CSRA, 1988).

Roads with heavy traffic volumes require a higher percentage of passing sight distance than roads with a light traffic volume to provide the same level of service and safety when overtaking (CSRA, 1988). The passing sight distances used on Namibian rural roadways as determined on South African road conditions are provided in Table 2.17.

Table 2.17 Passing sight distance on level roads (CSRA, 1988)

Design Speed (km/h)	Passing Sight Distance (m)
60	420
70	490
80	560
90	620
100	680
110	740
120	800

2.7.6.3 Decision Sight Distance

While the concept of the SSD and the PSD are the main Sight Distances to influence road safety, the Decision Sight Distance (DSD) is a third important element. SSDs are sufficient for reasonably competent and alert drivers to come to sudden stops under ordinary situations, but greater distances are needed for drivers to take complex decisions.

The DSD is the distance needed for a driver to detect an unexpected or otherwise difficult to perceive information source in a roadway environment; to recognise its potential threat to safety; to select an appropriate speed and path; and to initiate and complete a safe manoeuvre. The DSD provides drivers additional margins for errors whenever there is a likelihood for errors in information reception, decision making and actions by the drivers. The DSD, as provided in Table 2.18, is related to the reaction time involved in a complex driving task. The reaction time selected for this purpose is 7.5 seconds, which is roughly the mean of values as provided by American practices (American Association of State and Transportation Officials (AASHTO), 2001). The calculated values in Table 2.18 are thus based on SSD to allow for the condition where the decision is to bring the vehicle to a stop. The TRH 17 (CSRA, 1988) reports that this has the effect of increasing the normal reaction time of 2.5 seconds by a further 5 seconds of travel at the design speed of the road, which has an adverse effect on road safety. The DSD is measured from an eye height of 1.05m to the road surface.

Table 2.18 Decision sight distance on level roads (CSRA, 1988)

Design Speed (km/h)	Decision Sight Distance (m)
40	130
50	160
60	190
70	215
80	240
90	270
100	300
110	325
120	350
130	380
140	410

2.7.7 Access management

Access management is the concept that access-related vehicular manoeuvres and volumes can have serious consequences on the performance of traffic operations and road safety (Ahmed, 2013; Alsubeai, 2017; Jaiswal & Bhatore, 2016). Access management complements geometric design by reducing the likelihood of access-related conflicts, by minimising the frequency of major conflict movements and reducing the severity of crashes due to such conflicts (Ahmed, 2013; Karlaftis & Golias, 2002).

Several studies have concluded that higher access density leads to more road crashes (Ahmed, 2013; Mitra *et al.*, 2017; Jinghui & Xuesong, 2018). A study by Ahmed (2013) in Malaysia, indicated that the doubling of access point frequency from 10 to 20 per kilometre increased the crash rates by roughly thirty percent. Poor access controlled highways have much greater road crash rates than the well-controlled highways (Mitra *et al.*, 2017). Figure 2.20 shows the impact of access points per km on crash rates on the roads in Tennessee (Mitra *et al.*, 2017). Using macro level analysis method, Jinghui & Xuesong (2018) also found a linear relationship between access density and road crash rates on road sections in Shanghai.

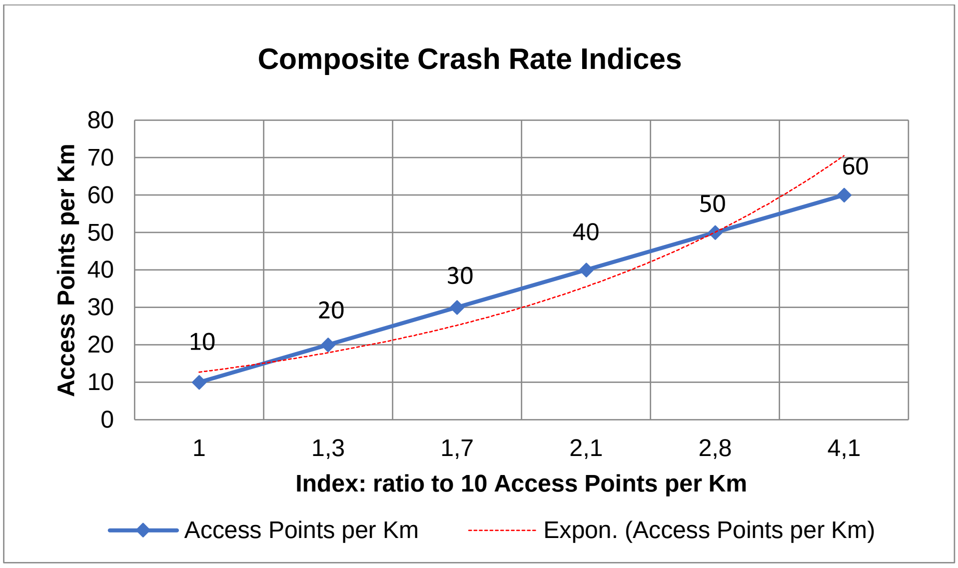


Figure 2.20 Impact of road access points per km on crash rates (Mitra, Haque and King, 2017)

2.7.8 Pavement condition

High traffic volumes, appalling weather conditions and bad ground conditions expose the road surface to wear and tear (rutting, cracks, and road unevenness) and create hazardous surface conditions that reduce riding comfort and can consequently lead to road crashes (Mohammed *et al.*, 2017). The pavement surface is often described using several key pavement surface condition indicators; International roughness index (IRI); pavement serviceability index (PSI); condition score; and ride score (Ghanbari, 2017).

IRI relates road roughness to the overall road surface condition (Titi *et al.*, 2018). Several thresholds for overall pavement condition in terms of IRI have been recommended, with Table 2.19 showing thresholds recommended by the Federal Highway Association.

Table 2.19 Thresholds for pavement condition using IRI (Federal Highway Administration, 2014)

Road classification	IRI unit	Category			
		Poor	Fair	Good	Excellent
All roads	m/km	IRI> 2.68	1.50<IRI≤ 2.68	IRI≤1.50	

The pavement Serviceability Index (PSI) is defined as a numerical index computed from objective measurements of certain types of pavement surface characteristics and indicative of the pavements ability to safely serve traffic at any particular point in the pavements service life (AASHTO, 2010). The PSI scale has a rating ranging from 0 to 5, with 0 to 1 rated as very poor and 4 to 5 rated as very good. Chan et al 2009 notes that the minimum acceptable level of PSI ranges from 2.5 to 3.

Li & Huang (2015) state that the condition score describes the average person’s opinion of the condition of the pavement by combining several factors into a single value; measurements of ride quality; average daily traffic; distressing ratings; and speed limit. The score attributes for the different pavement conditions are shown in Table 2.20.

Table 2.20 Condition score categories (Li & Huang, 2015)

Pavement condition	Condition score scale
Very poor	1-49
Poor	50-69
Fair or good	70-89
Very good	90-100

The ride score pavement condition method describes the overall ride quality of the road section (Li & Huang, 2015). The ride scores rating ranges between 0.1 (rough) and 5.0 (smooth) calculated as the length-weighted average of the raw serviceability index values measured from road data. The ride score pavement condition categories are depicted in Table 2.21.

Table 2.21 Ride score pavement condition scale (Li & Huang, 2015)

Pavement condition	Ride score scale
Rough	0.1-2.5
Fair	2.6-3.5
Smooth	3.6-5

Several studies have quantitatively investigated the impact of pavement condition on road crashes using pavement condition indicators (Ghanbari, 2017; Cenek *et al.*, 2012; Tehrani & Falls, 2015). A study by Ghanbari (2017) found that the roughness of the road affects the riding quality experienced by drivers. It can also lead to hazardous situations: Pavement roughness[23] has been found to influence driver steering capabilities by changing the normal forces that act at the tire-pavement interface, therefore negatively affecting the lateral forces required to control a vehicle (Chan *et al.*, 2008).

A study by Cenek *et al.* (2012) found that road roughness can also cause significant loss of braking force or slip resistance on a vehicle. As the impact of road roughness can vary on the wheels of the vehicles, this exposes the vehicle to different levels of friction on each side. Differential friction significantly affects vehicle braking and can lead to incongruous conditions for all road users (Ghanbari, 2017).

Ghanbari (2017) concluded that attempts to execute a turn having a small radius by vehicles traveling at high speeds on a rough road leads potentially dangerous safety situations. Such

[23] Pavement roughness is defined in accordance with ASTM E867 as the deviation of the surface from the true planar surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamic loads and drainage (ASTM Standard E867-06, 2017 cited in Federal Highway Administration, 2014).

dangerous conditions exist on a straight road section when a vehicle attempting to overtake at a high speed suddenly has to return to its original lane due to oncoming traffic. The driver attempting to overtake may potentially lose control of the vehicle.

Similarly, a study by King (2014), investigating the effect of road roughness on traffic speed and road safety in Australia, found that a statistically strong relationship exists between increased pavement roughness and higher crash rates and severity levels on road sections as illustrated in Figure 2.21. King (2014) also observed that passenger vehicles experienced a higher likelihood of being involved in road crashes than heavy commercial vehicles when pavement roughness increase. Moreover, a reduction in operating speeds when the pavement roughness increased was detected on some road sections (King, 2014). Li *et al.* (2013) suggested purposefully laying down rougher pavements on high speed roadways as a potential solution to address higher severity crashes.

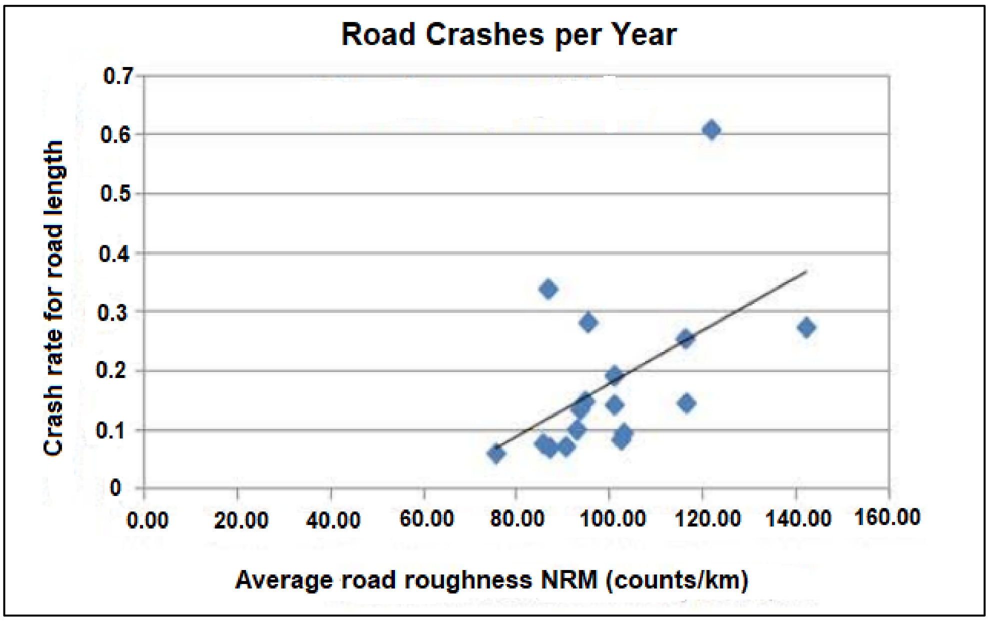


Figure 2.21 Pavement roughness vs road crash rates (King, 2014)

Cairney & Bennet (2008) found a good correlation between crash rates and pavement roughness following a polynomial relationship. However, no clear relationship was found between road rutting and road crash rates. A study by Li *et al.* (2013) in Texas found that relatively higher severity crashes occurred on roads with very good pavement conditions as illustrated in Figure 2.22. The higher severity crashes were attributed to the higher speed impact crashes on roads with very good pavement conditions (Li *et al.*, 2013).

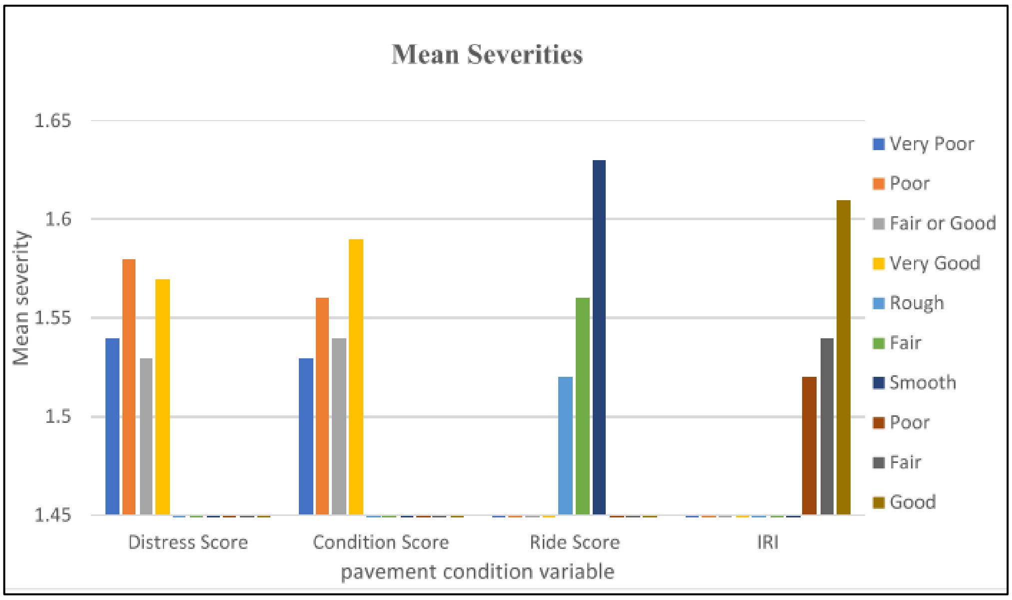


Figure 2.22 Mean severities for several pavement indicator groups (Li *et al.*, 2013)

A study by Tehrani & Falls (2015) investigated the relationship between IRI values and road safety in Canada. Road sections with high IRI values were observed to have a higher crash probability to those with low IRI values. In addition, a statistically significant correlation was found between the crash frequency and rut depth. In contrast, a study Cenek *et al.* (2012) found no significant relationship between IRI values and the likelihood of crashes occurring.

Several studies in the literature have concluded that the condition of the pavement significantly contributes to the safety of road users. Significant correlations were found to exist between pavement roughness and crash rates, while the contribution of rut depth to road safety was not well defined in literature.

2.8 Geographical Information System (GIS) tools for analysing road crashes and road design

Areas with concentrated crashes are often referred to as crash hotspots (Toran & Moridpour, 2015; Thakali *et al.*, 2015). The detection of clustering patterns of road traffic crashes has been enabled by both the effective application of Geographical Information System (GIS) in transportation research areas and by the opportunity given by Global Positioning System (GPS) with regard to spatial accuracy localisation of road traffic crashes (Hashimoto *et al.*, 2016; Satria and Castro, 2016; Ghadi & Török, 2017). The primary reason behind employing spatial techniques for the detection of road crash hotspots rather than classical statistical techniques, is that road crashes are a spatial phenomenon (Yalcin, 2013; Choudhary *et al,* 2015).

Spatial methods employed for the identification of road traffic crashes clustering patterns produce two kinds of results. The first one is identifying the global clustering tendency of road crashes within a road section, which includes the Quadrat methods (Ouni and Belloumi, 2019), the Nearest Neighbour methods (Satria and Castro, 2016; Afghari, 2018) and the K-function (Shafabakhsh *et al.*, 2017; Ouni & Belloumi, 2018). The second result is identifying the local cluster tendency of the crashes within a road section, which includes Kernel Density Estimation (KDE) (Kundakci, 2014; Toran & Moridpour, 2015; Pljakić *et al.,* 2019) and spatial autocorrelation approaches such as local Moran (Getis & Ord, 2010; Pirdavani *et al.,* 2014) and Getis-ord indices (Songchitruksa & Zeng, 2010; Aghajani *et al,* 2017).

In this section, spatial techniques in GIS employed to analyse road crashes are presented. Spatial analysis is used to geographically specify the road crash locations and to assess specific patterns of crash distribution through the visualisation of raster maps.

2.8.1 Kernel density

Kernel Density Estimation (KDE) is a spatial data analysis method in QGIS (Satria and Castro, 2016; Pljakić *et al.,* 2019). KDE is employed to determine the risk spread of road crashes by computing the number of crash incidents in a defined region or road network (Kundakci, 2014; Hashimoto *et al.,* 2016). The spread of crash risk can be defined as the area around the cluster where crash risk may increase due to a road crash. KDE is considered in two forms: (i) planar Kernel density Estimation (PKDE) and (ii) network Kernel Density Estimation (NKDE), an extension of the standard KDE (Ouni & Belloumi, 2018; Pljakić *et al.,* 2019).

The standard KDE applies the Euclidian distance measure in a continuous planar space by analysing hotspot locations (Thakali, Kwon and Fu, 2015). A study by Truong & Somenahalli (2011) employed KDE and spatial autocorrelation approach to identify and rank pedestrian-vehicle crash locations and unsafe bus stops in Adelaide, Australia. The study identified 3 and 10 pedestrian-vehicle hotspots at intersections and mid-block locations respectively.

NKDE employs network distance measure along a roadway while analysing hotspot locations (Ouni and Belloumi, 2018). A study by Benedek *et al.* (2016) employed the NKDE to identify vulnerability areas for road crashes in Cluj-Napoca in Romania. The results indicated that the majority of the vulnerable areas for road crashes were located at the entrances and exits of the city. While there are a variety of KDE features to choose from, several studies (Toran & Moridpour, 2015; Ghadi & Török, 2017; Ouni & Belloumi, 2018) have suggested that the Kernel function has no significant impact on the density pattern. The density pattern is influenced by the choice of bandwidth, with several optimal bandwidth variation intervals ranging from 200m to 1000m applied depending on the aim of the study (Hashimoto *et al.,* 2016; Cheng *et al.,* 2018; Pljakić *et al.,* 2019). A bandwidth between 200m and 400m is recommended for urban road networks (Kundakci, 2014; Shafabakhsh *et al.*, 2017), while a bandwidth between 600 to 1000m is recommended for rural road networks to produce a raster output (Toran & Moridpour, 2015; Pljakić *et al.,* 2019).

2.8.2 Moran's Index Statistic

Moran's Index (MI) is a statistical tool measuring the spatial dependence of the road crash location (Moran, 1948). Moran's index method is based on the covariance relationship of the statistical correlation coefficient (Satria and Castro, 2016; Cheng *et al.,* 2018). Moran's Index can be described with Equation [2.5].

$$I = \frac{n}{S_0} \cdot \frac{\sum_i^n \sum_{j=1}^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i^n (x_i - \bar{x})^2} \qquad [2.5]$$

Where, x_i, x_j denote the ith and jth spatial observed value respectively, $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i w_{ij}$ represents the elements of a spatial binary contiguity matrix and computes whether neighbourhood relationships exist between location i and its adjacent location j. S_0 refers to the summation of all elements of w_{ij}.

A single value for the spatial correlation and checking the clustering of the road crash spatial pattern is provided by Moran's Index (Satria and Castro, 2016). The statistical inference on Moran's Index applies the calculated value and both z-score and p-value to evaluate if the spatial road crash pattern clusters observed are dispersed or random and determines the level of concentration (Songchitruksa & Zeng, 2010; Satria and Castro, 2016), as illustrated in Figure 2.23.

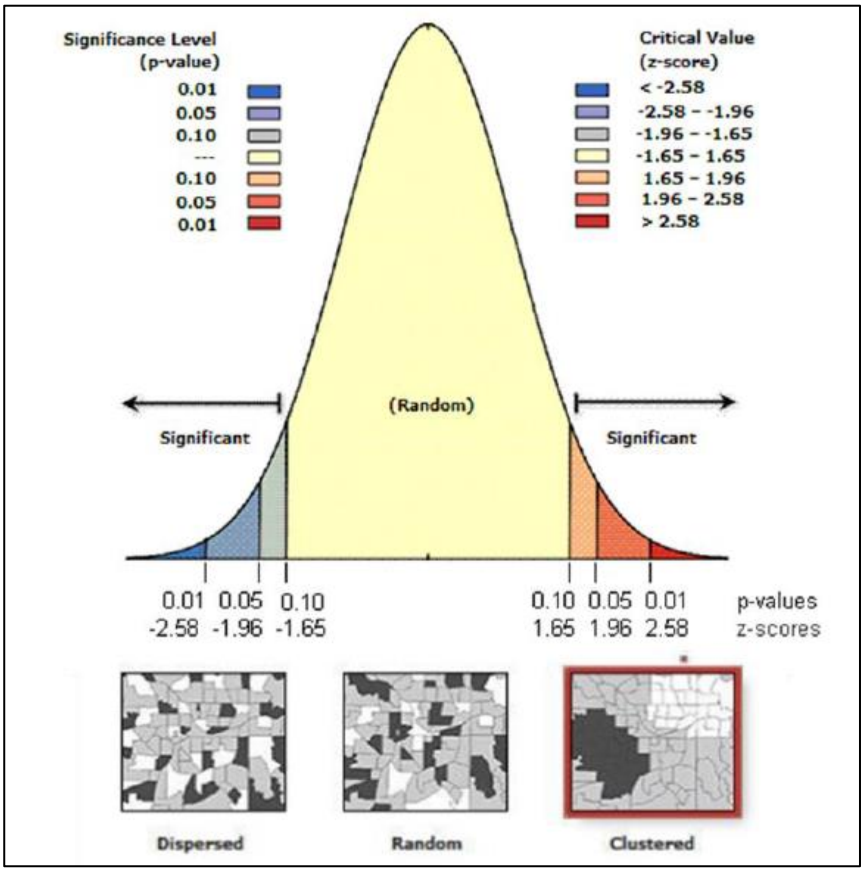


Figure 2.23 Z-scores and p-values interpretation by Moran's Index (Moran, 1948)

Dense locations of the proximity between two points are often defined as the inverse of the distance between them (Satria and Castro, 2016; Gomes *et al.,* 2017). The difference between each value and the value of the global Moran's Index average is known as the attribute similarity severity index of the two points (Truong and Somenahalli, 2011). A study by Pirdavani *et al.* (2014) developed road crash prediction models (CPMs) using geographically weighted regression. The CPMs were developed by computing Moran's Index for the dependent and selected explanatory variables. The results illustrated the necessity of considering spatial correlation when developing CPMs, as this provides an insight of the spatially varying relationship between crashes and related factors through the CPM estimated values.

2.8.3 Getis-Ord

G statistics are a family of statistics with a number of attributes that make them attractive to measure the inter-dependence of spatially distributed variables, especially when applied in conjunction with Moran's Index (Songchitruksa & Zeng, 2010; Aghajani *et al.,* 2017). G Statistics deepen the knowledge and understanding of the process that leads to spatial dependency and improve the detection of local 'pockets' dependence that may not appear using global statistics (Arthur, 1995; Getis & Ord, 2010). The Getis-Ord statistics are utilised to identify road crash hotspot locations. The Getis-ord statistics for each feature in the data set are calculated by the hotspot analysis

(Songchitruksa & Zeng, 2010; Saha & Ksaibati, 2016). A high value of the Getis-ord statistic represents a group of high index values (hotspots), while a low value represents a low index value (Getis & Ord, 2010; Satria and Castro, 2016).

2.9 Review of statistical modelling tools

Crash prediction is a crucial step in the management of road safety processes. The Highway Safety Manual suggests the use of safety performance functions (SPF) while predicting crash frequencies on different types of roads and also considers that the SPFs will vary significantly with the change of road environment – the road geometry, road traffic, road side environment (AASHTO, 2010). However, the manual is appropriate only for road segments of homogenous characteristics, expressed in terms of traffic volumes and roadway design characteristics. As a result, there is a need to develop indigenous crash prediction models, aimed at predicting crashes in developing countries where heterogeneity in traffic composition is prevalent (Basu and Saha, 2017; Ambros *et al.*, 2018).

Two potential events are likely to result from a road crash: a non-zero event (fatality) or a zero-event (non-fatality). In some cases, road crash may result in zero fatalities, hence this can result in an excess number of zeros in a crash dataset (Imprialou *et al.*, 2016). The Poisson regression model is the simplest model applied to count data. As count data may exhibit over-dispersion (or in instances zero-inflated data -excess zeros), Poisson regression models are limited by the assumption that data exhibits equal mean and variance. In such a case, this shortcoming is addressed by applying Negative Binomial regression (NB), which largely belongs to a family of Generalised Linear Models (GLMs) (Mannering and Bhat, 2014; Kiranet *et al.*, 2017). Even though NB models are capable of handling over-dispersion quite well, they may not be sufficient in addressing zero-inflated data. The issue of captured excess zeros is addressed through using zero-augmented models (zero-inflated models) and Hurdle models (Ridout *et al.*, 1998; Imprialou *et al.*, 2016). Zero-inflated models are a mixture of models that combine a count component and a point mass at zero, while Hurdle models combine a left-truncated count component with a right-censored hurdle component (Saffari & Adnan, 2011; Saffari *et al.*, 2012). Robust multiple linear regression models (MLR) have also been found to accommodate both crash rates and crash count data. The MLR approach involves data aggregation to satisfy linear regression assumptions; namely error structure normality and homoscedasticity. The robust MLR technique has been found to generate crash predictions consistent with traditional NB and zero-augmented NB GLMs (Rakha *et al.*, 2010; Mohammed *et al.*, 2018; Islam *et al.*, 2019). These models are described in detail in this section.

2.9.1 Poisson regression

Poisson regression is a traditional basic count model on which numerous other count models are based (Montgomery and Runger, 2014). Poisson models are some of the most popular when modelling count data. The Poisson distribution is the starting point for Poisson regression, shown by Equation [2.6].

$$H_{Y_i}(y_i) = \frac{e^{\mu_i}\mu_i^{y_i}}{y_i^!} \qquad [2.6]$$

The logarithm of the mean of Poisson distribution (μ_i) is assumed to be a linear function of the independent variable x_i given by Equation [2.7].

$$\log(\mu_i) = x_i\beta \qquad [2.7]$$

Where: y_i denotes the dependent variable having a Poisson distribution

x_i denotes the independent variables

Suppose the dependent variable (Y_i) is a count response variable that follows Poisson distribution, the probability of Y_i can be modelled as detailed by the Equation [2.8].

$$f_i(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)\frac{(1+\alpha y_i)^{y_i^{-1}}}{y_i^!}exp\left(\frac{-\mu_i(1+\alpha y_1)}{1+\alpha\mu_i}\right) \qquad [2.8]$$

Where: y_i = 0,1, 2,…, n

$\mu_i = \mu_i(X) = e^{XB}$, where X is a $(k-1)$ dimensional vector of covariates and B is a $k-$ dimensional vector of regression parameters.

α = is the dispersion parameter

The dispersion parameter is observed in three dispersion cases; case (1) equi-dispersion; when $\alpha = 0$. Hence Equation [2.8] reduces to PR, case (2) over-dispersion; when $\alpha > 0$. Equation [2.8] thus adds to one and case (3) under-dispersion; when $\alpha < 0$. Here, Equation [2.8] gets truncated and may not add up to one. Therefore, the variance and mean of the response variables in Poisson regression are given by Equation [2.9] and equation [2.10] respectively.

$$V(Y_i|x_i) = \mu_i(1+\alpha\mu_1)^2 \qquad [2.9]$$

$$(\mu_i) = E(Y_i|x+i) \qquad [2.10]$$

2.9.2 Negative Binomial regression

The crash counts are mostly characterised by a Negative Binomial (NB) distribution. The NB distribution is a core part of Negative Binomial regression models, given by Equation [2.11].

$$f(y_i) = p(Y_i = y_i) = \frac{\Gamma(\theta+y_i)}{\Gamma(\theta)y_i}\left(\frac{\theta}{\theta+\mu_i}\right)^{\theta}\left(\frac{\mu_i}{\mu_i+\theta}\right)^{y_i}, y_i = 0,1,2, \dots, n \qquad [2.11]$$

Where; $\theta = \frac{1}{\alpha}$. α is the dispersion parameter. $\Gamma(.)$ is the gamma function. The dependant variable denoted by Y_i has a Negative Binomial distribution with two parameters $\mu_i \geq 0$ and $\theta \geq 0$, with the mean and variance denoted by Equation [2.12] and Equation [2.13] respectively.

$$E(Y_i) = \theta\mu_i \qquad [2.12]$$

$$var(Y_i) = E(Y_i)(1+\mu_i) = \theta\mu_i(1+\mu_i) \qquad [2.13]$$

2.9.3 Generalized Linear Models (GLMs)

GLMs are a set of statistical modelling tools used when the dependant variable violates the integral assumption of linearity (Ridout *et al.*, 1998; Bagha & Madisetti, 2019). In that event, the dependant variable does not follow a normal distribution. Hence, GLMs that assume a link linear relationship based on a chosen link function are utilised to complete analyses (Montgomery & Runger, 2014; Bruce & Bruce, 2017). Statistical GLMs are a vital member of the exponential family, which take the form shown in Equation [2.14].

$$f(y_i;\theta;\phi) = exp\left[\frac{(y\theta - b(\theta))}{a(\phi)} + c(y,\phi)\right] \qquad [2.14]$$

Therefore, statistical GLMs can be written as shown in Equation [2.15]

$$\mathrm{y} = \mathrm{g}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + e \qquad [2.15]$$

Where; y is the vector of the dependant variables counts; x_i are linearly associated covariates; β_i represents the regression coefficients; e is the error variable unaccounted for by the covariates x_i; g is a monotonic function linking the mean of the dependant variable to linear covariates and other functions.

The values of the regression coefficients β_{i-n} , related to the covariates through Equation [2.16] and Equation [2.17] respectively, are estimated by the Maximum Likelihood (ML) estimations.

$$E(Y_i) = \mu_i \qquad [2.16]$$

$$\mathrm{g}(\mu_i) = X^T\beta \qquad [2.17]$$

For each Y_i, the log-likelihood function is given by Equation [2.18].

$$l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i) \qquad [2.18]$$

Where; the functions for b, c and d are known and linked to Equation [2.16], Equation [2.17] and Equation [2.18], through Equation [2.19], Equation [2.20] and Equation [2.21].

$$E(Y_i) = \mu_i = C'(\theta_i)/b'(\theta_i) \qquad [2.19]$$

$$Var(Y_i) = [b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)]/[b'(\theta_i)]^3 \qquad [2.20]$$

$$\text{g}(\mu_i) = X^T\beta = \eta_i \qquad [2.21]$$

Where; X is a vector with elements $x_{ij}, j = 1,2, \dots, n$. Hence, the function for the log-likelihood for all the Y_i variables is given by Equation [2.22].

$$l = \sum_{i=1}^{n} l_i = \sum y_i\, b(\theta_i) + \sum c\,(\theta_i) + \sum d\,(y_i) \qquad [2.22]$$

To obtain the maximum likelihood estimate for parameter β_j in Equation [2.23], the chain rule for differentiation is used by considering each term on the right-hand side to obtain Equation [2.24]. The variance-covariance of the U_j matrix portrays the terms in Equation [2.25].

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{n}\left[\frac{\partial l_i}{\partial \beta_j}\right] = \sum_i^n \left[\frac{\partial l_i}{\partial \theta_i}.\frac{\partial \theta_i}{\partial \mu_i}.\frac{\partial \mu_i}{\partial \beta_j}\right] \qquad [2.23]$$

$$U_i = \sum_i^N \left[\frac{(y_i - \mu_i)}{var(Y_i)}.x_{ij}.\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\right] \qquad [2.24]$$

$$\mathfrak{I}_{jk} = E\left[U_j.U_k\right] \qquad [2.25]$$

The formula for the maximum likelihood estimation is thus given by Equation [2.26]

$$b^{(m)} = b^{(m-1)} + \left[\mathfrak{I}^{(m-1)}\right]^{-1} U^{(m-1)} \qquad [2.26]$$

where the difference between $b^{(m)}$ and $b^{(m-1)}$ is considered to be insignificant.

2.9.3.1 Generalized Poisson Regression Model

The study will use the GLMs for the estimation of the dependant variable. The parameter μ will be used to express the mean instead of the parameter λ, which is used in most literature to express the mean (Barua *et al.*, 2016). The Poisson regression given by Equation [2.27] is used to model the relationship between the dependant variable and explanatory variables.

$$H_{Y_i}(y_i) = \frac{e^{\mu_i}\mu_i^{y_i}}{y_i^!};\ y_i = 0,1,2,\ldots.,n; \mu_i > 0 \quad [2.27]$$

Where; $\mu_i = \exp(X'\beta)$ is the models fitted mean; X is the vector of the covariates and Y_i is the dependant variable counts. The Poisson distribution assumes equal mean and variance. The dependant variable Y_i is modelled as shown in Equation [2.28] should the count data follow Poisson distribution.

$$f_i(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right).\frac{(1+\alpha y_i)^{y_i-1}}{y_i^!}.exp\left(\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right), y_i = 0,1,2,\ldots,n \quad [2.28]$$

Where; $\mu_i = \mu_i(X) = e^{XB}$; X is a $(k-1)$ dimensional vector of covariates; B is a $k-$ dimensional vector of the regression parameters and α is the dispersion parameter.

The dispersion parameter α occurs in three observed forms; Case (1) Equi-dispersion, with $\alpha = 0$ and Equation [2.28] reduced to PR; Case (2) Over-dispersion, with $\alpha > 0$ and Equation [2.28] always adding up to one; and Case (3) Under-dispersion, with $\alpha < 0$ and Equation [2.28] getting truncated, therefore, may not add up to one. The variance and mean of the dependant variable Y_i are given by Equation [2.29] and Equation [2.30] respectively.

$$V(Y_i|x_i) = \mu_i(1+\alpha\mu_i)^2 \quad [2.29]$$

$$\mu_i = E(Y_i|x+i) \quad [2.30]$$

2.9.3.2 Generalized Negative Binomial Regression Model

In cases where the count data exhibits significant differences between variables, causing the variance to be greater than the mean over-dispersion) or less than the mean (under-dispersion), models such as the Generalized Negative Binomial Regression Model (NBR) are preferred because of their accuracy (Ridout *et al.*, 1998; Montgomery & Runger, 2014). Poisson regression models, which may exhibit severe drawbacks limiting their use (mean assumed to be equal to the variance) are often shunned in this case. Moreover, Poisson distribution has one variable parameter, compared to the Negative Binomial distribution with two parameters (Ho, 2006; Montgomery & Runger, 2014). Hence, the Negative Binomial regression is considered more flexible than Poisson regression. The count data response variable Y_i is determined by the Negative Binomial Regression model using Equation [2.31].

$$f(\gamma_i) = p(Y_i = \gamma_i) = \left(\frac{\Gamma(\theta+\gamma_i)}{\Gamma(\theta)\gamma_i}\right).\left(\frac{\theta}{\theta+\mu_i}\right)^{\theta}.\left(\frac{\mu_i}{\mu_i+\theta}\right)^{\gamma_i}, \gamma_i = 0,1,2,\ldots,n;\ n < \infty \quad [2.31]$$

Substituting $v_i = \frac{\theta}{\theta+\mu_i}$, Equation [2.31] is thus replaced by Equation [2.32].

$$f(\gamma_i) = p(Y_i = \gamma_i) = \left(\frac{\Gamma(\theta+\gamma_i)}{\Gamma(\theta)\gamma_i}\right).(v_i)^{\theta}.(-(v_i-1))^{\gamma_i} \qquad [2.32]$$

Where; $\theta = \frac{1}{\alpha}$; α is the dispersion parameter; $\Gamma(.)$ is the gamma function.

The dependant variable parameter Y_i has a Negative Binomial distribution with two parameters, $\mu_i \geq 0$ and $\theta \geq 0$. Therefore, the mean and variance are given by Equation [2.33] and Equation [2.34] respectively.

$$Mean = E(\gamma_i) = \theta\mu_i \qquad [2.33]$$

$$Variance = Var(Y_i) = E(Y_i)(1+\mu_i) = \theta\mu_i(1+\mu_i) \qquad [2.34]$$

Although Poisson and NBR models are recommended for count data modelling. In instances were a dataset has an inflated occurrence of zeros, Zero-Inflated models are recommended for analyses.

2.9.3.3 Zero-Inflated Poisson Regression Model

The presence of excess zero cases can result in an over-representation of these cases in estimated models. Zero-Inflated models that address excess zeros in count datasets are recommended as one of the alternatives for a better goodness-of-fit (Field, 2013; Mannering and Bhat, 2014). Using the Zero-Inflated Poisson Regression model (ZIP), the dependant variable $(Y_i = 0)$ with probability γ_i is assumed to follow a Poisson distribution with mean μ_i and probability $1-\gamma_i$. The variable had a distribution with two components; a zero $(\gamma_i = 0)$ and non-zero component $(\gamma_i \neq 0)$, given by Equation [2.35] and Equation [2.36] respectively.

$$\Pr(Y_i = 0) = \gamma_i + (1-\gamma_i)^{-\mu_i} \qquad [2.35]$$

$$\Pr(Y_i = r) = (1-\gamma_i)e^{\frac{(-\mu_i)(\mu_i)^r}{r!}}, r = 1,2,3,\ldots,n \qquad [2.36]$$

The a mean and variance for the dependant variable can be determined using Equation [2.37] and Equation [2.38] respectively.

$$E(\gamma_i|x_{i,}z_i) = \mu_i\theta, i = 1,2,3,\ldots,n \qquad [2.37]$$

$$V(\gamma_i|x_{i,}z_i) = \mu_i(1-\gamma_i)(1+\mu_i\gamma_i), i = 1,2,3,\ldots,n \qquad [2.38]$$

γ_i and μ_i are expressed explicitly as functions of the explanatory variables to assess the extent of the link between the covariates and the dependant variable in ZIP. Therefore, the logistic regression model given by Equation [2.39] is applied as the standard method to model the probability of excess zeros in the count data.

$$\log it(\gamma_i) = XB \quad [2.39]$$

Where; X is the covariates (x_i) vector and B is a vector is a parameter β vector.

The effect of the explanatory variables on the dependant variable, excluding the excess zeros in the count data, can be modelled using Poisson distribution given in Equation [2.40].

$$log(\mu_i) = Z\delta \quad [2.40]$$

Where; parameters Z and X are $s-$ and $w-$ dimensional explanatory variable vector, whereas δ and B are corresponding regression coefficient vectors.

2.9.3.4 Zero-Inflated Negative Binomial Regression Model

Zero-Inflated Negative binomial (ZINB) models are a combination of distributions assigning the mass of $1-y$ and y to a Negative Binomial distribution and excess zeros respectively, with $0 \leq \gamma \leq 1$ (Saffari & Adnan, 2011; Kiran *et al.*, 2017). The ZINB distribution is given by Equation [2.41].

$$P(Y_i = r) = \begin{cases} \gamma + (1-\gamma)\left(\frac{\theta}{\theta+\mu}\right)^{\theta}, r = 0 \\ (1-y).\left(\frac{\Gamma(\theta+r)}{r!\Gamma(\theta)}\right).\left(\frac{\theta}{\theta+\mu}\right)^{\theta}.\left(\frac{\mu}{\mu+\theta}\right)^{r}, r = 1,2,3,\ldots,n \end{cases} \quad [2.41]$$

The dependant variable mean and variance are determined using Equation [2.42] and Equation [2.43] respectively.

$$E(Y) = (1-\gamma)\mu \quad [2.42]$$

$$Var(Y) = (1-\gamma)\mu(1+\gamma\mu+\frac{\mu}{\theta}) \quad [2.43]$$

When $\frac{1}{\theta} \approx 0$ and $\mu \approx 0$ Equation [2.41] reduces to the Poisson distribution. Equation [2.41] also approaches the Zero-Inflated Poisson as $\theta \to \infty$. Similarly, Equation [2.41] approaches the Negative Binomial distribution as $\gamma \to 0$. Parameters γ and μ related to the explanatory variables by the ZINB regression model through Equation [2.44] and Equation [2.45].

$$\log(\mu_i) = XB \quad [2.44]$$

$$\text{logit}(\gamma_i) = Z\delta \quad [2.45]$$

Where parameters Z and X are $s-$ and $w-$ dimensional vectors of the explanatory variables, while δ and β are corresponding vectors of regression coefficients.

2.9.3.5 Generalized Poisson Hurdle Model

The flexibility of the Generalized Poisson Hurdle Model (PHM) enables it to be applied to either over or under-dispersed count data. The PHM is considered valuable due its application of the hurdle at zero (Cameron & Trivedi, 1998; Mannering and Bhat, 2014). Similarly, it is suitable for application on positive dichotomous response variables, considering zero $(Y_i = 0)$ and non-zero cases $(Y_i \neq 0)$. The probability function of the PHM is given by Equation [2.46].

$$P(Y_i = k)\ for\ i = 1,2,3, \dots, n = \begin{cases} h_1(0), \mathrm{k} = 0 \\ (1 - h_1(0))h_2(k), k \geq 1 \end{cases} \quad [2.46]$$

Where; $h_1(0)$ is the probability value when a zero count exists and $h_2(k)$, for $k \geq 1$, is the probability value when a non-zero count exists. Should a significantly higher zero case be observed in the dataset than can be modelled by Equation [2.46], Equation [2.47] and equation [2.48] are recommended.

$$P(Y_i = 0) = 1 - q_i;\ 0 \leq q_i \leq 1 \quad [2.47]$$

$$P(Y_i = r) = q_i . \left(\frac{\mu^r e^{-\mu}}{r!(1 - e^{-\mu})}\right), r = 1,2,3, \dots, n; 0 < n < \infty \quad [2.48]$$

Where; q_i is the element that models all zero cases and μ represents the mean of the truncated Poisson distribution. In addition, the probability of the zero count cases can be modelled by applying the logistic regression model given in Equation [2.49].

$$\log q_i = XB \quad [2.49]$$

2.9.3.6 Hurdle Negative Binomial model

The Hurdle Negative Binomial Model (HNB) is a two-part model applied in breaking down the dependant variable Y_i into two observed random regression components, given as $y_i > 0$ and $Y_i|y_i > 0$ (Saffari *et al.*, 2012; Mannering and Bhat, 2014). The HNB is structured as shown in Equation [2.50] and Equation [2.51].

$$P(Y_i = 0) = 1 - q_i; 0 \leq q_i \leq 1 \quad [2.50]$$

$$P(Y_i = r) = q_i \left(\frac{\Gamma(r + \theta)}{r!\Gamma(\theta)}\right)\left(1 + \frac{\mu}{\theta}\right)^{-r}; r = 1,2,3, \dots, n, 0 \leq \mu < \infty \quad [2.51]$$

Where; μ is the mean parameter and θ represents over-dispersion.

2.9.4 Robust Multiple Linear Regression Modelling Approach

The linear regression model development approach is considered in this section. Using the crash rate as the dependant variable, the variable model can be written as shown in Equation [2.52].

$$W(CR) = \frac{(365 \times 5)^p}{10^6} \cdot \exp(\beta_0 + \beta_1 L_1 + \cdots + \beta_n L_n + E) \qquad [2.52]$$

Where $W(CR)$ is the dependant variable and represents the Winsorized crash rate. The regression constant in the model equation is represented by β_0. The terms β_1 to β_n represent the model coefficients for the respective covariates. The terms L_1 to L_n represent the independent variables. Here, E is a random error term that accounts for the error that is not captured in the model (Rakha *et al.*, 2010; Schmidt *et al.*, 2012; Mohammed *et al.*, 2018).

The analysis using crash rates ensures that the data are normalised across the different road sections. The development of the MLR using the least squares approach requires that the data follow a normal distribution (Karlaftis and Golias, 2002). The approach for applying the robust MLR to the data involves sorting data based on one of the variables and then aggregating the data using a variable bin size to ensure that another variable remains constant across the variable bins (Hicks and Fetter, 1991; Schmidt *et al.*, 2012). Data transformation can then be applied to the data to ensure normality and equal variance.

2.10 Road crash modelling and analyses techniques

In recent years, numerous studies have been conducted with the goal of developing crash predictive tools for roadway facilities for rural highways. Statistical road safety modelling is defined by Hauer (2014) as the fitting of a statistical model to data; namely crash data and characteristics of roadways and traffic. A wide range of statistical models frequently applied include multiple linear regression, negative binomial, Poisson, binomial, zero-inflated Poisson (ZIP) and negative binomial ZINB), and geographically weighted regression (GWR) models (Ayati & Abbasi, 2014; Arani *et al.*, 2017; Mohanty & Gupta, 2015). Notably, the ZIP and ZINB models are applied to account for the preponderance of excess zero's observed in crash data (Miranda-Moreno *et al.*, 2007). The Highway Safety Manual (AASHTO, 2010) notes that identifying the appropriate statistical model for the type of crash data is vital to addressing the road safety issues by estimating consistent and representative parameter estimates.

2.10.1 Crash modelling: Global perspective

Statistical relationships between road crashes, design elements and traffic conditions on the roadways have been extensively modelled and evaluated in recent years. El-basyouny and Sayed (2009) state that the application of crash prediction models in assessing the safety of road infrastructure has become a standard practice among road safety stakeholders globally. Dwikat (2014) mentions that the development and use of crash prediction models in identifying crash hotspots has been crucial in improving the road safety condition of roads worldwide. Studies by Rakha *et al.* (2010) and Rogers (2003) have respectively investigated the use of robust multiple linear regression (MLRs) and generalized linear regression models (GLM) to quantify associations between explanatory variable and road crashes, with all models adopted showing an acceptable level of goodness of fit and over-dispersion.

Recent studies have challenged the underlying statistical assumptions adopted in popular models for road crash modelling (Lord and Ivan, 2006; Miaou and Lord, 2007; Saha and Ksaibati, 2016). First, the assumption that the dispersion parameter is a fixed parameter across sites and time periods is challenged (Miranda-Moreno *et al.*, 2007). Second, an examination of the mathematical limitations of some functional forms and their properties at the boundaries demonstrated that for a given set of data, a large number of plausible functional forms with almost the same overall statistical goodness of fit is possible (Murthy and Rao, 2015). This allows for an alternative class of logical formulations that enable a richer interpretation of the data to be introduced (Miaou and Lord, 2007).

A distinction is made between the crash prediction models that use multivariate explanatory variables to predict a univariate dependant variable and those that involve multivariate dependant and independent variables (El-basyouny & Sayed, 2009; Ho, 2006). The former are termed as univariate crash prediction models, while the latter are termed multivariate crash prediction models. Kockelman

(2006) states that it should be noted that both univariate and multivariate crash prediction models use multiple covariates. Univariate analysis methods are used to study trends in a data set by determining the central tendency measures and dispersion values (Bruce & Bruce, 2017; Saxena *et al.*, 2006). Multivariate analysis methods, particularly Generalized Linear Models (GLM), are widely used in the context of road safety (Songpatanasilp *et al.*, 2015). A GLM usually comprises three components: a random component, a linear function of the regression variables and an invertible link function (Oppong, 2012; Sisiopiku, 2011; Songpatanasilp *et al.*, 2015).

The standard Poisson regression model has been applied for modelling crash data, with the model assuming that the number of road crashes over a period of time are independently Poisson distributed (Mohanty and Gupta, 2015). The Poisson regression model is often restrained by the assumption that the mean and variance of the predicted variable are equal (Taylor, Lynam and Baruya, 2000). Miranda-Moreno et al. (2007) explains that the shortcoming of the Poisson regression models is caused by the vector covariates often not explained completely due to the conditional mean and omitted exogenous variables or randomness. This leads to a problem of over-dispersion caused by unmeasured heterogeneities (Cameron and Trivedi, 1998). Over-dispersion is addressed through capturing the random variables in the conditional mean of the Poisson model by introducing a random effect term in a multiplicative way (Deublein *et al.*, 2013). This leads to the development of mixed Poisson models, such as Poisson-lognormal and Poisson gamma (Negative Binomial) models (Miranda-Moreno *et al.*, 2007).

Similarly, Miaou and Lord (2007) note that the popular univariate approach for developing crash prediction models uses the Poisson-gamma hierarchy, which leads to the Negative Binomial regression model. The Poisson lognormal regression represents a viable alternative for modelling the extra–Poisson variation. Even though the majority of crash prediction models are developed using models with fixed dispersion parameters, Miranda-Moreno *et al.* (2007) challenged the use of fixed dispersion parameters by examining various dispersion parameter relationships in crash prediction models.

Lord and Ivan (2006) state that extensive research has been carried out to address the problem of observing the excessive zeroes in road crash data, in addition to the development of crash prediction models from data characterised by a low sample mean, especially if combined with a small sample size. Ayati and Abbasi (2014) note that various modelling techniques have been proposed in advocating the use of random parameter negative binomial regression models.

Several covariates have exhibited spatial dependency such as road and environmental characteristics across geographical areas (Satria & Castro, 2016). These covariates showed spatial heterogeneity and significantly influenced the estimation of model parameters. The spatial variations were addressed in the analyses by using geographically weighted Poisson and negative binomial regression models (GWPR and GWNBR) to investigate the influence of spatial dependent covariates on road crashes (Zheng *et al.*, 2011). The limited ability of traditional generalized linear models to take spatial effects into consideration can be overcome through the use of spatial regression techniques such as GWR models (Li *et al.*, 2013). Pirdavani *et al.* (2014) argue that despite GWR models addressing the spatial dependency of the covariates, they fail to account for the possible over-dispersion that can be found in events that occur independently and randomly over time such as road crashes.

Othman and Thomson (2007) state that model validation is an important step in crash prediction model development. The goal of model validation is not only to compare the accuracy of the different models developed, but also to evaluate the overall accuracy of crash prediction models for use in road safety (Semar, 2003). Butchart and Mikton (2014) stated that crash prediction model validation is required to demonstrate that a model is appropriate, meaningful and useful for the purpose it is intended. Road crash prediction models can be used as a quantitative tool to evaluate the impact of road design and traffic conditions on road safety (Al-Matawah, 2009).

2.10.2 Crash modelling: Namibian perspective

Only a very limited literature scope exists that explores the relationship between road design and traffic characteristics, and road crashes, using crash modelling techniques in Namibia. A study by Ambunda & Sinclair (2019) on the effect of two-lane two-way rural roadway design on road safety represents one of the first attempts to quantify the extent of the link between road crashes and road design on Namibian roads.

The study used multivariate analyses techniques, namely Poisson gamma (Negative binomial) regression models to statistically investigate the road safety relationships. Similar to Taylor *et al.* (2000), Ambunda & Sinclair (2019) determined that the dependent variable variance and mean were not equal, which violated the Poisson model condition that requires the mean and variance to be equal. This consequently invalidated the t-test parameter estimates. Poisson gamma models overcame this restriction caused by over-dispersion and provided the functional form crucial to linking road crashes to the investigated covariates (Ambunda and Sinclair, 2019).

2.11 Key conclusions from the literature

Improving road safety is one of the most vital objectives for transportation stakeholders. In order for safety to improve on roadways effectively and efficiently, stakeholders need to understand how the various complex factors are linked and affect road safety.

Existing literature that has attempted to establish and quantify the associations between road elements and road safety was reviewed in this Chapter, with a focus on traffic and road characteristics on single carriageways. Several conclusions were established on the varying extent of influence of road design elements on the safety of road users on the roadway. The following key conclusions were made on the impact of two-lane roadway elements on national rural roadways:

a) Uniform road design positively influences road safety on the roadway through communicating information needed for drivers to safely traverse road sections and to safely interact with other road users.
b) Stopping, passing and decision sight distances were reported as key safety components, as driver's ability to see ahead resulted in safe vehicle operations.
c) Speed and speed variations have a significant influence on the occurrence of road crashes. Higher speeds and speed variations were associated with sections with higher crash rates.
d) Mixed conclusions were found on the influence of lane widths and shoulder characteristics on road safety. Roads with narrower lane widths were association with lower driver speed selections and safer driving behaviour, while also having the highest risk for head-on crashes and single vehicle crashes. Narrower ground shoulder widths were associated with lower driver speed selections as the visual cues they communicated gave drivers the perception of a narrow road, which meant that space to correct driver errors was limited, leading to safer driving behaviour.
e) High vertical grades were found to result in higher crash incidences, especially on roads with higher heavy traffic composition.
f) Horizontal curve radii below the critical radius of 350m were strongly associated with high crash rates. The safety of the horizontal curves was found to increase with increasing curve radius. Curves length greater than 1000m were not recommended due to limited passing sight distance issues experienced by drivers.
g) The roughness of the pavement surface was found to contribute to the safety of drivers on the roadway. Good correlations were identified between higher surface roughness and higher road crash rates. Poor road surface conditions as compared to good road conditions were associated with a higher crash risk for road users.
h) Several statistical methods have been employed to investigate the significance of the association between road elements and road safety, influenced by the type of crash data and study location.

A summary of several of the key empirical studies along with the variables and modelling techniques are presented in Table 2.22.

Table 2.22 Previous literature on associations between rural roadway elements and road safety

Authors / Variables	Vayalamkuzhi & Amirthalingam (2016)	Maji *et al.* (2018)	Ghanbari (2017)	Mohammed *et al.* (2017)	Abele & Møller (2011)	Ben-Bassat & Shinar (2011)	Elvik *et al.* (2004)	Kockelman *et al.* (2008)	Aghayari *et al.* (2017)	Mitra *et al.* (2017)	Arani *et al.* (2017)	Rakha *et al.* (2010)
Covariates												
Geometric factors												
Lane width	✓	✓	✓	✓	✓		✓	✓				
Segment length		✓	✓	✓	✓					✓	✓	✓
Road alignment	✓	✓	✓		✓	✓		✓	✓	✓	✓	
Sight distance					✓	✓	✓					
Shoulder		✓		✓				✓				
Number of lanes	✓			✓					✓		✓	
Road side and environmental factors												
Access				✓				✓	✓	✓		
Pavement condition			✓				✓		✓			
Traffic factors												
Traffic volume	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Speed	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	
Percentage of heavy vehicles			✓		✓							
Dependant variables												
Crash frequency		✓	✓						✓	✓		
Crash occurrence	✓			✓		✓	✓	✓	✓	✓	✓	
Crash severity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Crash rate	✓				✓	✓	✓					
Model type												
Poisson regression			✓					✓				
Poisson gamma (NB)	✓		✓	✓						✓		✓
Robust Multiple Linear Regression												✓
Power model							✓					
Geographically weighted regression									✓		✓	
Others (Hurdle, etc.)		✓			✓	✓						

Chapter 3: Methodology

3.1. Introduction

This chapter discusses the methodology that was used to develop road crash prediction models and investigate the combination of effects of the road design and traffic environment on road crashes in the national rural road environment of Namibia. It provides an outline of the approach used to achieve the objectives of the study; first, through the collection and processing of quantitative and qualitative rural road traffic crash data and road characteristics information; and second, through analysing the data to identify the extent of the relationships between road crashes on the national rural road network and the road design environment. The chapter also describes the study instruments and software packages utilised in collecting, processing and analysing the study database. The chapter is outlined below:

1. Data collection
2. Data processing and study database
3. Research instruments
4. Data analysis
5. Project timeline
6. Ethics

3.2. Data collection

Data collection plays a crucial role in statistical analysis. In research, various methods are utilised to gather data, which fall into two categories; primary data and secondary data (Ajayi, 2017). Primary data was collected from all the national rural roads using the process described in Section 3.2.2. Using multiple forms of equipment, the primary data was used to supplement data sourced from different institutions. The institutions include: the Namibian National Road Safety Council (NRSC); Motor vehicle Accident Fund of Namibia (MVA); the Namibian Police and the Roads Authority of Namibia. Secondary data was collected from these institutions as the primary source of input data for the study.

3.2.1. Data collection study area

Fatal and serious injury crash data was sourced for the Namibia national rural road network. The national road network is divided into several classes according to the functions of the roads and traffic volumes experienced on these roads. The national rural roads span across all the fourteen regions in Namibia and are maintained by the Namibian Roads Authority, through subsidies provided by the Namibian Government and road user taxes and other fees collected by the Road Fund Administration (RFA).

The study focused on fatal and serious injury crashes on trunk and main roads on the national rural road network as shown in Figure 3.1. Hence, crash data was sourced from the Namibian National Road Safety Council (NRSC), Motor Vehicle Accident Fund of Namibia (MVA) together with Namibian police forms for the aforementioned road classes. On a similar note, data on roadway design and conditions was sourced from the Roads Authority of Namibia (RA). This focused mainly on traffic volumes, speeds (operational, design and posted), road lane characteristics, road shoulder characteristics, road alignment, sight distances, access density and pavement conditions. Collection of roadway data also involved onsite data collection on the rural roads to supplement data sourced from the relevant authorities.

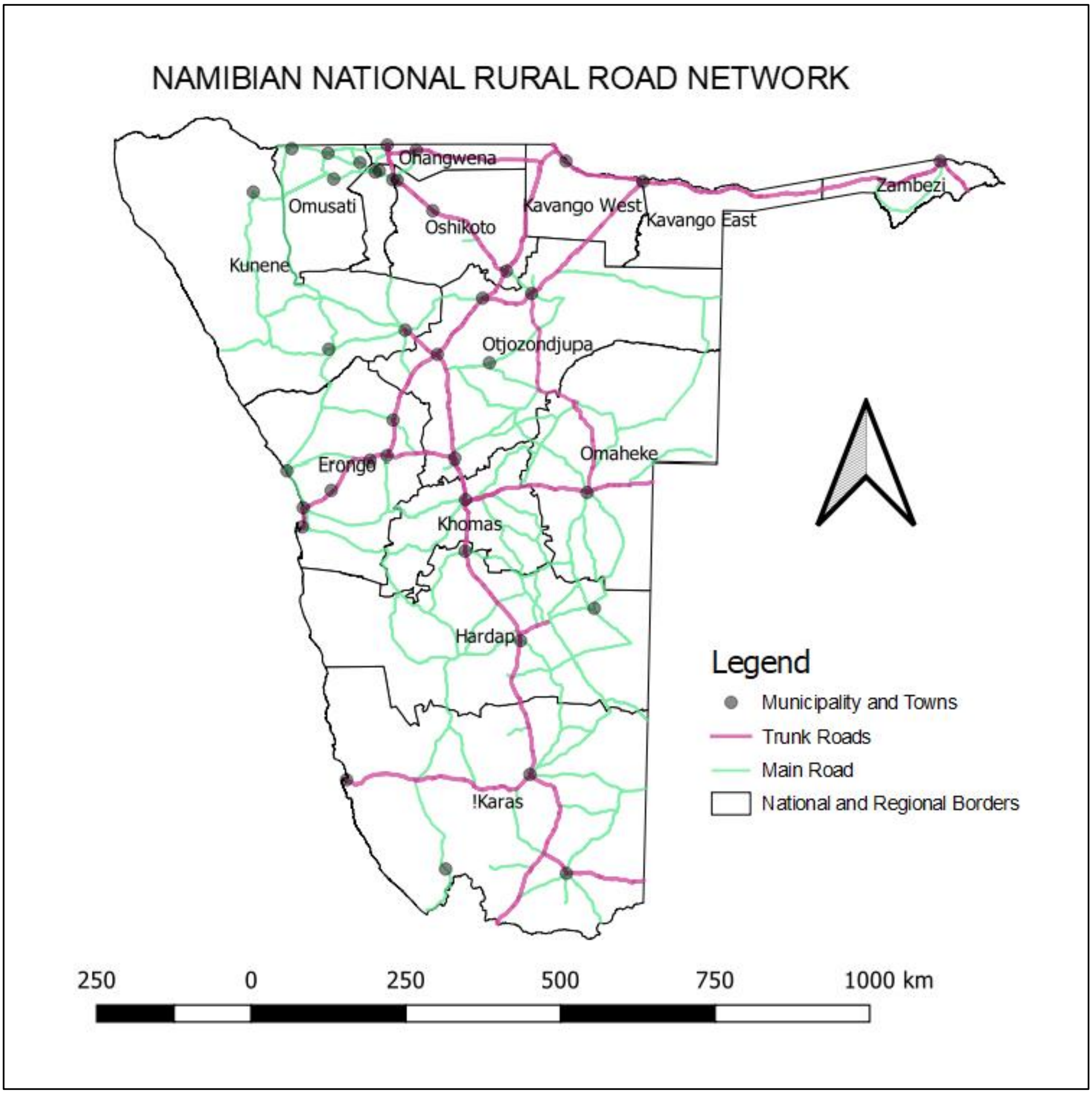


Figure 3.1 National rural road classes in Namibia

3.2.2. Primary data collection

Primary data is defined as data collected for the first time by the researcher (Ajayi, 2017). At locations on road segments where variable data was not available on the RMS or was not accurately recorded, the researcher carried out site observations and measurements to supplement secondary data and ensure the quality of data used in the analyses.

3.2.2.1. Road crash data

The location of the road traffic crash data recorded by the road safety stakeholders in Namibia is not geo-coded. The crash locations are described in text format using the km markers and landmarks close to the crash locations. The researcher had to determine the geographical coordinates (longitude and latitude) of the crash data using google satellite images and aerial photographs to identify the exact location of the road crash as shown in Figure 3.2.

I	J	K	L	M	N	O
Year	Location	Ylat	Xlong	Built_	Region	Crash_Ty
2012	Uis	-20.98230	15.08626	0	Erongo	Single veh
2012	Tses	-25.80833	18.08805	0	!Karas	Other/unk
2012	Oshivello	-18.52204	17.10689	0	Oshikoto	Single veh
2012	Otavi	-19.45336	17.58863	0	Otjozondj	Sideswipe
2012	Otjinene	-21.37257	18.54517	0	Omaheke	Other/unk
2012	Windhoek	-21.91199	16.58441	0	Otjozondj	Head rear
2012	Otjiwarong	-20.81251	16.80452	0	Otjozondj	Head on c
2012	Rundu	-17.93134	20.31272	0	KE	Pedestrian
2012	Okahao	-17.89038	15.13147	0	Omusati	Single veh
2012	Opuwo	-18.56102	14.30312	0	Kunene	Single veh
2012	Divundu	-18.07054	21.39157	0	KE	Pedestrian
2012	Otjiwarong	-20.71568	16.79328	0	Otjozondj	Single veh
2012	Otjiwarong	-20.34806	16.44422	0	Otjozondj	Other/unk
2012	Omega	-17.83078	23.03615	0	Zambezi	With anim
2012	Leonardville	-22.93411	18.74937	0	Omaheke	With anim
2012	Oshivello	-18.49812	17.03019	0	Oshikoto	Single veh
2012	Ohangwena	-17.49390	15.89937	0	Ohangwer	Head on c
2012	Walvisbay	-22.89819	14.54434	0	Erongo	Single veh
2012	Grootfontei	-19.28711	18.44352	0	Otjozondj	Single veh
2012	Okahandja	-21.65478	16.87170	0	Otjozondj	Head on c
2012	Gobabis	-23.57550	18.29459	0	Omaheke	Single veh
2012	Oshakati	-17.77034	15.69062	0	Oshana	Sideswipe
2012	Grootfontei	-19.32211	18.39594	0	Otjozondj	Head on c
2012	Okahao	-17.88437	15.15607	0	Omusati	Pedestrian
2012	Eenhana	-17.52053	16.03519	0	Ohangwer	Head on c
2012	Eenhana	-17.50326	16.18348	0	Ohangwer	Turn right
2012	Otjiwarong	-20.88299	16.79783	0	Otjozondj	With anim
2012	Omaruru	-21.68701	15.97656	0	Erongo	Single veh

Figure 3.2 Geo-coded NRSC data

Identifying the exact crash location was vital in determining the road and traffic conditions on the road segment on which the crash occurred. Inaccurate information on crash locations limited the collection of vital roadway geometric data. Therefore, a data collection process illustrated in Figure 3.3 was developed and used to collect and code data related to crash locations to supplement data collected from authorities.

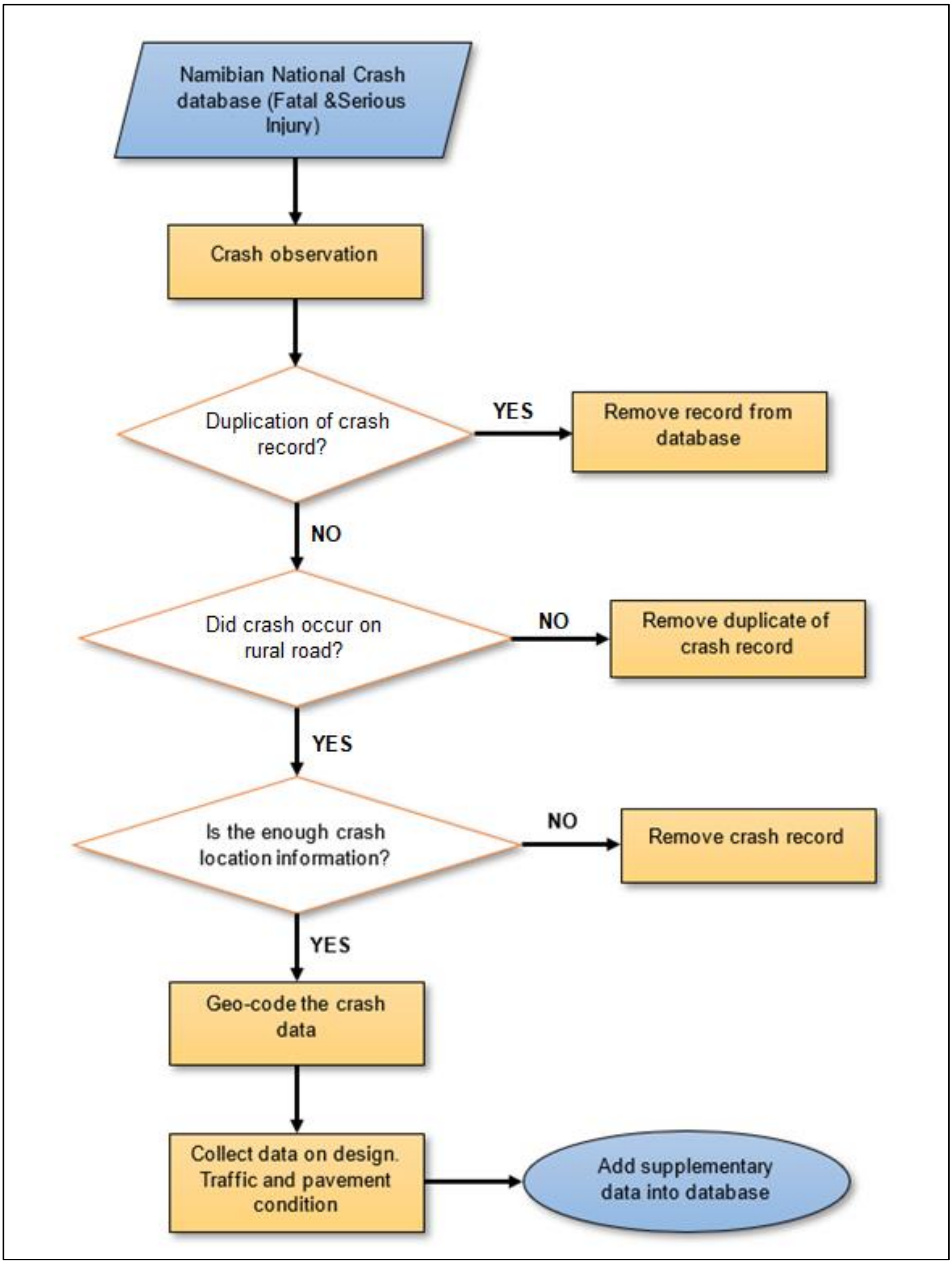


Figure 3.3 Data collection process used for collecting supplementary information for dataset

3.2.2.2. Road design characteristics, pavement and traffic conditions

It was important to establish the roadway condition (design, pavement and traffic condition) on the national rural road way network to determine the level of compliance of the road design variables that were used in the study with the Technical Recommendations for Highways 17 on the Geometric Design of Rural Roads – TRH 17 (Committee of State Road Authorities, 1988), the Technical Recommendations for Highways 20 on the Structural Design, Construction and Maintenance of Unpaved Roads - TRH 20 (Comittee for State Road Authorities, 1990) and the Technical Recommendations for Highways 26 on Road Classifications and Access Management – TRH 26 (Committee of State Road Authorities, 1988), used for road classification and alignment designs in Namibia. Persia *et al.* (2016) note that establishing roadway conditions through road safety management systems allows for a set of procedures that support road authorities in decision making, related to the improvement of safety on a road network. Variable data collected on the roadway parameters was tested in the crash prediction models developed, with results compared to the crash prediction models test on the road design standards (TRH 17, TRH 20 and TRH 26) used on the roads for a safety analysis. In essence, the safety analysis was vital in investigating the compliance of the road conditions and the validity of the current road design standards on the safety of the road.

For that reason, in this study, for road segments on which volume data was not available on the RMS, Average Annual Daily Traffic (AADT) was measured in one location only. That single value was used as the AADT of the control section. For road segments on which AADT was measured on two or more locations, the average AADT was calculated and used as the AADT for the segment. The vehicle population and types were also determined from observation on the study segments. The traffic speeds on the study segments were measured by the researcher on segments where no information is provided, to determine the 85th percentile operational speeds of the traffic.

Information on the geometric characteristics of the segments was measured by the researcher on site at the study locations and used to supplement, improve the quality and accuracy of the information provided by the Roads Authority of Namibia. The pavement conditions were determined by the researcher using the condition score index provided in Table 2.20 together with road surface condition information given in the NRSC dataset.

a) Selection of study variables

The study road characteristic variables discussed and revisited in this section were chosen for inclusion in the analyses due to their relevance to road safety. The quality of data available on variables in the RA database and the ability to access the study locations to supplement data provided by the RA and NRSC was also considered. The study used the data on the following road variables to inform on the existing road conditions of national rural roads and to develop crash prediction models.

1. Traffic crash rates: The crash rate technique improves upon the average crash frequency method by normalising the frequency of road crashes with road user exposure (AADT and section length) (Cenek *et al.*, 2012). In analysing the safety performance of a road segment, fatal and serious injury severities are important in identifying road segments with the highest severity risks and in need of safety interventions (Ambunda and Sinclair, 2019). Traffic crashes are key in determining the crash rates on the road segments, enabling the comparison and ranking of road segments according to their safety performance (Bamdad Mehrabani and Mirbaha, 2018).
2. Traffic volume: The number of vehicles crossing a particular point on the study section per hour was instrumental in determining the Average Annual Daily Traffic (AADT), which was vital in calculating the road crash rates on the study sections. Duivenvoorden (2010) confirms that a statistically significant correlation exists between road crashes of different levels of severity and multiple covariates, including the AADT.
3. Design, posted and operating speeds: Porter *et al.* (2012) state that design speed is a tool used to develop the geometric features of a road during the road design stage. The posted speed regulates the speeds that road users should adhere to when traversing a road section. The operating speed is the speed at which road users generally operate on a particular road. Wang *et al.* (2009) assert that speed is an important factor in road safety. It does not only affect crash severity levels but is also related to the risk of being involved in a road crash. Deller (2013) states that speed and excessive speed remains one of the most vital contributing factors to road crashes.
4. Lane width and surface type: Lane width is defined as the width of the roadway available for drivers to travel. Deller (2013) notes that drivers tend to speed on roads with greater lane widths compared to roads that are narrow. A study by Dong *et al.* (2015) found that a reduction in the lane width resulted in an increase in injury severity and in the likelihood of a road crash.
5. Section length: This represents the section of the road along which vehicles travel. A study by Ahmed (2013) found that as the length of the road section increases, drivers tend to increase their speed and make risky manoeuvres. The opposite happens on shorter road segments, where decelerations to bring a vehicle to a sudden halt can impact the steering capabilities of drivers (Chan *et al.*, 2008). The length of the road section together with the volume of the road section determine the level of exposure for the road users (Chen *et al.*, 2007).
6. Number of road lanes: A road lane is defined as the portion of the roadway designated for use by a single line of vehicles in a single direction. Ahmed (2013) notes that road lanes help to control, guide drivers and reduce traffic conflicts. A study by Yang *et al.* (2017) found that the number of road lanes available to the road users influences drivers' tendency to make risky overtaking manoeuvres, which impacts the safety situation on the roadway.
7. Shoulder widths and type: The American Association of State Highways and Transportation Officials (2011) (AASHTO) define a road shoulder as the width of the roadway adjacent to the traffic lanes. Ben-Bassat and Shinar (2011) state that roadway paved shoulders have several

functions, including stop and pull off, and recovery area for driver errors. Karlaftis and Golias (2002) state that narrow shoulders can create a dangerous situation where the driver will not have a recovery area in case of lane deviation and they therefore increase the likelihood of off-road crashes. However, wide shoulders may also create a dangerous road situation due to higher driver speed selections, as drivers feel they have enough space to correct errors (Liu *et al.*, 2016).

8. Horizontal and Vertical curvature: These curves facilitate the smooth transition of a vehicle when there is a change of direction or elevations. Turner *et al.* (2015) note that roadway curves are a necessary and important element of nearly all highways. Several studies have indicated that highway curves exhibit higher road crash rates than tangent sections, and that crash rates increase as the degree of the curves increase (Chen *et al.*, 2007; Othman and Thomson, 2007; Hassan and Easa, 2003).
9. Sight distances: The alignment of the roadway has a great impact on road safety because a drivers' ability to see ahead is necessary for the safe operation of the vehicle and thus for the overall safety of the road system (Ahmed, 2013). A stopping sight distance of sufficient length is necessary so that a driver can safely stop a vehicle to avoid hitting an unexpected object, while a passing sight distance of sufficient length is necessary to allow for safe overtaking manoeuvres (Bassan, 2016).
10. Access Density: Alsubeai (2017) defines access density as the number of access points on the roadway per km. Turner *et al.* (2015) affirm that access density impacts safety on roadways. Ahmed (2013) notes that increasing the number of accesses per km to a highway increases the likelihood of access related road crashes and reduces the operational efficiency of the roadway.
11. Pavement condition: The condition of the pavement is affected by traffic volumes, weather and ground conditions, which potentially expose the road surface to wear and tear (Mohammed *et al.*, 2017). Several studies have found that increasing the roughness of the pavement surface resulted in poor road safety conditions; The impact of road roughness varies on the vehicles wheels, which exposes the vehicle to different levels of friction on each side, resulting in poor steering capabilities (Ghanbari, 2017; Chan *et al.*, 2008; King, 2014).

Table 3.1 shows a list of codes and formats used in the collection of roadway data related to the crash locations used in the study.

Table 3.1 List of codes and formats for variables

<table>
<tr><th>Variable</th><th>Description</th><th colspan="2">Code</th><th>Format</th><th>Unit</th></tr>
<tr><td>Traffic volume</td><td>Number of vehicles crossing a particular point on the study section per hour</td><td colspan="2">Value</td><td>Numeric</td><td>Vehicles/hr</td></tr>
<tr><td>Posted speed limit</td><td>Posted speed limited to which users should adhere to</td><td colspan="2">Value</td><td>Numeric</td><td>Km/h</td></tr>
<tr><td>Operating speed</td><td>Speeds at which road users operate at when using the road section</td><td colspan="2">Value</td><td>Numeric</td><td>Km/h</td></tr>
<tr><td>Lane width</td><td>Width of the roadway facility available to drivers</td><td colspan="2">Value</td><td>Numeric</td><td>m</td></tr>
<tr><td rowspan="2">Surface type</td><td rowspan="2">The type of road lane surface (paved/unpaved)</td><td colspan="2">Paved (0)</td><td rowspan="2">Text</td><td>-</td></tr>
<tr><td colspan="2">Unpaved (1)</td><td>-</td></tr>
<tr><td>Section length</td><td>Length of study section</td><td colspan="2">Value</td><td>Numeric</td><td>m</td></tr>
<tr><td>Number of road lanes</td><td>Number of lanes in both directions available to drivers</td><td colspan="2">1 to 4 lanes</td><td>Numeric</td><td>-</td></tr>
<tr><td>Shoulder width</td><td>Width of the roadway adjacent to the road lanes</td><td colspan="2">Value</td><td>Numeric</td><td>m</td></tr>
<tr><td rowspan="2">Shoulder type</td><td rowspan="2">The type of ground shoulder surface (paved or gravel) adjacent to road lane</td><td colspan="2">Paved (0)</td><td rowspan="2">Text</td><td>-</td></tr>
<tr><td colspan="2">Unpaved (1)</td><td>-</td></tr>
<tr><td>Horizontal curvature</td><td>Rate of change of horizontal alignment per road length (Curvature coefficient)</td><td colspan="2">Value</td><td>Numeric</td><td>-</td></tr>
<tr><td rowspan="2">Vertical Curvature</td><td rowspan="2">The maximum and minimum slopes on the study road (Elevation profile)</td><td colspan="2">Flat (0)</td><td rowspan="2">Text</td><td rowspan="2">-</td></tr>
<tr><td colspan="2">Sloped (1)</td></tr>
<tr><td>Sight distances</td><td>Distance provided to drivers</td><td colspan="2">Value</td><td>Numeric</td><td>m</td></tr>
<tr><td>Access density</td><td>Number of access points provided to traffic to joining study road</td><td colspan="2">Value</td><td>Numeric</td><td>Access/km</td></tr>
<tr><td rowspan="4">Pavement condition</td><td rowspan="4">The riding comfort of the pavement</td><td>Very poor</td><td rowspan="2">Poor (1)</td><td rowspan="4">Text</td><td>-</td></tr>
<tr><td>Poor</td><td>-</td></tr>
<tr><td>Fair or good</td><td rowspan="2">Good (0)</td><td>-</td></tr>
<tr><td>Very good</td><td>-</td></tr>
</table>

3.2.3. Secondary data collection

Secondary data is defined as data collected or produced by investigator agencies and organisations earlier (Ajayi, 2017). The researcher collected data from the numerous road safety stakeholders to be utilised in the study.

3.2.3.1. Road Crash data

Information on road crashes are captured by the Namibian Police, the Motor Vehicle Accident Fund of Namibia (MVA) and the National Road Safety Council of Namibia (NRSC). A crash record consists of information on the date, time, location, number and types of vehicles, weather, number and types of injury severity, road surface condition, lighting and the type and cause of the road crash. Existing road crash data was collected in Microsoft Excel format from the NRSC, MVA databases and the Namibian Police road crash report forms (shown in ***Appendix A***), considering data quality and availability from 2012 to 2016.

1. **Sampling of crash counts for statistical analysis**

The target population for this study are drivers who were involved in fatal and serious injury traffic crashes in the national rural road spaces. National rural roads of various classifications in Namibia were chosen as focus areas for the study based on the scale of safety concerns over the high number of fatal and serious injury crashes. Several reports and studies have confirmed that Namibian roads are some of the most dangerous regionally and globally, with fatal and serious injury crash rates above the average value for the African continent (Amweelo, 2016; Nambahu, 2018). The choice of the study was further motivated by access to road crash data for the study area.

The dataset for fatal and serious injury crashes comprised 3 192 casualties on the Namibian national rural roads from 2012 to 2016. Therefore, it was important to establish the appropriate sample size to draw correct inferences on the study population and for sound statistical results. Due to the relatively random nature of traffic crashes, a high crash rate in any given year may simply be a random fluctuation around a much lower long-term crash rate average on the study segments, leading to regression towards the mean (Choi *et al.*, 2019). Therefore, a study period of 3-5 years is recommended to minimise the effects of the regression to the mean phenomenon (Demissie, 2017). Considering the recommendations made, the study focused on road crash data for a period of five years, from 2012 up until 2016.

The study period was chosen due to the high quality of the data available for this study period. Despite the good quality of the data, all the records lacked appropriate location information or had vague location descriptions. The researcher was required to cross-analyse multiple databases from the Namibian Police and identify the locations of the road crashes in the Namibian National Road

Safety Council database. Moreover, additional road information on the crash locations was required for the database and therefore required researcher to go out on site to undertake these remedial measures.

In order to determine whether the 3 192 crash counts were sufficient to draw statistical inferences from the study population, a minimum required sample size (number of crash records) was calculated and compared with the number of available crash observations. This allowed the researcher to determine whether the minimum sample size requirement criteria was complied with, and whether inferences drawn would be representative of the entire population.

The statistical power method was used in STATISTICA to evaluate the minimum sample size required to detect statistically significant relationships at a desired level of confidence. In inferential statistics, the probabilities of a type I error and a type II error are determined (Elviket *et al.*, 2004; Field, 2013). A type I error is referred to as an alpha error (α) and a type II error is referred to as a beta error (β) (Ali and Bhaskar, 2016). The type I error value (α) is the probability that the null hypothesis H_0 will be rejected when in fact it is true. In essence, a difference that does not exist is being investigated, committing a type I error (Elviket *et al.*, 2004). The alpha value is often simple to determine, as it is can be specified in the model, usually set at 5 percent (0.5) (Ali and Bhaskar, 2016; Cohen, 1992).

The Type II error value (β) is not specified, rather the sample size (*N*), significance level (α) and the effect size (ES) influence the Type II error value (β), and similarly, they influence the power, which is equal to 1-β (Gogtay, 2010). Power is the probability that a difference that exists will be detected. The β value is the probability of a type II error, and a type II error is when the researcher fails to reject a false null hypothesis (Ali and Bhaskar, 2016). In essence, the model states that no difference exists when in fact it exists. Cohen (1988) illustrates the statistical decision matrix used in hypothesis testing in Table 3.2.

Statistical power analysis deals with a type II error, estimating the power as 1-β as illustrated in Table 3.2. The analysis can be interpreted as the probability that a statistical test will correctly reject a false null hypothesis (Elvik *et al.*, 2004). Cohen (1992) suggested that the maximum acceptable *p* value of a type II error should be 20 percent (0.2), implying that to detect reasonable effects, the power of a statistical test (1- β) should be at least 80 percent (0.8).

Table 3.2 Statistical test decision matrix (Cohen, 1988)

Test decision	True state of population	
	Effect absent H_0 is true	Effect present H_0 is false
Test result: $p< \alpha$ Test decision: reject H_0 Conclusion: "effect exists"	Type I error $p= \alpha$[i]	Power $p= 1-\beta$
Test result: $p\geq \alpha$ Test decision: accept H_0 Conclusion: "effect absent"	Correct decision $p= 1-\alpha$	Type II error $p= \beta$[ii]

---

[i] α is the probability (p) of a type I error, which rejects the null hypothesis (H_0) when true

[ii] β is the probability of a type II error, which fails to reject the null hypothesis (H_0) when false

The statistical power analysis method in STATISTCA was applied to test whether the crash count records used in the study were sufficient enough to record a statistical effect and the size of that effect. The α-level for the power analysis was set at 0.05 with a desired power of 0.9 (90 percent chance) of detecting a statistical effect should one exist). The analysis determined the required sample as 2 931 crash counts, indicated in Table 3.3, which is slightly lower than the actual sample size of 3 192 used in the study.

Table 3.3 Summary output of the statistical power analysis

	Sample Size Calculation **One Proportion, Z, Chi-Square Test** **H0: Pi = Pi0**
	Value
Null Proportion (Pi0)	0.5000
Population Proportion (Pi)	0.5299
Alpha (Nominal)	0.0500
Actual Alpha (Exact)	0.0502
Power Goal	0.9000
Actual Power (Normal Approx.)	0.8997
Actual Power (Exact)	0.9000
Required Sample Size (N)	**2 931**

The results of the power analysis suggest that the power of the study sample size of 3 192 is greater than 90 percent (0.90). To detect accurate and reliable inferences, the power of a statistical test (1-β) should be at least 80 percent (0.80). It can be observed from the plot of power goal against sample size (exact) in Figure 3.4 that the minimum sample size at the power of 80 percent is smaller than the actual sample size used in the study.

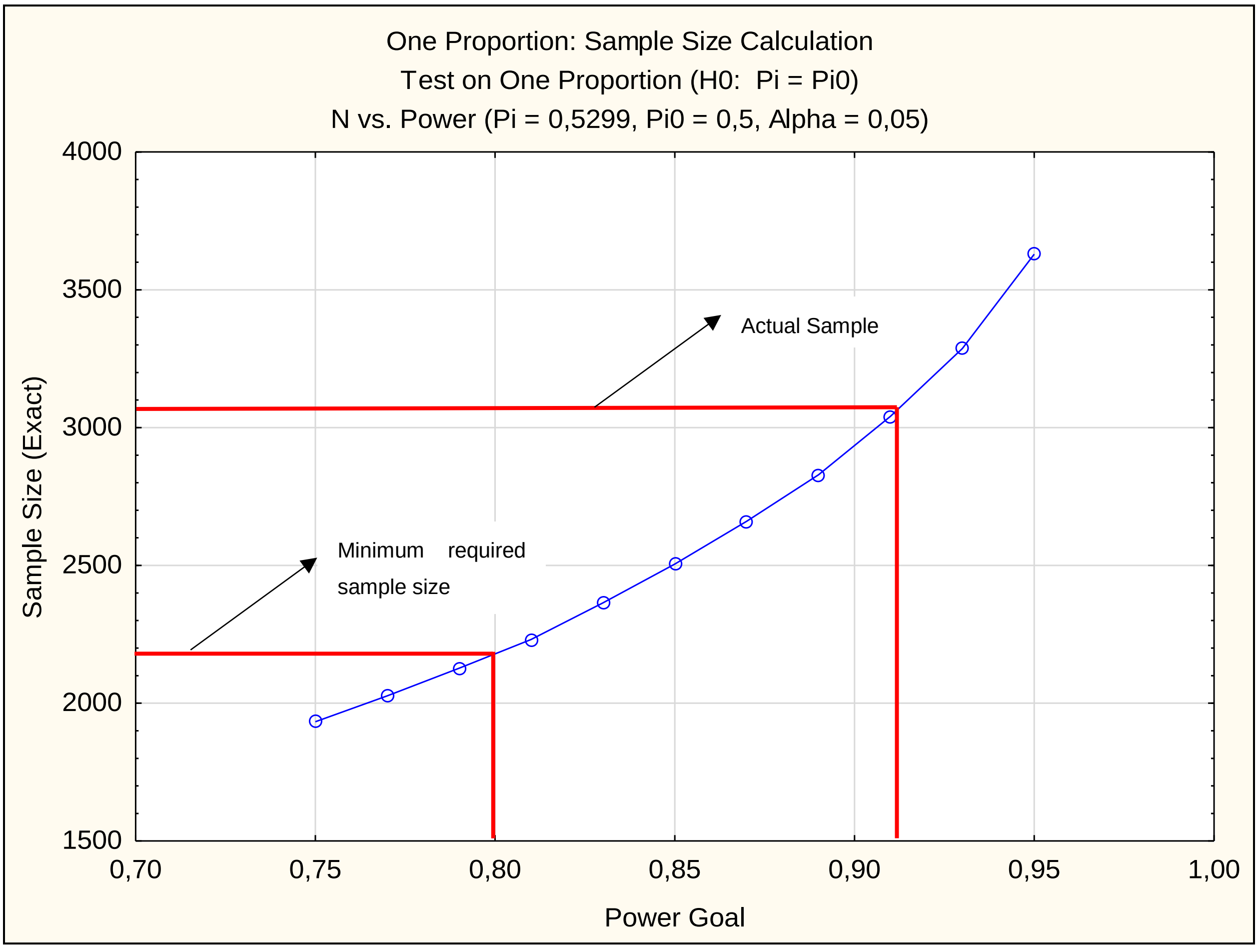


Figure 3.4 Statistical power analysis Power goal (minimum sample) vs sample size (actual)

2. Identification of variables currently informing crash risk

Identifying the variables currently informing crash risk in the crash record information is one of the first vital steps to understand and determine the extent to which the roadway and environment impact national rural road crashes in Namibia, which may occur unaccompanied or in combination with human or vehicle related factors.

The variables in Table 3.4 were identified from the NRSC crash dataset and are employed in determining the national extent to which roadway and environmental factors were involved in rural road crashes. Moreover, they provided a basis to understand the circumstances and context in which fatal and serious injury crashes occurred. It is important to note that these variables have been used for many decades by the police; and that they are completed by the police officers and not by people with specialist knowledge about road design. As such, most of the roadway factors that are currently recorded are inevitably generic and relate only to obvious issues, not to the relationship between design features of the roadway and crash risk.

Table 3.4 Data variables in crash dataset

Column name	Variable	Format
day_of_week	The day of the week	Text
time_of_day	The time of day	Time
year	Year	Numeric
no_vehicle	Number of vehicles	Numeric
no_fatality	Number of fatalities	Numeric
seriously_injured	Number of serious injuries	Numeric
slightly_injured	Number of slight injuries	Numeric
not_injured	Number of no injury cases	Numeric
vehicle_damage_only	Number of property damage only cases	Numeric
Operating_speed_on_road	85^{th} percentile operating speed	Numeric
junction_type	Junction type	Text
road_type	Road type	Text
weather	Weather	Text
severe_wind	Whether there were severe winds	Text: TRUE or FALSE
light_condition	Lighting conditions	Text
road_surface	Road surface	Text
road_surface_type	The type of road surface	Text
road_surface_quality	The quality of road surface	Text
road_mark_type	Road mark type	Text
road_direction	Horizontal alignment of the road	Text
road_shape	Vertical alignment of the road	Text
traffic_control_type	The type of traffic control	Text
road_signs_visible	Whether the road sign was visible	Text: TRUE or FALSE
obstructions	The type of obstruction observed	Text
accident_type	Accident type	Text
built_up_area	Whether the accident location is in the built-up area	Text: TRUE or FALSE
y_lat	Latitude coordinate of crash location	Numeric
x_long	Longitude coordinate of crash location	Numeric
road_name	The name of the road on which crash occurred	Text/ Numeric
driver_action_A/ driver_action_B	Actions of motorists involved before the accident	Text
personaccidentcount_P1	Number of the persons involved in the accident	Numeric
person_gender_P1	The gender of the persons involved	Text
person_age_P1	The age of the persons involved in the accident	Numeric
Is person_1_P1_driver	Whether the persons in consideration were driving	Text: Yes or No
person_accident_description_P1/ person_accident_description_P2	Description of the accident by the persons involved in the accidents Or Description of the	Text: natural language text

	accident by the officer who reported/investigated the accident	
person_alcohol_drug_test_confirmed_P1/ person_alcohol_drug_test_confirmed_P2	Whether the persons involved were tested positive for alcohol or drugs	Text: TRUE or FALSE
person_seatbelt_helmet_used_P1/ person_seatbelt_helmet_used_P2	Whether the persons involved were using seat belts or helmets	Text: TRUE or FALSE

3. Identification of crash risk and risk factor categorisation

Any road traffic system is highly complex and is influenced by a multitude of factors, including road users, road environment and vehicles. To identify and address the hazards on a roadway requires a systems approach, where interactions between different interlinked factors are considered (Hughes *et al.*, 2015; Adanu *et al.*, 2020). Traditionally, crash risk factors analyses have examined the human, roadway and environment and vehicle separately. Building on Haddon's insights discussed in Section 1.8, the study used a systems approach to define pre-crash risk types, informed by crash descriptions from crash victims. This allowed for the categorisation of crash records by main error categories; human, roadway and environmental and vehicle factors discussed in Table 1.1. The human-related factors in road crashes are shown in Table 3.5.

Table 3.5 Human related risk factors

Main error category	Risk factor
Recognition error	Inadequate surveillance Internal distractions Inattention Confusion over the road environment Visual impairments Complex environments/overestimation Response delays
Decision error	Too fast for conditions Too fast for a curve False assessment if another's actions Misjudgement of gap or other's actions Failure to use passive safety features Swerve in front of other traffic Unsafe passing
Performance errors	Overcompensation Poor directional control Panic/freezing General driving ability/skills Other performance error
Intentional risk	Fatigue Alcohol/drugs Aggression Dangerous manoeuvres Traffic violations Following too close

	Speeding Too fast for conditions
Physiological conditions	Physical impairments Heart attack Eyesight Medications Age -senior driver/ped (>65) Age – young driver/ped (<25) Age – child ped (<15) Blackouts

Table 3.6 displays roadway factors and vehicles factors. The study is specifically focused on identifying the context and extent to which roadway factors play a role in crash occurrence, as a pre-cursor to developing crash prediction models for the national rural roads.

Table 3.6 Roadway and vehicle related risk factors

Main error category	**Risk factors**
Roadway Factors	Potholes Animals Obstructions Work zone Faulty traffic lights Weather Poor visibility/ night/glare/dawn/dusk Road surfaces Stone projected by another vehicle Stone Speed differentials/ congestion Road geometry: Curve/slope
Vehicles Factors	Tyre burst Defective lights or indicators Defective brakes Missing or defective mirrors Defective steering or suspension Overloaded or poorly loaded vehicle or trailer Tyre hooked off the vehicle Other

Other road user factors that that were identified to play a role in crashes are listed in Table 3.7.

Table 3.7 Risk factors related to other road users

Main category	Risk factors
Other road user factors	Cyclist unsafe riding Intoxicated cyclist Cycling in darkness Cyclist distractions Jaywalking Traffic light violations Unsafe crossing/signalised crossing Crossing between parked cars Pedestrian using roadway Child running after car Intoxicated pedestrian

4. **Application of risk factors to crash records**

The study investigated the descriptions provided by the person(s) involved in road crashes and investigators together with other provided information to identify risk factors for the crash records. Therefore, risk factors were classified into several levels below according to the information provided:

a) Level 1: Primary risk factor
b) Level 2: Secondary risk factor
c) Level 3: Other possible risk factor

3.2.3.2. Road design characteristics, pavement and traffic conditions

Road geometry and traffic data on the rural roads was retrieved from the Road Management System (RMS) and Road Referencing System (RRS) of the Roads Authority of Namibia (RA) in PDF format as shown in Figure 3.5 and Figure 3.6. Road geometry data includes data on lane width, hard-shoulder width, horizontal and vertical curves characteristics, segment length, road access density and sight distance. The study also used the geometric standards stipulated in the Technical Recommendations for Highways 17 (TRH 17) on the Geometric Design of Rural Roads in the development of crash prediction models to evaluate the performance of existing road geometry.

The traffic conditions related to the study location, namely the average annual daily traffic (AADT), traffic composition (Percentage of heavy vehicles) and traffic speeds were sourced in Excel format from the RA as shown in Figure 3.7. In addition, the condition of the pavement (riding comfort index (RIC)) for each segment during the study period 2012 to 2016 was determined from the road data provided by the RA, in combination with information from the dataset provided by the NRSC.

Using these factors, road crash prediction models were developed to estimate the effect of these factors on fatal and serious road crash counts. The traffic condition and crash data on the rural roads to be analysed was examined to ensure that no changes in either the layout or major traffic volumes due to alignment changes occurred over the course of the study period.

Surfaced Cross Sections

Road Referencing System

System: Road Referencing System **Network:** 11 - [New network based on network 10]

Surfaced Road Cross Sections Report

			FORWARD							MEDIAN	*BACKWARD*								
			SHOULDER		LANES			SHOULDER			SHOULDER		LANES			SHOULDER			
Road	Begin KM	End KM	GSF1	SSF1	F3	F2	F1	SSF2	GSF2		GSB2	SSB2	B1	B2	B3	SSB1	GSB1	Tot Surfaced	Tot Unsurfaced
T0103	0.00	5.00	3.10	0.00	0.00	0.00	3.41	0.00	0.00	0.00	0.00	0.00	3.41	0.00	0.00	0.00	3.10	6.82	6.20
T0103	5.00	10.00	2.50	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	3.30	0.00	0.00	0.00	2.50	6.4	5.00
T0103	10.00	15.00	2.70	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	2.70	6.3	5.40
T0103	15.00	20.00	2.20	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	2.50	6.3	4.70
T0103	20.00	25.00	1.50	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	3.10	0.00	0.00	0.00	2.50	6.2	4.00
T0103	25.00	30.00	1.90	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	3.10	0.00	0.00	0.00	1.90	6.2	3.80
T0103	30.00	35.00	2.20	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	3.10	0.00	0.00	0.00	2.20	6.2	4.40
T0103	35.00	40.00	2.10	0.00	0.00	0.00	3.00	0.00	0.00	0.00	0.00	0.00	3.10	0.00	0.00	0.00	2.10	6.1	4.20
T0103	40.00	45.00	2.10	0.00	0.00	0.00	4.10	0.00	0.00	0.00	0.00	0.00	4.10	0.00	0.00	0.00	2.10	8.2	4.20
T0103	45.00	50.00	2.00	0.00	0.00	0.00	4.00	0.00	0.00	0.00	0.00	0.00	4.00	0.00	0.00	0.00	2.10	8	4.10
T0103	50.00	55.00	2.30	0.00	0.00	0.00	4.00	0.00	0.00	0.00	0.00	0.00	4.00	0.00	0.00	0.00	2.30	8	4.60
T0103	55.00	60.00	1.70	0.00	0.00	0.00	3.40	0.00	0.00	0.00	0.00	0.00	3.40	0.00	0.00	0.00	1.70	6.8	3.40
T0103	60.00	65.00	1.70	0.00	0.00	0.00	3.40	0.00	0.00	0.00	0.00	0.00	3.40	0.00	0.00	0.00	1.70	6.8	3.40
T0103	65.00	70.00	2.30	0.00	0.00	0.00	3.20	0.00	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	2.30	6.4	4.60
T0103	70.00	75.00	3.20	0.00	0.00	0.00	3.40	0.00	0.00	0.00	0.00	0.00	3.30	0.00	0.00	0.00	2.30	6.7	5.50
T0103	75.00	80.00	2.50	0.00	0.00	0.00	3.60	0.00	0.00	0.00	0.00	0.00	3.30	0.00	0.00	0.00	2.50	6.9	5.00
T0103	80.00	85.00	1.60	0.00	0.00	0.00	3.60	0.00	0.00	0.00	0.00	0.00	3.60	0.00	0.00	0.00	1.60	7.2	3.20
T0103	85.00	90.00	2.50	0.00	0.00	0.00	3.20	0.00	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	2.50	6.4	5.00
T0103	90.00	95.00	2.50	0.00	0.00	0.00	3.60	0.00	0.00	0.00	0.00	0.00	2.80	0.00	0.00	0.00	2.50	6.4	5.00
T0103	95.00	100.00	2.30	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	2.50	6.3	4.80
T0103	100.00	105.00	2.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	2.00	6.4	4.00
T0103	105.00	110.00	2.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	2.00	6.4	4.00

RMS: Road Referencing System *Page 1 of 30* *Printed on 26 March 2018*

Figure 3.5 Cross sections report from Roads Authority Namibia

Road Log

Road Referencing System

System: Road Referencing System **Network:** 11 - [New network based on network 10]

Road Log Report

T0103

Direction	Begin KM	End KM	Feature Category	Feature Type	Position	Description
Single	0.00	0.00	Intersection \ Junction	T Junction Start	ON	Keetmanshoop town at Intersection with T0102/T0401 (B4). Center Traf
Single	0.67	0.67	Intersection \ Junction	T Junction Left	ON	T1001: deproclaimed
Single	2.48	2.48	Intersection \ Junction	T Junction Left	ON	KEETMANSHOOP NORTH: M0088
Single	4.02	4.02	Intersection \ Junction	T Junction Right	ON	M0027
Single	41.06	41.06	Intersection \ Junction	Crossing	ON	LEFT D3906 RIGHT D3911
Single	80.21	80.21	Intersection \ Junction	T Junction Left	ON	M0098
Single	81.29	81.29	Intersection \ Junction	T Junction Right	ON	D0619
Single	87.55	87.55	Intersection \ Junction	T Junction Left	ON	D3921
Single	102.80	102.80	Intersection \ Junction	T Junction Right	ON	D3908
Single	129.39	129.39	Intersection \ Junction	T Junction Right	ON	D3919
Single	131.23	131.23	Intersection \ Junction	T Junction Right	ON	D1068
Single	131.32	131.32	Intersection \ Junction	T Junction Left	ON	D1077
Single	167.79	167.79	Intersection \ Junction	T Junction Left	ON	M0032 (GIBEON)
Single	168.59	168.59	Intersection \ Junction	T Junction Right	ON	M0032
Single	225.35	225.35	Intersection \ Junction	T Junction Left	ON	M0034
Single	227.95	227.95	Intersection \ Junction	T Junction Right	ON	D1098 , Hendrik Witbooi St, Mariental
Single	229.67	229.67	Intersection \ Junction	T Junction End	ON	T0104: Mariental town at Intersection with M0029, Van Niekerk St.

Figure 3.6 Road log report from Road Authority Namibia

L/H/Volume Report

Date	01 01 2016										
Site	014 - 014										
Description	From Rehoboth to Windhoek										
Date	Time	Dur.	Lane 1			Lane 2					
			Light	Hvy	Total	Light	Hvy	Total	Light	Hvy	Total
01 01 2016	00:00:00	28103280	688361	106408	794769	697148	108972	806120	1385509	215380	1600889
			688361	106408	794769	697148	108972	806120	1385509	215380	1600889
	Days	325,2694						**ADT**	**4260**	**662**	**4922**

Figure 3.7 Traffic volume report from Roads Authority Namibia

3.3. Data quality and limitations

The quality of analyses and decision making in road safety is highly dependent on the quality of the data on which analyses are based (Montella *et al.*, 2012; Abdulhafedh, 2017). Quality data results in a better understanding of road traffic operational problems, locating of hazardous road sections, identifying risk factors, developing accurate diagnosis and remedial measures, and evaluating the effectiveness of road safety programs (Mannering and Bhat, 2014; Abdulhafedh, 2017).

Road crash investigations require a comprehensive, accurate and up to date database for an analysis to provide sound and accurate inferences. Therefore, the study required quality information pertaining to the road crashes, traffic, and roadway design and condition for a reliable statistical analysis. The quality of the data used is discussed in Section 3.3.1 and Section 3.3.2.

3.3.1. Road crash data

Road crash data focused on fatal and serious injuries only on national rural road network in Namibia was obtained from the NRSC in Microsoft Excel format for the period 2012 to 2016. The NRSC comprised 98 894 crash records. Approximately 34 percent (33 471) of these road crashes occurred on the national rural road network, with 37 field columns of information on each crash observation. The data set also included highlighted information of the roadway infrastructure and condition of the crash locations. The study identified 742 duplicated crash records, which reduced the crash dataset to 32 729 crash observations. The study focused on fatal and serious injury crashes only, which led to a crash dataset of 3 192 crash observations on the national rural road network. A summary on the quality of information of the variables in the crash dataset is detailed in Table 3.8.

Table 3.8 Quality of road crash data on the national rural road network for period 2012 to 2016

Field name	Number of variables	Variables not reported/ Unknown	Completeness
Date	3 192	0	100.00%
Weekday	3 192	0	100.00%
Time	3 192	0	100.00%
Lighting condition	2 956	236	92.61%
Visibility	2 701	491	84.62%
Weather	3 192	0	100.00%
Month	3 192	0	100.00%
Year	3 192	0	100.00%
Police station	3 192	0	100.00%
Latitude	16	3 176	0.50%
Longitude	16	3 176	0.50%
Km marker	1 244	1 948	38.97%
Location description	1 948	1 244	61.03%
Built up area (False)	3 192	0	100.00%
Crash type	3 192	0	100.00%
Crash cause	3 192	0	100.00%
Fatal injuries	3 192	0	100.00%
Serious injuries	3 192	0	100.00%
Number of vehicles involved	3 192	0	100.00%
Vehicle type	3 181	11	99.66%
Is person_1 driver?	3 192	0	100.00%
Person_1 gender	3 085	107	96.65%
Posted speed limit	2 865	327	89.76%
Road type	3 122	70	97.81%
Road number	3 122	70	97.81%
Road direction	3 122	70	97.81%
Junction type	3 122	70	97.81%
Surface type	3 192	0	100.00%
Surface quality	3 011	181	94.33%
Surface condition	3 011	181	94.33%
Road marking type	2 926	266	91.67%
Road marking condition	3 001	191	94.02%
Terrain	3 173	19	99.44%
Road sign type	2 942	250	92.17%
Road sign condition	2 942	250	92.17%
Traffic control	3 012	180	94.36%
Obstruction	3 192	0	100.00%

The crash locations in the dataset are described using text and km markers on the national rural road network. Only 0.5 percent of the crash observations had coordinates in the dataset. Information on the roadway facilities, traffic and surface condition, which was crucial for analyses in the study, had a completeness ranging from approximately 89 percent to100 percent in the dataset. The data quality deficiencies in the NRSC dataset were addressed by comparing crash records against MVA crash records. The crash sample size of 3 192 records was tested using the power analysis in Section 3.2.3 to test whether it was sufficient to conduct sound statistical analyses and to make inferences on the study population.

3.3.2. Road design characteristics, pavement and traffic conditions

Data on road design characteristics, pavement and traffic conditions was sourced in PDF format from the Roads Authority of Namibia (RA) and was also found in the NRSC crash dataset as detailed in Table 3.9. The RA dataset includes 14 fields of road design and traffic condition information for multiple rural roads on the national road network. Table 3.9 shows a summary of the attributes of the crash location design, condition and traffic data provided by the RA and used in the study. Additional information on national rural roads with zero crashes over the study period was also collected for the development of the crash prediction models.

Table 3.9 Quality of roadway design, condition and traffic data from the Roads Authority of Namibia

Field name	Number of variables	Unknown variables	Completeness
AADT	3105	87	97%
Posted Speed	3192	0	100%
85th Percentile operating speeds	2922	270	92%
Lane width	1629	1563	51%
Road lane surface type	1629	1563	51%
Section length	2847	345	89%
Number of road lanes	2847	345	89%
Shoulder width	1629	1563	51%
Shoulder type	1629	1563	51%
Horizontal curvature	0	3192	0%
Vertical curvature	0	3192	0%
Sight distances	1811	1381	57%
Access density	1629	1563	51%
Pavement conditions	2943	249	92%

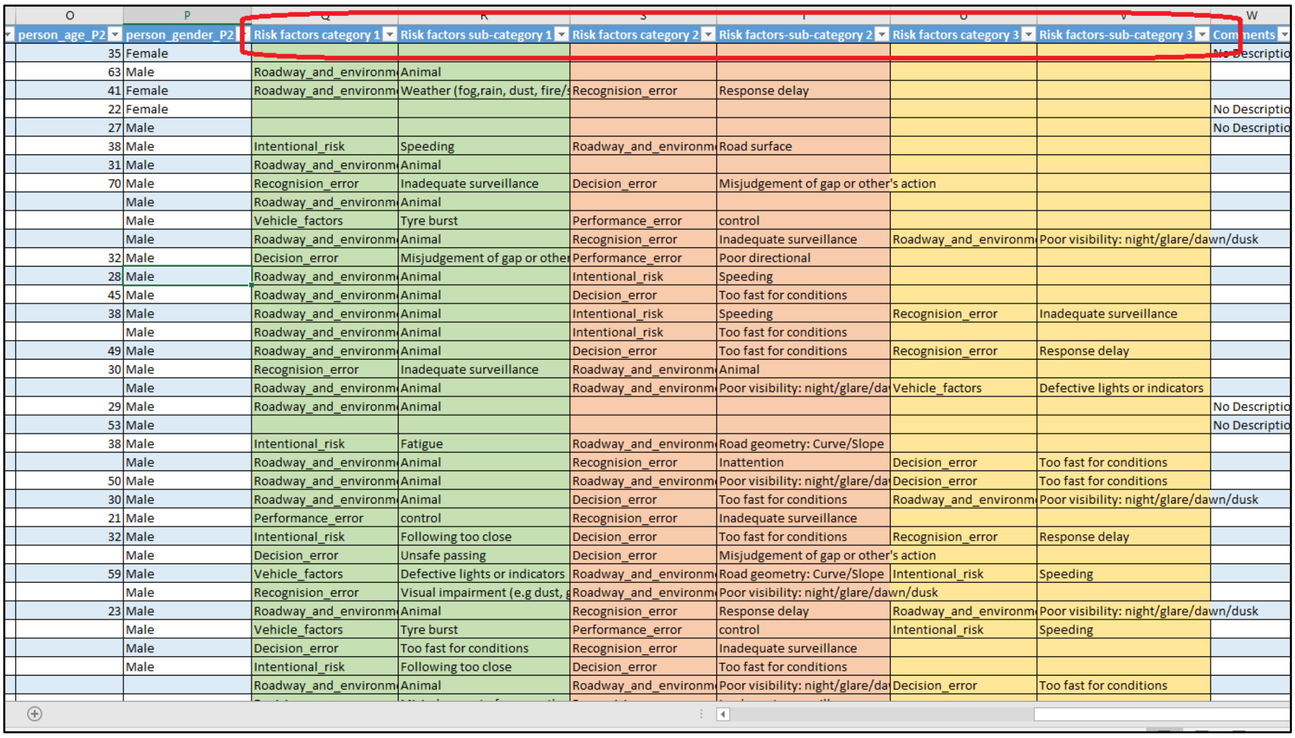
Data on posted speed limits of the various rural roads was the only fully complete (100 percent) variable in the dataset. No data was found in the RA road management system on the horizontal and vertical curvature variables. Data on lane and shoulder characteristics, roadway access and sight distances variables were slightly above 50 percent complete. Data on traffic volumes, operating speeds, section lengths, lane numbers and pavement conditions had a completeness ranging between 89 and 97 percent.

3.4. Data processing

After obtaining the raw data from the various road safety stakeholders discussed in Section 3.2, data processing was performed. Data processing initially involved cleaning the raw data and performing initial screening in order to make the data useful for performing further statistical analyses. As the study data was collected from various sources, linking all the available data and compiling the data (road crash rates, road geometric characteristics, road traffic and pavement condition information) into one dataset was vital in ensuring the quality of the data before proceeding with further analysis.

3.4.1. Road crash data processing

The road crash data had to be processed to determine the extent to which roadway factors were involved as risk factors in the occurrence of road crashes on rural roads. This involved examining the victim/ witness descriptions of the crashes and determining the various levels described in Section 3.2.3, at which the different road crash risk factors were involved in the crashes as illustrated in Figure 3.8.



person_age_P2	person_gender_P2	Risk factors category 1	Risk factors sub-category 1	Risk factors category 2	Risk factors-sub-category 2	Risk factors category 3	Risk factors-sub-category 3	Comments
35	Female							No Descriptio
63	Male	Roadway_and_environm	Animal					
41	Female	Roadway_and_environm	Weather (fog,rain, dust, fire/s	Recognision_error	Response delay			
22	Female							No Descriptio
27	Male							No Descriptio
38	Male	Intentional_risk	Speeding	Roadway_and_environm	Road surface			
31	Male	Roadway_and_environm	Animal					
70	Male	Recognision_error	Inadequate surveillance	Decision_error	Misjudgement of gap or other's action			
	Male	Roadway_and_environm	Animal					
	Male	Vehicle_factors	Tyre burst	Performance_error	control			
	Male	Roadway_and_environm	Animal	Recognision_error	Inadequate surveillance	Roadway_and_environm	Poor visibility: night/glare/dawn/dusk	
32	Male	Decision_error	Misjudgement of gap or other	Performance_error	Poor directional			
28	Male	Roadway_and_environm	Animal	Intentional_risk	Speeding			
45	Male	Roadway_and_environm	Animal	Decision_error	Too fast for conditions			
38	Male	Roadway_and_environm	Animal	Intentional_risk	Speeding	Recognision_error	Inadequate surveillance	
	Male	Roadway_and_environm	Animal	Intentional_risk	Too fast for conditions			
49	Male	Roadway_and_environm	Animal	Decision_error	Too fast for conditions	Recognision_error	Response delay	
30	Male	Recognision_error	Inadequate surveillance	Roadway_and_environm	Animal			
	Male	Roadway_and_environm	Animal	Roadway_and_environm	Poor visibility: night/glare/da	Vehicle_factors	Defective lights or indicators	
29	Male	Roadway_and_environm	Animal					No Descriptio
53	Male							No Descriptio
38	Male	Intentional_risk	Fatigue	Roadway_and_environm	Road geometry: Curve/Slope			
	Male	Roadway_and_environm	Animal	Recognision_error	Inattention	Decision_error	Too fast for conditions	
50	Male	Roadway_and_environm	Animal	Roadway_and_environm	Poor visibility: night/glare/da	Decision_error	Too fast for conditions	
30	Male	Roadway_and_environm	Animal	Decision_error	Too fast for conditions	Roadway_and_environm	Poor visibility: night/glare/dawn/dusk	
21	Male	Performance_error	control	Recognision_error	Inadequate surveillance			
32	Male	Intentional_risk	Following too close	Decision_error	Too fast for conditions	Recognision_error	Response delay	
	Male	Decision_error	Unsafe passing	Decision_error	Misjudgement of gap or other's action			
59	Male	Vehicle_factors	Defective lights or indicators	Roadway_and_environm	Road geometry: Curve/Slope	Intentional_risk	Speeding	
	Male	Recognision_error	Visual impairment (e.g dust, g	Roadway_and_environm	Poor visibility: night/glare/dawn/dusk			
23	Male	Roadway_and_environm	Animal	Recognision_error	Response delay	Roadway_and_environm	Poor visibility: night/glare/dawn/dusk	
	Male	Vehicle_factors	Tyre burst	Performance_error	control	Intentional_risk	Speeding	
	Male	Decision_error	Too fast for conditions	Recognision_error	Inadequate surveillance			
	Male	Intentional_risk	Following too close	Decision_error	Too fast for conditions			
		Roadway_and_environm	Animal	Roadway_and_environm	Poor visibility: night/glare/da	Decision_error	Too fast for conditions	

Figure 3.8 Crash risk factors levels in crash dataset

This step also allowed for the creation of georeferenced crash data aggregated at the national road level. Crash rates were combined together with roadway design and condition characteristics for the national rural road network and normalised to allow for a sound comparison over the whole network. The combined Excel file with crash rates was then imported into QGIS, converted to a QGIS shapefile and mapped to visualise the level of distribution of road crashes on the road network.

3.4.2. Road design characteristics, pavement and traffic conditions processing

Roadway data was collected from the on-site observations, PDF format and Excel format data from authorities. The raw data was then aggregated at the national level using Excel. The national rural road network was provided by the RA in shapefile format and used to identify all rural roads on which data was available and on which data had to be collected on site to augment the data collected from authorities in the excel sheet. Information on the road design standards (TRH 17) was also added to the Excel spreadsheet to augment all crash data and roadway data required for the analyses on a national dataset level. The steps taken to process roadway data are illustrated in Figure 3.9.

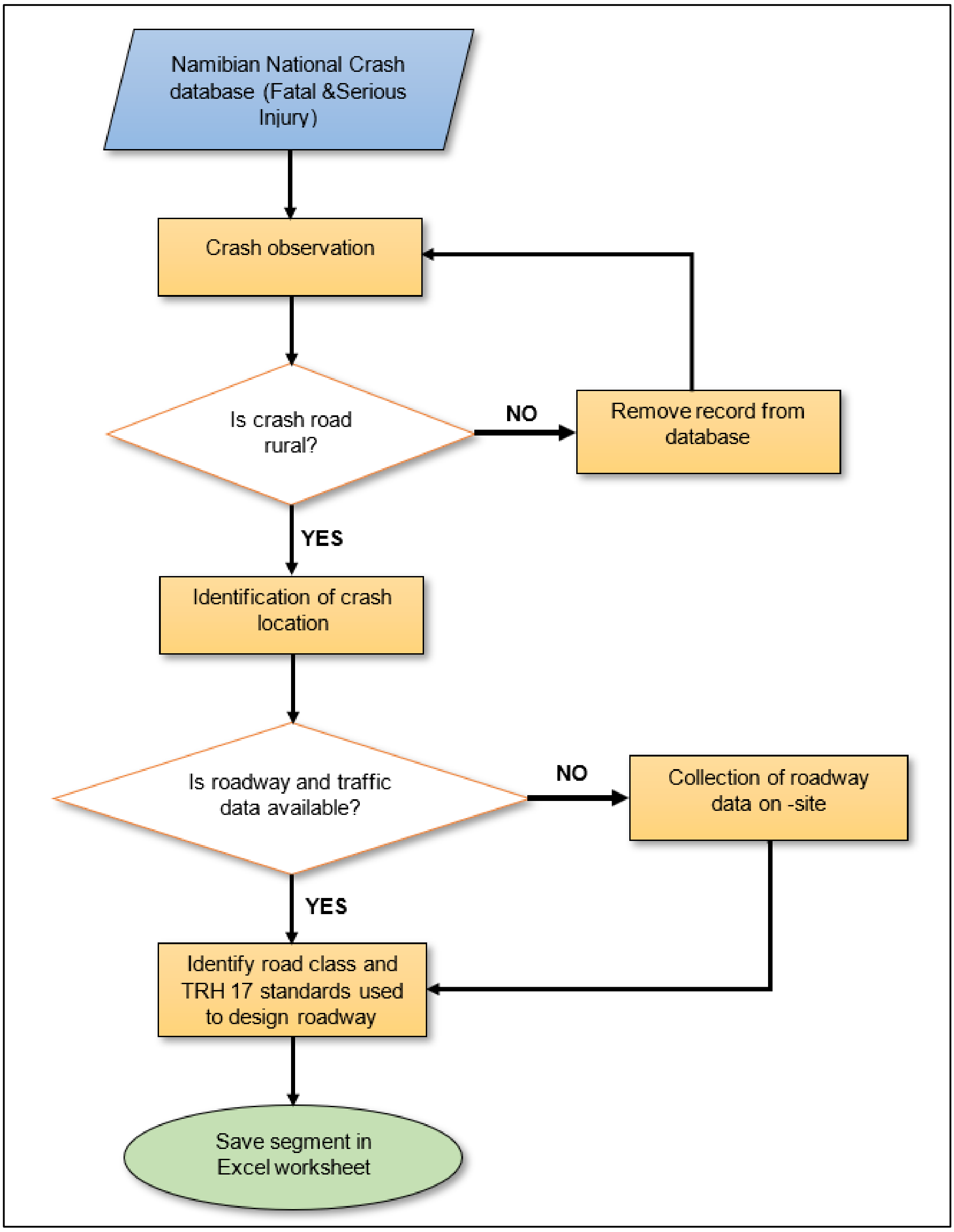


Figure 3.9 Roadway data processing steps

3.5. Research instruments

The following software packages and study equipment were used to collect, manage and analyse research data:

3.5.1. Data collection tools

a) Radar speed gun: The device was used to measure the speeds of moving vehicles on the national rural roads.
b) Measuring (Trundle) wheel: The device was used to measure the width of lanes on the national rural roads.
c) Google maps: The web-based application was used in determining the coordinates of the crash locations on the national rural road network.

3.5.2. Data Management tools

d) Mendeley: The application enabled the creation of the reference database and as a means to organise and manage the study material (journals, reports and other research studies).

3.5.3. Data processing and analysis tools

e) QGIS: The Geographical Information System tool was used to develop heat maps, which provided a visual summary of the road crash clusters of multiple severity on the road network.
f) IBM SPSS Statistics 25, STATISTICA and Microsoft Excel 2019: These software applications provided a comprehensive set of data processing and statistical tools to clean, aggregate and process study data. These software’s were also used as data management tools (Section 3.5.2)

3.6. Data analysis

3.6.1. Road crash, driver risk factors and behavioural aspects analyses (Univariate and Bivariate analyses)

The statistical package STATISTICA, IBM SPSS 25 and Excel were used to analyse data in this section. The packages were used to run univariate and bivariate analyses. The statistical analyses were based on a 95 percent confidence level. The data was coded accordingly and dummy variables for the independent variables namely: Gender, Weekday, Week, Month and Week of the month and Region were created for the analyses, using 1 and 0 for the variable under consideration. The dependant variable was the fatal and serious crash counts. Table 3.10 describes the categorisation of key variables used in the analyses.

Table 3.10 Categorisation of demographic variables

Independent variables	Categorisation
Gender	Gender of driver involved in crash {0 = Male; 1 = Female}
Week of the month	The week of the month when the accident took place: {1 = 1st week of the Month, 2 = Second Week of the Month, 3 = Third Week of the Month, 4 = 4th Week of the Month, 5 = 5th Week of the Month}
Weekday	Whether the accident took place during the weekday or weekend day {0= Week day: [Monday, Tuesday, Wednesday, Thursday & Friday]} and 1= Weekend day: [Saturday & Sunday]}
Month	Month in which the accident took place {1 = January; 2 = February; 3= March; 4= April; 5 = May; 6 = June; 7 = July; 8 = August; 9 = September; 10 = October; 11 = November and 12 = December}
Region	The region in which the road accident took place: { 1 = Erongo, 2 = Caprivi, 3 = Hardap, 4 = Karas, 5 = Kavango West, 6 = Kavango East, 7 = Khomas, 8 = Ohangwena, 9 = Omaheke, 10 = Omusati, 11 = Oshana, 12 = Oshikoto, 13 = Otjozondjupa and 14 = Kunene}

3.6.1.1. Univariate analysis (Descriptive)

The univariate analysis method was used to describe and observe the trends of the historic crash data by reviewing the distribution of the crash records and determining the central tendency measures (mean, mode and median) and dispersion values (standard deviation, range, quartiles, variance , minimum and maximum values) (Bruce and Bruce, 2017). Further, the extent to which roadway factors shown in Figure 3.10, at different levels of significance and combinations, were involved in crash observations was determined using univariate analysis methods.

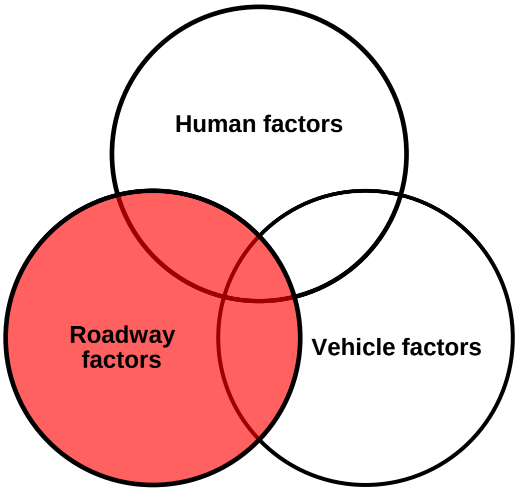


Figure 3.10 Crash risk factors combination

3.6.1.2. Bivariate analysis

Bivariate analysis refers to the process of investigating associations between two variables with the aims of describing the data set and drawing inferences from the association of those variables (Ali and Bhaskar, 2016). Although univariate analysis (descriptive statistics) is vital to describe a set of data in terms of the frequency of occurrence, central tendencies and the dispersion of the data, it is not sufficient to address the statistical queries that arise from reviewing data (Saxena *et al.*, 2006). Inferential statistics is a type of bivariate statistical analysis used to arrive at conclusions beyond sample statistics, with the aim of hypothesis testing (Bruce and Bruce, 2017).

Confidence intervals and hypothesis testing are dependent on whether the statistical test is parametric or non-parametric (Ali and Bhaskar, 2016). The underlying assumptions of parametric tests restrict its application to a "*normally distributed population, a data set with a homogeneity of variance and a dataset where all the observations are independent of each other"* (Montgomery and Runger, 2014).

Several assumptions were made in numerous steps shown in Figure 3.11 to ensure that the most appropriate statistical tests were used in this study. Firstly, the distribution of the data was analysed for outliers using Excel and STATISTICA. This was done through determining Cook's distance and the development of box-whisker diagrams and histograms to display percentiles and outlier summaries. The lower quartile (Q1) (25^{th} percentile) and upper quartile (Q3) (75^{th} percentile) indicators were determined to apply the outlier labelling rule. The interquartile range (IQR) (difference between the upper and lower quartiles) was calculated for the distribution. For accurate outlier spotting, values lower than the lower limit (Q1 – (2.2(IQR)) and values greater than upper limit (Q3 + (2.2(IQR)) were labelled as outliers.

Secondly, linearity – which assumes a linear relationship between the independent and dependant variables - was evaluated using scatter plots. A linear relationship is assumed if the scatter plot follows a linear pattern, otherwise a non-linear relationship is assumed.

Thirdly, the distribution of the data was assessed for normality by visually inspecting whether the histograms were symmetrical or not (if bell-shaped or not). The normal distribution test can also be done using the probability-probability plot (P-P plot). For a normal distribution the data points are expected to be as close as possible to the ideal diagonal line in the plot. Should the data points significantly deviate from the diagonal line, the normal distribution is not appropriate to describe the distribution.

Lastly, the assumption of homogeneity of variance of the data set was tested using Levene's test (Neill, 2006). Levene's test investigates the null hypothesis that different data groups have an equal variance at an alpha level of 5 percent (0.05) (Gastwirth *et al.*, 2010). The *p*-value determined by Levene's test confirms whether the assumption is approved or negated. For *p*-values greater than 0.05, the assumption that the variance is equal across the data groups is accepted. For *p*-values less than 0.05, the alternative hypothesis that variance is different across the data groups is accepted.

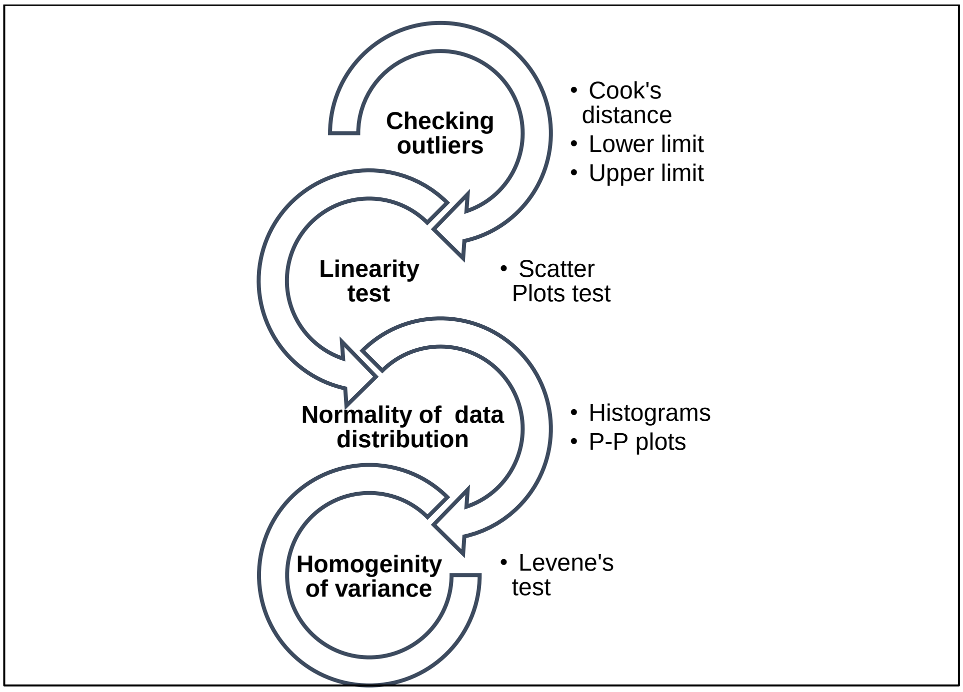


Figure 3.11 Assumptions applicable to Bivariate analyses

In carrying out inferential statistic tests on the data set, an extension of the independent t-test called the one-way analysis of variance (ANOVA) was used to test whether statistically significant differences exist between the means of two or more data groups (Al-Matawah, 2009). As ANOVA does not indicate which data groups are different, follow-up post hoc tests were used to identify the specific differences alluded to by the ANOVA test (Ho, 2006; Field, 2013). Various post-hoc tests (in red) shown in Figure 3.12 were used to investigate the differences in the data groups, based on assumptions on variances and data groups sample sizes. Therefore, the post-hoc tests used in the study were determined by the variance tests from Levene's test and the sample sizes of the crash data groups.

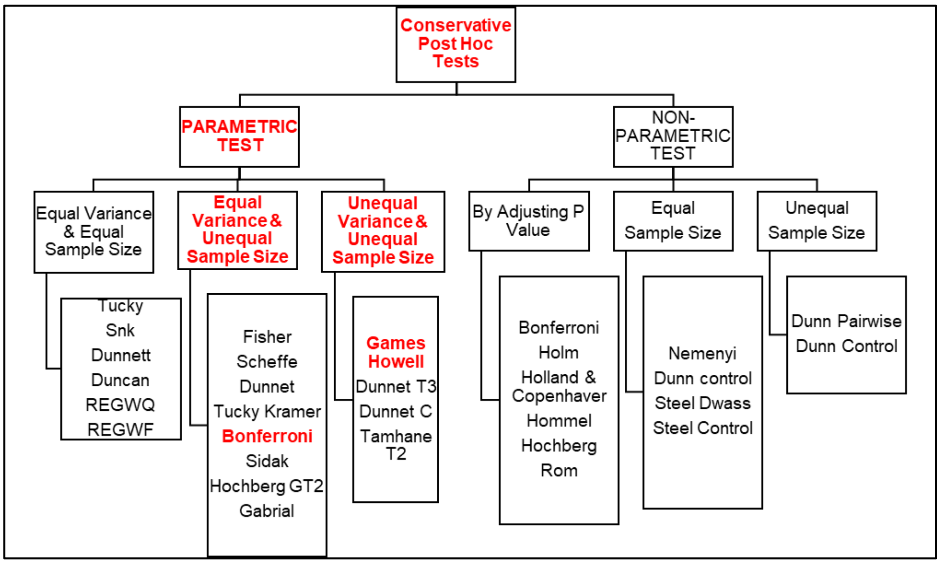


Figure 3.12 Post-hoc tests application

3.6.2. Determining road segment crash rates

Crash frequency is a useful tool to compare the temporal differences in the number of crashes occurring at a given location and observing trends (Vinayakamurthy *et al.*, 2017). Crash frequency is often inadequate to compare the occurrences of road crashes on multiple road locations as it does not consider road user exposure (Demissie, 2017). The crash rate method improves on the crash frequency by normalising the frequency of the road crashes with exposure, as measured by the traffic factors and the length of the study links(Garber and Hoel, 2009). The crash rate method also allows for a direct comparison of the road safety condition of multiple road segments (Cenek *et al.*, 2012).

In order to identify road segments with the highest severity risks, the crash rate considered the fatal and serious crash information from the database and the rest of the rural road network, using Equation [3.1] presented as crashes per million vehicle kilometres.

$$CR = \frac{Crashes \times 10^6}{AADT \times 365 \times T \times L} \qquad [3.1]$$

Where; *CR* = Crashes per million vehicle kilometres

AADT = Average Annual daily Traffic

L = Length of road segment (km)

T = Length of study period (years)

365 = Number of days in a year

3.6.3. Road crash geospatial analysis (Crash distribution on road network))

In road safety, road crash hot spots refer to a location with a record of large number of road crashes or crashes with high severity (Toran and Moridpour, 2015). Geographic Information System (GIS) is one of the useful tools in crash hot spot analysis. Using GIS, it is possible to join road crash dataset to the road network and other variables (Lloyd, 2010; Choudhary *et al.,* 2015). GIS in spatial data analysis is used to analyse road crash hot spots in road networks (Ouni and Belloumi, 2019). In this study, an analytical procedure proposed by Mitchel (2005) was adopted to carry out the geospatial analysis with the use of QGIS. The analytical procedure is presented in Figure 3.13.

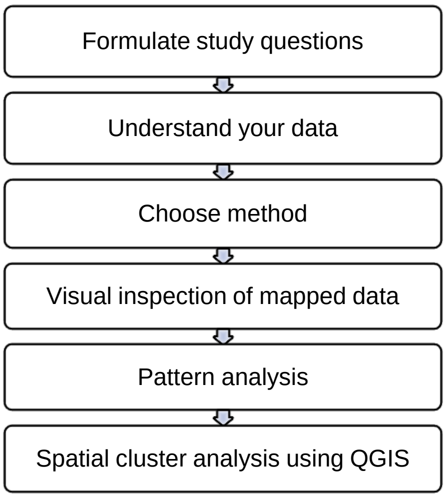


Figure 3.13 Geospatial analytical procedure analysis (Mitchel, 2005)

1. Formulation of study questions

The geospatial analysis process was performed with the intention of addressing specific study questions as formulated in Chapter one of the study. Geospatial analysis enabled the identification and examination of the location of road crashes to study the combination effects of road design and traffic conditions on road crashes. The geospatial analytical procedure addressed the following research questions: (i) Where are the road crash hotspots on the Namibian national rural road network? (ii) What are the characteristics of the road crashes on the identified study sections? (iii) Do the design variables on the identified study sections comply with road design standards in Namibia? Addressing the formulated study questions in a spatial context generated an understanding of the relationship between the road crashes and the national rural road environment.

2. Understanding of data

The type of data and its features help determine the specific method to use in geospatial analyses (Mitchel, 2005; Smith *et al.*, 2009). Features can be represented in GIS using two models of the world; vector and raster (Farkas *et al.*, 2016). With the *vector model*, each feature is a row in a table and feature shapes are defined by x and y locations in space (GIS connects the dots to draw lines and outlines) (Lloyd, 2010). Features can be discrete locations or events, lines or areas (Satria and Castro, 2016; Dereli and Erdogan, 2017).

Locations - such as the precise location of a road crash - are represented as points having a pair of geographic coordinates (Lloyd, 2010), as shown in Figure 3.14.

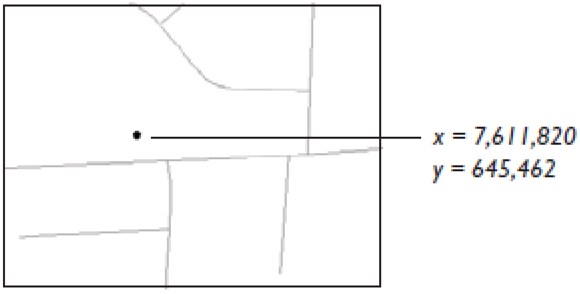


Figure 3.14 Representation of a location in GIS

Lines - such as roads, streams or pipelines - are represented as a series of coordinate pairs (Lloyd, 2010), as shown in Figure 3.15.

Figure 3.15 Representation of a line in GIS

Areas are defined by borders and are represented as closed polygons (Taha, 2016; Lloyd, 2010), as shown in Figure 3.16. Areas can be defined as administrative; such as regions or provinces, or naturally occurring boundaries; such as watersheds (Câmara *et al.*, 2002; Farkas *et al.*, 2016). When analysing vector data, much of the analysis involves working with the attributes in the layers data table (Lloyd, 2010).

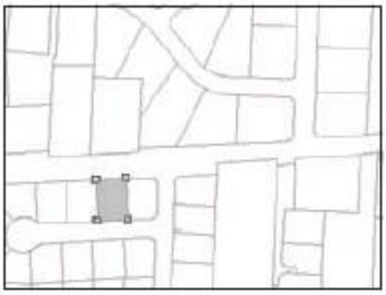


Figure 3.16 representation of an area in GIS

With the *raster model*, features are represented as a matrix of cells in continuous space (Sommer and Wade, 2006). Each layer represents one attribute (other attributes can be attached); associated with a numerical value or class and positional information (Lloyd, 2010). Most analysis occur by combining the layers to create new layers with new cell values.

The type, quality, strengths and weaknesses of the spatial data are significant in determining the analytic tasks and statistical techniques applicable. Therefore, understanding the type of data and its aspects is an important step in the overall process of geospatial analysis. The spatial data for the study comprised different types of data, i.e. points, lines, polygons and raster data.

3. Choosing Geospatial analyses method

A wide variety of approaches can be performed in geospatial analysis to study spatial locations and investigate the distribution of a phenomena. These approaches are known as spatial statistics or geostatistics as they apply a range of statistical techniques designed to analyse and predict the values attached to spatial phenomena (Sommer and Wade, 2006). Geostatistics makes use of standard statistical techniques such as exploratory data analysis, descriptive and inferential statistics, and modelling techniques to analyse the spatial data (Câmara *et al.*, 2002).

The geospatial analysis in the study encompassed two main tasks; the creation and manipulation of map layers and running of exploratory spatial data analysis (ESDA). The first task includes activities such as the creation of map layers in QGIS from the spatial data, reviewing the created maps and checking attribute data connected to the map features, editing attribute tables, aggregating data and performing spatial queries. The second task - exploratory spatial data analysis (ESDA) - involves a range of techniques to (i) visualise spatial data in a spatial framework using maps and other graphics; (ii) identify patterns of spatial clustering and association through spatial correlation and regression analysis; (iii) detect significant patterns; and (iv) recommend different forms of spatial heterogeneity (de Smith *et al.,* 2009). Descriptive statistics and feature clustering to quantify spatial patterns are utilised by this approach. The spatial autocorrelation techniques used in the study are discussed below.

4. Visual inspection of mapped data

The visualisation of data was the starting point for ESDA after the creating of maps and performance of spatial queries in QGIS. Several techniques are involved in the visualisation of data, including data graphing and mapping using a combination of visual elements; heatmaps, choropleth maps, scatter plots, graphs and 3D maps (de Smith *et al.,* 2009). A heat map is a graphical representation of data where the different individual values contained in a matrix are represented using a colour coding system to provide a visual summary (Sommer and Wade, 2006; de Smith *et al.,* 2009).

The study applied spatial visualisation GIS-related techniques, including categorising of spatial data and designing map symbology for each category, controlling selected values to be displayed, addressing spatial outliers, creating a map series and mapping density values, creating map layouts, adding graphs and printing map outputs.

5. Pattern analysis

Spatial pattern analysis is used to geographically specify the locations where road crashes occurred and to assess the specific patterns of distribution through map visualisation (Toran and Moridpour, 2015; Kundakci, 2014). Kernel Density Estimation (KDE) is one of the most significant spatial data analysis techniques. Several studies in the literature have employed the KDE technique to analyse road traffic crashes. The main reason for employing this method is that hotspots in KDE are based on an area with crash risk rather than a certain point. This is because the real position of the crash is dependent on the accuracy of a GPS device. Identifying the exact position of a road crash is not always easy. For instance, the point of cause of a road crash may be different from the position of the crash, thus the location of the crash reported by the police officer is different from the exact point of the crash. In this study, KDE was applied to identify road crash hotspots on the different national rural road classifications.

Kernel Density Estimation

A kernel distribution is a nonparametric representation of the Probability Density Function (PDF) for a random variable (Satria and Castro, 2016; Ghadi and Török, 2017). Kernel distribution is used when a parametric distribution cannot properly describe the data. Also, kernel distribution is utilised to avoid making assumptions about the distribution of spatial data (Pljakić *et al.*, 2019). Kernel distribution is defined by a smoothing function and a bandwidth value which control smoothness of the resulting density curve and affect the results of the hotspot analysis (Toran and Moridpour, 2015; Shafabakhsh *et al.*, 2017).

KDE involves placing a symmetrical surface over each variable and evaluating the distance from a point to a reference location based on a mathematical function (Toran and Moridpour, 2015; Hashimoto *et al.*, 2016). The values of all surfaces related to each variable are accumulated for the reference location and this procedure is repeated for all reference locations in this estimation. In the kernel method, a study area is divided into a number of predetermined cells (Toran and Moridpour, 2015). Hence, the kernel method draws a circular neighbourhood around each feature point (each road crash). Subsequently Equation [3.2] is used, which goes between 1 at the position of the crash and 0 at the neighbourhood boundary as illustrated in Figure 3.17.

$$f(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} K(\frac{d_i}{h}) \quad [3.2]$$

Where; $f(x, y)$ = Density estimation at location (x, y)

n = Number of observations

h = bandwidth or kernel size

K = Kernel function

d_i = Distance between the location (x, y) and the location of ith observation

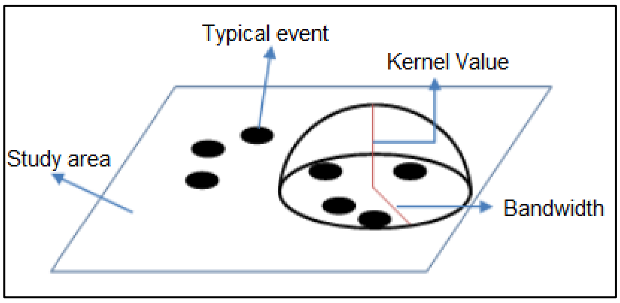


Figure 3.17 Kernel function

There are different types of kernel functions, such as Quartic, Conic, Gaussian, Negative exponential and epanichnekok (Toran and Moridpour, 2015; Satria and Castro, 2016; Pljakić *et al.*, 2019). The choice of the kernel function K is less important than the impact of the bandwidth r in planar KDE (Toran and Moridpour, 2015). The study applied the specific form of the Quartic kernel function (QKF) shown in Equation [3.3]. The QKF was applied as it provides the best approximation of the true density of the variables in the study area.

$$k\left(\frac{d_i}{h}\right) = K\left(1 - \frac{d_i^2}{h^2}\right) \text{ when } 0 < d_i \leq h \quad [3.3]$$

$$k\left(\frac{d_i}{h}\right) = 0 \text{ when } d_i > h$$

Where; k = Kernel function

d_i = Distance between the location (x, y) and the location of the ith observation

K = Scaling factor (To ensure the total volume under Quartic curve is 1)

6. Spatial cluster analysis using QGIS

The planar kernel density estimation (KDE) was applied to visualise where the clusters of road crashes appear. The bandwidth and the grid size are two key parameters that influence the results of the hotspot analysis (Satria and Castro, 2016; Saha and Ksaibati, 2016). Five different bandwidth values were tested (200 m, 400 m, 500 m, 800 m and 1 000 m) to achieve the best visualisation of road crash hotspots on a grid cell size of 30 m by 30 m, given the size of the study area and the processing time required for hotspot identification in QGIS. The bandwidth values were applied in several previous studies and were adopted in this study, to allow for a comparison between the study results and previous study findings.

The KDE tool produces a raster map where the density of the road crashes is displayed by continuous surfaces (Hashimoto *et al.*, 2016; Pljakić *et al.*, 2019). Lighter shades on the raster map represent locations with lower road crash intensity, while darker shades indicate areas with higher road crash densities. The study classified the surfaces into four equal intervals according to their density as shown in Figure 3.18. The top density locations in the classification are defined for the highest 25 percent of total density and lowest density sections for the lowest 25 percent of total density in each location.

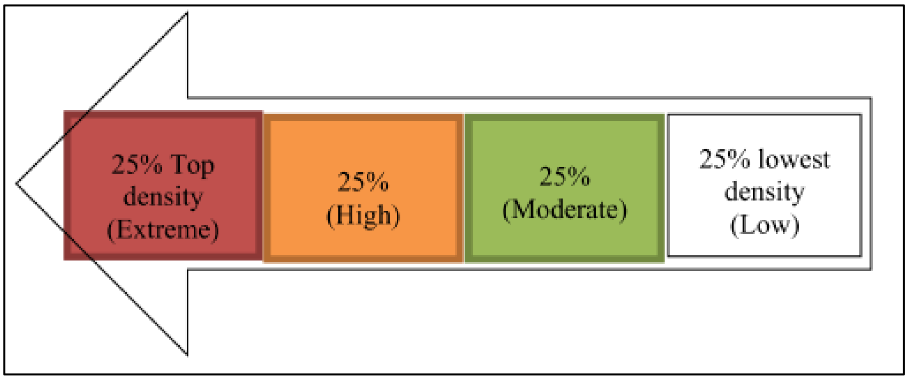


Figure 3.18 Classification of road crash hotspots

3.6.4. Road Crash Prediction Model Development (Multivariate analysis)

The study developed a General Regression Model – multiple linear regression (MLR) model approach that can be applied to predict rural road crash rates and investigate the combinational effects of geometric and traffic characteristics on road safety. The approach involved the aggregation of design and traffic factors detailed in Section 3.2.2, and fatal and serious injury (FSi) to satisfy the linear regression assumptions – namely error structure normality and homoscedasticity. The modelling approach was tested and validated using data from three datasets, representing FSi crashes on all rural roads, higher and lower order rural roads. Through the use of data manipulation, it was possible to satisfy the assumptions of the GRMs and thus develop robust crash prediction models (CPMs). The study produced and compared the CPMs using the base mean multiple linear regression models and the robust winsorised and transformed CPMs to determine the best performing model described in Section 2.9.4. The approach that was taken in the development of the crash prediction tool incorporates principles from Safe System and Sustainable Safety approaches to road safety. This section provides a description of the crash prediction model development process and the goodness-of-it measures of the model.

3.6.4.1. Model development

General Regression Model – best subsets multiple linear regression (MLR) is an exploratory model building regression analysis approach that was used to perform and build a correlation analysis between FSi rural road crashes (independent variable) the various geometric and traffic characteristics (covariates) (Rakha *et al.*, 2010; Islam *et al.*, 2019). The best subsets MLR compared all possible models using a specified set of predictors (geometric and traffic variables) and displayed the best-fitting model (Denis, 2021). The model function took the form shown by Equation [3.4].

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \cdots + B_mX_m \qquad [3.4]$$

Where: Y = Dependant variable,

B_0 = Regression constant,

$B_1, B_2, B_{3....}\ B_m$ = Regression coefficients of respective m dependant variables,

$X_1, X_2, X_{3....}\ X_m$ = Covariates.

A statistical test of the model was done, which included determining the following: (1) the coefficient tests (R^2 test), (2) the significance test of the regression coefficient (t-test), and (3) the significance test of regression equation (F-test) (Field, 2013). In any case were the significant test of regression equation failed, it was possible that important factors were missing during the selection of covariates or the relationship between the independent variable and the covariates was found to be non-linear (Field, 2013; Gupta, 2017). In such a case the CPM is rebuilt.

The MLR analysis assumed and examined several key assumptions for the developed crash prediction models (Rakha *et al.*, 2010). These assumptions comprised:

- Linear relationship between model variables
- Error structure normality of the model variables
- Multicollinearity, independence and homoscedacity between the model variables

3.6.4.2. Testing autocorrelation and variable selection

I. Factor Analysis

The factor analysis is a statistical technique applied to reduce a large number of variables into a fewer regression factors – latent variables, based on shared variance. The Factor Analysis method is part of the General Linear Models (GLM) and assumes several key assumptions. These assumptions include: (I) a linear relationship and no multicollinearity between relevant variables included in analysis, and (II) a true correlation between the tested variables and factor (Qian and Künsch, 1996; Rohe and Zeng, 2020).

The factor analysis technique extracts the maximum common variance from all variables and places them under a common score. The total variance of a particular variable consists of three components:

1. Variance that is shared with other variables (common variance)
2. Variance that is specific to that variable (unique variance), and
3. Error or random variance (referred to as unreliability of variance)

The proportion of common variance present in a variable is referred to as "communality". As a result, a variable with no unique variance and error variance would have a commonality of one (1) while a variable that shares none of its variance with other variables would have a commonality of zero (0). Communality is a key concept in factor analysis as the approach is oriented towards finding common variance between the analysis variables (Achcar *et al*., 2013). For this reason, variables with low communalities (less than 0.20 o that 80 percent of variance is unique) are eliminated from the CPM analysis.

The study thus applied the common factor analysis method to extracts common variance and reduce the large number of variables into smaller set of factors. This method does not include the unique variance of all the variables and is applied in Structural Equation Modelling (SEM) (Gargoum and El-Basyouny, 2016); (Bamdad Mehrabani and Mirbaha, 2018). The factors analysis had the following components:

a) Factor loading

Factor loading is the correlation coefficient for the analysis variable and factor. Factor loading indicates the variance explained by the variable on that particular factor. In the SEM approach, as a rule of thumb, 0.7 or the highest factor loading selected allows that the factor extracts sufficient variance from that variable, while ensuring that the variables are not duplicated in the factor rows (Rohe and Zeng, 2020). Table 3.11 shows an extract of the higher order rural roads variance at a factor loading of 0.58, resulting in the reduction of the variables into a set of five factors. The factor loading for all and lower rural road datasets are given in Appendix B.

Table 3.11 Principle factor components from factor loadings-Varimax normalised for High Order Rural Roads

Variable	Factor Loadings (Varimax normalized) (High Order Rural Roads Extraction: Principal components (Marked loadings are >.58)				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
AADT_Heavy	0,184	0,862	0,154	0,011	-0,108
AADT_Light	0,136	0,894	-0,063	0,019	0,141
85th Percentile Speed (Ops)	0,113	-0,047	-0,099	0,774	0,181
Lane_Width	0,021	0,380	0,611	-0,058	-0,182
No_Lanes	-0,449	0,570	-0,214	0,169	0,078
Shoulder_type	0,855	0,057	-0,161	0,001	0,151
Surface_SW	-0,881	-0,025	0,086	0,038	-0,007
Ground_SW	0,725	0,169	0,190	0,101	-0,239
Horizontal_(Curves/Length)	0,074	0,261	-0,794	-0,104	-0,105
Terrain_Vertical	-0,129	0,208	0,227	0,663	-0,210
Access_Density	-0,127	0,098	-0,083	0,074	0,615
Pavement_Condition	0,094	-0,001	0,111	-0,102	0,740
SSD	-0,010	0,181	0,286	-0,304	-0,414
Expl.Var	2,346	2,200	1,308	1,202	1,337
Prp.Totl	0,180	0,169	0,101	0,092	0,103

b) Eigenvalues

Eigenvalues are referred to as characteristic roots. The Eigenvalues showed variance explained by each particular factor out of the total variance. The commonality column explains how much variance is explicated by the first factor out of the total variance (Walker and Maddan, 2009; Daniel, 2016). The Eigenvalues were used to determine the best number of variables that can be applied to develop the best CPM.

Criteria for determining the number of factors

Eigenvalues are a good criterion for determining factors according to the Kaiser Criterion. The study also applied the scree plot (a line of eigenvalues of factors) as an indicator of the number of factors to retain in the principal component analysis (Ho, 2006). The cut-off for the principle factors was determined according to where the “elbow” formed on the scree plot, which represented the point where a smaller number of interpretable factors explain the maximum amount of variability in the data. Figure 3.19 shows the eigenvalues and scree plot applied in the selection of factors in the higher order rural roads dataset. The eigenvalues and scree plots for all and lower order rural roads are given in Appendix B.

	Eigenvalues (High Order Rural Roads) Extraction: Principal components			
Factors	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,469948	18,99960	2,46995	18,99960
2	2,087724	16,05942	4,55767	35,05901
3	1,626941	12,51493	6,18461	47,57394
4	1,188430	9,14177	7,37304	56,71571
5	0,993876	7,64520	8,36692	64,36091
6	0,917453	7,05733	9,28437	71,41824
7	0,842470	6,48054	10,12684	77,89878
8	0,760974	5,85364	10,88782	83,75243

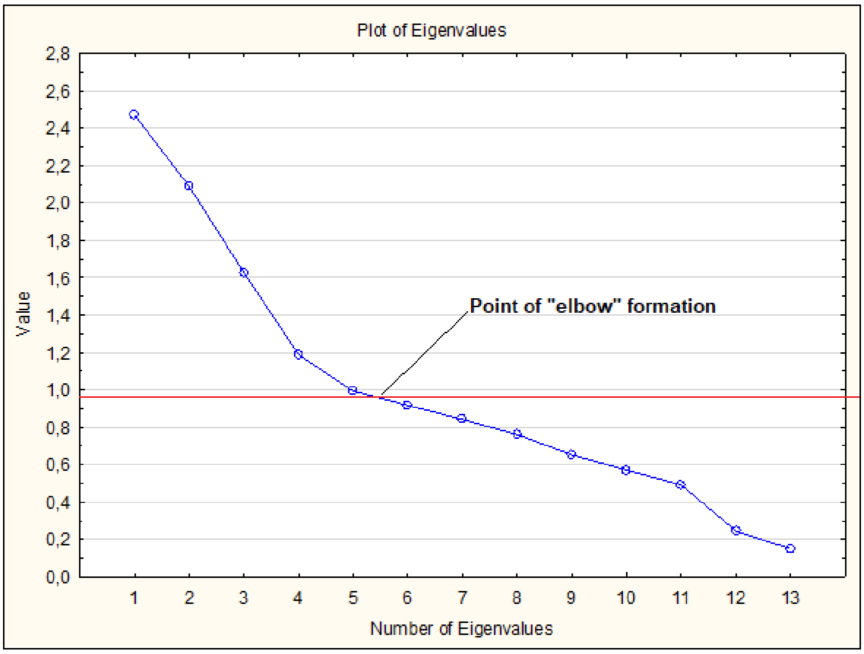


Figure 3.19 Eigenvalues and Scree plot for High Order Rural Roads

Rotation Method

Rotation method improves the reliability and understandability of the Factor Analysis output. The rotation method affects the percentage of variance extracted from the factors (Field, 2013; Daniel, 2016). The study applied the Kaiser-Varimax Rotation (KVR). The KVR maximises the sum of the variance of the squared loadings. This process results in high factor loadings for a smaller number of variables and low factor loadings for the rest (Daniel, 2016).

Key summaries for Factor Analysis

The Factor Analysis technique assumes and tests the following assumptions about the dataset:

i. No outlier: Assumes that there are no outliers in the dataset.
ii. Adequate sample size: The case must be greater than the factor.
iii. No perfect multicollinearity: Factor analysis is an interdependency technique. There should be seamless multicollinearity between the dataset variables.
iv. Homoscedasticity: Factor analysis does not require homoscedasticity between variables since it is a linear function between measured variables.
v. Linearity: Factor analysis assumes of linearity. Non-linear variables can also be used after it has been transferred into the model and converted into a linear variable.
vi. Interval data: Interval data is assumed in factor analysis.

II. Durbin Watson Test

The Durbin Watson (DW) test is measure of autocorrelation (serial correlation) in residuals from a regression analysis (Maxwell and David, 1995). Autocorrelation is the similarity of a time series over successive time intervals. Autocorrelation can lead to underestimates of the standard error and can misidentify predictors as statistically significant (Alexopoulos, 2010; Field, 2013). The study applied the Durbin Watson statistic to test the assumption that the error terms used in the CPM are independent of each other. The DW test statistic was calculated using Equation [3.5]

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \qquad [3.5]$$

Where, DW = the Durbin Watson value

$e_i = y_i - \hat{y}_i$ are the residuals

n = the number of elements in the sample

k = the number of independent variables

The DW test reports a test statistic value between zero (0) and four (4), where:

- A DW value equal to two (2) means no autocorrelation
- A DW value from 0 to < 2 means positive autocorrelation
- A DW value > 2 to 4 means negative correlation

A rule of thumb is that DW values in the range of 1.5 to 2.5 are relatively normal. However, values outside of this range could be a cause of concern as they suggest that the data elements being either too close (positive autocorrelation) or too far (negative autocorrelation) from the subsequent data element (Field, 2013). Table 3.12 shows the results of the Durbin Watson residual test on the higher order rural road crash prediction model developed. A DW value of 2.009569 which is closer to two indicates that no autocorrelation exists in the model and a very low serial correlation of -0.005469 also supports the conclusion made by the DW value.

Table 3.12 Durbin-Watson Test for High Order Rural Roads CPM

Durbin-Watson d (CR Model and Serial Correlation of Residual)		
	Durbin-Watson d	Serial Corr.
Estimate	**2.009569**	**-0.005469**

3.6.4.3. Outlier analysis

Outliers are defined as data points that are different from the rest of the data (Chambers *et al.*, 2000; Achcar *et al.*, 2013) . The identification of outliers is vital as the results of statistical analyses should not be highly influenced by errant data points (Field, 2013). The study applied 2D box plots as a diagnostic tool for detecting outliers and data influential points, and ultimately used the Winsorization technique to address detected outliers in the dataset. Winsorizing is the process of replacing a specified set of extreme values of a given variable in a set of sample data with specified values computed from the data. The 2D Box Plots of the crash rate distribution are shown in Figure 3.20 before and after the Winsorization process. A pre-defined rule is used to adjust an outlying (positive) value Y_i of the dataset variable Y downwards, leaving the remaining values unchanged (Hicks and Fetter, 1991; Reifman and Keyton, 2010). The value of the adjusted variable is denoted Y_i^* and the corresponding winsorised estimator adjusted to a fixed cut-off is represented by Equation [3.6].

$$\hat{Y}_t = \sum_{j=1}^{n} adj w_j^t y_j \qquad [3.6]$$

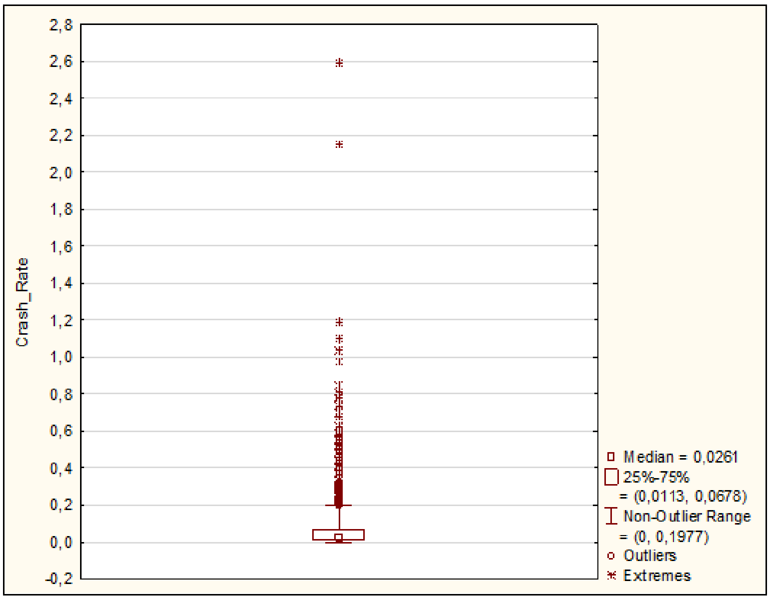
Where, t = truncation level

y_j = reported crash rate for the j^{th} unit

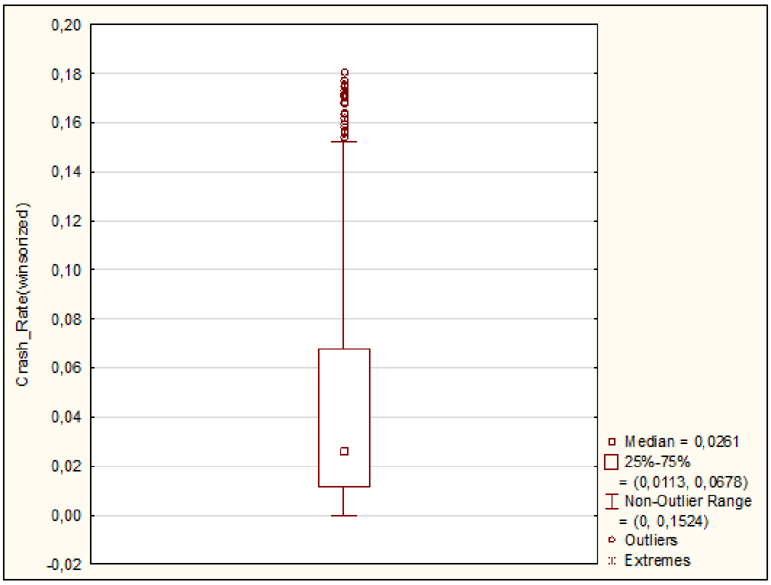
$$adj = \frac{\sum_n w_j}{\sum_n w_j^t}$$

$$w_j^t = \begin{matrix} w_j, \ if \ w_j y_j \le t \\ \frac{t}{y_j}, \ if \ w_j y_j > t \end{matrix}$$

The weights of the observations whose expanded weighted value is larger than t are truncated so that the expanded value now equals to t. The truncated portions are then smoothed over for all observations.



(i) 2D Box Plot of crash rate before Winsorization



(ii) 2D Box Plot of crash rate after Winsorization

Figure 3.20 2D Box Plots of the crash rate distribution before and after Winsorization

3.6.4.4. Crash model Biplots

Biplots are a graphical representation of information in a $n \times p$ data matrix, with information in rows representing samples and information in columns representing covariates. In the Principal Component (PC) analysis a plot can be obtained by graphing the first two principal components of the units (Gower *et al.*, 2011). In biplots the idea is to add information about the covariates to the PC graph.

Construction of Biplots

The best two-dimensional approximation of data in a $n \times p$ matrix is determined by approximating the j^{th} observation vector $\underline{x}_j$ in terms of the sample values of the first two PC's. The approximation is given by Equation [3.7].

$$\underline{x}_j = \underline{\overline{x}} + \hat{y}_{j_1}\hat{e}_1 + \hat{y}_{j_2}\hat{e}_2 \qquad [3.7]$$

Where; $\underline{\hat{e}}_1$ and $\underline{\hat{e}}_2$ are the first two eigen vectors of $(n-1)S = x_c' x_c$. Where, x_c is equal to the mean corrected data with row vectors ($\underline{x}_j - \underline{\overline{x}}$)'.

On the biplot, the eigen vectors $\underline{\hat{e}}_1$ and $\underline{\hat{e}}_2$ define plane. The coordinates $\hat{y}_{j_1}$ and $\hat{y}_{j_2}$ for $j = 1, \dots, n$, define the n units in that plane - Principal Component scores. The variables $x_1 \dots x_p$ are positioned on the graph by the row vectors of $\hat{E} = [\underline{\hat{e}}_1, \underline{\hat{e}}_2]: p \times 2$, since:

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \underline{\hat{e}}_1' \\ \underline{\hat{e}}_2' \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \qquad [3.8]$$

The lengths of the vectors from x_1 to x_p can be adjusted to ensure that all the variables are plotted on the same graph as the points $(\hat{y}_{1j}, \hat{y}_{2j}); j = 1, \dots n$.

3.6.4.5. Assessment of goodness-of-fit

Two goodness -of-fit statistic tests were used to evaluate the fit of the crash prediction models developed for the rural road network – R-Squared and the overall F-test. The model fit tests are based on the two sums of squares theories: Sum of Squares Total (SST) and Sum of Squares Error (SSE) (Alexopoulos, 2010; Field, 2013). The SST measures how far the data points are from the mean and the SSE measures how far the data points are from the crash predictions model's predicted values (Field, 2013). Different combination of the SSE and SST values provide different information about how the crash models compare to the base mean models (Field, 2013).

I. The R-Squared and Adjusted R-Squared

The difference between the SST and SSE is the improvement in prediction from the regression model developed, compared to the mean model. The R-squared value is then determined by dividing the difference between the SST and SSE by the SST (Maydeu-Olivares and Garcia-Forero, 2010; Field, 2013). The R-squared value represents the proportional improvement in prediction from the regression model compared to the mean model, and indicates the goodness of fit of the crash model to the crash dataset.

The R-squared statistic has the useful property that it is intuitive: it ranges from zero to one. An R squared statistic value of zero indicates that the proposed crash model does not improve predictions over the base mean test model, while a statistic value of one indicates a perfect prediction characteristic from the crash model (Alexopoulos, 2010; Field, 2013). Therefore, improvements in the crash model result in proportional increases in the R-squared statistic.

One pitfall of the R-squared statistic is that it can only increase as predictors are added to the crash prediction model. This increase in the statistic is artificial when predictors are not actually improving the model's goodness-of-fit to the crash data. To remedy this, a related statistic, the Adjusted R-squared, incorporates the crash model's degrees of freedom in the test (Field, 2013). The Adjusted R-squared will decrease as predictors are added if the increase in the model fit does not make up for the loss of degrees of freedom. In the same way, the Adjusted R-squared statistic will increase as predictors are added if the increase in the model fit is improving. The adjusted R-squared should always be used with models with more than one covariate. In summary, The Adjusted R-squared statistic is interpreted as the proportion of the total variance explained by the model in the outcome variable (Montgomery and Runger, 2014).

II. The F-Test

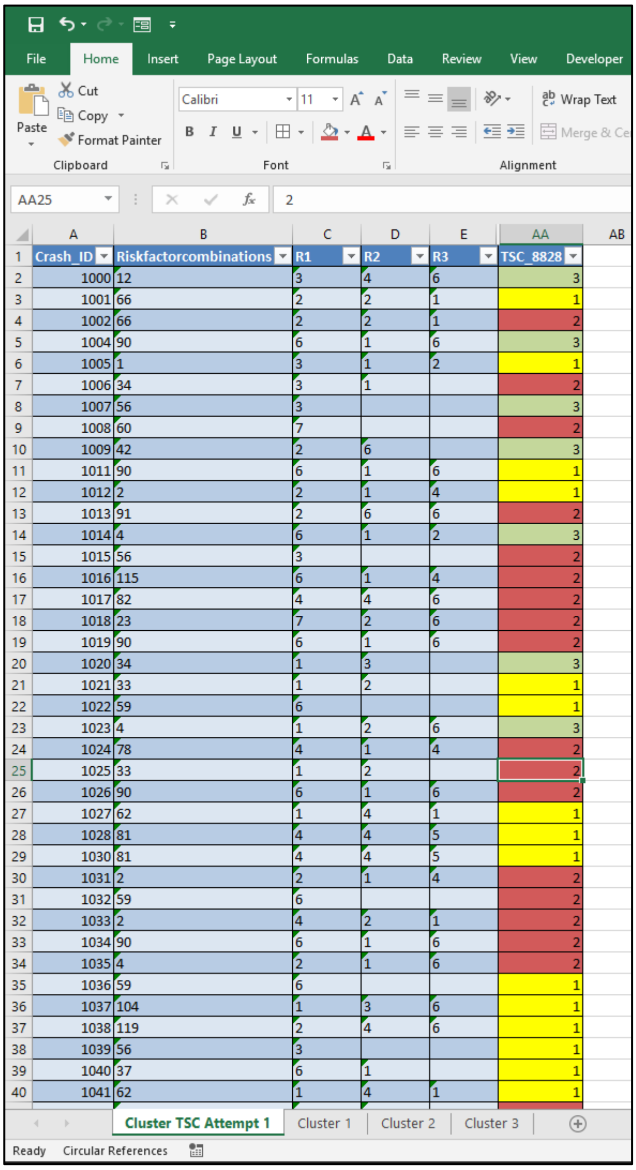
The F-test statistic evaluates the null hypothesis that all regression coefficients are equal to zero versus the alternative that at least one coefficient is not equal to zero (Field, 2013; Niewiadomska-Bugaj and Bartoszynski, 2021). An equivalent null hypothesis is when the R-squared statistic is equal to zero. A significant F-test indicates that the observed R-squared I reliable and is not a spurious result of oddities in the crash dataset. Thus, the F-test determines whether the proposed relationship between the outcome variable and the set of covariates is statistically reliable and can be useful when the objective is prediction and correlation.

3.6.4.6. Model Benchmarking

Benchmarking is aimed at demonstrating and improving the performance of crash prediction models by means of utilising a more diverse dataset, including some potential explanatory variables. To this end, benchmarking was carried out based on the macro CPMs developed with available road crash information from countries or regions (Northern Cape, Chile, Australia) with similar road conditions; to test the applicability of the crash prediction models in these countries. Gomes *et al.* (2019) notes the importance of addressing the strong dependence of CPMs on suitable and diverse input information, enabling these models to perform as “powerful tools” in road safety.

3.6.5. The Two-Step Cluster Analysis

The Two-Step Cluster (TSC) Models - A hybrid approach which first uses a distance measure to separate groups and then a probabilistic approach to choose the optimal subgroup models. Using the driver risk factors identified in Section 3.2.3, the study coded and grouped, through the TSC technique, all the risk factor combinations (see Figure 3.21) for each crash record. This allowed for combinations to be applied and tested in the TSC against explanatory factors explored in the study – demographic, temporal, and roadway and environmental factors. The development of the TSC models is discussed in Section 3.6.5.1.



Crash_ID	Riskfactorcombinations	R1	R2	R3	TSC_8828
1000	12	3	4	6	3
1001	66	2	2	1	1
1002	66	2	2	1	2
1004	90	6	1	6	3
1005	1	3	1	2	1
1006	34	3	1		2
1007	56	3			3
1008	60	7			2
1009	42	2	6		3
1011	90	6	1	6	1
1012	2	2	1	4	1
1013	91	2	6	6	2
1014	4	6	1	2	3
1015	56	3			2
1016	115	6	1	4	2
1017	82	4	4	6	2
1018	23	7	2	6	2
1019	90	6	1	6	2
1020	34	1	3		3
1021	33	1	2		1
1022	59	6			1
1023	4	1	2	6	3
1024	78	4	1	4	2
1025	33	1	2		2
1026	90	6	1	6	2
1027	62	1	4	1	1
1028	81	4	4	5	1
1030	81	4	4	5	1
1031	2	2	1	4	2
1032	59	6			2
1033	2	4	2	1	2
1034	90	6	1	6	2
1035	4	2	1	6	2
1036	59	6			1
1037	104	1	3	6	1
1038	119	2	4	6	1
1039	56	3			1
1040	37	6	1		1
1041	62	1	4	1	1

Figure 3.21 Coding and grouping of risk factor combinations by TSC technique

3.6.5.1. Development of the Two-Step Cluster analysis model

The Two-Step Cluster technique is an explanatory tool designed to reveal natural groupings (clusters) within a dataset that would otherwise not be apparent. The algorithms employed by the TLC have several desirable features that differentiate it from traditional clustering techniques. These features are:

- The ability to create clusters based on both categorical and continuous variables
- Automatic selection of the number of clusters
- The ability to analyse large data files efficiently

1. Clustering principles

In order to handle categorial and continuous variables, the Two-Step Analysis procedure uses a likelihood distance measure which assumes that variables in the cluster models are independent. Further, each continuous variable is assumed to have a normal (Gaussian) distribution and each categorical variable is assumed to have a multinomial distribution (Bacher *et al.*, 2004).

The TSC technique can be summarised as follows:

Step 1 – Pre-clustering of cases: The TSC tool technique begins with the construction of a Cluster Features (CF) Tree. The tree begins by placing the first case at the root of the tree in a leaf node that contains variable information about that case. Each successive case is then added to an existing node or forms a new node, based upon its similarity to existing nodes and using the distance measure as the similarity criterion. Two distance measures are available: Euclidean distance and a log-likelihood distance (Bacher, 2000; Chiu *et al.*, 2001). The log-likelihood distance can handle mixed type attributes. The log-likelihood distance between two clusters i and s is defined in Equation [3.9].

$$d(i,s) = \xi_i + \xi_s - \xi_{(i,s)} \qquad [3.9]$$

Where;

$$\xi_i = -n_i \left(\sum_{j=1}^{p} \frac{1}{2} \log(\hat{\sigma}_{ij}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^{q} \sum_{l=1}^{m_j} \hat{\pi}_{ijl} \log(\hat{\pi}_{ijl}) \right) \qquad [3.10]$$

$$\xi_s = -n_s \left(\sum_{j=1}^{p} \frac{1}{2} \log(\hat{\sigma}_{sj}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^{q} \sum_{l=1}^{m_j} \hat{\pi}_{sjl} \log(\hat{\pi}_{sjl}) \right) \qquad [3.11]$$

$$\xi_{(i,s)} = -n_i \left(\sum_{j=1}^{p} \frac{1}{2} \log(\hat{\sigma}_{ij}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^{q} \sum_{l=1}^{m_j} \hat{\pi}_{ijl} \log(\hat{\pi}_{ijl}) \right) \qquad [3.12]$$

ξ_v can be interpreted as a kind of dispersion (variance) within cluster $v(v = i, s, (i, s))$. ξ_v consists of two parts. The first part $-n_v \sum \frac{1}{2} \log(\hat{\sigma}_{vj}^2 + \hat{\sigma}_j^2)$ measures the dispersion of the continuous variables x_j within cluster v. If only $\hat{\sigma}_{vj}^2$ would be used, $d(i, s)$ would be exactly the decrease in the log-likelihood function after merging cluster i and s. The term $\hat{\sigma}_j^2$ is added to avoid the degenerating situation for $\hat{\sigma}_{vj}^2 = 0$. The entropy $-n_v \sum_{j=1}^{q} \sum_{l=1}^{m_j} \hat{\pi}_{vjl} \log(\hat{\pi}_{vjl})$ is used in the second part as a measure of dispersion for the categorical variables.

Similar to agglomerative hierarchical clustering, those clusters with the smallest distance $d(i, s)$ are merged in each step. The log likelihood function for the step with k clusters is computed as shown in Equation [3.13].

$$l_k = \sum_{v=1}^{k} \xi_v \qquad [3.13]$$

The function l_k is not the exact log-likelihood function. The function can be interpreted as dispersion within clusters. If only categorical variables are used, l_k is the entropy within k clusters.

Step 2 – Clustering of cases: A model based hierarchical technique is applied here. This means the leaf nodes of the CF tree are grouped using an agglomerative clustering algorithm. The agglomerative clustering can be used to produce a range of solutions. To determine the best number of clusters, each of the cluster solutions are compared using Schwarz's Bayesian Criterion (BIC) or the Akaike Information Criterion (AIC) as the clustering criterion (Chiu *et al.*, 2001). Using the two-phase estimator to automatically determine the number of clusters, the AIC is computed as shown in Equation [3.14].

$$AIC_k = -2l_k + 2r_k \qquad [3.14]$$

Where r_k is the number of independent parameters. The BIC is computed as shown in Equation [3.15].

$$BIC_k = -2l_k + r_k \log n \qquad [3.15]$$

Step 3 – Cluster membership assignment: Each object is assigned deterministically to the closest cluster according to the distance measure used to find the clusters. The deterministic assignment may result in biased estimates of the cluster profiles if the clusters overlap (Bacher, 2000). The importance measures of the assigned covariates are standardized so that they range from 0 to 1. This measure is set to range from 0 to 1, with the maximum value for any predictor set to 1. The use of p values as a beginning was designed to allow some comparability of categorical and scale or "continuous" predictors. The base 10 logarithmic transformation was chosen for utility in spreading out the p values. The negative is then required to make resulting raw values positive, though if it were neglected it would cancel out in the numerator and denominator of the ratios used to calculate the final values.

Step 4 – Modification: The modification procedure allows for the defining of an outlier treatment. The researcher specified a value for the fraction of noise (5 percent). A leaf (pre-cluster) is considered as a potential outlier cluster if the number of cases is less than the defined fraction of the maximum cluster size. Outliers are ignored in the second step (Chiu *et al.*, 2001; Bacher *et al.*, 2004).

3.7. Ethics

At the University of Stellenbosch, ethical considerations are guided by the Policy for Responsible Research Conduct at Stellenbosch University (SU) (Stellenbosch University, 2013). The main guiding values Policy for Research and Conduct at SU are:

a) Transparency;
b) Mutual respect;
c) Scholarship (scientific and academic professionalism); and
d) Responsibility.

Chapter 4: Results of the study

4.1 Introduction

The majority of road crashes are caused by a combination of interrelated factors. Although human related factors are a significant contributor to road crashes, direct control and prediction of human factors is difficult. For that reason, human factors can be indirectly controlled and predicted through investigations of roadway and environmental factors, particularly roadway traffic characteristics and geometric design. For that reason, a mixed analysis method was used in the study to understand the types of crashes on the national rural roads, examine their relationship with interrelated factors and attempt to mitigate their occurrence through developing crash predictive models factoring in road characteristics.

This chapter presents the results of the study done using mix analysis methods discussed in Chapter 3.

The Chapter is outlined below:

1. Road crash univariate and bivariate analyses
2. Road crash geospatial analyses
3. Road crash prediction model results
4. Driver characteristics and risk factors – roadway condition analyses models

4.2 Road crash univariate and bivariate analyses

This section provides a univariate analysis of the crash data used in the study. It is important to contextualise the fatal and serious injury (FSI) crash data collected and used in the study as it is a key aspect in carrying out the study through determining and analysing driver risk factors and behaviour and subsequently the development of the crash prediction models (CPMs) and models investigating the combinational effect of national rural road conditions on driver risk factors.

4.2.1. Road crash frequency analysis

4.2.1.1. Temporal variation of road crashes

The crash dataset analysed in the study comprises 3 190 road crashes involving fatal and/ or serious injuries collected by the Namibian National Road Safety Council (NRSC), Namibian Motor Vehicle Accident Fund (MVA) and Namibian Police Authorities over a period between 2012 and 2016. An analysis of the crash data depicted in Figure 4.1, found that 493 (15 percent) of the crashes were recorded in 2012, 691 (22 percent) of the crashes reported in 2013, 701 (22 percent) FSI crashes in occurred in both 2014 and 2015, while 604 (19 percent) crashes were reported in 2016. The annual frequency determined from the crash data for the study period indicate that an average of 638 fatal and serious injury crashes were recorded on the national rural road network in Namibia between 2012 and 2016. It is evident from the temporal analysis that the frequency of FSI crashes does not vary significantly over the study period. Using the Namibian population to determine the exposure of road users, the study found that 21.3 FSI crashes per 100 000 population on national rural roads in the study area.

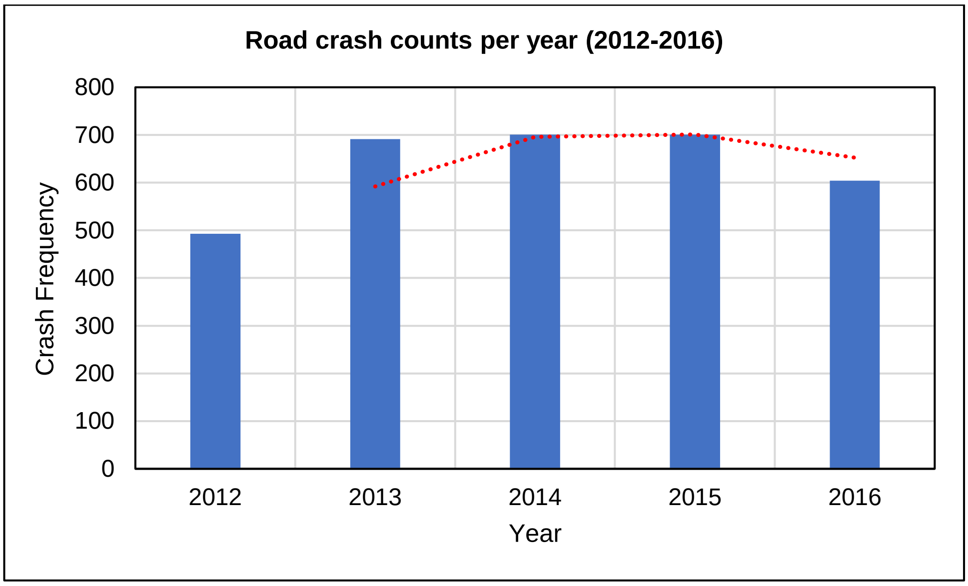


Figure 4.1 Frequency of road crashes per year

The analysis of the road crash counts by month of the year indicates that the drivers are at the highest risk of being involved in fatal and serious injuries during the peak holiday months. The highest number of crash incidence is observed in December as illustrated by Figure 4.2. December is normally the festive period and traffic on the national rural roads tends to peak during this period. Another peak is observed during May and August. These months are filled with public holidays in Namibia. Because of this, the traffic peaks on rural roads as holiday makers begin to travel. It is also observed in Figure 4.2 that the lowest road crash incidences were reported in February and October over the calendar years.

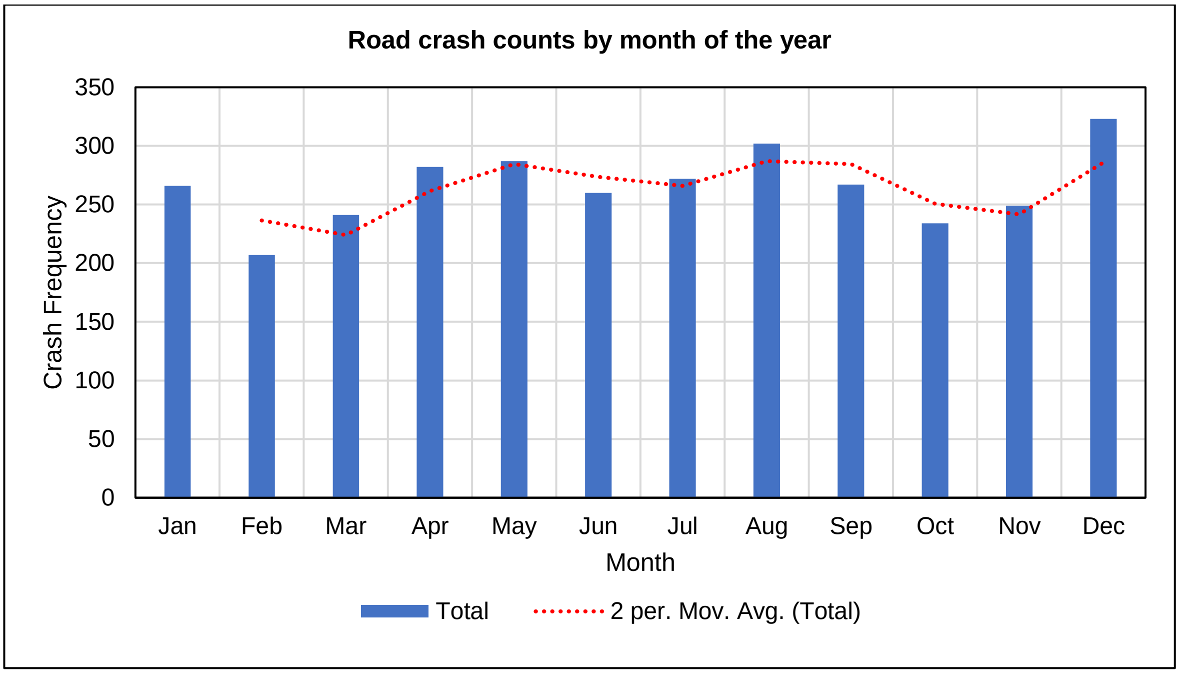


Figure 4.2 Frequency of road crashes per month from 2012 to 2016

4.2.1.2. Road crash frequency by yearly quarters

The 5-year road crash weekly incidence dataset on national rural roads was divided up into four quarters of the calendar year to assess seasonal trends in the crash frequencies as shown in Figure 4.3. Each year comprises four quarters, and each calendar year quarter consists of 13 weeks, with each quarter in the 5-year crash sample size comprising 65 weekly crash count variables.

Figure 4.3 illustrates the road crash count weekly fluctuations of five plots over the four quarters of a calendar year for each year included in the analysis. It can be observed that higher weekly road crash counts are more pronounced in the third quarter of the calendar year. More marked weekly differences between the highest and lowest crash frequencies are observed in the second and third quarter of the calendar year. A further detailed year by year observation indicates noticeably high weekly crash count fluctuations for the year 2014 in the last quarter of the calendar year.

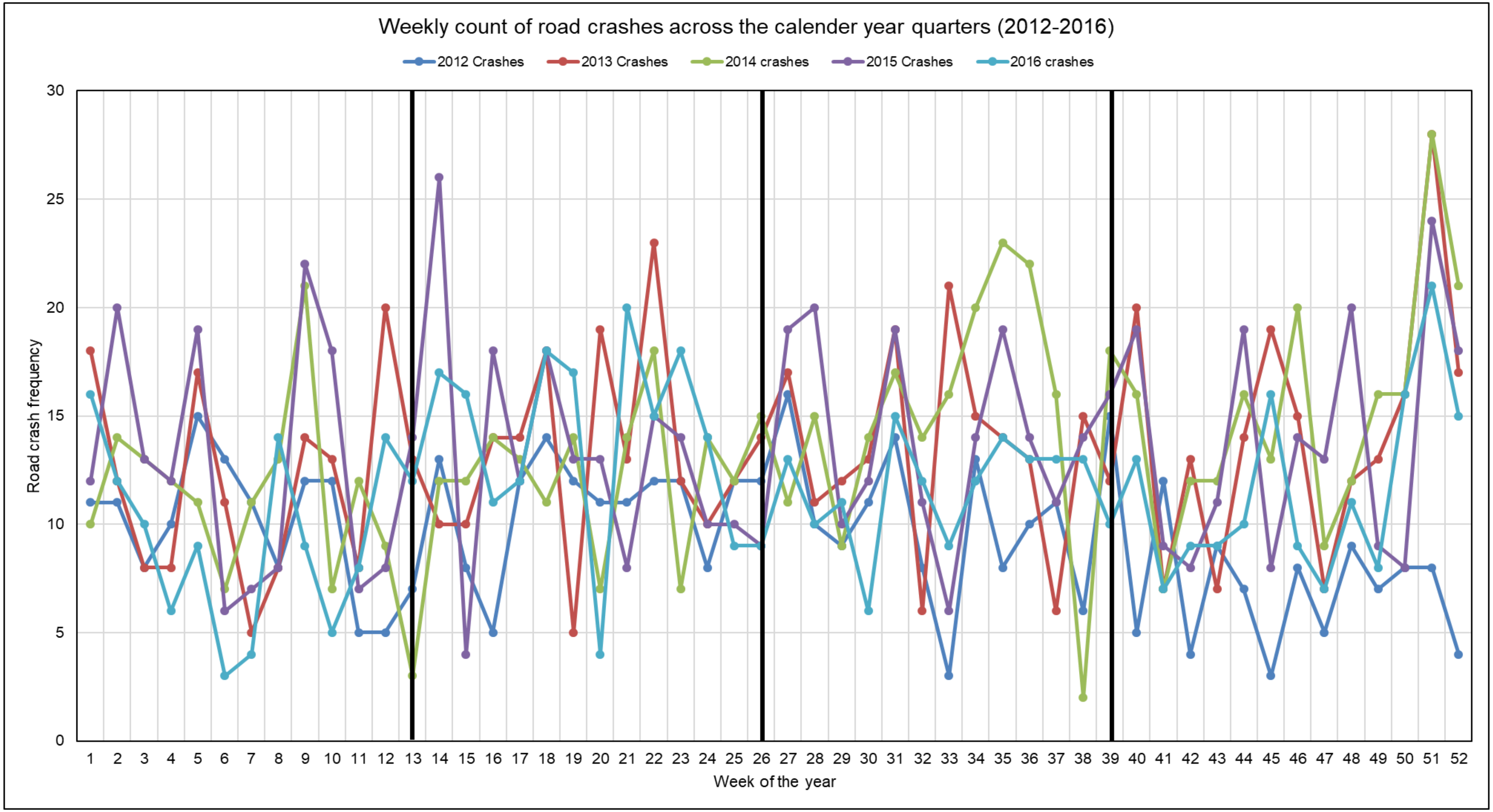


Figure 4.3 Weekly road crash count across calendar year quarters

The quarterly frequencies and means of weekly road crash counts were generated for the 5-year crash dataset after identifying a disproportionate distribution of road crash incidences over the calendar year. Table 4.1 and Figure 4.4 present a descriptive analysis of the quarterly road crash dataset.

Table 4.1 Descriptive statistics of yearly quarterly road crash counts

Dependant variable: Weekly count of road crash incidences			
Quarter of Calendar Year	**Mean**	**Std. Dev**	**N**
Quarter 1	10.98	4.414	65
Quarter 2	12.75	4.187	65
Quarter 3	12.94	4.419	65
Quarter 4	12.40	5.656	65
Total	12.27	4.669	260

From Table 4.1, it is evident that the highest mean weekly crash count is observed over the third quarter of the calendar year. This peak is recorded over the winter season months in Namibia. From Figure 4.4, on the temporal variation of the estimated marginal mean of weekly crash counts, slight temporal variations are observed over the quarters of the calendar year, with a peak mean value in the third quarter and a minimum mean value in the first quarter of the year.

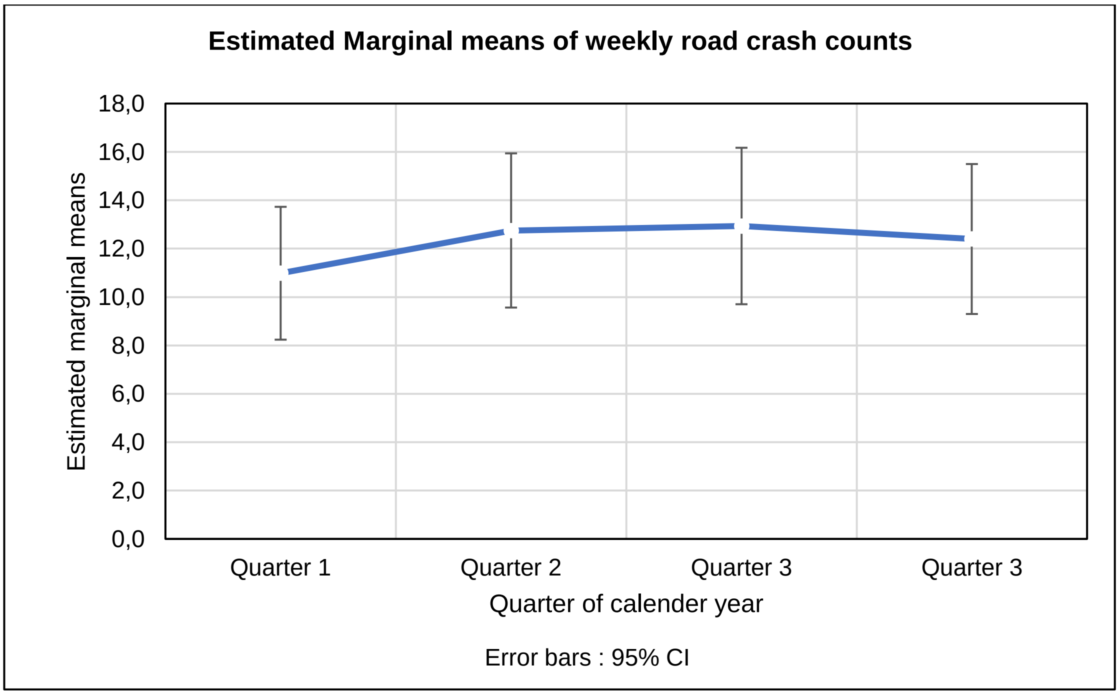


Figure 4.4 Estimated marginal means of weekly (quarterly) road crash counts across yearly quarters

To determine whether statistically significant mean differences exist among the mean values of the quarterly weekly crash counts, the individual mean differences were tested using the Analysis of Variance (ANOVA) test. The two underlying assumptions of ANOVA were tested and the results determined the type of Post-hoc test applied in the analysis. These ANOVA assumptions are: (1) normality of distributions; and (2) homogeneity of variance. The Post-hoc test provided detailed information on where the statistically significant means exist between the means. The results of the ANOVA test scores and Levene's test for homogeneity of variance are presented in Table 4.2 and Table 4.3 respectively.

Table 4.2 Results of ANOVA Test Scores on Weekly (quarterly)) road crash counts

Tests of Effects Between-Subjects					
Dependent Variable: Weekly road crash counts (quarterly)					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.(p)**
Corrected Model	152.754[a]	3	50.918	2.300	0.048
Intercept	39138.846	1	39138.846	1768.238	0.000
Group	152.754	3	50.918	2.300	0.048
Error	5666.400	256	22.134		
Total	44958.000	260			
Corrected Total	5819.154	259			

a. R Squared = 0.026 (Adjusted R Squared = 0.015)

The ANOVA test scores presented in Table 4.2 indicates that the mean of the sums of squares (variance estimate) between the calendar year quarters of 50.918. (i.e. the mean difference between the quarters of the calendar year). The variance estimate within the quarters of the calendar year is indicated as 22.134. The F value (F ratio) of 2.300 for this test is then calculated by dividing the variance estimate between groups by the variance estimate within groups. The F ratio indicates that the variance estimate between the groups (quarters) is about 2 times greater than the amount of error variance (within subjects' variance) that has been accounted for. The results presented in Table 4.2 also indicate that the test is significant at 5 percent level ($p=0.048< 0.05$) which implies that the null hypothesis (the assumption that the means between the groups for the dataset are equal) is rejected and the assumption of homogeneity of variance is invalid. The R-squared gives an indication of how much variance in the dependant variable is accounted for by the covariates. An adjusted R-squared value if 0.015 indicates that 1.5 percent of crash incidence variance is explained by the predictors (calendar year quarters) at 95 percent confidence level.

In the same way, the results of Levene's test for homogeneity of variance for weekly counts over calendar year quarters (presented in Table 4.3) indicate that the test is significant at 95 percent confidence level ($p_{mean}=0.019<0.05$). The Levene's test results reject the null hypothesis which implies that the variance is equal across the calendar year quarters. Levene's test is crucial in determining the appropriate Post-hoc test as explained in Section 3.7.1.2 and Figure 3.12. The choice of Post-hoc test is contingent on the assumptions of equal variances and equal group sample sizes. The Games-Howell Post-hoc test was determined as the appropriate technique to assess the mean differences as a result of unequal variance and group sample sizes.

Table 4.3 Results of Levene's Test for Homogeneity of variance for weekly (quarterly) road crash counts

Levene's Test of Equality of Error Variances[a,b]					
		Levene Statistic	df1	df2	Sig.
Count	Based on Mean	3.377	3	256	0.019
	Based on Median	3.140	3	256	0.026
	Based on Median and with adjusted df	3.140	3	249.199	0.026
	Based on trimmed mean	3.193	3	256	0.024

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.[a,b]

a. Dependent variable: Quarterly weekly crash count

b. Design: Intercept + Group

The results from the Games-Howell Post-hoc test applied are presented in Table 4.4. The Games-Howell Post-hoc procedure compared the means of all calendar year quarters (groups) with each other. The values "Sig (p)" values shown in red are statistically significant ($p<0.05$) at 95 percent confidence interval.

The Games-Howell test results found the mean values of the quarterly weekly crash counts to be consistent over the 2nd and 4th quarters of the calendar year. The mean differences of these quarters (2nd and 4th) were not significant ($p<0.05$) at 95 percent confidence level. Statistically significant mean differences were identified between the mean values of the 1st and 3rd quarters of the calendar year, which are also visually evident in Figure 4.4.

Table 4.4 Results of Games-Howell Post Hoc Test on weekly (quarterly) road crash counts

Post Hoc test: Games Howell						
Multiple Comparisons						
Dependant variable: Yearly Weekly (Quarters) count of road crashes						
(I) Quarter	(J) Quarter	Mean Difference (I-J)	Std. Error	Sig. (p)	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-1.77	0.755	0.093	-3.73	0.20
	3	-1.95*	0.775	**0.041**	-3.97	0.06
	4	-1.42	0.890	0.388	-3.73	0.90
2	1	1.77	0.755	0.093	-0.20	3.73
	3	-0.18	0.755	0.995	-2.15	1.78
	4	0.35	0.873	0.977	-1.92	2.63
3	1	1.95*	0.775	**0.041**	-0.06	3.97
	2	0.18	0.755	0.995	-1.78	2.15
	4	0.54	0.890	0.930	-1.78	2.86
4	1	1.42	0.890	0.388	-0.90	3.73
	2	-0.35	0.873	0.977	-2.63	1.92
	3	-0.54	0.890	0.930	-2.86	1.78

Based on observed means.

The error term is Mean Square (Error) = 22.134.

*. The mean difference is significant at the 0.05 level.

4.2.1.3. Road crash frequency by week of the month

The amount of traffic on the national rural; roads is affected by the week of the month during which road users are paid. This ultimately impacts the level of safety on the roads due to higher exposure levels. The univariate analysis included three categorical covariates according to the week of the month to investigate the trend of road crash incidents according to the financial state of the drivers. These covariates are termed: (1) Pay week; (2) 2^{nd} week after pay week; and (3) Other weeks. The pay week represents the week that contains the first date of the month (e.g. 1^{st} June). The second week after pay week represent the week following the pay week and other week denotes the remaining week of the month. Table 4.5 presents the descriptive statistics for the weekly fatal and serious injury 5-year road crash counts.

Table 4.5 Descriptive statistics of weekly road crash counts

Dependant variable: Weekly road crash counts			
Weekly financial status	**Mean**	**Std. Deviation**	**N**
Pay week	11.78	4.244	65
2nd week after pay week	12.60	4.620	65
Other weeks	12.35	4.990	130
Total	12.27	4.731	260

The mean differences across the three covariates are illustrated in Figure 4.5. It is visually apparent from Figure 4.5 that the weekly road crash counts peak over the second week after pay week and the mean value differences are lowest during the pay week.

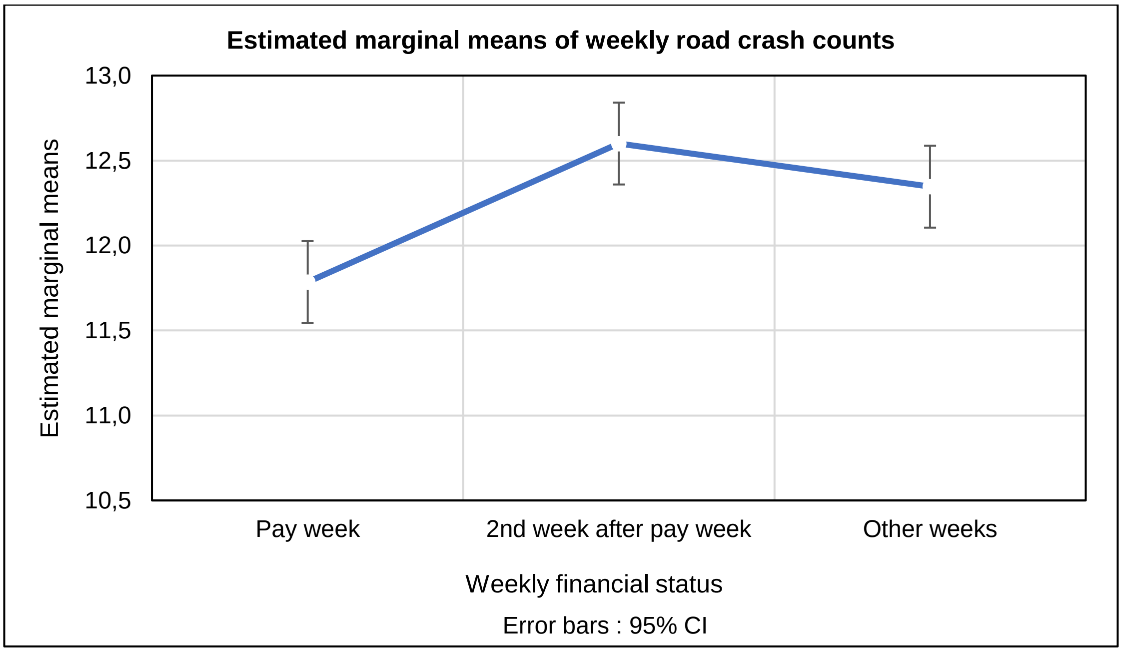


Figure 4.5 Estimated marginal means of weekly road crash counts

The study tested the mean differences among the three covariates using the ANOVA test and Levene's test of homogeneity of variance. The ANOVA test scores are presented in Table 4.6. The ANOVA test scores reveal that the individual mean differences between the weeks of the month are not statistically significant (p=0.599> 0.05) at 95 percent confidence level. For this reason, the null hypothesis that equal variance exists across the study groups is accepted.

Table 4.6 Results of ANOVA Test Scores on weekly road crash counts

Tests of Effects Between-Subjects					
Dependent Variable: Weekly road crash counts					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Corrected Model	23.146[a]	2	11.573	0.513	0.599
Intercept	35077.885	1	35077.885	1555.384	0.000
Week of the month	23.146	2	11.573	0.513	0.599
Error	5796.008	257	22.553		
Total	44958.000	260			
Corrected Total	5819.154	259			

a. R Squared = 0.004 (Adjusted R Squared = -0.004)

The Levene's test for homogeneity of variance results are presented in Table 4.7. Levene's test also demonstrates that the no statistically significant (p=0.256>0.05) difference exists between the means of the weeks of the month at 95 percent confidence interval. Therefore, the test results suggest that the null hypothesis (equal variance across the test groups) is valid.

Table 4.7 Results of Levene's Test for Homogeneity of variance for weekly road crash counts

Levene's Test of Equality of Error Variances[a,b]					
		Levene Statistic	**df1**	**df2**	**Sig.**
Count	Based on Mean	1.369	2	257	0.256
	Based on Median	1.263	2	257	0.284
	Based on Median and with adjusted df	1.263	2	255.078	0.284
	Based on trimmed mean	1.262	2	257	0.285

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.[a,b]

a. Dependent variable: Weekly road crash counts

b. Design: Intercept + Weekly road crash counts

The procedure defined in Figure 3.12 identified the Bonferroni Post-hoc test as the most suitable test to assess the individual mean difference across the weeks of the month (test group) with equal variance and unequal sample sizes. The results of the Bonferroni Post-hoc test are presented in Table 4.8. The Bonferroni test indicates that a statistically significant ($p<0.05$) difference exists between the means of the weekly crash count of the "pay week" and the "2nd week after the pay week" at 95 percent confidence interval. This suggests that the fatal and serious injury crashes occurred more frequently over the pay week s compared with the second week after the pay week. This conclusion is illustrated by the significant mean differences between the pay weeks and second week after the pay weeks covariates shown in Figure 4.5.

Table 4.8 Results of Bonferroni Post Hoc Test on weekly road crash counts

Post Hoc test: Bonferroni						
Multiple Comparisons						
Dependant variable: Weekly road crash counts						
(I) Week of the month	(J) Week of the month	Mean Difference (I-J)	Std. Error	Sig.(p)	95% Confidence Interval	
					Lower Bound	Upper Bound
Pay week	2nd week after pay week	-0.82*	0.833	**0.002**	-2.82	1.19
	Other weeks	-0.56	0.721	1.000	-2.30	1.18
2nd week after pay week	Pay week	0.82*	0.833	**0.002**	-1.19	2.82
	Other weeks	0.25	0.721	1.000	-1.48	1.99
Other weeks	Pay week	0.56	0.721	1.000	-1.18	2.30
	2nd week after pay week	-0.25	0.721	1.000	-1.99	1.48

Based on observed means.

The error term is Mean Square (Error) = 22.553.

*. The mean difference is significant at the 0.05 level.

4.2.1.4. Road crash frequency by day of the week

The distribution of the 5-year daily road crash counts over a week on rural roads is presented in Figure 4.6. The univariate analysis identified a peak occurrence of road crashes over weekends, on Friday (548 road crashes), Saturday (663 road crashes) and Sunday 551 road crashes). The weekend road crashes represent a majority (55 percent) of all crashes over the week. The lowest occurrence of road crashes was observed during the week, on Tuesday (313 road crashes) and Wednesday (315 road crashes).

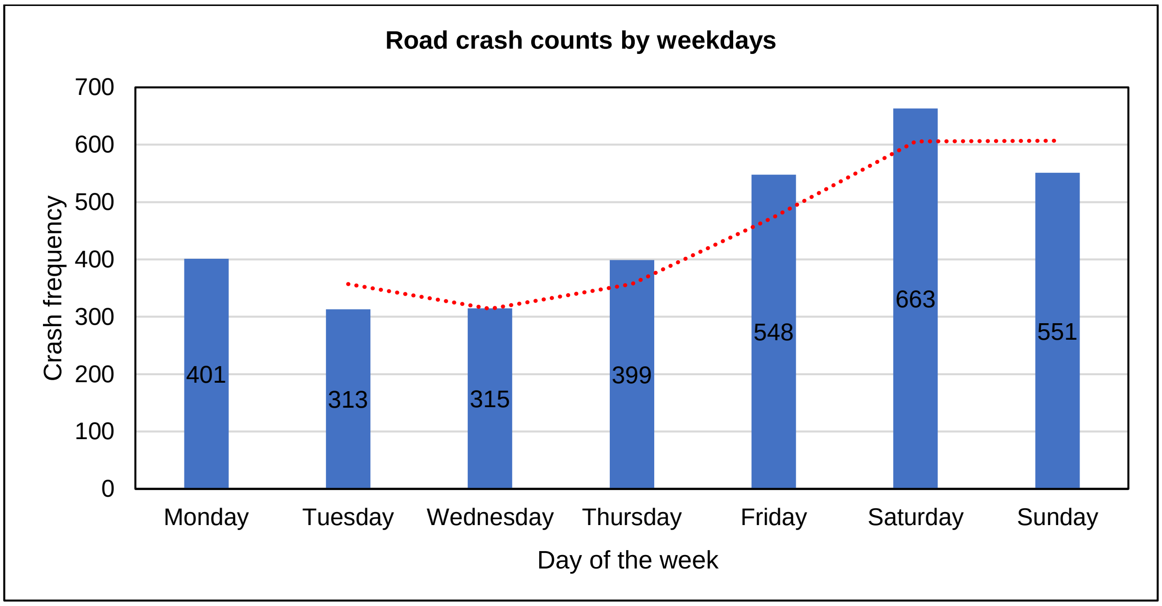


Figure 4.6 Road crash frequency by weekdays

The analysis identified a disproportionate distribution of road crashes over the week. For this reason, the daily frequencies and mean daily frequencies for the 5-year national road crash dataset were computed. The results for descriptive analysis are presented in Table 4.9 and illustrated in Figure 4.7.

Table 4.9 Descriptive statistics of weekdays road crash counts

Dependant variable: Weekdays road crash counts			
Weekday	**N**	**Mean**	**Std. Dev**
Mon	381	76.2	15.401
Tue	295	59	9.028
Wed	292	58.4	8.989
Thu	374	74.8	12.296
Fri	518	103.6	11.803
Sat	640	128	31.757
Sun	535	107	16.956
Holiday	155	31	5.612
Total	3190	638	90.730

Holiday dates were included in the analysis for more insights on crash risk on these certain days. 31 national public holidays for each year (2012-2016 period). Data on holidays were collected from the official website of the Namibian Government (Government of Namibia, 2020) and corroborated using the www.timeanddate.com website.

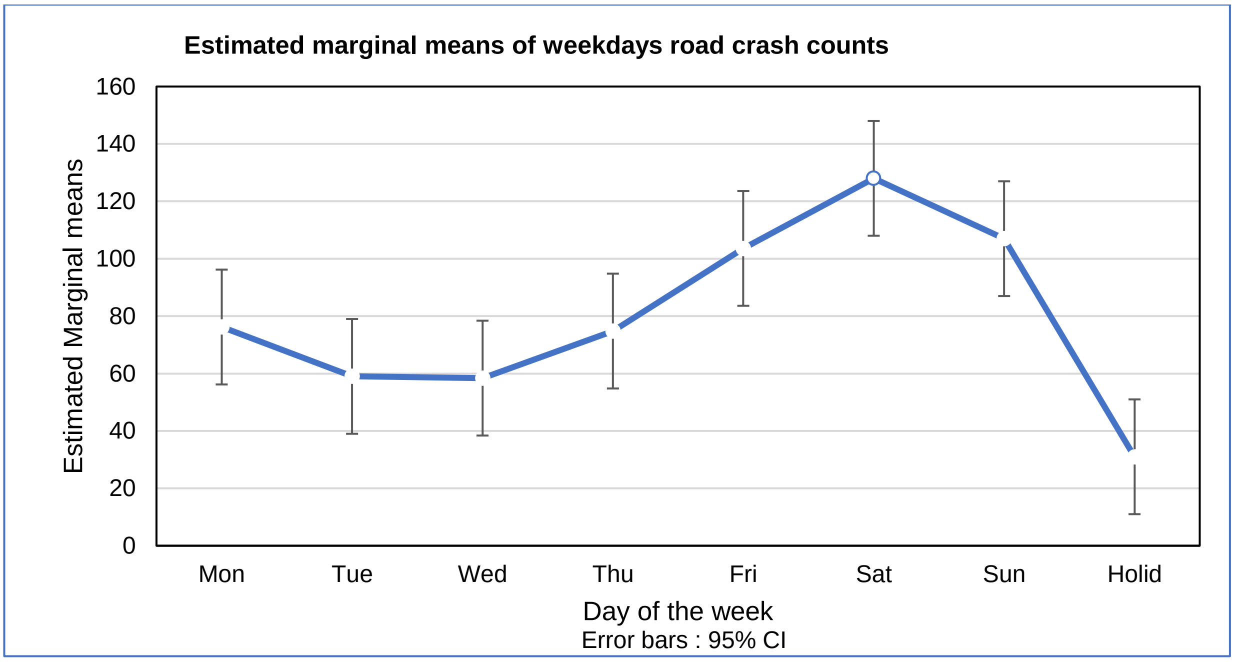


Figure 4.7 Estimated marginal means of weekday road crash counts

The ANOVA test was used to ascertain whether a statistically significant difference exists between the individual means of the weekday's road crash incidences. The results of the ANOVA test are presented in Table 4.10. The ANOVA test scores indicate that the test identified a statistically significant (p=0.000<0.05) difference between the weekday road crash counts (test group) at 95 percent confidence interval. As a result, the null hypothesis that the equal variance exists across the weekday is termed invalid.

Table 4.10 Results of ANOVA Test Scores on weekdays road crash counts

Tests of Effects Between-Subjects					
Dependent Variable: Weekday count of road crashes					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig. (p)**
Corrected Model	34697.500[a]	7	4956.786	19.655	0.000
Intercept	254402.500	1	254402.500	1008.783	0.000
Weekday	34697.500	7	4956.786	19.655	0.000
Error	8070.000	32	252.187		
Total	297170.000	40			
Corrected Total	42767.500	39			

a. R Squared = 0.811 (Adjusted R Squared = 0.770)

Similar to the ANOVA test, Levene's test was used to assess the homogeneity of variance across the test group (weekday road crash count). In addition, Levene's test results determined the choice of Post-hoc test applied in the study. The results of Levene's test are given in Table 4.11. The results indicate that the test is statistically significant (p=0.034<0.05) at 95 percent confidence level, which implies that the null hypothesis (equal variance across the test group) is invalid.

Table 4.11 Results of Levene's Test for Homogeneity of variance for weekdays road crash count

Levene's Test of Equality of Error Variances[a,b]					
		Levene Statistic	**df1**	**df2**	**Sig. (p)**
Count	Based on Mean	2.533	7	32	0.034
	Based on Median	1.334	7	32	0.267
	Based on Median and with adjusted df	1.334	7	16.777	0.295
	Based on trimmed mean	2.433	7	32	0.041

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.[a,b]

a. Dependent variable: Road crash count

b. Design: Intercept + Weekday

The approach illustrated by Figure 3.12 was followed in identifying the appropriate Post-hoc test after assessing the results of Levene's test. For this analysis, the test identified unequal variances between the study groups and the sample sizes (N) differ across the week days (test groups) as shown in Table 4.9. As a result, the Games-Howell Post-hoc test was suitable to assess the mean differences for unequal variance and sample size. Table 4.12 presents the results of the Games-Howell Post-hoc procedure, with statistically significant (p<0.05) probability values marked in red.

The Games-Howell results indicate that individual mean differences on Friday, Saturday and Sunday (weekend) are higher compared with other week days (probability values "Sig. (p)" are significant (p<0.05) at 95 percent confidence interval). Also Notable in Table 4.9, road crashes occurred more frequently over these days (weekends) compared to other days of the week. In the same way, the mean differences on holidays are higher compared with Friday, Saturday and Sunday (p<0.05). This suggests that a statistically significant relationship exists between road crash occurrence on national rural roads over holidays during weekends. In contrast, the mean differences were found to be consistent over Mondays, Tuesday, Wednesday and Thursdays (p>0.05).

Table 4.12 Results of Games-Howell Post Hoc Test on weekdays road crash counts

Post Hoc test: Games-Howell						
Multiple Comparisons						
Dependent Variable: Weekdays count of road crashes						
(I) Weekday	(J) Weekday	Mean Difference (I-J)	Std. Error	Sig.(p)	95% Confidence Interval	
					Lower Bound	Upper Bound
Mon	Tue	17.20	7.984	0.469	-16.49	50.89
	Wed	17.80	7.975	0.435	-15.88	51.48
	Thu	1.40	8.814	1.000	-33.94	36.74
	Fri	-27.40*	8.678	0.045	-62.37	7.57
	Sat	-51.80*	15.784	0.015	-121.06	17.46
	Sun	-30.80*	10.244	0.001	-71.44	9.84
	Hol	45.20	7.331	0.169	11.21	79.19
Tue	Mon	-17.20	7.984	0.469	-50.89	16.49
	Wed	0.60	5.697	1.000	-21.95	23.15
	Thu	-15.80	6.822	0.390	-43.45	11.85
	Fri	-44.60*	6.645	0.003	-71.38	-17.82
	Sat	-69.00*	14.765	0.016	-140.08	2.08
	Sun	-48.00*	8.591	0.015	-84.97	-11.03
	Hol	28.00	4.754	0.056	8.17	47.83
Wed	Mon	-17.80	7.975	0.435	-51.48	15.88
	Tue	-0.60	5.697	1.000	-23.15	21.95
	Thu	-16.40	6.812	0.354	-44.03	11.23
	Fri	-45.20*	6.635	0.003	-71.96	-18.44
	Sat	-69.60*	14.760	0.010	-140.69	1.49
	Sun	-48.60*	8.583	0.014	-85.57	-11.63
	Hol	27.40	4.739	0.054	7.65	47.15
Thu	Mon	-1.40	8.814	1.000	-36.74	33.94
	Tue	15.80	6.822	0.390	-11.85	43.45
	Wed	16.40	6.812	0.354	-11.23	44.03
	Fri	-28.80*	7.622	0.005	-58.98	1.38
	Sat	-53.20*	15.230	0.036	-123.06	16.66
	Sun	-32.20*	9.367	0.006	-70.24	5.84
	Hol	43.80	6.045	0.063	16.94	70.66
Fri	Mon	27.40*	8.678	0.045	-7.57	62.37
	Tue	44.60*	6.645	0.003	17.82	71.38
	Wed	45.20*	6.635	0.003	18.44	71.96
	Thu	28.80*	7.622	0.005	-1.38	58.98
	Sat	-24.40	15.151	0.735	-94.41	45.61
	Sun	-3.40	9.239	1.000	-41.16	34.36
	Hol	72.60*	5.845	0.000	46.85	98.35
Sat	Mon	51.80*	15.784	0.015	-17.46	121.06
	Tue	69.00*	14.765	0.016	-2.08	140.08
	Wed	69.60*	14.760	0.010	-1.49	140.69
	Thu	53.20*	15.230	0.036	-16.66	123.06
	Fri	24.40	15.151	0.735	-45.61	94.41
	Sun	21.00	16.100	0.870	-48.25	90.25
	Hol	97.00*	14.422	0.017	24.47	169.53
Sun	Mon	30.80*	10.244	0.001	-9.84	71.44

	Tue	48.00*	8.591	**0.015**	11.03	84.97
	Wed	48.60*	8.583	**0.014**	11.63	85.57
	Thu	32.20*	9.367	**0.006**	-5.84	70.24
	Fri	3.40	9.239	1.000	-34.36	41.16
	Sat	-21.00	16.100	0.870	-90.25	48.25
	Hol	76.00*	7.987	**0.002**	38.37	113.63
Hol	Mon	-45.20	7.331	0.169	-79.19	-11.21
	Tue	-28.00	4.754	0.056	-47.83	-8.17
	Wed	-27.40	4.739	0.054	-47.15	-7.65
	Thu	-43.80	6.045	0.063	-70.66	-16.94
	Fri	-72.60*	5.845	**0.000**	-98.35	-46.85
	Sat	-97.00*	14.422	**0.017**	-169.53	-24.47
	Sun	-76.00*	7.987	**0.002**	-113.63	-38.37

Based on observed means.
The error term is Mean Square (Error) = 252.187.
*. The mean difference is significant at the 0.05 level.

4.2.1.5. Road crash frequency by time of day

The analysis of road crash incidences by time of crash during the 5-year period is presented in Figure 4.8. The analysis indicates that the risk of a driver being involved in a road crash on national rural roads is highest during the peak hours of the day. From Figure 4.8, the highest road crash incidences are observed in the early evening hours, between 15h00 and 19h00, with a marked peak between 17h00 and 18h00. This stretch of high crash frequencies represents approximately 37 percent of all crashes throughout the day. Another peak is observed in the morning hours, starting from 06h00 to 8h00, as well as in the late morning to midday hours, occurring between 11h00 and 12h00. The lowest crash incidences were recorded early morning hours from 02h00 to 05h00.

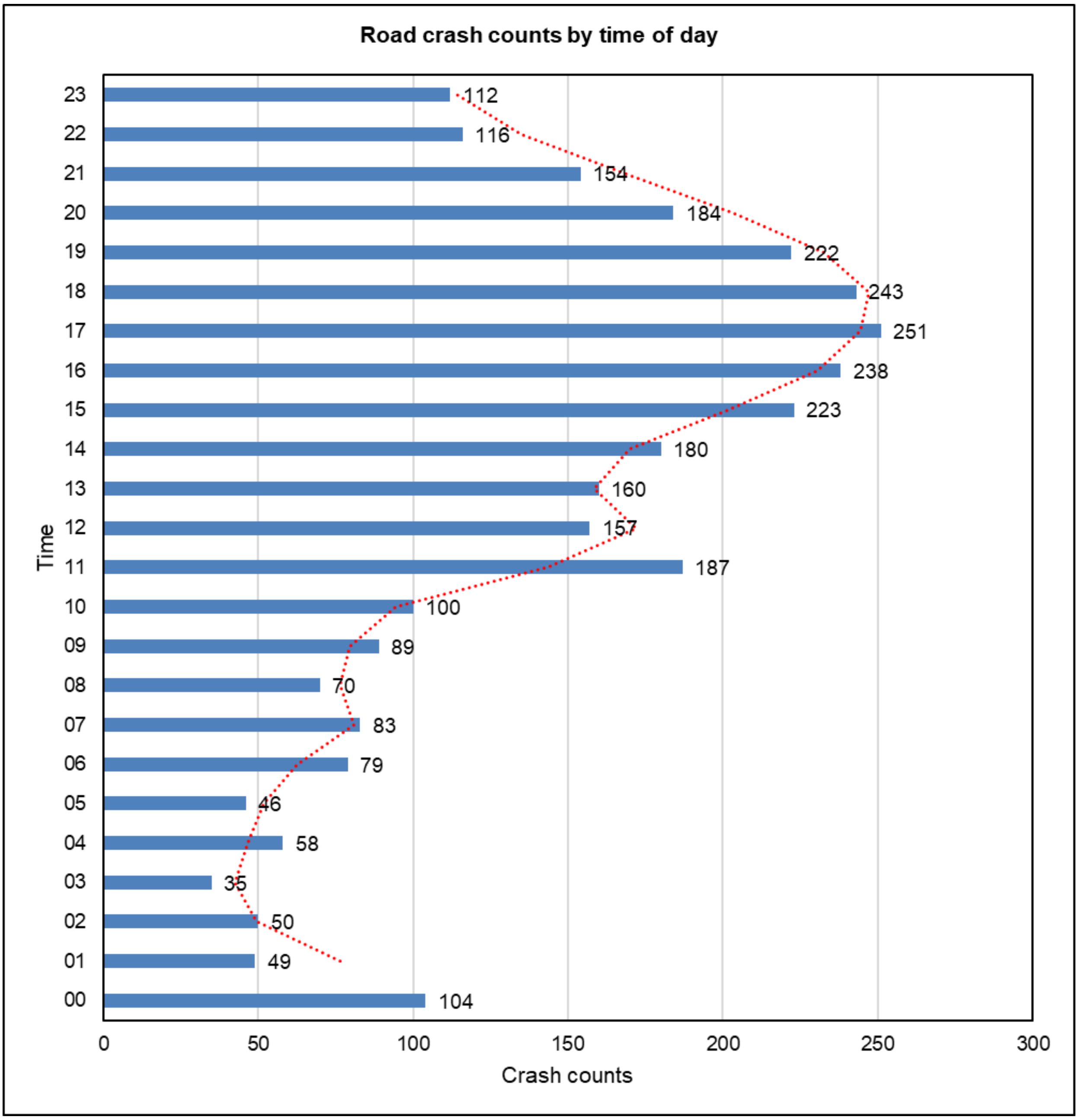


Figure 4.8 Road crash counts by time of day

Figure 4.9 illustrates a male to female driver ratio for road crashes across the time of day. It is observed that the risk for drivers to be involved in a road crash is higher for male drivers than female drivers throughout the whole day. Examining the ratios across the time of day, male drivers are at the highest road crash risk during the early morning hours compared to female drivers, with the crash risk (M: F=24) peaking between 02h00 to 03h00. This high crash risk occurs during the time period (02h00 to 05h00) in which the frequency of road crashes on the national rural roads is lowest. Another notable crash risk peak occurs between 22h00 and 23h00, with male drivers 15 times more likely to be involved in a road crashes than their female counterparts. The crash risk over time periods with higher road crash frequencies is observed to be lower compared with the lower frequency time periods, with a male to female driver ratio averaging six.

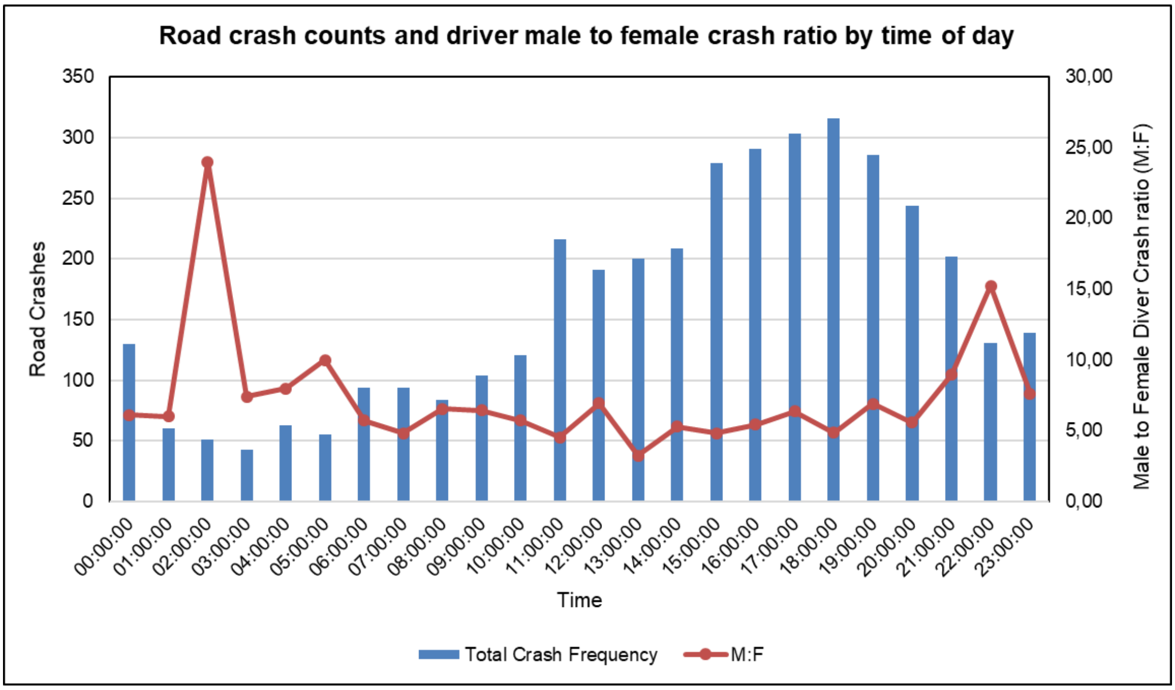


Figure 4.9 Road crash counts by time of day and driver gender ratio

4.2.1.6. Road crash frequency by driver age and gender

The 5-year national rural road crash dataset exhibited an overrepresentation of male drivers as expected (see Figure 4.10). The dataset comprised 3 320 (85 percent) male drivers and 567 (14.52 percent) female drivers involved in rural road crashes. The records indicate that the gender of the road crash casualty was indicated as “unknown” for 19 (0.49 percent) of the drivers involved in a crash. From the crash records, the male driver to female driver crash risk ratio (M: F) was computed as 5.86. Similar to the average male to female driver ratio (M: F= 5 to 6) seen in Figure 4.9. This suggest that male drivers are approximately six times more likely to be involved in a fatal or serious injury crash than female drivers on the national rural roads.

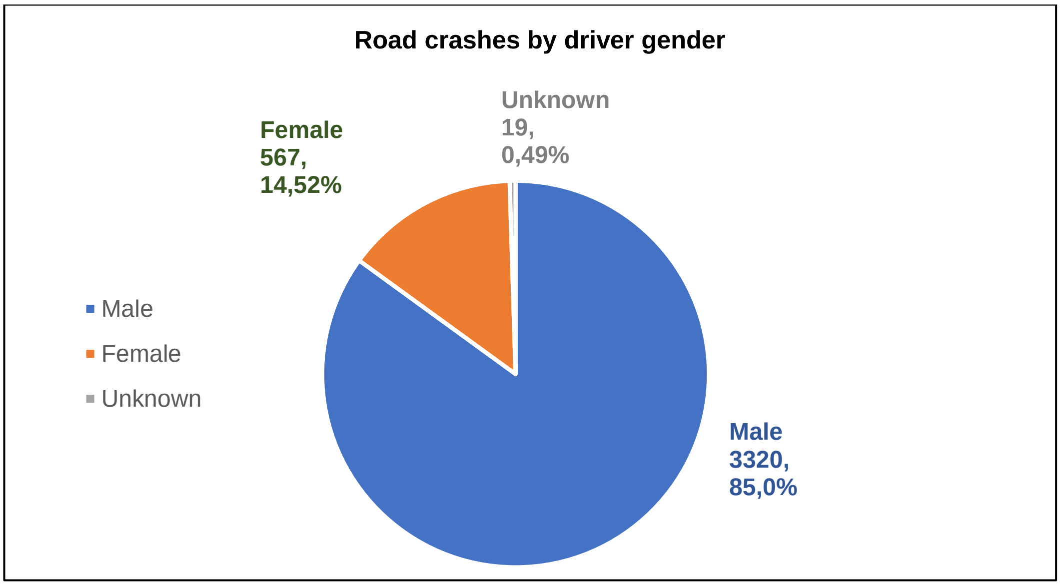


Figure 4.10 road crash counts by driver gender

The crash analysis distributed the driver road crash casualties among various age groups in the crash records as illustrated in Figure 4.11. The analysis was restricted to drivers involved in fatal and serious injury (FSI) only crashes on the national rural roads. The dataset comprised of 3 906 drivers, of which 19 cases were removed due to insufficient information on driver gender and age. For the remaining 3 887 cases, the computed mean driver age was 28.16 years, with a standard deviation (S.D) of 14.33. The highest observed age in the crash analysis was 85 years while the lowest observed age was 11 years.

As illustrated by Figure 4.11, the road crashes are disproportionately distributed across the various driver age groups. The highest frequency of FSI road crashes is observed in the driver age group of 31 to 35 years, closely followed by the driver age group of 26 to 30 years. From the driver age group of 21- 25 years, road crash frequencies rise notable for both genders. This can be potentially attributed to the high levels of exposure these drivers experience around that age. As expected, a considerable reduction in road crash frequencies is observed from driver age group of 46 to 50 years and older, due to lower risk exposure for these drivers.

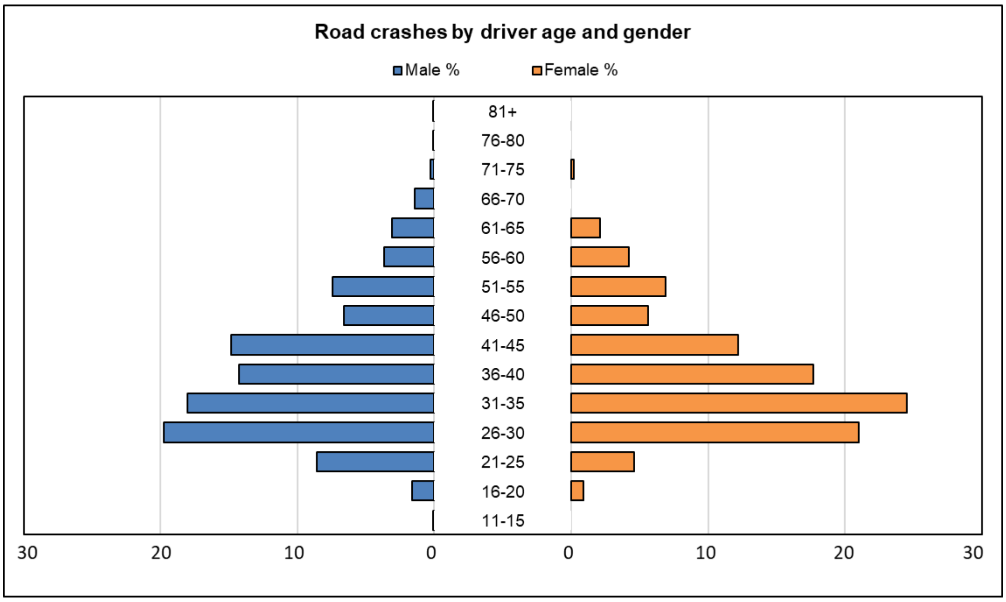


Figure 4.11 Road crash counts by driver age and gender

The disproportionate distribution of road crashes among the driver genders and across the various driver age groups is also evident in Table 4.13. In addition, the results in Table 4.13 and Figure 4.12 indicate the male to female driver ratio for road crash frequencies across the various age groups.

Table 4.13 road crash counts by gender and age

Age	Male	Female	Total Crash Frequency	M: F
11-15	1	0	1	-
16-20	54	5	59	10.80
21-25	286	26	312	11.00
26-30	656	119	775	5.51
31-35	600	139	739	4.32
36-40	475	100	575	4.75
41-45	494	69	563	7.16
46-50	219	32	251	6.84
51-55	248	39	287	6.36
56-60	122	24	146	5.08
61-65	102	12	114	8.50
66-70	48	0	48	-
71-75	9	1	10	9.00
76-80	4	0	4	-
81-85	3	0	3	-
86-90	0	0	0	-

Across all age groups, the crash analysis indicates that male drivers are at a much higher risk compared to female drivers on rural roads. Examining the top five age groups with the highest male to female driver crash risk ratio, male drivers are at the highest crash risk in the young adults (21-25 years) and teenager (16-20 years) age groups, with male drivers more than ten times likely to be involved in a road crash than females (M: F= 11.00 and M: F= 10.80 ratios respectively. Interestingly, male drivers are also at a higher risk in the 71-75 age group (M: F=9.00) and 61-65 (M: F=8.50) age group. However, these age groups recorded the lowest crash frequencies. Another age group that recorded higher male driver crash risk ratio is the 41-45 age group (M: F= 7.16). Notable observation from Figure 4.12, the male to female driver ratios were lower in the age groups with the highest road crash frequencies. This is in line with the expected higher crash risk exposure for both genders.

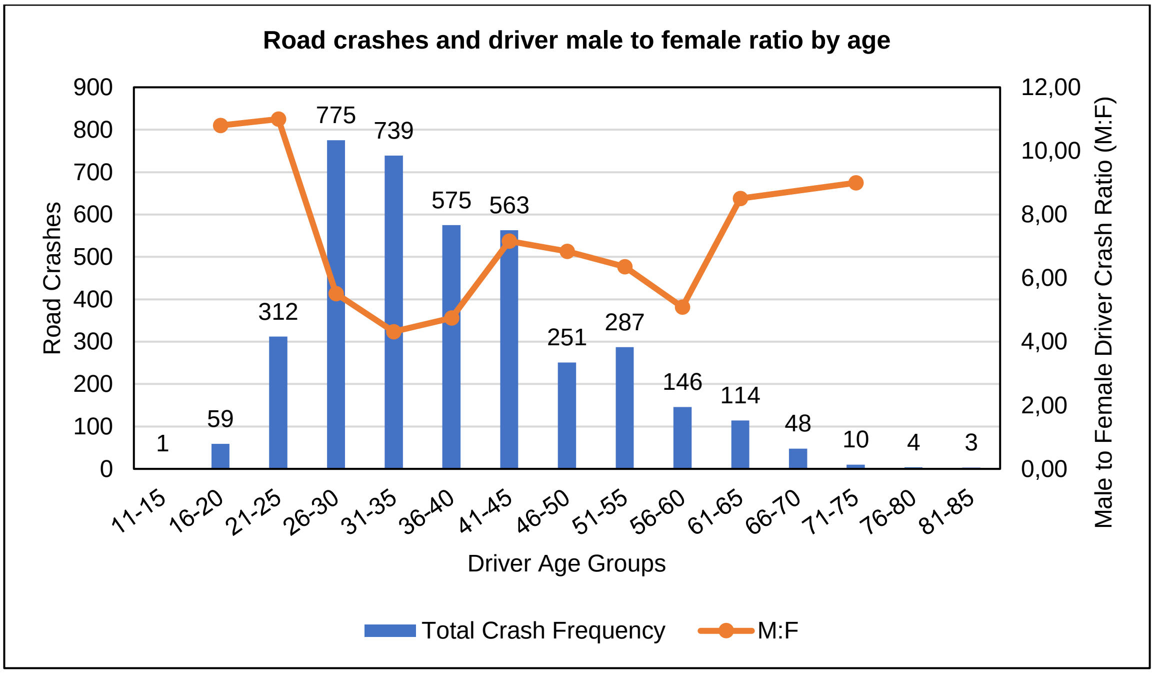


Figure 4.12 Road crash counts by gender (M: F) and age

4.2.2. Road crash analysis by fatal and serious injury (FSI) severities

This section provides further insights into the distribution of fatal and serious only injuries of all car occupants across the time of day, day of the week and month of the year over a 5-year period. The dataset comprises 6 712 cases, of which 4 644 (69 percent) are male and 2 068 (31 percent) are female road user casualties. As expected, the crash dataset comprised an overrepresentation of male casualties with an injury crash risk ratio (M: F=2.25) more than double that of female road users.

4.2.2.1. FSi occupants by time and gender

The distribution of road users fatal and serious injuries (FSI) only casualties by crash occurrence time and the corresponding male to female casualty ratios are presented in Figure 4.13 and Figure 4.14 respectively. It is evident from the analysis that fatal and/ or serious road crashes are more prevalent among male road users than among female road users.

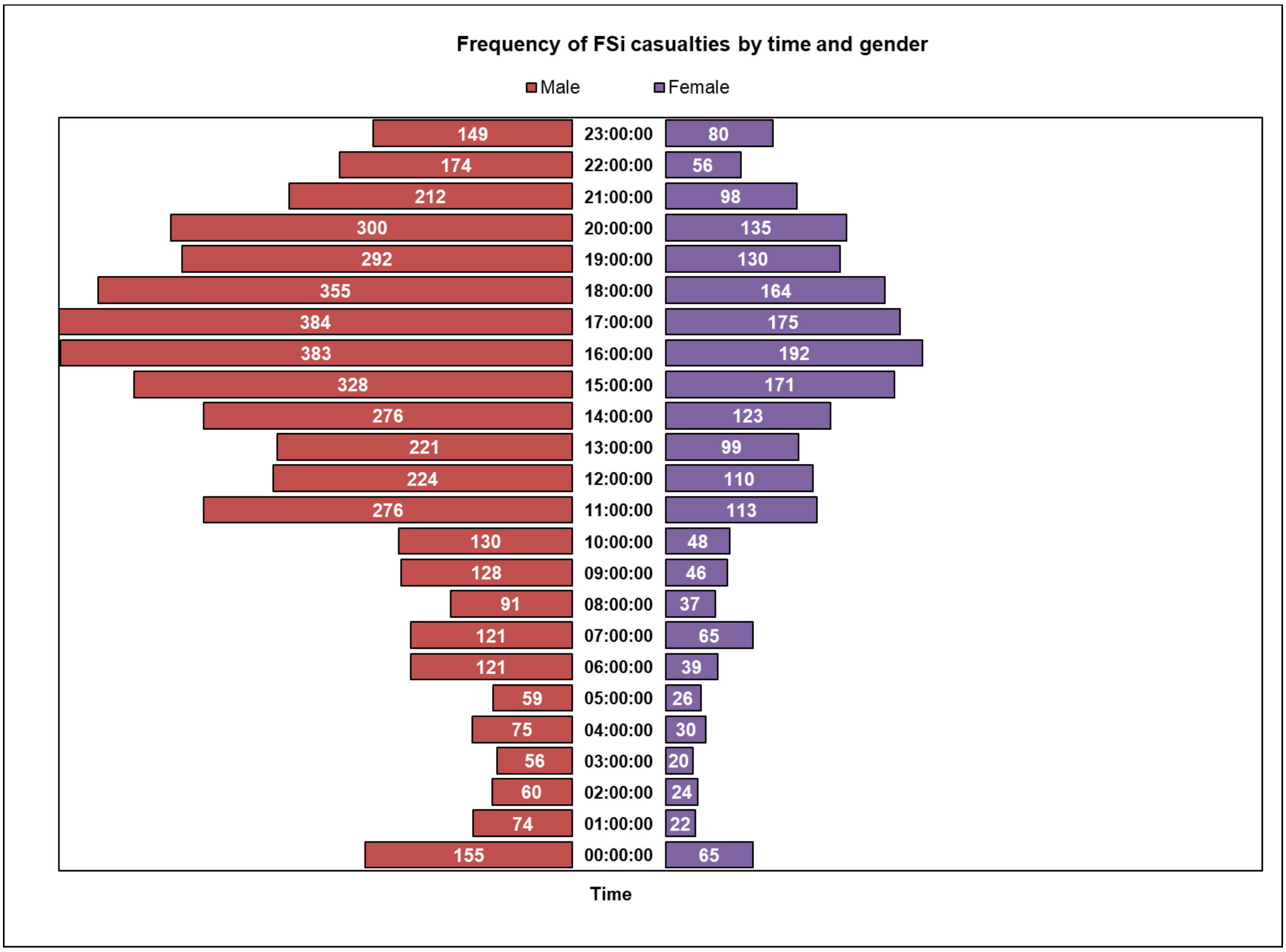


Figure 4.13 Distribution of FSI casualties by time and gender

It is observed from the crash analysis that FSI frequencies peak in the late afternoon to early evening, with higher injury frequencies stretching from 15h00 to 18h00. The FSI injury frequencies during this stretched peak represent approximately 32 percent of all injuries recorded across the day. The highest injury peak during this time period is notable from 16h00 to 17h00 (579 road user FSI casualties).

Examining the male to female injury ratios illustrated in Figure 4.14, male road users are at a highest FSI risk (M: F = 3.36) during the early morning hours (01h00 to 02h00) despite the lower FSI casualties recorded then. Male road users are also at a higher risk of sustaining fatal and/ or serious injuries (M: $F_{morning}$= 3.10 and M: $F_{evening}$= 3.11) in the morning hours (06h00 to 07h00) and late evening hours (22h00 to 23h00) respectively. The lowest gender injury ratios occurred during the high injury casualty time periods, which were observed in the late afternoon to early evening. It is evident that all the ratios are above one. As a result, it can be concluded that male road users are generally at a higher FSI risk than female road users across the day.

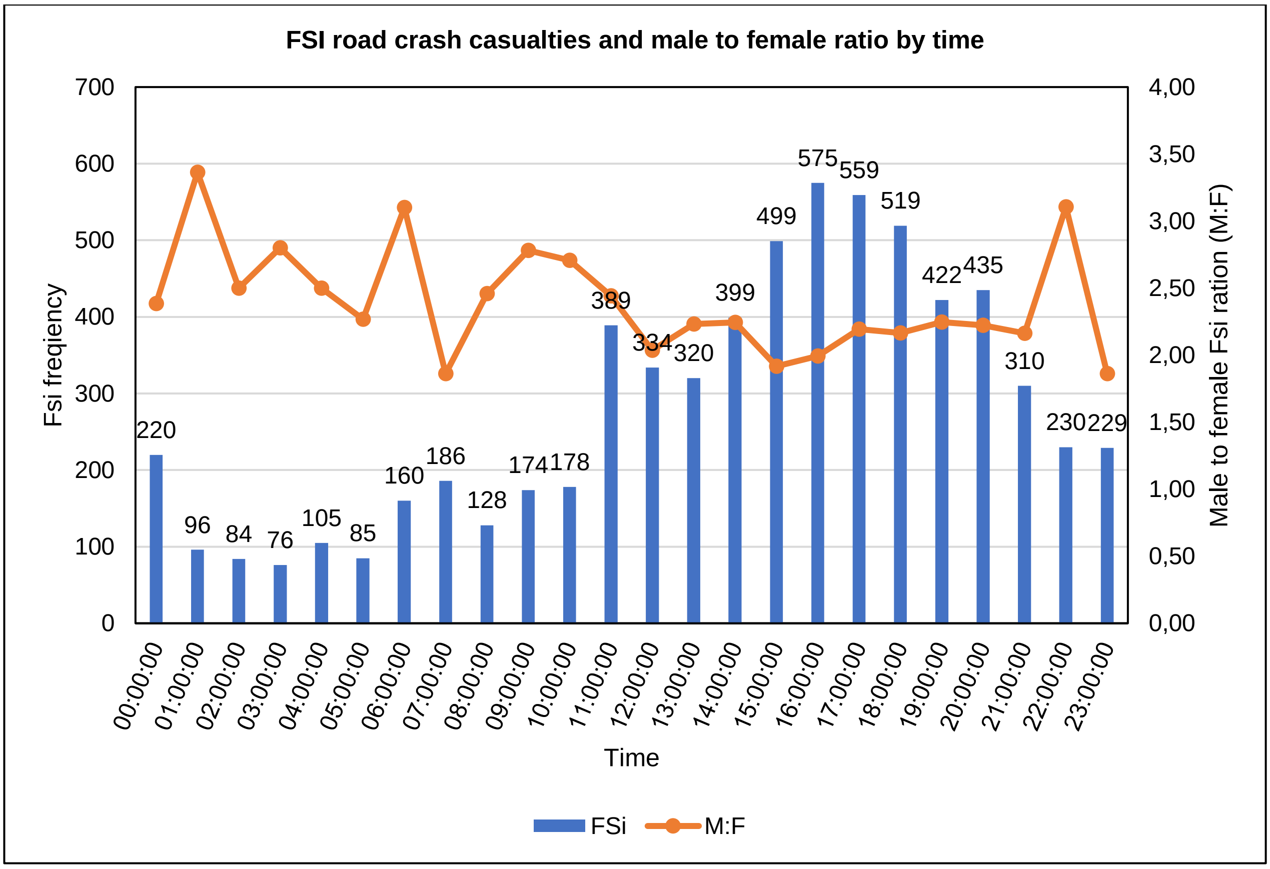


Figure 4.14 FSI road crash casualties and male to female ratio by time of day

4.2.2.2. FSI occupants by day of the week

The study assessed the temporal fatal and serious injury (FSI) road user casualties' temporal variations across the days of the week through descriptive and inferential statistics. The descriptive statistics for the daily national rural road causalities are presented in Table 4.14 and visually illustrated in Figure 4.15 and Figure 4.16.

Table 4.14 Descriptive statistics of weekdays FSI road crash casualties

Dependent Variable: FSI Casualty Count			
Weekday	**Mean**	**Std. Deviation**	**N**
Mon	155.20	39.271	5
Tue	119.80	32.889	5
Wed	121.80	25.223	5
Thu	149.80	20.117	5
Fri	215.00	47.207	5
Sat	268.80	86.085	5
Sun	235.80	39.047	5
Hol.	81.40	9.555	5
Total	168.45	72.813	40

From Table 4.14, it is observed that the highest frequency of fatal and serious road crash injuries occurred on Saturdays (1 427 FSI casualties), followed by Fridays (1 174 FSI causalities) and Sundays (1 233 FSI casualties). The lowest FSI frequencies over the day of the week were observed over Holidays (407 FSI casualties).

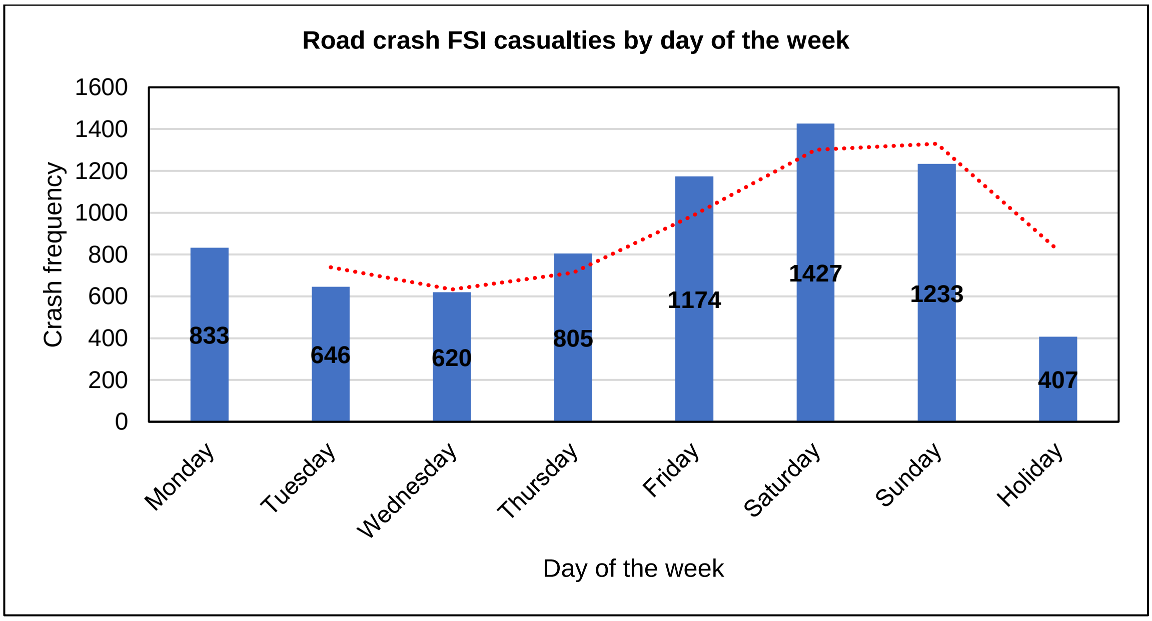


Figure 4.15 Distribution of FSI road crash casualties by day of the week

Similar to Figure 4.15, the estimated marginal means for the week day casualties illustrated in Figure 4.16 indicate that higher casualties were observed over the weekend days (Friday, Saturday and Sunday) and were lower during the week days (Monday to Thursday) and over holidays.

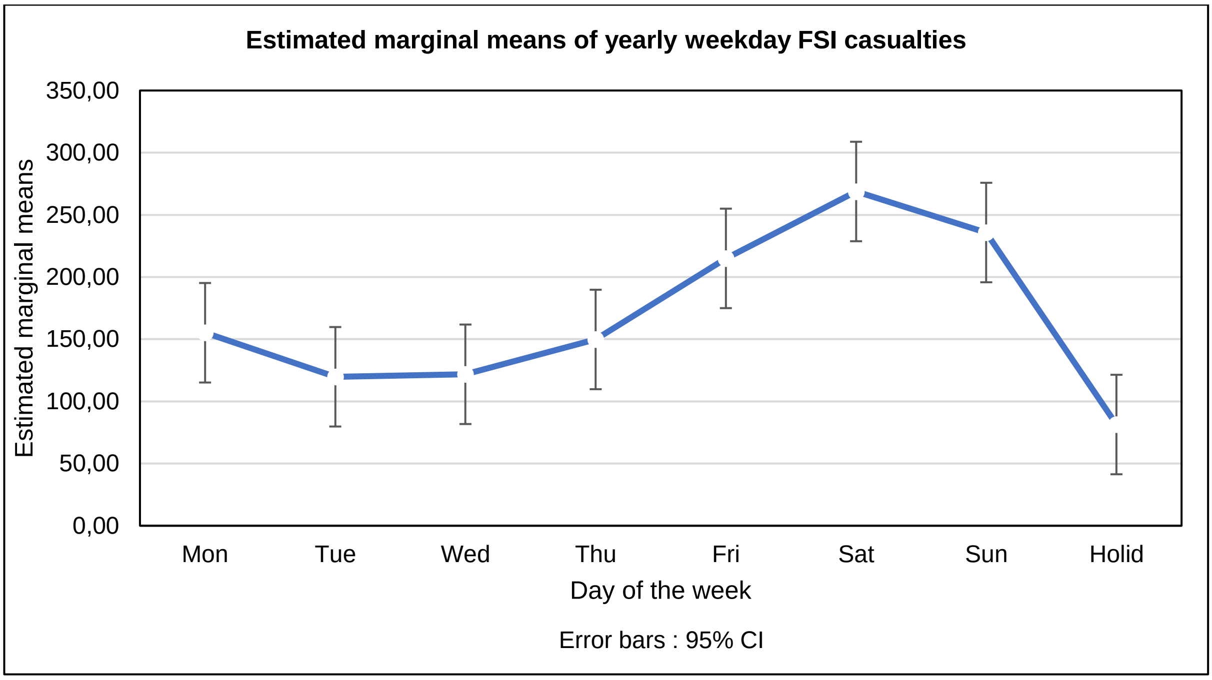


Figure 4.16 Estimated FSI road crash casualties means by day of the week

Using the ANOVA test, the daily FSI road casualties' individual mean differences were evaluated at 95 percent confidence interval. The ANOVA test scores indicated in Table 4.15 show that the mean differences between the test groups (weekdays) are statistically significant ($p=0.000<0.05$). This suggest that the assumption that variances of casualties are equal across the days of the week is invalid. The test scores also indicate that 64.8 percent of the variance in the week day variances is predicated on the influence of the predictors.

Table 4.15 Results of ANOVA Test Scores on weekdays FSI road crash casualties

Tests of effects between subjects					
Dependent Variable: FSI road crash casualties					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Corrected Model	147085.900[a]	7	21012.271	11.267	0.000
Intercept	1135016.100	1	1135016.100	608.588	0.000
Weekday	147085.900	7	21012.271	11.267	0.000
Error	59680.000	32	1865.000		
Total	1341782.000	40			
Corrected Total	206765.900	39			

a. R Squared = .711 (Adjusted R Squared = .648)

In the same way, Levene's test for homogeneity of variance (see Table 4.16) was applied to test the assumption that road FSI casualties' variances are equal across the days of the week. Further, Levene's test results give an indication of the appropriate Post-hoc procedure to apply in assessing the individual mean differences between the test groups (days of the week). Levene's test results presented in Table were found to be statistically significant ($p=0.022<0.05$) at 95 percent confidence interval. This suggests that the equal variance assumption across the test group is negated. Owing to the statistical significance of the Levene's test results and following the procedure illustrated in Figure 3.12, the Games-Howell Post-hoc procedure was identified as the appropriate test to assess individual mean differences.

Table 4.16 Results of Levene's Test for Homogeneity of variance for weekdays FSI road crash casualties

Levene's Test of Equality of Error Variances[a,b]					
		Levene Statistic	**df1**	**df2**	**Sig.**
FSI Cases	Based on Mean	2.788	7	32	0.022
	Based on Median	1.596	7	32	0.172
	Based on Median and with adjusted df	1.596	7	15.573	0.209
	Based on trimmed mean	2.683	7	32	0.026

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.[a,b]
a. Dependent variable: FSI Cases
b. Design: Intercept + Weekday

The Games-Howell Post-hoc test results are presented in Table 4.17. The Games-Howell test was used to assess the differences in mean values across the days of the week. The relationships found statistically significant ($p<0.05$) at 95 percent confidence level are marked in red.

From Table 4.17, the differences in weekdays casualty mean values emerged statistically significant ($p<0.05$) between:

- Tuesday and Sunday
- Wednesday and Sunday
- Thursday and Sunday and holidays
- Friday and holidays; and
- Sunday and holidays

No statistically significant differences in mean values was identified between:

- Monday and all days of the week; and
- Saturday and all the days of the week.

Table 4.17 Results of Games-Howell Post Hoc Test on weekdays FSI road crash casualties

Post Hoc test: Games Howell						
Multiple Comparisons						
Dependent Variable: FSI Weekday causalities						
(I) Weekday	(J) Weekday	Mean Difference (I-J)	Std. Error	Sig.(p)	95% Confidence Interval	
					Lower Bound	Upper Bound
Mon	Tue	35.40	22.908	0.768	-56.00	126.80
	Wed	33.40	20.873	0.740	-53.13	119.93
	Thu	5.40	19.733	1.000	-80.21	91.01
	Fri	-59.80	27.462	0.448	-169.44	49.84
	Sat	-113.60	42.315	0.288	-301.63	74.43
	Sun	-80.60	24.767	0.123	-178.60	17.40
	Hol	73.80	18.075	0.094	-14.79	162.39
Tue	Mon	-35.40	22.908	0.768	-126.80	56.00
	Wed	-2.00	18.536	1.000	-76.69	72.69
	Thu	-30.00	17.242	0.670	-102.14	42.14
	Fri	-95.20	25.730	0.079	-200.36	9.96
	Sat	-149.00	41.213	0.122	-338.50	40.50
	Sun	-116.00*	22.832	**0.014**	-207.05	-24.95
	Hol	38.40	15.317	0.362	-35.12	111.92
Wed	Mon	-33.40	20.873	0.740	-119.93	53.13
	Tue	2.00	18.536	1.000	-72.69	76.69
	Thu	-28.00	14.428	0.564	-85.86	29.86
	Fri	-93.20	23.936	0.076	-196.14	9.74
	Sat	-147.00	40.117	0.128	-339.35	45.35
	Sun	-114.00*	20.789	**0.012**	-200.09	-27.91
	Hol	40.40	12.062	0.157	-15.15	95.95
Thu	Mon	-5.40	19.733	1.000	-91.01	80.21
	Tue	30.00	17.242	0.670	-42.14	102.14
	Wed	28.00	14.428	0.564	-29.86	85.86
	Fri	-65.20	22.949	0.250	-168.57	38.17
	Sat	-119.00	39.536	0.237	-313.56	75.56
	Sun	-86.00*	19.644	**0.048**	-171.12	-.88
	Hol	68.40*	9.960	**0.006**	24.51	112.29
Fri	Mon	59.80	27.462	0.448	-49.84	169.44
	Tue	95.20	25.730	0.079	-9.96	200.36
	Wed	93.20	23.936	0.076	-9.74	196.14
	Thu	65.20	22.949	0.250	-38.17	168.57
	Sat	-53.80	43.907	0.898	-241.61	134.01
	Sun	-20.80	27.398	0.991	-130.24	88.64
	Hol	133.60*	21.540	**0.022**	26.28	240.92
Sat	Mon	113.60	42.315	0.288	-74.43	301.63
	Tue	149.00	41.213	0.122	-40.50	338.50
	Wed	147.00	40.117	0.128	-45.35	339.35

	Thu	119.00	39.536	0.237	-75.56	313.56
	Fri	53.80	43.907	0.898	-134.01	241.61
	Sat	33.00	42.274	0.988	-155.06	221.06
	Hol	187.40	38.735	0.061	-11.16	385.96
Sun	Mon	80.60	24.767	0.123	-17.40	178.60
	Tue	116.00*	22.832	**0.014**	24.95	207.05
	Wed	114.00*	20.789	**0.012**	27.91	200.09
	Thu	86.00*	19.644	**0.048**	.88	171.12
	Fri	20.80	27.398	0.991	-88.64	130.24
	Sat	-33.00	42.274	0.988	-221.06	155.06
	Hol	154.40*	17.978	**0.005**	66.34	242.46
Hol	Mon	-73.80	18.075	0.094	-162.39	14.79
	Tue	-38.40	15.317	0.362	-111.92	35.12
	Wed	-40.40	12.062	0.157	-95.95	15.15
	Thu	-68.40*	9.960	**0.006**	-112.29	-24.51
	Fri	-133.60*	21.540	**0.022**	-240.92	-26.28
	Sat	-187.40	38.735	0.061	-385.96	11.16
	Sun	-154.40*	17.978	**0.005**	-242.46	-66.34

Based on observed means.

The error term is Mean Square (Error) = 1865.000.

*. The mean difference is significant at the 0.05 level.

4.2.2.3. Distribution of FSi occupants by month of the year

The distribution of fatal and serious injury casualties across the months of the year is illustrated in Figure 4.17. On the whole, the trend of fatal injuries causalities was found to be consistent across the months of the year, with the highest casualties recorded over December (256 fatal road casualties). In contrast, serious injury road casualties had three separate peaks across the months of the year. The highest serious injury peak is observed in December (467 serious injuries). The other peaks are observed over May (418 serious injuries) and August (394 serious injuries). These distinct peaks coincide with the holiday seasons in Namibia, when the traffic load on national rural roads is high. As a consequence, a high exposure for road users over these holiday months.

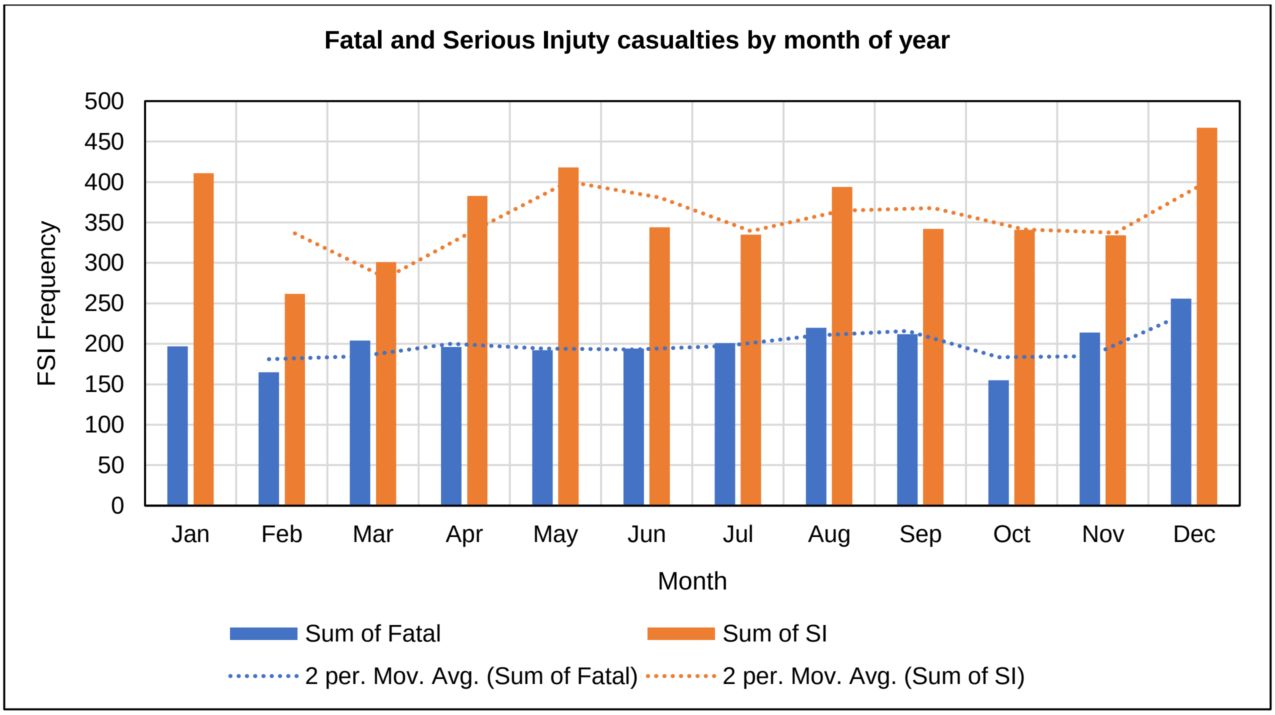


Figure 4.17 FSI road crash casualties by month of the year

4.2.3. Analysing driver risk factors and behavioural aspects

Road crashes and the consequences arising from them can be represented by a system of interlinked factors. However, the traditional crash causation model illustrated in Figure 1.3, lean towards placing the fault of road crashes on individual road users, with all other factors being perfect. As a consequence, human factors have become an overarching category of blame, rather than a vital source of information in identifying the multiplicity of factors that coupled together represent potential crash risks. The reality is that human factors are commonly involved in road crashes. In order to try and minimise some of the human led causation factors, it is important to provide a road environment (design characteristics) that does not provide the driver with too much or too little information at a single time, as this can cause confusion. The relationship between driver performance and environmental demand is clearly summed up in Blumenthal's early work (illustrated in Figure 4.18), which remains relevant today. Blumenthal's findings (cited in Shinar, 2017)) show that increasing the demands of a driver led to an increase in the likelihood of a crash occurring.

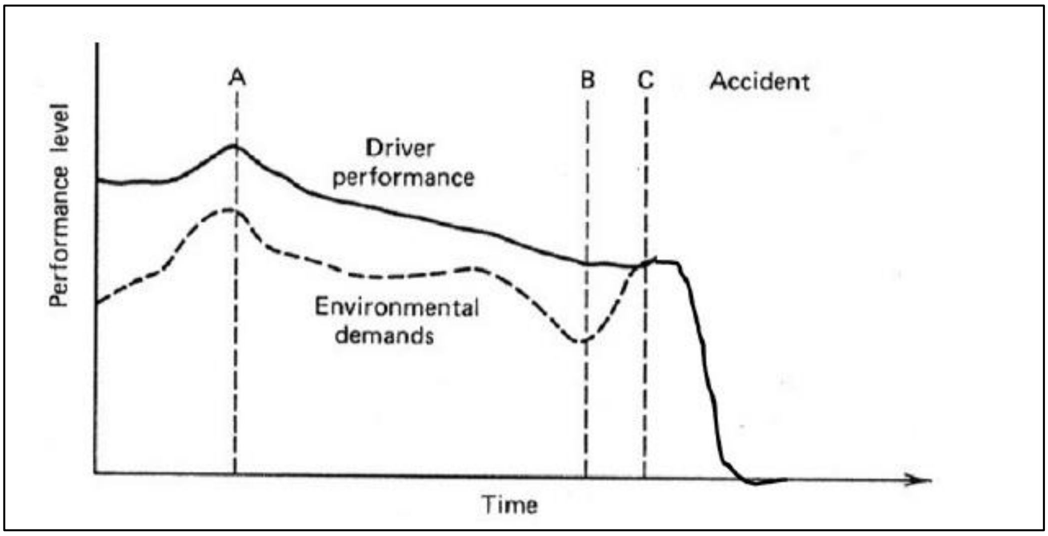


Figure 4.18 Blumenthal environmental demand and performance model (1968) (Shinar, 2017)

The developed localised road crash predictive models (CPMs) are important tools working to identify hazardous areas on the road. It is, however, important to recognise that models cannot work in isolation and identifying driver factors on national roads will play a crucial role in understanding how human factors intersect with road environment factors. CPMs are a crucial tool in tackling the frequency of fatal and serious injury crashes, moreover coupled with identified driver risk factors in the study areas to develop appropriate remedial measures. It is important to understand the full crash causation process, as it provides vital information and almost always leads to a wide scope of possible areas of preventive and remedial actions. This section of the study assesses the role of driver behaviour and risk factors and attempts to understand the extent to which crash risk factors, including the road environment (traffic and design characteristics), impact crash risk on national rural roads in Namibia.

4.2.3.1. Driver gender-based crash risk analysis

The driver gender-based crash risk analysis described in Section 3.4.1 and Section 3.7, was carried out using the crash datasets for the 5-year period. The results of the analysis are presented in Table 4.18. The dataset comprised 3 325 drivers, with a gender breakdown showing that 2 629 (79 percent) were male drivers while 696 (21 percent) were female drivers. This indicates an overrepresentation of male driver (M: F=3.78) in the crash dataset. The male to female driver ration in the dataset is higher compared with the international male to female ratio (M: F= 3).

The driver risk analysis clearly indicates that human factors were predominant crash factors for both genders, followed by roadway and environmental factors, vehicle and other factors respectively. A detailed breakdown of human factors (contributing 67 percent to male driver crash causation) in male drivers indicates that intentional risks were the highest contributing factors (23 percent), followed by recognition errors (20 percent). In comparison, the human factors (contributing 70 percent to female driver crash causation) for female drivers indicate that recognition errors were the predominant factors (25 percent), followed by intentional risks (23 percent) in the dataset. The composition of the top three primary risk factors in both genders is presented below:

- Both driver genders were found to exhibit inadequate surveillance on the rural roads. This risk factor was more marked in females (9 percent) than in males (7 percent).
- Inattention among the drivers was observed a significant risk factor – more in females (8 percent) than in males (6 percent).
- Misjudgement of gaps was notable in both driver genders, with the risk factor slightly more dominant in males (6 percent) than females (5 percent).
- Dangerous manoeuvres (M=4 percent; F=6 percent) and following too close (M=4 percent; F=5 percent) risk factors were also identifiable in both driver genders. More unexpected, they were more marked in female drivers than in male drivers on national roads.
- Traffic violations were equally noticeable in both driver genders, accounting for 4 percent as a primary risk factor in each gender.

For both male and female drivers, roadway and environmental risk factors were the second highest (25 and 26 percent respectively) contributor to crash occurrence, after human related crash factors. The following primary factors were of interest:

- Both driver genders were found to have significantly higher encounters with animals compared with other roadway and environmental factors. These encounters were higher for males than females, which can possibly be attributed to the higher exposure/ overrepresentation of male than female drivers.

- Poor visibility and weather were equally experienced as a primary risk factor by both driver genders.
- Speed differential was more marked as a primary risk factor in female drivers than in males.

The study identified vehicles factors as more of a primary risk factor in male drivers compared with female drivers.

Table 4.18 Driver gender and risk factors

Risk factors		As a primary contributing factor (Level 1) (n)		As a primary contributing factor (Level 1) (%)	
		Male	Female	Male	Female
Recognition error	Inadequate surveillance	191	62	7%	9%
	Internal distraction	29	9	1%	1%
	Inattention	165	53	6%	8%
	Confusion over the road environment	26	11	1%	2%
	Visual impairment	60	12	2%	2%
	Complex environment: overestimation	10	17	0%	2%
	Response delay	48	12	2%	2%
	Sub-total	529	176	20%	25%
Decision error	Too fast for conditions	16	18	1%	3%
	Too fast for a curve	38	4	1%	1%
	False assumption of other's action	92	15	3%	2%
	Misjudgement of gap or other's action	162	32	6%	5%
	Failure to use passive safety features	2	1	0%	0%
	Swerve in front of other traffic	11		0%	0%
	Unsafe passing	3		0%	0%
	Sub-total	324	70	12%	10%
Performance error	Overcompensation	38	11	1%	2%
	Poor directional control	92	21	3%	3%
	Panic/Freezing	20	27	1%	4%
	Other performance error	81	14	3%	2%
	General driving ability: Skills	72	9	3%	1%
	Sub-total	303	82	12%	12%
Intentional risk	Fatigue	109	26	4%	4%
	Alcohol	44	12	2%	2%
	Drugs	0		0%	0%
	Aggression	44		2%	0%
	Dangerous manoeuvre	113	44	4%	6%
	Traffic violation	93	31	4%	4%
	Following too close	101	34	4%	5%
	Speeding	21		1%	0%
	Too fast for conditions	84	14	3%	2%
	Sub-total	609	161	23%	23%

Physiological risk	Physical impairment			0%	0%
	Heart attack			0%	0%
	Eyesight			0%	0%
	Medications			0%	0%
	Age Senior driver/ped (<65)			0%	0%
	Age Young driver (<25)			0%	0%
	Age Child ped (<15)			0%	0%
	Blackout	3		0%	0%
	Sub-total	3	0	0%	0%
Roadway and environmental	Potholes	18	11	1%	2%
	Animal	305	73	12%	10%
	Obstructions			0%	0%
	Work zones			0%	0%
	Faulty traffic light			0%	0%
	Roadblock			0%	0%
	Weather	74	25	3%	4%
	Poor visibility: night/glare/dawn/dusk	87	19	3%	3%
	Road surface	62	23	2%	3%
	Stone projected by another car	11	1	0%	0%
	Stone	12		0%	0%
	Speed differentiation: Congestion	21	29	1%	4%
	Road geometry: Curve/slope	59	3	2%	0%
	Sub-total	649	184	25%	26%
Vehicle factors	Tyre bust	53	7	2%	1%
	Defective lights or indicators			0%	0%
	Defective brake	18		1%	0%
	Missing or defective mirrors			0%	0%
	Defective steering or suspension	16		1%	0%
	Overloaded or poorly loaded vehicle or trailer	29		1%	0%
	Other	14	3	1%	0%
	Tyre hooked off the vehicle	22		1%	0%
	Sub-total	152	10	6%	1%
Other road user error	Cyclist unsafe riding	4		0%	0%
	Bicycle equipment malfunction			0%	0%
	Cycling without helmet			0%	0%
	Intoxicated cyclist			0%	0%
	Unsafe riding environment			0%	0%
	Cycling in darkness	9		0%	0%
	Cyclist distraction			0%	0%
	Traffic light violation	9	2	0%	0%
	Pedestrian using the roadway	29	11	1%	2%
	Intoxicated pedestrian	9		0%	0%
	Sub-total	60	13	2%	2%

4.2.3.2. Driver-age based crash risk analysis

This analysis probed the relationship between the driver ages and primary risk factors in crash occurrences between 2012 and 2016. The results of the driver-age based crash risk analysis on national rural roads are presented in Table 4.19. Several points of interest across the driver age groups are discussed below:

- Drivers in the adolescent age group (less than 18 years) were found to be more prone to human errors (86 percent) than other road user errors in the risk analysis. Of the human errors, response delay (29 percent), inadequate surveillance (21 percent) and driving too fast for curves (21 percent) were found as the most notable primary contributing factors in the crash occurrences.
- For the young adults (18 to 25 years) in the crash dataset, as expected, a majority (75 percent) of the primary risk factors involved in the crashes were deemed to be human-related errors. The majority of the primary factors were found to be intentional (33 percent of human errors), with traffic violations (11 percent of intentional risks) playing a major role. Roadway and environmental risk factors (25 percent) played the second highest impact of crash occurrence in this age group, with a marked contribution to crashes by animals (19 percent) on the national rural roads.
- The driver gender analysis described and illustrated in Figure 4.11 shows that the age group 26 to 35 years represents the largest population (approximately 40 percent of driver population) of drivers on the national rural roads. In the same way as the previously discussed age groups, though lower, human related errors contributed to the highest number of crashes on the roads, representing 64 percent of all risk factors. Inattention (10 percent), inadequate surveillance (8 percent) and dangerous manoeuvres (8 percent) were identifiable primary human risk factors in this age group. Roadway and environmental risk factors were found to account for 33 percent of all risk factors attributed to drivers in the 26 to 35 years age group. Animals (17 percent of roadway and environmental risk factors) on the rural roads were identified as a significantly high primary crash risk factor in this grouping. Vehicle and other road user factors contributed 4 percent to crashes in the aforementioned age group.
- The study found that human-related risk factors also played the highest role in crash occurrences among the 36 to 65 years age group in the crash dataset. The crashes that occurred in the age group comprised 70 percent human-related risk factors. Roadway and environmental, vehicle and other road user factors represented approximately 22 percent, 5 percent and 3 percent respectively, of all primary risk factors in the age group. Of contrast to other age groupings, where animals are the highest contributing risk factor in the roadway and environmental category, the road surface (10 percent) was identified as the highest contributing risk factor.

- As expected, the highest human related risk factors in crash occurrences among all the age groups was identified in the elderly (greater than 65 years). These can possibly be attributed to reduced physiological processes as the aging process occurs. The human related errors contributed approximately 95 percent in all crashes were elderly drivers were involved during the period 2012 to 2016. The highest primary risk factors identified in the human-related errors by the elderly were confusion over the road environment, a false assumption of other road users' action and panic/ freezing in complex situations on the road. All these primary risk factors equally accounted for 11 percent of the human-related factors.

Table 4.19 Analysis of road crash risk by driver age using primary contributing risk factors

Risk factors		As a primary contributing factor (Level 1) (n)					As a primary contributing factor (Level 1) (%)				
		Adolescent (<18 years)	Young adults (18-25 years)	Adults (26-35 years)	Middle Age (36-65 years)	Elderly (>65 years)	Adolescent (<18 years)	Young adults (18-25 years)	Adults (26-35 years)	Middle Age (36-65 years)	Elderly (>65 years)
Recognition error	Inadequate surveillance	3	33	63	49	3	21%	10%	8%	7%	6%
	Internal distraction		2	1	15		0%	1%	0%	2%	0%
	Inattention		16	81	34	4	0%	5%	10%	5%	8%
	Confusion over the road environment		2	3	7	6	0%	1%	0%	1%	11%
	Visual impairment					3	0%	0%	0%	0%	6%
	Complex environment: overestimation				6		0%	0%	0%	1%	0%
	Response delay	4	1		22	5	29%	0%	0%	3%	9%
	Sub-total	**7**	**54**	**148**	**133**	**21**	**50%**	**17%**	**19%**	**19%**	**40%**
Decision error	Too fast for conditions		13	2		1	0%	4%	0%	0%	2%
	Too fast for a curve	3		14	21		21%	0%	2%	3%	0%
	False assumption of other's action	1	11	41	37	6	7%	3%	5%	5%	11%
	Misjudgement of gap or other's action		4	17	26	5	0%	1%	2%	4%	9%
	Failure to use passive safety features		1				0%	0%	0%	0%	0%
	Swerve in front of other traffic		7	2			0%	2%	0%	0%	0%
	Unsafe passing						0%	0%	0%	0%	0%
	Sub-total	**4**	**36**	**76**	**84**	**12**	**29%**	**11%**	**10%**	**12%**	**23%**

Performance error	Overcompensation		9	15	2	1	0%	3%	2%	0%	2%
	Poor directional control		18	10	53		0%	6%	1%	7%	0%
	Panic/Freezing		7	2	14	6	0%	2%	0%	2%	11%
	Other performance error	1		31	51		7%	0%	4%	7%	0%
	General driving ability: Skills		12	11	17	2	0%	4%	1%	2%	4%
	Sub-total	**1**	**46**	**69**	**137**	**9**	**7%**	**14%**	**9%**	**19%**	**17%**
Intentional risk	Fatigue		4	44	31		0%	1%	6%	4%	0%
	Alcohol		13	11	3		0%	4%	1%	0%	0%
	Drugs						0%	0%	0%	0%	0%
	Aggression		14	19	6		0%	4%	2%	1%	0%
	Dangerous manoeuvre		21	62	16	1	0%	6%	8%	2%	2%
	Traffic violation		35	19	16	2	0%	11%	2%	2%	4%
	Following too close		19	27	41	4	0%	6%	3%	6%	8%
	Speeding		1	10	2		0%	0%	1%	0%	0%
	Too fast for conditions			13	28		0%	0%	2%	4%	0%
	Sub-total	**0**	**107**	**205**	**143**	**7**	**0%**	**33%**	**26%**	**20%**	**13%**
Physiological risk	Physical impairment						0%	0%	0%	0%	0%
	Heart attack						0%	0%	0%	0%	0%
	Eyesight						0%	0%	0%	0%	0%
	Medications						0%	0%	0%	0%	0%
	Age Senior driver/ped (<65)					1	0%	0%	0%	0%	2%
	Age Young driver (<25)						0%	0%	0%	0%	0%
	Age Child ped (<15)						0%	0%	0%	0%	0%
	Blackout				1		0%	0%	0%	0%	0%
	Sub-total	**0**	**0**	**0**	**1**	**1**	**0%**	**0%**	**0%**	**0%**	**2%**

Roadway and environmental	Potholes						0%	0%	0%	0%	0%
	Animal		62	133	49	3	0%	19%	17%	7%	6%
	Obstructions						0%	0%	0%	0%	0%
	Work zones						0%	0%	0%	0%	0%
	Faulty traffic light						0%	0%	0%	0%	0%
	Roadblock						0%	0%	0%	0%	0%
	Weather		16	41	23		0%	5%	5%	3%	0%
	Poor visibility: night/glare/dawn/dusk		5	26	11		0%	2%	3%	2%	0%
	Road surface			49	70		0%	0%	6%	10%	0%
	Stone projected by another car						0%	0%	0%	0%	0%
	Stone						0%	0%	0%	0%	0%
	Speed differentiation: Congestion			14	6		0%	0%	2%	1%	0%
	Road geometry: Curve/slope						0%	0%	0%	0%	0%
	Infrastructure						0%	0%	0%	0%	0%
	Sub-total	**0**	**83**	**263**	**159**	**3**	**0%**	**25%**	**33%**	**22%**	**5%**
Vehicle factors	Tyre bust			5	14		0%	0%	1%	2%	0%
	Defective lights or indicators						0%	0%	0%	0%	0%
	Defective brake				9		0%	0%	0%	1%	0%
	Missing or defective mirrors						0%	0%	0%	0%	0%
	Defective steering or suspension						0%	0%	0%	0%	0%
	Overloaded or poorly loaded vehicle or trailer			17	11		0%	0%	2%	2%	0%
	Other						0%	0%	0%	0%	0%
	Tyre hooked off the vehicle			3	1		0%	0%	0%	0%	0%

	Sub-total	**0**	**0**	**25**	**35**	**0**	**0%**	**0%**	**3%**	**5%**	**0%**
Other road user error	Cyclist unsafe riding						0%	0%	0%	0%	0%
	Bicycle equipment malfunction						0%	0%	0%	0%	0%
	Cycling without helmet						0%	0%	0%	0%	0%
	Intoxicated cyclist						0%	0%	0%	0%	0%
	Unsafe riding environment						0%	0%	0%	0%	0%
	Cycling in darkness						0%	0%	0%	0%	0%
	Cyclist distraction						0%	0%	0%	0%	0%
	Obstructions						0%	0%	0%	0%	0%
	Traffic light violation						0%	0%	0%	0%	0%
	Pedestrian using the roadway	2		2	11		14%	0%	0%	2%	0%
	Child running after the car						0%	0%	0%	0%	0%
	Sub-total	**2**	**0**	**6**	**23**	**0**	**14%**	**0%**	**1%**	**3%**	**0%**

Total	**14**	**326**	**792**	**715**	**53**
Total (Percent)	**1%**	**17%**	**42%**	**38%**	**3%**
Total	**1900**				

4.2.3.3. Determination of crash risk factors and relationship between risk factors

A second level crash risk factor analysis was carried out to determine the probable relationship between the primary risk factors and other risk factors (level 2 and level 3) to have possibly influenced the occurrence of crashes. The results of this analysis are presented in Table 4.20. Several relationships were noticeable between the different levels of risk factors. Of interest are the following relationships:

- Level 2 and level 3 crash risk factors contributed the highest secondary risk (49 percent) to road crashes were intentional risks were identified as the leading risk factor preceding a crash.
- In road crashes were recognition errors were identified as the leading primary risk factor, level 2 and level 3 contributed slightly higher than three quarters (29 percent) of all the secondary risk identified to have preceded the crash. Recognition errors were identified as primary risk factors in 17 percent of all risk factors in the risk analysis.
- In road crashes were the leading risk factors were roadway and environment related, the results indicated that level 2 and level 3 risk factors contributed approximately 27 percent of all secondary risk factors. Also, roadway and environmental risk factors were identified as the second highest (25 percent) primary risk factors in the crash risk analysis, only after intentional risks (human-related factor).
- As expected on Namibian national rural roads, animals were identified as the highest (17 percent) individual primary risk factor for crashes. In the same way, they were also recorded as the highest (19 percent) level 2 and level 3 possible contributing factor when a primary factor was identified.
- Also identifiable from the crash risk analysis, the following risk factors were individual high level 2 and level 3 contributors: (1) dangerous road manoeuvres (15 percent); (2) the misjudgement of gaps or other driver’s road actions (14 percent); (3) traffic violations (12 percent); and (4) drivers following too close (11 percent).

Table 4.20 Analysis of road crash risk at Level 1,2 and 3 risk factors

Risk factors		As a primary contributing factor (Level 1)		As a possible contributing factor (Level 2 and Level 3)								
				Driver Human factors					Roadway and environmental factors	Vehicle factors	Other road user factors	Total (Percent)
		Frequency	Percentage	Recognition error	Decision error	Performance error	Intentional risk	Physiological conditions				
Recognition error	Inadequate surveillance	236	8%	109	91	17	72	3	19	2	0	10%
	Internal distraction	9	0%	0	9	0	10	0	9	0	0	1%
	Inattention	136	4%	53	64	82	51	5	44	0	0	10%
	Confusion over the road environment	21	1%	24	10	11	1	0	10	0	0	2%
	Visual impairment	51	2%	16	17	11	10	0	15	0	0	2%
	Complex environment: overestimation	18	1%	11	9	14	4	0	14	2	0	2%
	Response delay	61	2%	39	10	1	20	2	22	0	0	3%
	Sub-total	**532**	**17%**									**29%**
Decision error	Too fast for conditions	101	3%	74	21	6	13	2	23	0	0	4%
	Too fast for a curve	16	1%	11	7	2	5	1	8	0	0	1%
	False assumption of other's action	63	2%	51	12	5	11	3	3	1	1	3%
	Misjudgement of gap or other's action	271	9%	147	87	55	62	9	69	2	0	14%
	Failure to use passive safety features	4	0%	0	9	0	11	0	14	0	2	1%
	Swerve in front of other traffic	16	1%	17	11	4	6	1	1	0	0	1%
	Unsafe passing	5	0%	13	3	0	3	0	0	0	0	1%
Sub-total		**476**	**15%**									**25%**

Performance error	Overcompensation	22	1%	7	9	6	2	7	11	0	0	1%
	Poor directional control	105	3%	72	31	18	21	4	31	0	0	6%
	Panic/Freezing	25	1%	14	8	11	23	7	8	0	0	2%
	Other performance error	57	2%	32	26	1	16	8	15	0	0	3%
	General driving ability: Skills	50	2%	34	11	18	16	10	2	0	0	3%
	Sub-total	**259**	**8%**									**15%**
Intentional risk	Fatigue	101	3%	76	0	6	6	0	11	0	0	3%
	Alcohol	44	1%	31	20	9	26	14	1	0	0	3%
	Drugs	6	0%	2	7	0	5	2	0	0	0	1%
	Aggression	47	2%	19	18	24	38	0	1	0	0	3%
	Dangerous manoeuvre	225	7%	93	126	92	109	14	31	0	0	15%
	Traffic violation	223	7%	125	57	64	81	14	22	0	0	12%
	Following too close	195	6%	108	97	28	52	13	40	0	0	11%
	Speeding	28	1%	7	18	3	0	0	0	2	21	2%
	Sub-total	**869**	**28%**									**49%**
Physiological risk	Physical impairment	0	0%	0	0	0	0	0	0	0	0	0%
	Heart attack	0	0%	0	0	0	0	0	0	0	0	0%
	Eyesight	0	0%	0	0	0	0	0	0	0	0	0%
	Medications	0	0%	0	0	0	0	0	0	0	0	0%
	Age Senior driver/ped (<65)	0	0%	0	0	0	0	23	0	0	0	1%
	Age Young driver (<25)	0	0%	0	0	0	0	0	0	0	0	0%
	Age Child ped (<15)	0	0%	0	0	0	0	1	0	0	0	0%
	Blackout	14	0%	0	0	0	6	0	1	0	0	0%
	Sub-total	**14**	**0%**									**1%**
Roadway and	Potholes	6	0%	0	17	4	1	0	11	0	0	1%
	Animal	519	17%	293	121	9	23	6	127	0	0	19%
	Obstructions	43	1%	0	0	15	18	0	1	0	0	1%
	Work zones	6	0%	0	1	1	0	0	0	0	0	0%

	Weather	30	1%	1	26	8	0	0	38	0	0	2%
	Poor visibility: night/glare/dawn/dusk	85	3%	0	2	0	4	0	0	0	0	0%
	Road surface	36	1%	0	23	0	11	0	14	0	0	2%
	Stone projected by another car	29	1%	0	0	0	18	0	1	0	0	1%
	Stone	4	0%	0	0	0	0	0	1	0	0	0%
	Speed differentiation: Congestion	13	0%	19	1	0	0	0	0	0	0	1%
	Road geometry: Curve/slope	17	1%	0	1	7	0	0	10	0	0	1%
	Sub-total	**788**	**25%**									**27%**
Vehicle factors	Tyre bust	42	1%	0	6	18	8	0	27	0	0	2%
Vehicle factors	Defective lights or indicators	3	0%	0	0	0	2	0	3	0	0	0%
Vehicle factors	Defective brake	26	1%	0	0	3	4	0	14	0	0	1%
Vehicle factors	Defective steering or suspension	7	0%	0	0	0	3	0	10	0	0	0%
Vehicle factors	Overloaded or poorly loaded vehicle or trailer	30	1%	0	9	5	10	0	0	11	0	1%
Vehicle factors	Other	10	0%	0	0	2	0	0	0	0	0	0%
Vehicle factors	Tyre hooked off the vehicle	21	1%	0	4	6	0	0	7	0	0	1%
	Sub-total	**139**	**4%**									**5%**
Other road	Cyclist unsafe riding	3	0%	9	0	0	0	0	0	0	0	0%
Other road	Jaywalking	22	1%	11	2	1	0	8	15	0	0	1%
Other road	Pedestrian using the roadway	11	0%	19	0	0	0	0	24	0	0	1%
	Sub-total	**36**	**1%**									**3%**
	Total	**3113**	**100%**									

As discussed at the beginning of Section 4.2.3, the analysis of crash risk has traditionally examined the road user, vehicle and road environment separately. Moreover, researchers tend to look for one or a few crash risk factors, while in actual fact they should be analysing multiple causation factors. The systems approach taken in this study has attempted to build on Haddon's insights discussed in Section 1.1. This approach seeks to identify and rectify the major sources of error and design weaknesses that contribute to fatal and serious injuries on roads. For the Namibian national rural roads, the distribution of primary road crash risk factors was determined for national rural roads. The summary of this distribution is illustrated in Figure 4.19.

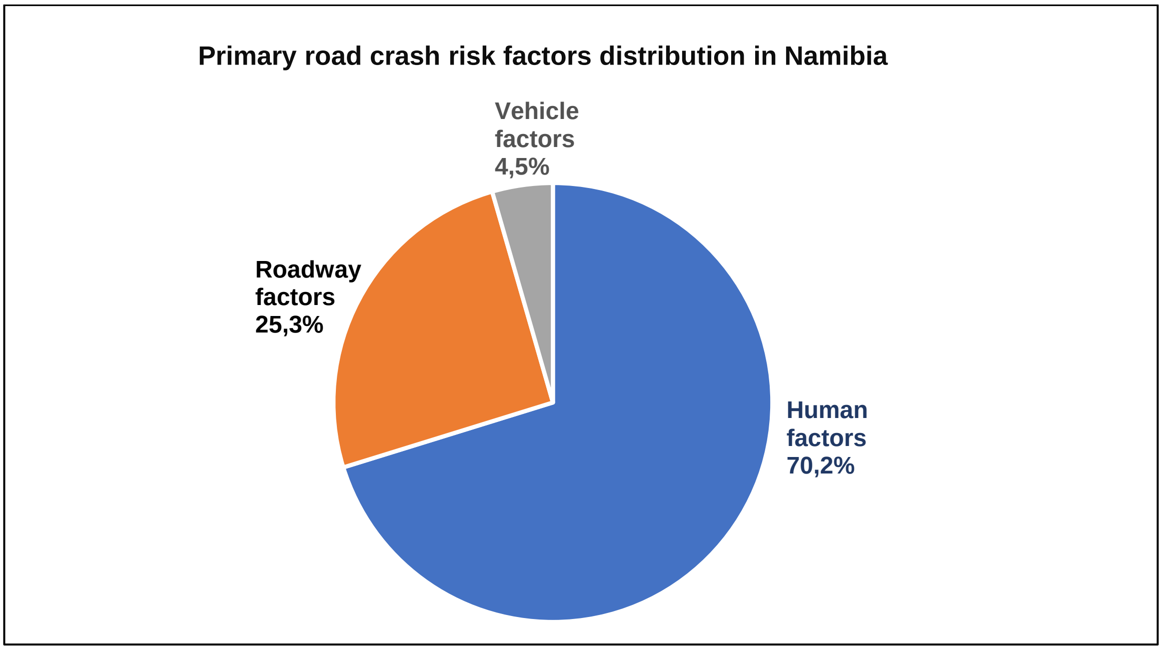


Figure 4.19 Distribution of road crash risk factors in Namibia

4.3 Road crash geospatial analyses

The distribution of road crashes was tested for all rural roads, high order (HORR) and low order rural roads (LORR) according to their functional classes detailed by the TRH 26 on Road Classification and Access Management in Table 2.1. The spatial distribution of fatal and serious injury (FSI) crashes was visualised by applying the planar Kernel Density Estimation (KDE) method to generate the raster maps for the three datasets. The FSI crash densities are classified into four classes of equal intervals showed in Figure 3.18. These classes are: (1) extreme; (2) high; (3) moderate; and (4) low crash intensities.

4.3.1. Distribution of road crashes on All Rural Roads

The FSI crash densities for crashes on all national rural roads are represented by the raster map in Figure 4.20. The raster map shows extreme clusters of FSI crashes in the central and northern parts of the high order rural road network. The raster map indicates higher crash densities in the western part of the HORR network and moderate to lower crash densities across the rest of the HORR network.

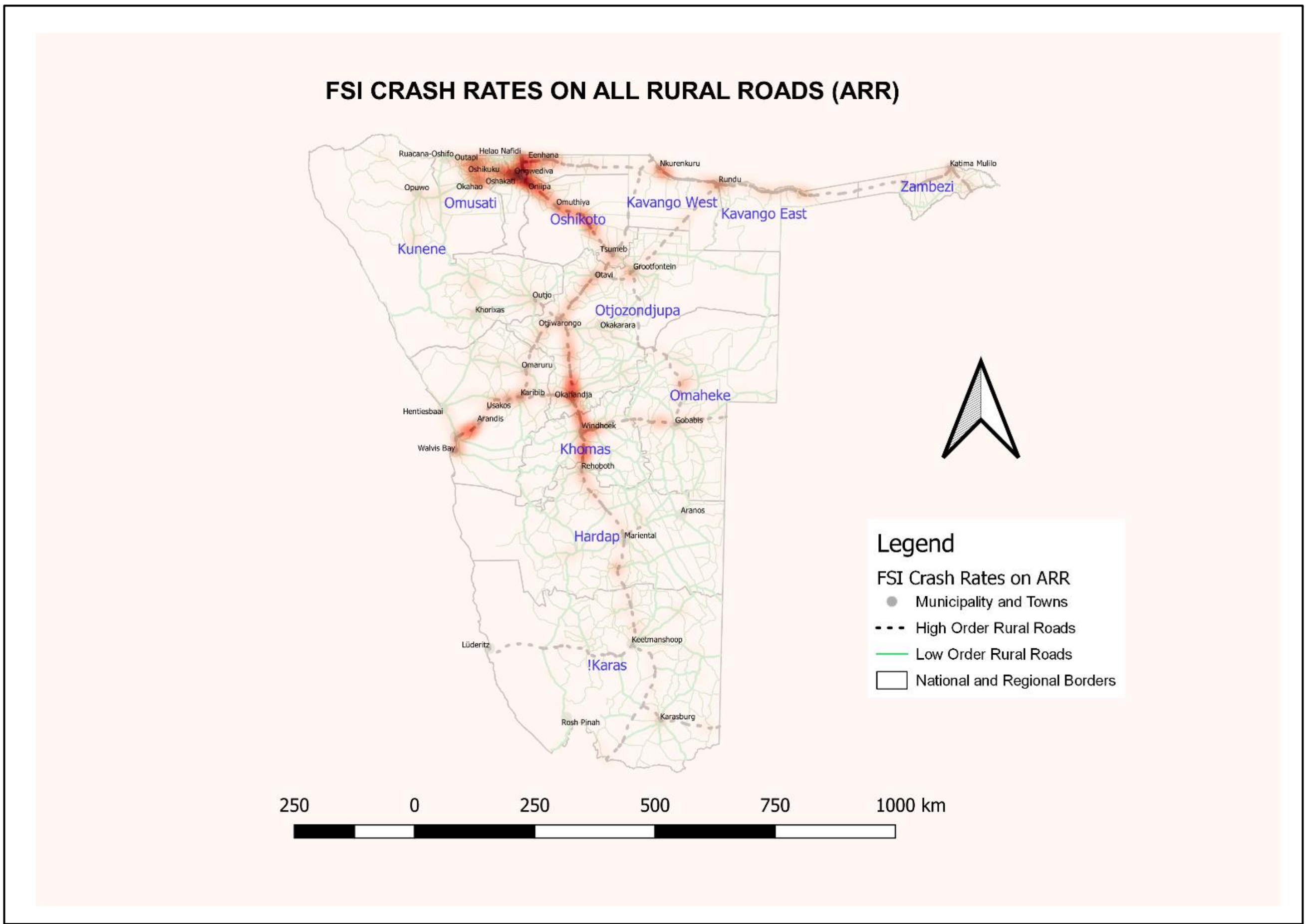


Figure 4.20 FSI crash rate distribution on all rural roads

4.3.2. Distribution of crashes on High Order Rural Roads

The map in Figure 4.21 presents the distribution of crashes that occurred on roads classified as high order (R1-R3) by the TRH 26. The KDE analysis indicates extreme crash densities in the northern, central and western parts of the HORR network. Higher to moderate crash densities were identified on the HORR network between the northern and central parts of the road network. The KDE analysis also revealed lower crash densities in areas towards the north-eastern and slightly south of the HORR network.

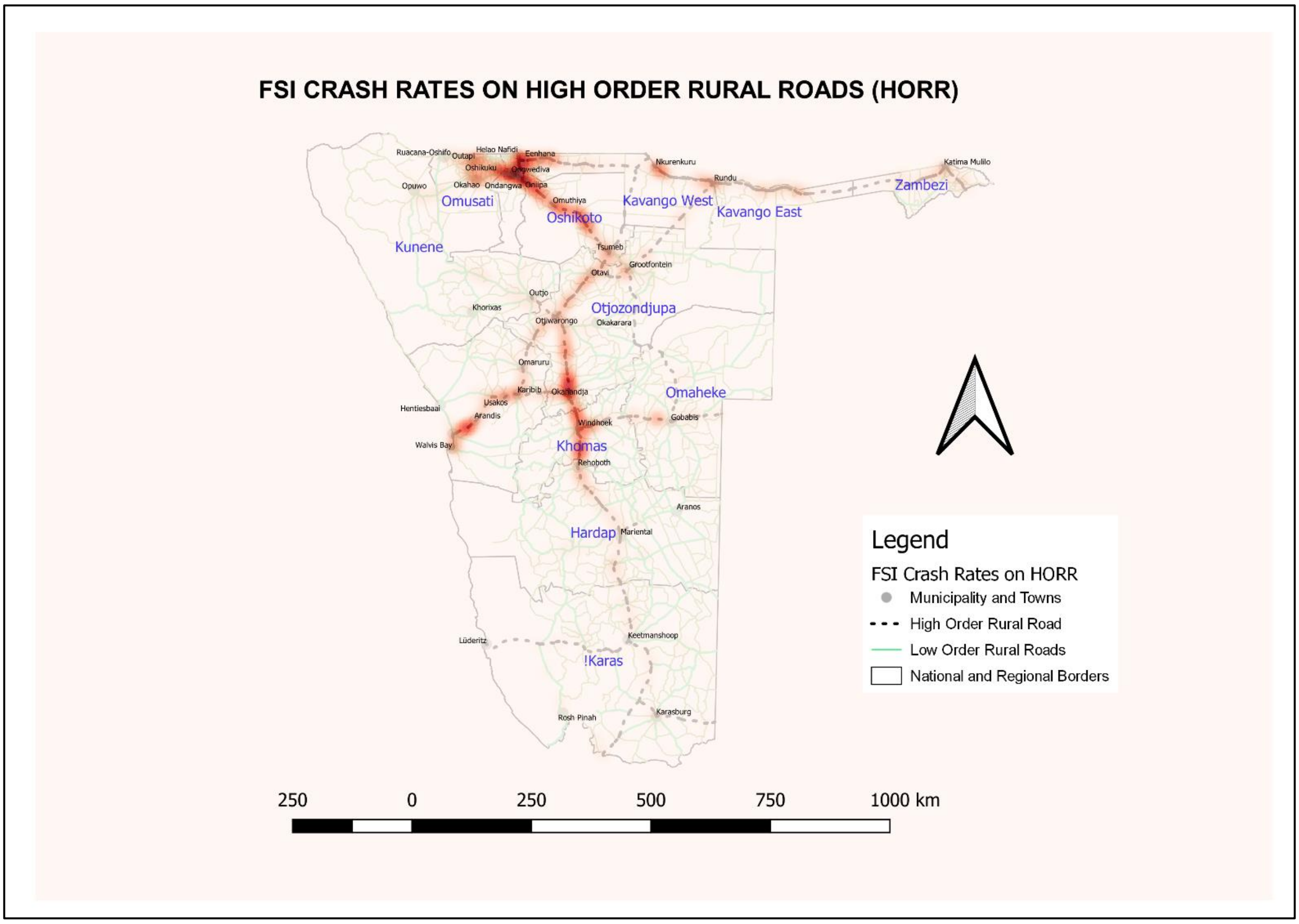


Figure 4.21 FSI crash rate distribution on high order rural roads

4.3.3. Distribution of crashes on Low Order Rural Roads

Figure 4.22 presents the distribution of FSI crash densities on roads classified as low order rural roads on the Namibian national road network. Extreme crash densities were identified towards the north, north western and eastern parts of the low order rural road network. Moderate to higher crash densities were identified in the north western part of the network while moderate to lower crash densities were identified on LORRs across the whole road network.

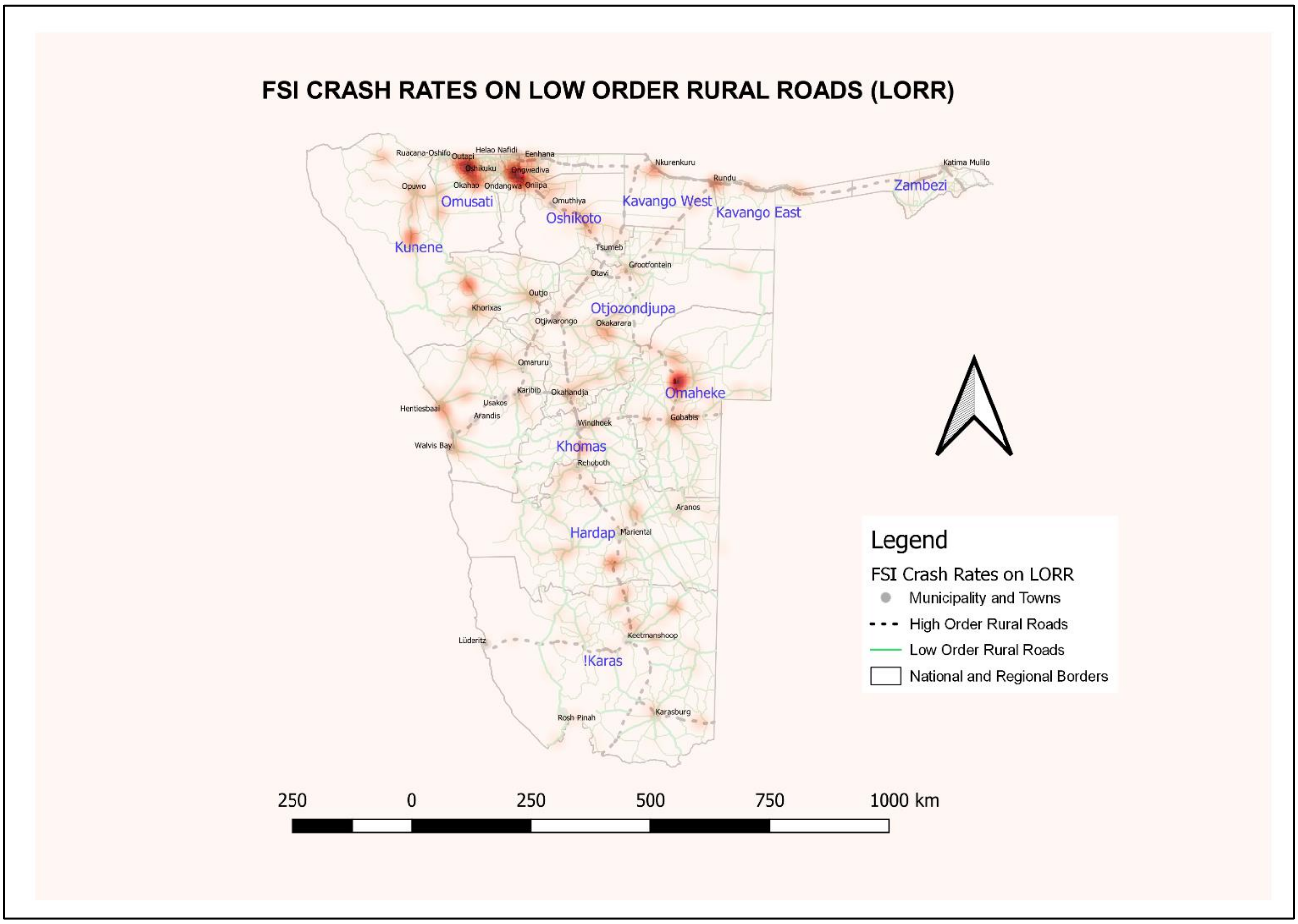


Figure 4.22 FSI crash rate distribution on low order rural roads

4.4 Compliance of National Rural Roads Design Environment with TRH 17 & TRH 26 Guidelines

This section explores the compliance of road design characteristics in the three study datasets with the Technical Recommendations for Highways 17 on the Geometric Design of Rural Roads (TRH 17), the Technical Recommendations for Highways 20 on the Structural Design, Construction and Maintenance of Unpaved Roads (TRH 20) and the Technical Recommendations for Highways 26 on Road Classification and Access Management (TRH 26). For this purpose, it is important to acknowledge that some of the roads investigated were designed for traffic conditions that have been far exceeded by current traffic conditions. Assessing the level of compliance on national rural roads with design guidelines is important to understanding the road environment on which the crash prediction models (CPMs) are developed and operating. For this reason, the level of compliance of model covariates will affect the parameter estimates and undoubtedly influences the type and magnitude of mediating effects that can be undertaken by authorities on the covariates and their impact on road safety.

4.4.1. Compliance Summary

Table 4.21 present the results of the road design covariates compliance assessment on the national rural roads to the TRH 17 and TRH 26, for the three datasets used in the study. These datasets are: (1) All national rural roads irrespective of the classification (ARR); (2) the High Order Rural Roads (HORR), class R1 to class R3; and (3) the Low Order Rural Roads (LORR), class R4 to R6 as shown in Table 2.1. The classification of the rural roads on the latter two datasets were according to a) the size and importance of the trip generator, b) reach connectivity and c) the travel stage. Some indication is also given by the traffic volumes, but that should not entirely be used in establishing the road classes.

The compliance assessment on the datasets was carried out on six geometric covariates with minimum design requirements stipulated in the aforementioned design guidelines. Covariates with a high level of non-compliance in the three datasets are highlighted in red. The lane width (LW) covariate demonstrated a higher level of compliance on paved roads in all datasets (LW_{ARR}= 65.25%; LW_{HORR}= 71.89%; LW_{LORR}= 62.11%). Of contrast, less than half of all lane widths on unpaved roads complied with the design requirements (LW_{ARR}= 47.81%; LW_{HORR}= 47.18%; LW_{LORR}= 48.46%). Significantly lower levels of compliance were shown by the surfaced shoulder width on all datasets (SSW_{ARR}= 14.20%; SSW_{HORR}= 21.24%; SSW_{LORR}= 8.22%)., with compliance levels below a quarter of the sample size. Similar to SSW, the proportion of road recommended to have paved shoulders in all three datasets also demonstrated compliance levels lower than a quarter (ST_{ARR}= 16.95%; ST_{HORR}= 24.40%; ST_{LORR}= 10.26%) of the sample size assessed.

The Ground Shoulder Width (GSW) demonstrated the highest level of compliance with design requirements on unpaved roads. All the three datasets proved compliance levels above eighty percent (GSW_{ARR}= 85.49%; GSW_{HORR}= 81.23%; GSW_{LORR}= 89.21%). However, the GSW compliance levels on paved roads were much lower than on paved roads. Less than half of all datasets (GSW_{ARR}= 30.21%; GSW_{HORR}= 43.12%; GSW_{LORR}= 22.45%) were found to comply with the design requirements for national rural roads. The compliance assessment also found significantly high levels of Stopping Sight Distance (SSD) compliance with design guidelines on paved national rural roads. More than ninety percent of the paved SSD datasets (SSD_{ARR}= 93.51%; SSD_{HORR}= 92.33%; SSD_{LORR}= 94.21%) proved to be compliant with the TRH 17. The SSD compliance levels on unpaved roads were evidently lower compared with SSDs on paved roads. Unpaved HORRs demonstrated a slight reduction in SSD compliance, while unpaved ARR and LORRs demonstrated higher reductions in compliance with design guidelines (SSD_{ARR}= 65.27%; SSD_{HORR}= 90.24%; SSD_{LORR}= 39.54%). The pavement conditions (PC) of all three datasets on paved and unpaved rural roads was evidently good on more than two thirds of all roads assessed. PC compliance on paved roads (PC_{ARR}= 79.24%; PC_{HORR}= 87.26%; PC_{LORR}= 71.82%) were slightly highly than PCs on unpaved roads (PC_{ARR}= 72.11%; PC_{HORR}= 74.11%; PC_{LORR}= 69.80%), with paved HORR PCs demonstrating a much higher compliance level.

Table 4.21 Geometric design and road characteristic compliance summary

	ARR			HORR			LORR		
	Average	**Compliance**	**TRH17**	**Average**	**Compliance**	**TRH17**	**Average**	**Compliance**	**TRH17**
LW	3,592m	65,25%	3,5m (Paved)	3,641m	71,89%	3,5m (Paved)	3,501m	62,11%	3,5m (Paved)
	8,789m	**47,81%**	9-11m (Unpaved)	8,811m	**47,18%**	9-11m (Unpaved)	8,748m	**48,46%**	9-11m (Unpaved)
SSW	0,179m	**14,20%**	1,5-3m (Paved)	0,356m	**21,24**%	2-3m (Paved)	0,020m	**8,22%**	1,5-2,5m (Paved)
	-	-	0m (Unpaved)	-	-	0m (Unpaved)	-	-	0m (Unpaved)
ST	Paved	**16,95%**	Paved	Paved	**24,40%**	Paved	Paved	**10,26%**	Paved
	-	-	Unpaved	-	-	Unpaved	-	-	Unpaved
GSW	1,532m	**30,21%**	1,5-3m (Paved)	1,633m	**43,12%**	2-3m (Paved)	1,402m	**22,45%**	1,5-2m (Paved)
	1,802m	85,49%	1,5-2,5m (Unpaved)	1,752m	81,23%	1,5-2,5m (Unpaved)	1,899m	89,21%	1,5-2,5m (Unpaved)
SSD	168,114m	93,51%	80-210m (Paved)	186,275m	92,33%	155-210m (Paved)	149,063m	94,21%	80-210m (Paved)
	108,580m	65,27%	80-155m (Unpaved)	138,211m	90,24%	80-155m (Unpaved)	77,214m	**39,54%**	80-155m (Paved)
PC	-	79,24%	Good (Paved)	-	87,26%	Good (Paved)	-	71,82%	Good (Paved)
	-	72,11%	Good (Unpaved)	-	74,11%	Good (Unpaved)	-	69,80%	Good (Unpaved)

4.4.2. Distribution of road crashes by design non-compliance

This section provides a description and spatial illustration of the crash distribution on national rural roads according to the non-compliance of design parameters assessed in the previous section (Section 4.4.1). Using the Kernel Density Estimation (KDE), the spatial analyses generated raster maps for the visualisation of the density of road crashes, with the classification intervals used shown by Figure 3.18.

The spatial distribution of road crashes by lane width (LW) non-compliance is shown in Figure 4.23. The extreme density of road crashes due to LW non-compliance is evident on the High Order and Low Order Rural Roads (HORR and LORR) in Northern part of the road network, with moderate to lower crash densities across the rest of the national rural road network.

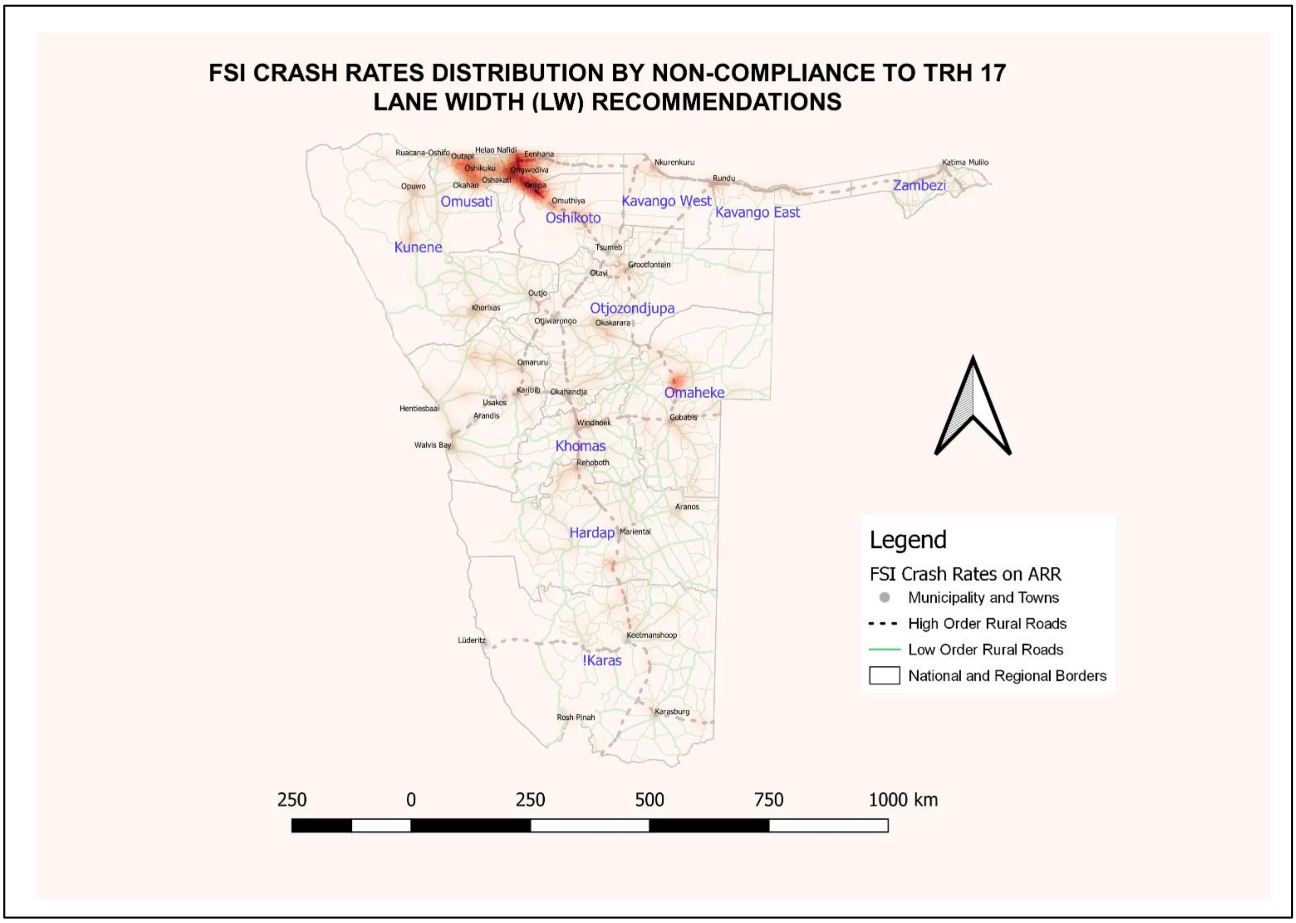


Figure 4.23 FSI crash rate distribution by non-compliance to TRH 17 LW design recommendations

Non-compliance to surfaced shoulder width (SSW) design requirements is shown to possibly have an influence on the occurrence of FSI crashes on High Order Rural Roads (HORR) in the northern, central and western part of the national rural road network in Figure 4.24. Moderate impact is shown in the north eastern section of the HORR with lower impacts towards the eastern and southern parts of the rural road network. This possible correlation is explored in Section 4.5.3.

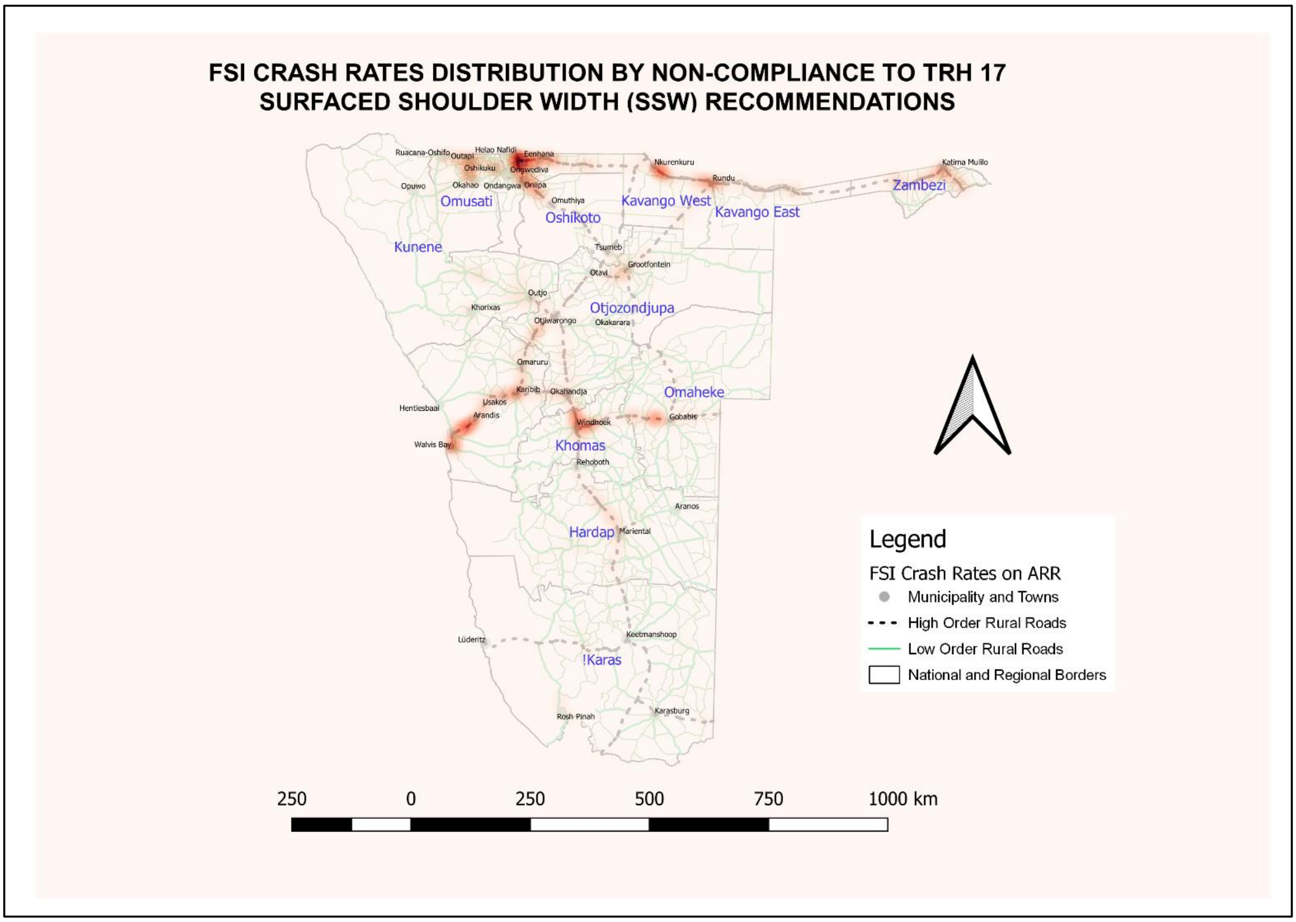


Figure 4.24 FSI crash rate distribution by non-compliance to TRH 17 SSW design recommendations

In the same way, the non-compliance of ground shoulder widths (GSW) on the national rural road network is notable on the Northern part of the road network, as shown by Figure 4.25, with extreme possible impacts on both LORRs and HORRs. This possible causal relationship between the crashes and GSWs is examined in Section 4.5.3. Slightly high to moderate impact on crash occurrence owing to GSW non-compliance is evident towards the central and western parts of the HORRs on the road network, with lower impacts visualised on the LORRs across the network.

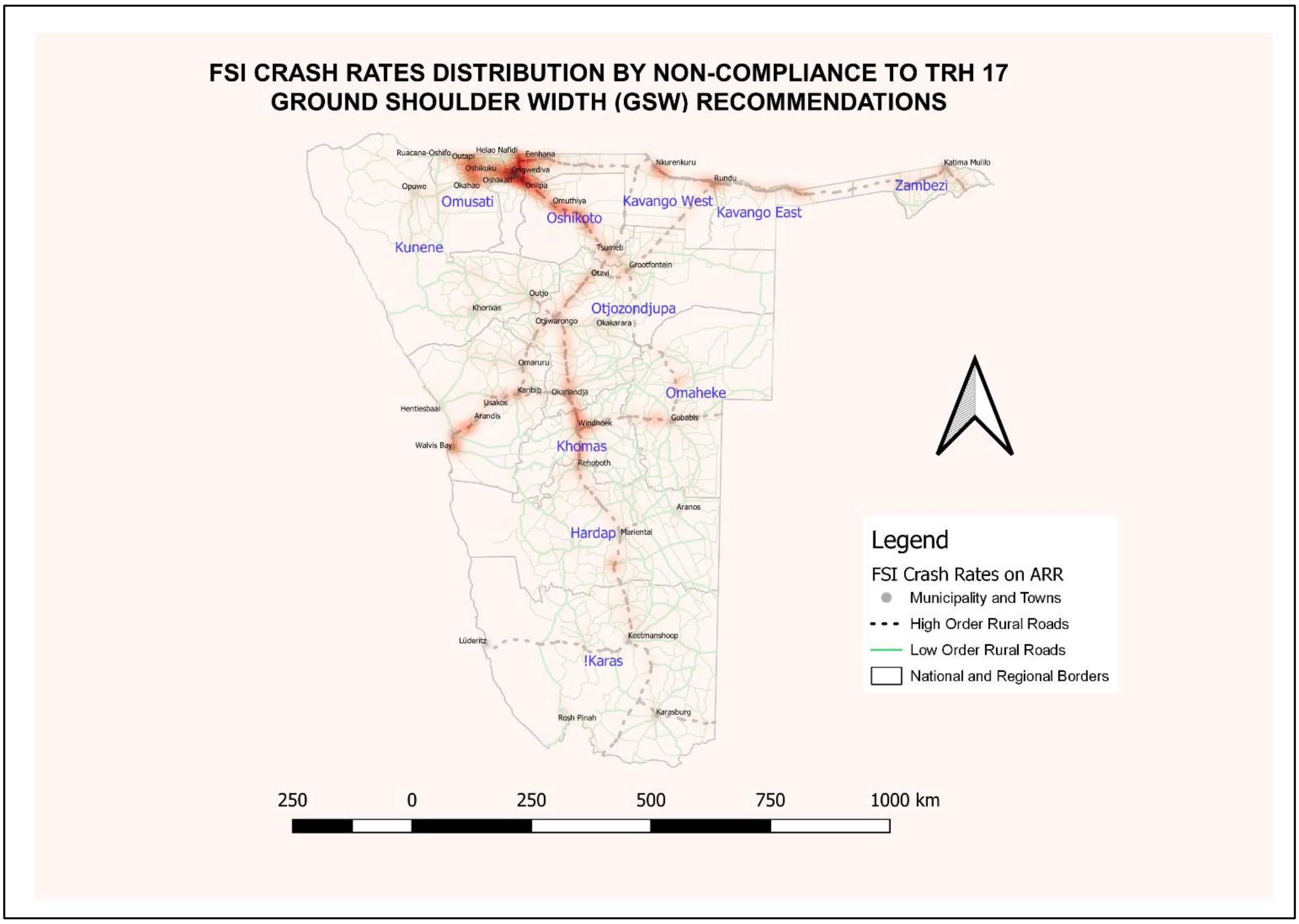


Figure 4.25 FSI crash rate distribution by non-compliance to TRH 17 GSW design recommendations

Figure 4.26 presents the impact of non-compliance of shoulder type (ST) on the occurrence of road crashes on both high and low order rural roads on the national road network. As a result of ST non-compliance, LORR and HORR road crash densities are shown to be extreme on the northern part of the road network. The central and western HORRs are shown exhibit higher densities of road crashes owing to ST non-compliance. LORRs across the national rural road network showed a low response to St non-compliance.

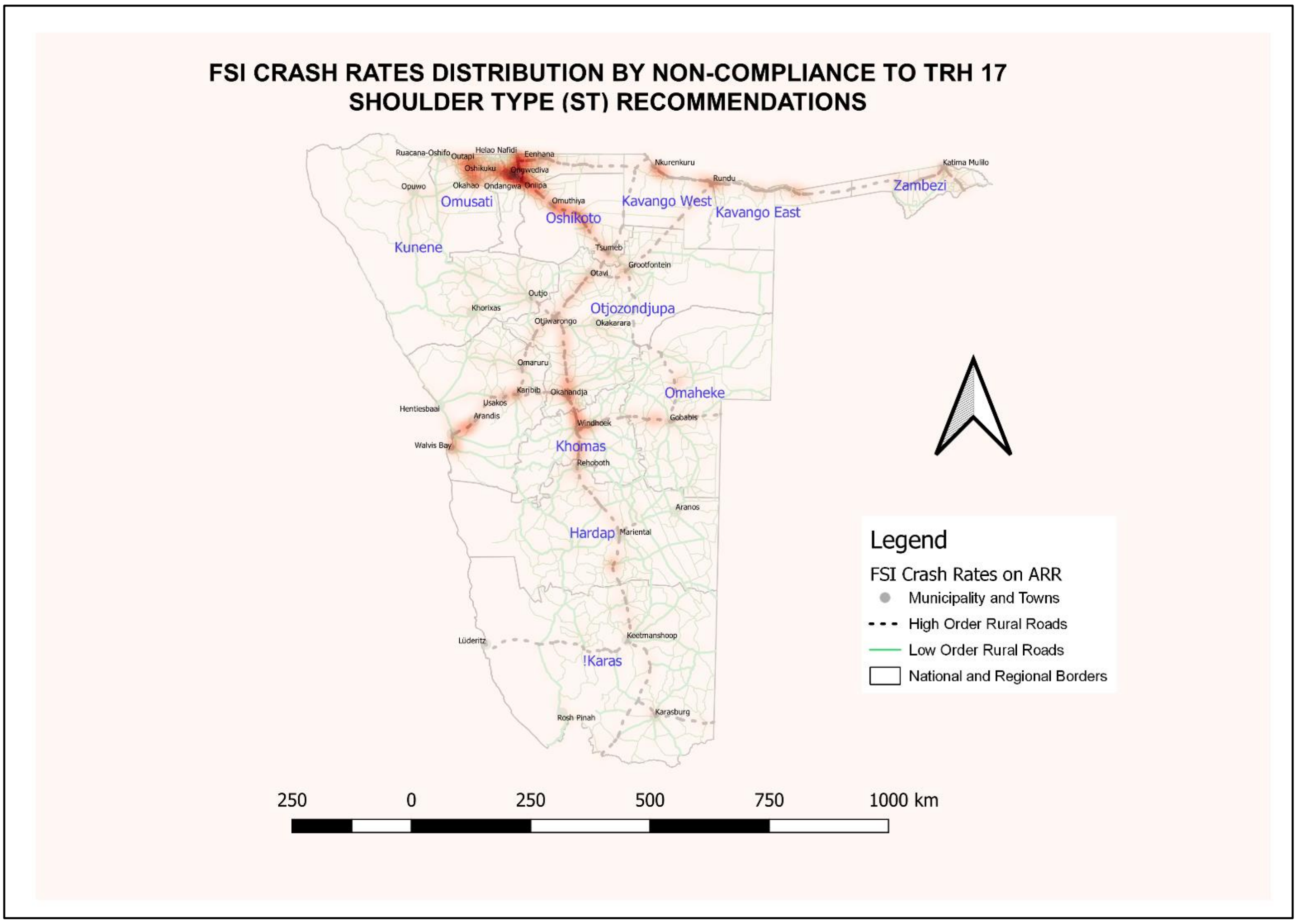


Figure 4.26 FSI crash rate distribution by non-compliance to TRH 17 ST design recommendations

The impact of stopping sight distance (SSD) non-compliance on crash distribution on the national rural road network is presented in Figure 4.27. In contrast to the impact of non-compliance of other design parameters on crash occurrence, SSD non-compliance is shown to cause extreme crash densities in the central part of the High Order Rural Road (HORR) network. The extreme densities due to SSD non-compliance are likely to stem from the fact that the central part of the HORR network is located in the hilliest part of the country. As a consequence, the terrain may have a high impact on the SSD available to drivers on the road network in the areas.

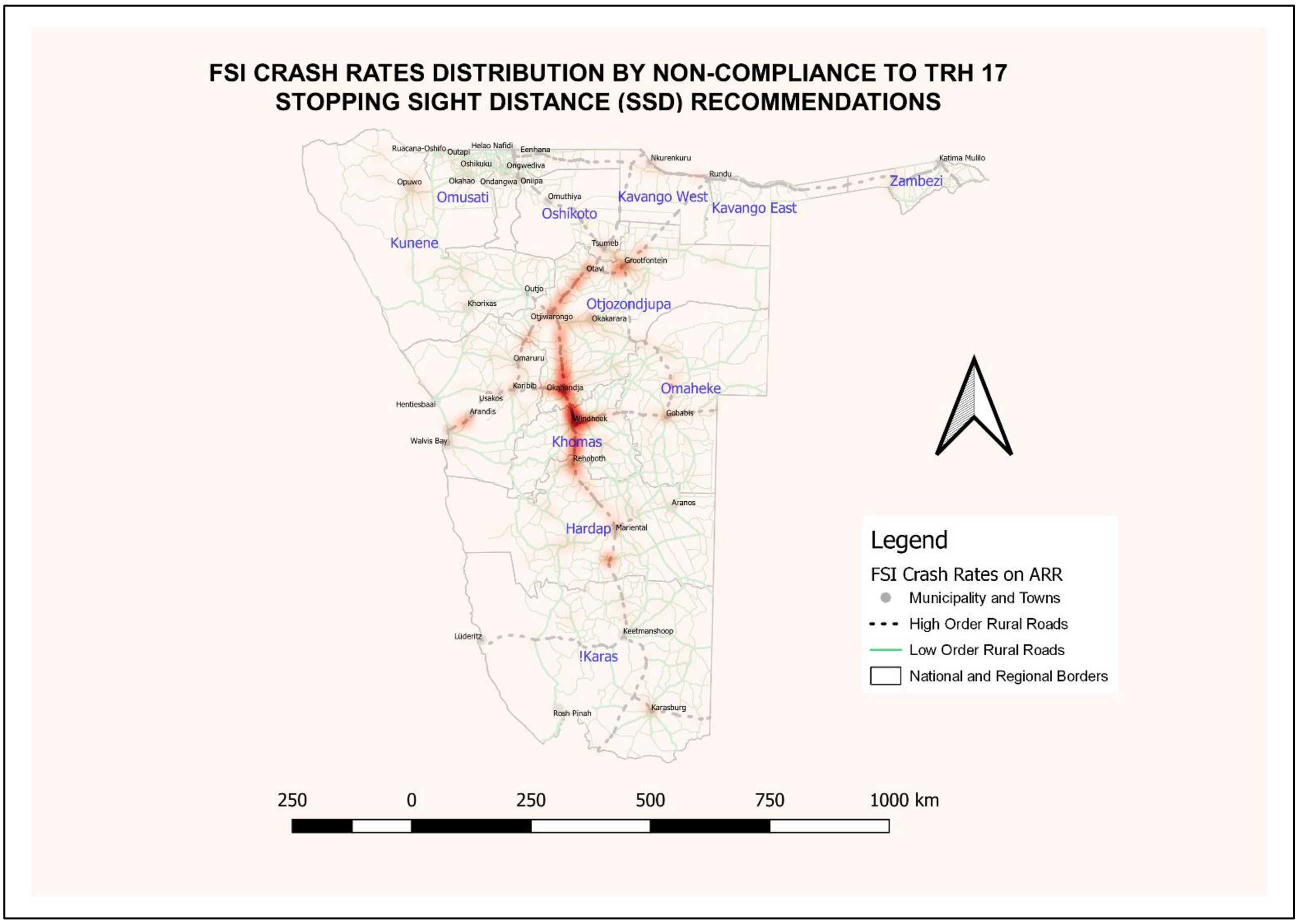


Figure 4.27 FSI crash rate distribution by non-compliance to TRH 17 SSD design recommendations

Figure 4.28 shows the impact of pavement condition (PC) non-compliance on road crash densities distribution across the national rural road network. The FSI crash densities demonstrate an extreme classification on the northern and central parts of the HORR network. Moderate to low crash densities are visible on the HORR network towards the western and northern parts from the centre. The non-compliance of the pavement condition to design guidelines was found to have a limited impact on crash occurrence on LORRs as depicted by Figure 4.28.

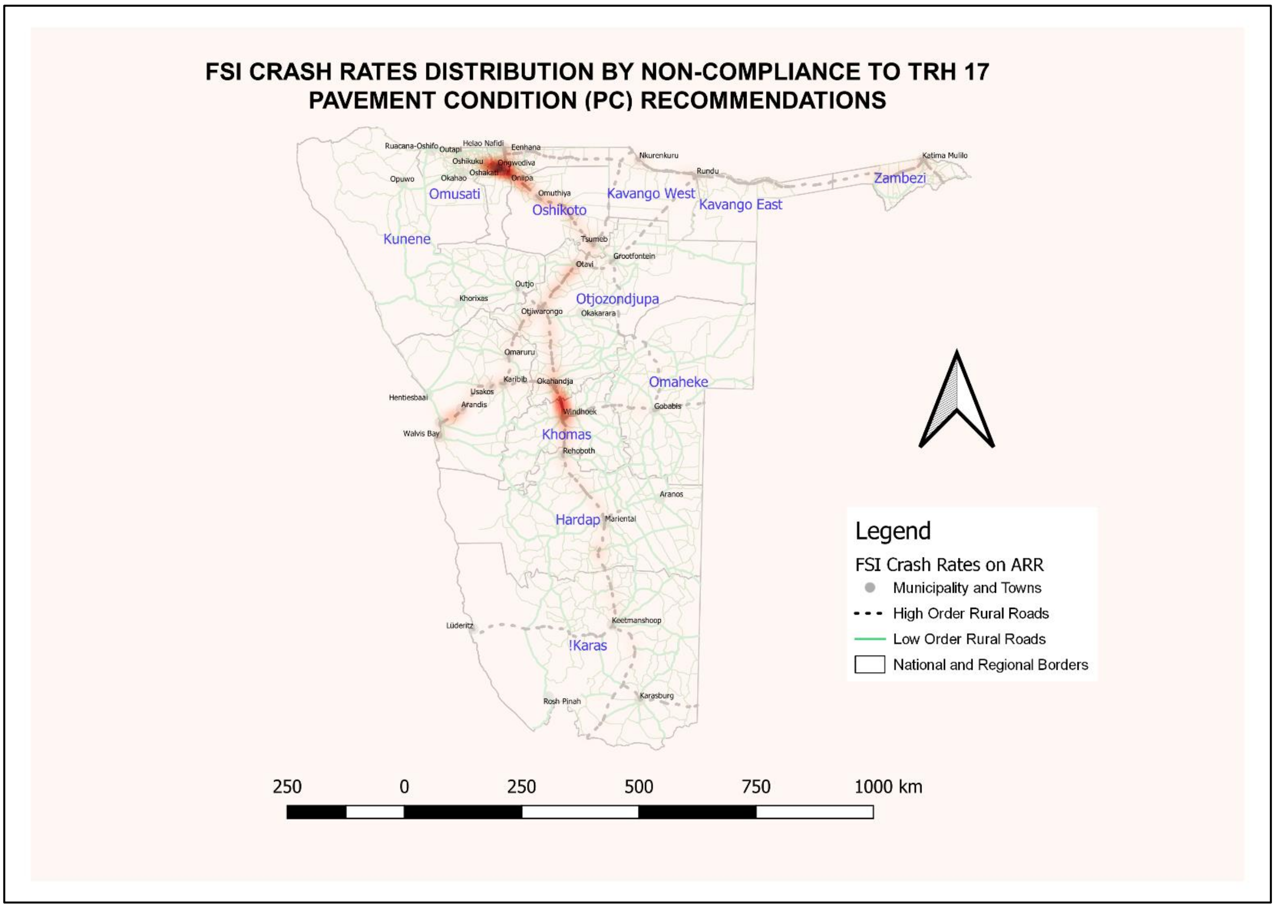


Figure 4.28 FSI crash rate distribution by non-compliance to TRH 17 PC design recommendations

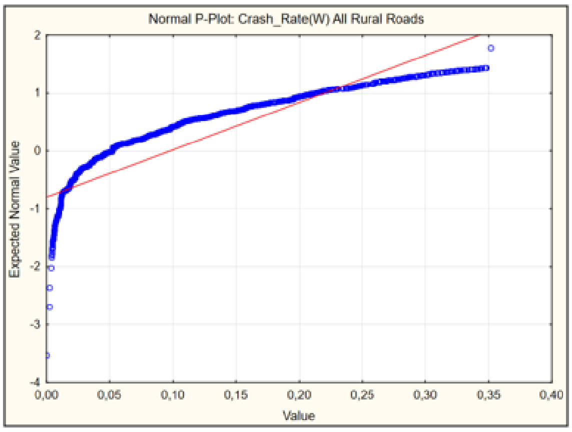
4.5 Road crash prediction model development

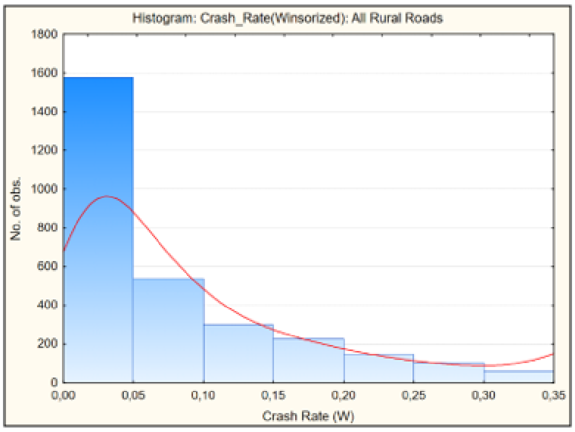
A way to improve road safety is by improving the road characteristics (design and traffic) to mitigate crash frequency and severity. In order to improve road characteristics, it is crucial to evaluate and define the relationships between road characteristics and road crashes.

For this reason, two road crash modelling techniques (General Regression Multivariate – Winsorized (MLR) and Simple Multivariate Regression – Base Mean Test Model (BMM) modelling approaches were applied in the study to develop crash prediction tools to find out the relationship between road characteristics (geometric design and condition) and road crash rates on national rural roads. The crash modelling techniques were applied on three rural road crash datasets. The first dataset encompassed all rural road (ARR) fatal and serious injury (FSIs) crashes, the second dataset encompassed FSIs on High Order Rural Roads (HORR R1-R3) and the third dataset comprised FSIs on Low Order Rural Roads (LORR R4-R6). The HORR and LORR classifications were carried out in the study using the Technical Recommendation for Highways 26 (TRH 26) on Road Classification and Access Management described in Table 2.1. The road crash models developed from the respective datasets are termed as Models 1-6 in the study.

4.5.1. Description of dependant variable

A combination of normality of distribution tests; box plots and the normal P-P plot, Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests and visual inspection of the histogram shapes were performed on the three dataset effect variables; ARR, HORR and LORR to test for normality. The results of the normality tests carried out and the descriptive statistics summaries for the three FSI datasets: ARR, HORR and LORR are described in Figures 4.29, Figure 4.30 and Figure 4.31 respectively. The normality tests indicated a normal distribution of the three datasets, confirmed by the values of the variance, which are higher than the mean values, implying that the three datasets are over-dispersed. The normality of the datasets is further confirmed by the K-S and S-W tests results. The K-S value further from one (1) is indicative of a normal distribution in the dataset ($K\text{-}S_{ARR}$ = 0.110, $K\text{-}S_{HORR}$ = 0.083 and $K\text{-}S_{LORR}$ = 0.142). Also, S-W statistic is indicative of a normal distribution when the test value is closer to one ($S\text{-}W_{ARR}$ = 0.925, $S\text{-}W_{HORR}$ = 0.927 and $S\text{-}W_{LORR}$ = 0.873). As a result, the Winsorized MLR modelling approach was identified as the appropriate method to handle the over-dispersed study datasets and to avoid overestimated standard errors and misleading inferences.





Tests of Normality: All Rural Roads						
	Kolmogorov-Smirnov (K-S)[a]			Shapiro Wilk		
ARR	Statistic	df	Sig.	Statistic	df	Sig.
	0,110	3191	0,000	0,925	3191	0,000

a. Lilliefors significance correction (p<0.01)

Summary Statistics (All Rural Roads)									
						Skewness		Kurtosis	
Valid N	Mean	Min	Max	Variance	Std. Dev	Stat	Std. Err	Stat.	Std. Err
3191	0,097	0,000	0,350	0,114	0,107	1,258	0,043	0,388	0,087

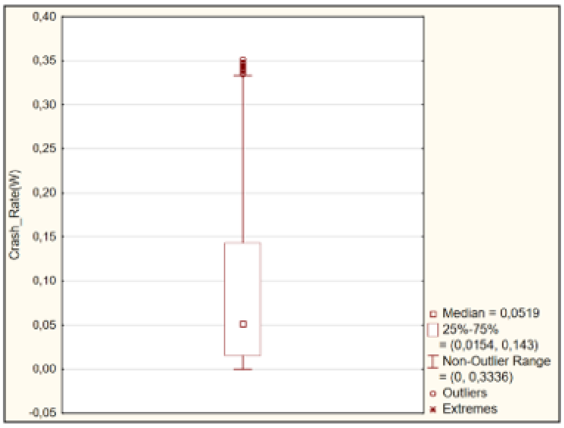
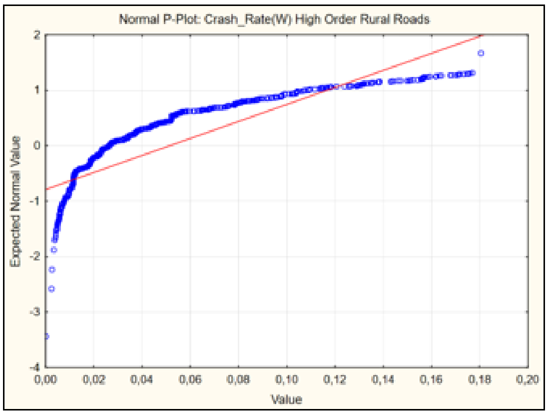
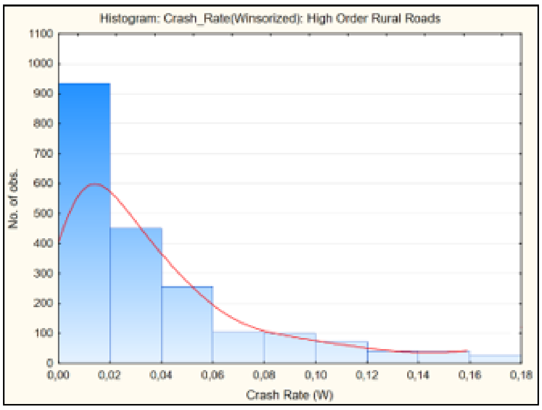


Figure 4.29 Description of All Rural Roads (ARR) output variables





Tests of Normality: High Order Rural Roads						
	Kolmogorov-Smirnov (K-S)[a]			Shapiro Wilk		
ARR	Statistic	df	Sig.	Statistic	df	Sig.
	0,083	2234	0,000	0,927	2234	0,000

a. Lilliefors significance correction (p<0.01)

Summary Statistics (High Order Rural Roads)									
						Skewness		Kurtosis	
Valid N	Mean	Min	Max	Variance	Std. Dev	Stat	Std. Err	Stat.	Std. Err
2234	0,051	0,000	0,180	0,083	0,055	1,365	0,0518	0,570	0,104

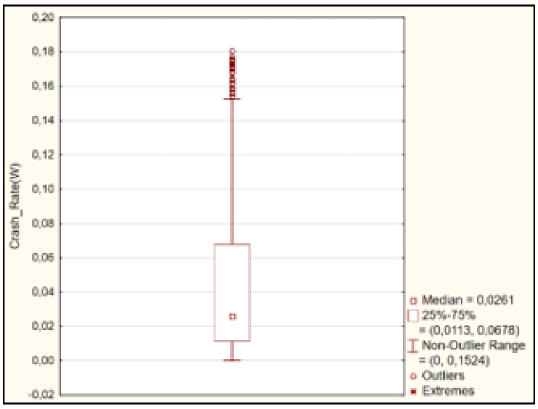
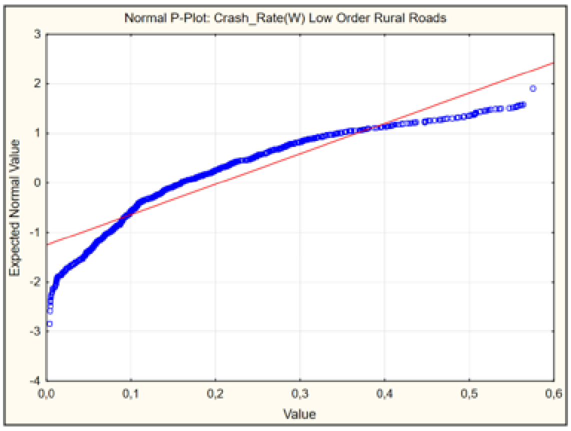
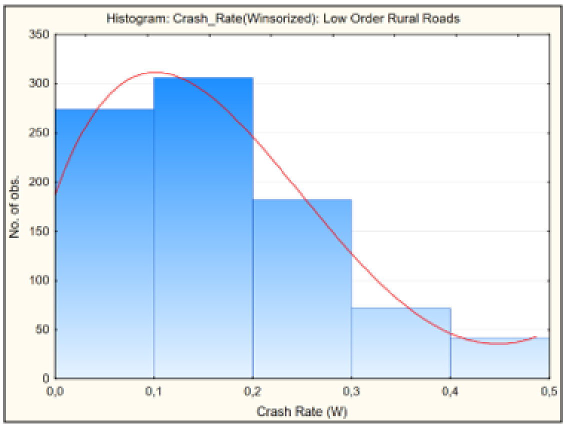


Figure 4.30 Description of High Order Rural Roads (HORR R1-R3) output variables





Tests of Normality: Low Order Rural Roads						
	Kolmogorov-Smirnov (K-S)ª			Shapiro Wilk		
ARR	Statistic	df	Sig.	Statistic	df	Sig.
	0,142	957	0,000	0,873	957	0,000

a. Lilliefors significance correction (p<0.01)

Summary Statistics (Low Order Rural Roads)									
						Skewness		Kurtosis	
Valid N	Mean	Min	Max	Variance	Std. Dev	Stat	Std. Err	Stat.	Std. Err
2234	0,203	0,0003 39	0,575	0,983	0,150	1,107	0,079	0,423	0,158

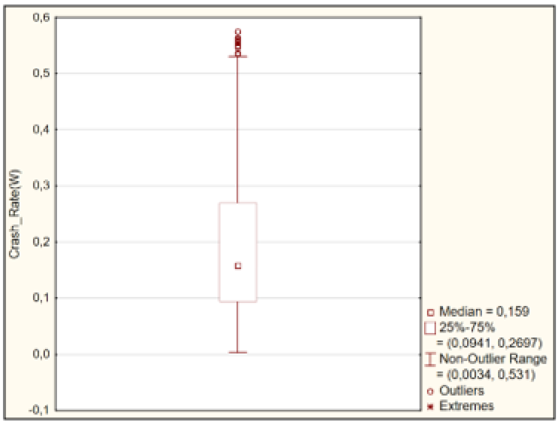


Figure 4.31 Description of Low Order Rural Roads (LORR R4-R6) output variables

4.5.2. Description of covariates

A summary of the covariates used in the model development procedure are presented in Table 4.22. A total of 16 variables were included in the model development process. The covariates are divided into two groups of variable types, numerical and categorical covariates. Of these variables, nine of the variables relate to the geometric characteristics of the rural roadway system. Seven of the variables relate to the characteristics of the rural roadway, describing the traffic modal split, terrain and roadway surface types and conditions.

Table 4.22 Summary statistics of all covariates

Descriptive statistic summary of covariates							
Covariate		Description of covariate	Min	Max	Mean	Std. Deviation	Variance
Numerical Covariates	AADT_Light	Light Vehicle Annual Average Daily Traffic of rural road section	85	14005	2328.44	2921.117	8532926.924
	AADT_Heavy	Heavy Vehicle Annual Average Daily Traffic of rural road section	2	1400	345.29	376.970	142106.055
	AADT_Total	Total Annual Average Daily Traffic (Heavy+Light Vehicles)	91	15362	2673.73	3231.762	10444282.947
	Operating_Speed (OS)	Operating Speeds on the rural road sections	0	120	44.02	53.010	2810.020
	Lane_Width (LW)	Width of rural road lanes (one-direction)	2.940	12.450	5.156	2.552	6.513
	No_Lanes (NL)	Number of lanes available to traffic on rural roads (Bi-direction)	1	6	1.79	.683	0.466
	Surface_SW (SSW)	Width of surfaced road shoulder	0.000	3.175	0.255	0.562	0.316
	Ground_SW (GSW)	Width of ground/ unsurfaced road shoulder	0.000	8.9900	1.713	0.652	0.425
	Horizontal_(Curves/Length) (Hor)	Horizontal curves per rural road km	0.000	0.709	0.176	0.143	0.020
	Access_Density (AD)	Access points per rural road km	0.000	0.409	0.121	0.086	0.007
	Section_Length (SL)	Length of rural road section	12.230	22.967	15.462	1.486	2.207
	SSD	Stopping sight distance on rural road section	15	2215	179.76	63.002	3969.274
Categorical Covariates	Surface_type (SurT)	Type of surface on road section (Paved/ Unpaved)	0	1	-	-	-
	Shoulder_type (ShoT)	Type of shoulder on road section (Paved/ Unpaved)	0	1	-	-	-
	Terrain_Vertical (TV)	Representative of vertical alignment (Flat/ Slope)	0	1	-	-	-
	Pavement_Condition	Condition of road surface	0	1	-	-	-

4.5.3. Crash Prediction Models (CPMs) results

This section presents and discusses the results of the road crash prediction models developed in the study, to uncover the relationship between the selected covariates and road crash rates on national rural roads.

4.5.3.1. Performance of CPMs: Goodness-of-fit measures

A well-fitting regression model results in predicted values close to the observed data values. The mean model, which uses the mean for every predicted value was used as the base test model (BMM). The fit of the proposed regression model should therefore be better than the fit of the mean model. Two statistics were used to evaluate the fit of the developed models: Adjusted R-squared and the overall F-test. The two goodness-of-fit tests are based on two sums of squares – Sum of Squares Total (SST) and Sum of Squares Error (SSE). SST measures how far the data is from the mean while SSE measures how far the data is from the model's predicted values. Different combinations of the SST and SSE provide different information on how the regression model compares to the mean model. A summary of the goodness-of-fit measures of the road crash models developed are presented in Table 4.23. The general regression multivariate (MLR) crash prediction models, developed using the winsorized (W) crash rate, were found to be the best fit for the datasets compared to the base models (mean models), due to the observed improvement (Adjusted R-squared) in the prediction of the crash models, and the higher statistically significant F-test value, which indicates that the observed R-squared is reliable and is not a spurious result of oddities in the study datasets. Further comparing the performance of the BMM and MLR models, the adjusted R-squared generated by the MLR models for CPM 1, CPM 2 and CPM 3 are 2.04 times higher, 2.01 times higher and 1.58 times higher than those generated by the BMMs respectively. These differences slightly increase for the F-test, with the with the MLR crash models exhibiting F-test values 2.13 times higher, 2.29 times higher and 1.61 times higher for the respective crash models. Furthermore, the full test and parameter estimates outputs for the Base Mean CPMs are provided in the Appendix C-1 (from Table C.1 to Table C.9) for comparison of the crash predictive models.

Table 4.23 Goodness-of-fit measures for all CPMs

Goodness of Fit[a]						
Parameter	Base Mean Models (BMM)			General Regression Multivariate (MLR)		
	CPM 1 (All Rural Roads)	CPM 2 (High Order Rural Roads)	CPM 3 (Low Order Rural Roads)	CPM 1 (All Rural Roads)	CPM 2 (High Order Rural Roads)	CPM 3 (Low Order Rural Roads)
Adjusted R-squared	0.21654	0.20950	0.10331	***0.44306***	***0.42078***	***0.16337***
F-test (p-value)	238.39 (0.000)	141.88 (0.000)	23.028 (0.000)	***508.22 (0.000)***	***325.14 (0.000)***	***37.142 (0.000)***

After the assessment of the performance of the base mean models (BMM) and the general regression multivariate model (MLR), the following analyses and discussions presented in the section are based on the results generated by the MLR road crash prediction models developed for the three datasets, towards revealing the relationship between national rural road characteristics and crash rates.

4.5.3.2. CPM 1 (Robust MLR) tests and parameter estimates (All Rural Roads)

The results of the Breusch-Pagan (BP) test are presented in Table 4.24. The BP test is a chi-squared test. The test statistic distributed $n\chi^2$ with k degrees of freedom. If the test statistic has a probability value (p value) below the alpha value of 0.05, that means the size of the error terms differ across the values of the model covariates. As a result, the null hypothesis of homoskedasticity is rejected and heteroskedasticity is assumed. The BP test results explain that a statistically significant ($p=0.000<0.05$) difference exists for the explanatory variables included in the crash prediction model (CPM 1) for the all rural roads dataset at 95 percent confidence interval.

Table 4.24 CPM 1 Breusch-Pagan test

Breusch-Pagan Test for heteroskedasticity		
BP	df	p-value
473.65	5	0.000

Figure 4.32 shows the plot of predicted model values with the observed dataset values. This demonstrates a visual representation of the assumed heteroskedasticity of the error terms of CPM 1.

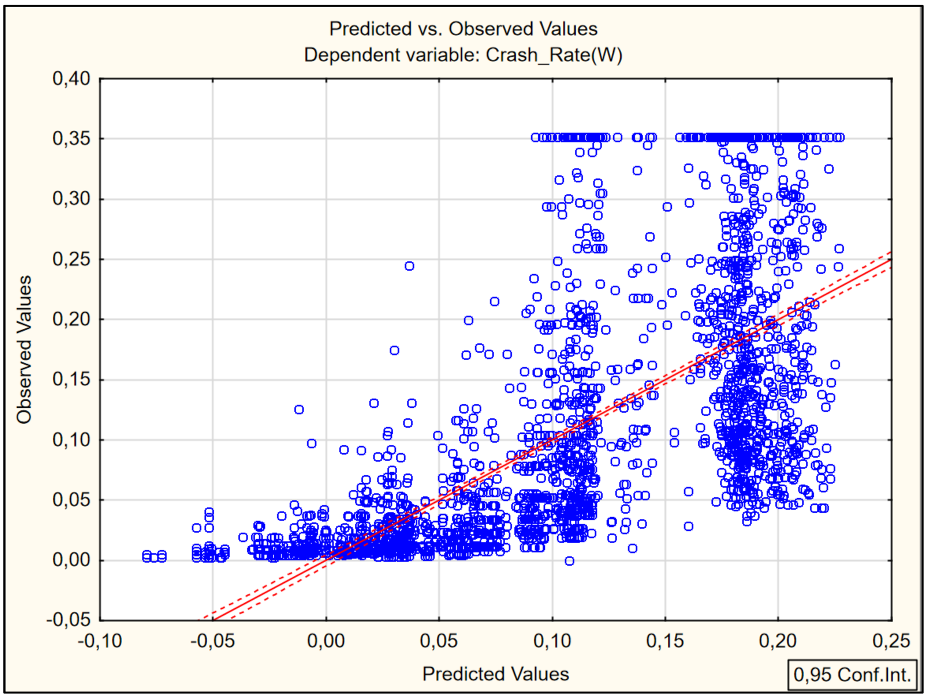


Figure 4.32 CPM 1 Predicted model values vs observed dataset values

The ARR biplot indicates the variance structure of the study variables for all the rural roads in the dataset. The biplot generated shows the projected observations (points) and the projected variables (vectors) approximated by the first two principal components (PCs) shown in Table 4.25.

Table 4.25 CPM 1 Principal Component summary

Principal Component	Eigenvalues (All rural Roads) Extraction: Principal components			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative (%)
1	3,676	26,257	3,676	26,257
2	2,240	15,999	5,916	42,256
3	1,337	9,549	7,253	51,805
4	1,221	8,721	8,474	60,525
5	1,104	7,885	9,577	68,411
6	0,865	6,180	10,443	74,590
7	0,859	6,138	11,302	80,728
8	0,714	5,102	12,016	85,830

The PCs in the biplot, graphically represented in Figure 4.33, explain the distribution and possible influence of the principle components on crash rates on all road classification – high order rural roads (HORRs) and low order rural roads (LORRs).

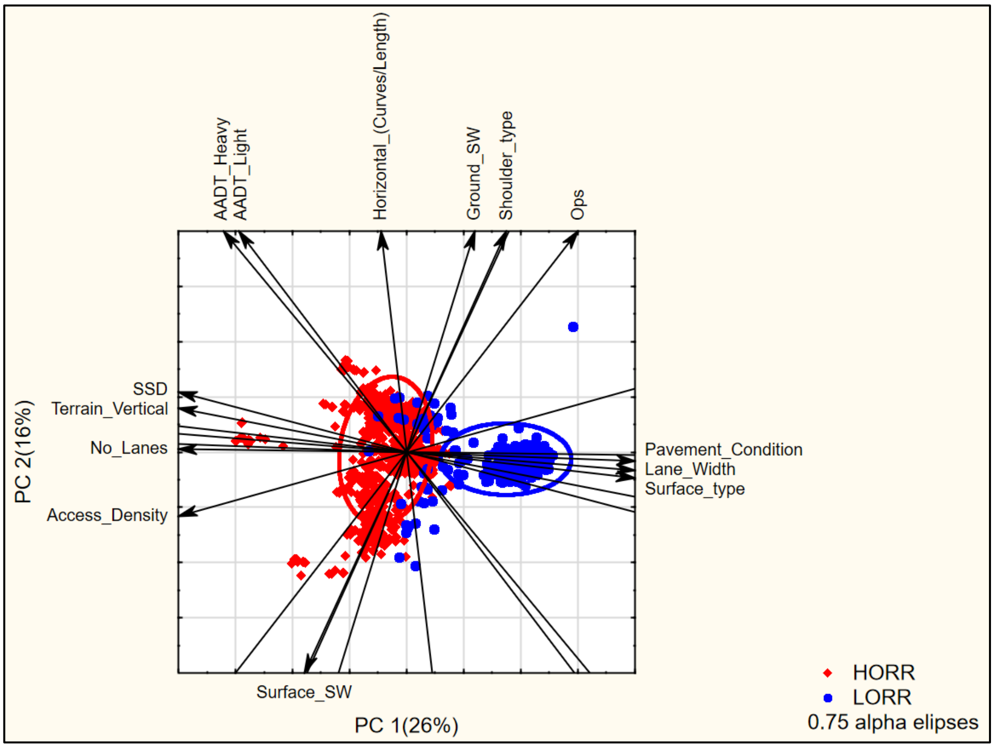


Figure 4.33 CPM 1 Principal Component biplot

As shown in Figure 4.33, the first two PCs explains 26 percent (PC1) and 16 percent (PC2) of the variance contributed by the variables on the different road classifications at an alpha elipses level of 0.75. Without factoring in autocorrelation, the biplot gives an indication of which covariates are likely to explain the correlation with crash rates. For the HORRs in the dataset, the model (CPM 1) variance in PC1 is potentially explained by:

- The widths of the surfaced shoulders

The model (CPM 1) variance for HORRs in PC 2 is potentially explained by:

- The access density
- The number of lanes on the HORR sections
- The hilliness of the vertical alignment
- The stopping sight distance (SSD) available to drivers on HORRs
- The heavy and light vehicle annual average daily traffic, and
- The number of horizontal curves per km on the road sections.

For the LORRs in the dataset, the biplot in Figure 4.33 indicates that the model (CPM 1) variance in PC 1 is potentially explained by:

- The width of the unpaved shoulder (Ground_SW) on the LORRs
- The type of hard shoulder available of the LORR sections, and
- The 85th percentile operating speed (Ops)

The variance in the model (CPM 1) for LORR sections in PC 2 is potentially explained by:

- The condition of the pavement surface on the LORRs
- The width of the available lanes, and
- The type of the surface (paved or unpaved) on the LORRs.

A detailed crash prediction model analysis for the HORRs (CPM 2) and LORRs (CPM 3) is carried out in Section 4.5.3.3 and Section 4.5.3.4 respectively.

The study applied the "best regression" developed macro in the multivariate modelling approach (MLR) to generate the best-fit crash prediction model for CPM 1. The regression coefficients for the best twenty (20) tested sub models for CPM 1 are presented in Table 4.26.

Table 4.26 Summary of best subset models for CPM 1

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Winsorized) (All Rural Roads) R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_Light	AADT_Heavy	85th Operating Speed (Ops)	Lane_Width	No_Lanes	Surface_type	Shoulder_type	Surface_SW	Ground_SW	Horizontal (C/L)	Terrain_Vertical	Access_Density	Pavement_Condition	Sight SD
3	**0,435**	**5**	**-**	**0,491**	**0,080**	**0,270**	**-**	**-**	**-**	**-0,070**	**-**	**-**	**0,083**	**-**	**-**	**-**
9	0,432	5	-	0,482	0,084	0,281	-	-	-	-	-	-	0,077	-0,040	-	-
13	0,432	5	-	0,484	0,078	0,279	-	-	0,036	-	-	-	0,081	-	-	-
18	0,432	5	-	0,492	0,080	0,278	-	-	-	-	0,034	-	0,077	-	-	-
19	0,431	5	-	0,475	0,082	0,274	-0,028	-	-	-	-	-	0,080	-	-	-
20	0,431	5	-	0,484	0,084	0,294	-	-	-	-	-	-	0,075	-	-	0,021
22	0,431	5	-	0,480	0,082	0,288	-	-	-	-	-	-0,011	0,076	-	-	-
23	0,431	5	-	0,481	0,082	0,289	-	-	-	-	-	-	0,076	-	-0,002	-
26	0,430	5	-	0,495	-	0,269	-	-	-	-0,072	-	-	0,096	-0,033	-	-
27	0,430	5	-	0,478	0,094	0,274		-	-	-0,061	-	-		-0,036	-	-
30	0,430	5	-	0,498	0,103		-	0,254	-	-0,077	-	-	0,089	-	-	-
31	0,430	5	-	0,493	-	0,274	-	-	-	-0,075	-	-0,020	0,094	-	-	-
33	0,429	5	-	0,482	0,094	0,287	-	-	-	-0,064	-	-		-	-	0,027
36	0,429	5	-	0,496	-	0,278	-	-	-	-0,074	-	-	0,094	-	-	0,013
37	0,429	5	-	0,476	0,091	0,279	-	-	-	-0,065	-	-0,022		-	-	-
38	0,429	5	-	0,496	-	0,273	-	-	-	-0,070	0,007	-	0,094	-	-	-
39	0,429	5	-	0,495	-	0,276	0,004	-	-	-0,074	-	-	0,094	-	-	-
40	0,429	5	-	0,494	-	0,275	-	-	-	-0,073	-	-	0,095	-	-0,003	-
44	0,429	5	-	0,481	0,091	0,286	0,014	-	-	-0,066	-	-		-	-	-
45	0,429	5	-	0,478	0,092	0,284	-	-	-	-0,063	-	-		-	-0,010	-

The study results identified subset 3 as the best performing crash prediction sub model (SM) for all the rural roads dataset as shown in Table 4.26. At 95 percent confidence interval, the best performing sub model (SM 3) generated shows a R-squared value of 0.435. Five covariates with varying standardised regression coefficients (b*) were identified as “best” performers and selected in the best sub model. These covariates are:

1. The proportion of heavy vehicles in the annual average daily traffic (AADT_Heavy) on the rural road network. The AADT_Heavy was identified as the best performing covariate after being selected in all the 20 sub models tested for CPM 1.
2. The width of the lanes on the rural road sections. The lane width (LW) covariate was identified as one of the covariates best explaining the relationship between crash rates and the geometric and traffic characteristics on total roads. The LW covariate was selected in nineteen (19) of the best 20 sub models generated by the MLR modelling approach.
3. The hilliness of the vertical alignment (Vertical terrain) of the entire rural road network dataset. The MLR results identified the vertical terrain covariate as the third best covariate explaining correlation to rural road crash rates, as the vertical terrain covariate was selected in fifteen (15) of the 20 best CPM 1 sub models.
4. The operating speed (Ops) on the rural road network. The MLR model results identified the speed selected by drivers on rural road sections as a covariate in the best-performing sub model in CPM 1. The operating speed covariate was selected fourteen (14) time in the sub models generated by the MLR approach.
5. The surface shoulder width (SSW) covariate. The width of the paved hard shoulders on the rural road section was selected as one of the covariates best explaining the correlation to crash rates. The SSW covariate exhibited significant correlations with crash rates in thirteen (13) of the best 20 tested sub models.

The covariates identified in the best performing sub model (SM 3) developed for CPM 1 were further investigated using the MLR modelling approach. Table 4.27 presents the parameter estimates for the final MLR road rash prediction model developed for the study, based on the entire national rural road FSI crash dataset. The best fit model comprises five (5) covariates that were found to exhibit significant effects on national rural road crash rates. The effect that the covariate has on the outcome variable is indicated by the sign and magnitude of the coefficient estimate b*. A positive coefficient b* sign implies that the covariate is associated with an increase in the rural road crash rates while a negative coefficient b* is associated with a decrease in the crash rate. All the covariates in the final crash model have exhibited effects statistically significant at an alpha level of 0.05 (5%). The adjusted R-square for CPM 1 suggests that 44.3 percent of the variance in all the rural road crash rates is accounted for by the covariates in the model. In addition, the continuous variable summary for the explanatory variables used in CPM 1 is presented in Table C.7 in Appendix C.

Table 4.27 CPM 1 Parameter Estimates

N=3189	Regression Summary for Dependent Variable: Crash_Rate(W) (All Rural Roads) R= 0.66628116; R^2= 0.44393059; Adjusted R^2= 0.44305709; CV-R^2=0.44 F (5,3183) = 508.22; p<0.0000 Std. Error of estimate = 0.07956						
	b*	Std. Err. of b*	b	Std. Err. of b	t (3183)	p-value	No. of times in best 20 SM
Intercept			0,076	0,005	15,133	0,00000	
AADT_Heavy	0,464	0,015	0,000	0,000	30,331	0,00000	20
85th Percentile Speed (Ops)	0,028	0,014	0,000	0,000	2,040	0,04147	14
Lane Width	0,293	0,016	0,012	0,001	18,206	0,00000	19
Surface_SW	-0,069	0,014	-0,013	0,003	-4,995	0,00000	13
Terrain_Vertical	0,082	0,013	0,022	0,004	6,087	0,00000	15
AADT_Light	Excluded						0
No_Lanes	Excluded						3
Surface_type	Excluded						1
Shoulder_type	Excluded						1
Ground_SW	Excluded						2
Horizontal (Curves/ length)	Excluded						3
Access_Density	Excluded						3
Pavement _Condition	Excluded						3
SSD	Excluded						3

For CPM 1, the standardised regression coefficients b* generated by the MLR modelling approach are given in Table 4.27. Using the coefficient estimates, the following can be concluded on the five (5) covariates that exhibited statistically significant associations to national rural road crash rates in STATISTICA.

- The highest absolute influence on rural road crash rates was exhibited by the proportion of heavy vehicles in the traffic streams (AADT_Heavy) on the road sections. The AADT_Heavy covariate exhibited a positive correlation (b*= 0.464) to crash rates, meaning that an increase in the heavy vehicle volume on the roads would result in an increase in the FSI crash rate.
- The width of the rural roads lanes (LW) demonstrated the second highest absolute influence on crash rates. The LW coefficient estimates (b*= 0.293) explains that an increase in the width of the lanes on the sections would result in an increase in the crash rates.
- The hilliness of the vertical alignment (vertical terrain) covariate demonstrated the third highest absolute influence of the rural road crash rates. The MLR CPM 1 generated coefficient estimate (b*= 0.082) explains that the rural road crash rate would increase as a result of an increase in the hilliness of the vertical alignment.
- The surface shoulder width (SSW) covariate demonstrated the fourth highest absolute influence on crash rates in the novel final crash prediction model for all the rural roads. The SSW covariate demonstrated a negative association (b*= -0.069) to the crash rate, meaning

that an increase in the surface of the paved shoulder width on the road sections would result in a decrease in the crash rate.

Despite the operating speed covariate performing better in SM 3 (b*= 0.080), it was found to have the fifth highest absolute influence on crash rates in the final novel model. The final model coefficient (b*= 0.028) explains that an increase in the driver speed selections on the rural roads would result in an increase in the crash rates as well.

4.5.3.3. CPM 2 (Robust MLR) tests and parameter estimates (High Order Rural Roads)

The Breusch-Pagan (BP) test results for the High Order Rural Roads (HORR) crash prediction model, referred to as CPM 2, are presented in Table 4.28. The probability value (p) of the BP tests is significantly smaller than the alpha value at 5 percent (p=0.000<0.05). For that reason, the results prove that a statistically significant different exists between the error terms of the variables included in CPM 2. The assumption of homoskedasticity is thus rejected and heteroskedasticity is assumed.

Table 4.28 CPM 2 Breusch-Pagan test

Breusch-Pagan Test for heteroskedasticity		
BP	df	p-value
396.00	5	0.000

The assumed heteroskedasticity of the error terms of CPM 2 are also visually demonstrated in Figure 4.34, which indicates the plot of the predicted and observed model (CPM 2) values.

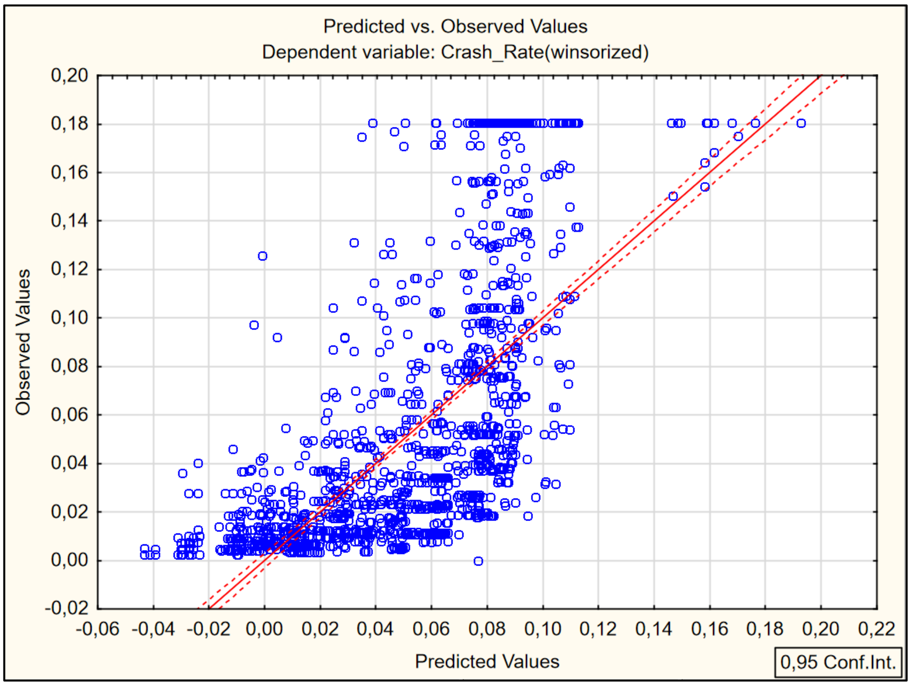


Figure 4.34 CPM 2 Predicted model values vs observed dataset values

The MLR modelling technique applied the “best regression” approach to determine the final crash prediction model. The regression coefficients for each tested sub-model for CPM 2 are presented in Table 4.29.

Table 4.29 Summary of best subset models for CPM 2

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Winsorized) (High Order Rural Roads) R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_Light	AADT_Heavy	85th Operating Speed (Ops)	Lane_Width	No_Lanes	Surface_type	Shoulder_type	Surface_SW	Ground_SW	Horizontal (C/L)	Terrain_Vertical	Access_Density	Pavement_Condition	Sight SD
12	**0,421**	**5**	**-**	**0,681**	**0,034**	**0,138**		**-**			**0,107**		**0,112**			**-**
13	0,421	5	-	0,682		0,135		-	-0,037		0,127		0,111			-
14	0,421	5	-	0,679		0,135		-		-0,032	0,093		0,119			-
15	0,421	5	-	0,689		0,139	0,030	-			0,116		0,111			
16	0,420	5	-	0,677		0,133		-			0,107	-0,021	0,114			-
17	0,420	5	-	0,681		0,135		-			0,109		0,115		-0,019	-
19	0,420	5	-	0,680		0,135		-			0,108		0,116	-0,004		-
40	0,416	5	-	0,658	0,038	0,140		-		-0,074			0,121			0,017
41	0,416	5	-	0,652		0,133		-		-0,078		-0,038	0,122			-
44	0,416	5	-	0,664		0,141	0,030	-		-0,084			0,120			-
45	0,415	5	-	0,657		0,135		-		-0,073			0,125	-0,024		-
46	0,415	5	-	0,657		0,136				-0,074			0,123		-0,019	-
78	0,412	5	-	0,645	0,037	0,139		-				-0,027	0,112			-
79	0,412	5	-	0,649	0,038	0,140		-					0,114	-0,026		-
80	0,411	5	-	0,667	0,045	0,147		-	-0,054		0,135					-
81	0,411	5	-	0,680	0,044	0,155	0,050	-			0,121					-
82	0,411	5	-	0,650	0,040	0,141		-					0,112		-0,021	-
83	0,411	5	-	0,650	0,038	0,142		-	0,018				0,116			-
84	0,411	5	-	0,679		0,149	0,044	-	-0,046		0,144					-
85	0,411	5	-	0,648	0,038	0,142	-0,000	-					0,114			-

As indicated in Table 4.29, the study results found subset 12 as the best performing sub model (SM) for CPM 2 with an R square of 0.421 at 95 percent confidence interval. Subset 12 was found to contain five (5) covariates with varying performances – standardised regression coefficients (b*). These covariates are:

1. The heavy vehicle annual average daily traffic (AADT_Heavy) on high order rural roads. The AADT_Heavy covariate performed well and exhibited statistical significance standardised regression coefficients at 95 percent confidence interval in all twenty tested CPM 2 sub models.
2. The width of the lanes (LW) on the high order rural roads sections. Similar to the AADT_Heavy, the LW covariate was also selected as a model predictor in all the 20 best performing sub models for CPM 2 at 95 percent confidence interval.
3. The vertical terrain characteristics on high order rural roads. The results indicate that the vertical terrain covariate was selected seventeen (17) times as one of the best performing predictors in the best CPM 2 test sub models developed.
4. The ground shoulder width (Ground_SW_ on the higher order road sections. The Ground_SW covariate was selected as one of the best performing predictors for the crash rate in ten (10) of the best 20 tested sub models for CPM 2.
5. The 85^{th} percentile operating speed (Ops) on higher order rural road sections. The Ops covariate was found to exhibit significant standard regression coefficients at 95 percent confidence interval in nine (9) of the best 20 tested CPM 2 sub models.

The standardised regression coefficients of the covariates in the best performing sub-model (subset 12) of CPM 2 shown in Table 4.29 were further investigated.

The parameter estimates for the crash prediction model (CPM 2) developed on the FSI dataset on High Order Rural Roads (HORR) are presented in Table 4.30. The F-test performed on the overall model linking the FSI crash rates on HORRs with the geometric design and traffic related covariates was found to be statistically significant ($p<0.05$) at 95 percent confidence interval. The adjusted R^2 for CPM 2 suggest that 42.2 percent of the variance in HORR crash rates is accounted for by the covariates in the model. The study found that five (5) of the fourteen (14) covariates tested in CPM 2 exhibited statistically significant ($p<0.05$) associations to rural road crash rates. All the five covariates in CPM 2 exhibited positive associations with crash rates on HORRs. Furthermore, the continuous covariate summary for CPM 2 is presented in Table C.8 in Appendix C.

Table 4.30 CPM 2 Parameter Estimates

N=2232	Regression Summary for Dependent Variable: Crash_Rate(W) (High Order Rural Roads) R= 0.64967267; R²= 0.42207458; Adjusted R²= 0.42077645; CV-R^2=0.42 F (5,2226) = 325.14; p<0.0000 Std. Error of estimate = 0.04218						
	b*	Std. Err. of b*	b	Std. Err. of b	t (2226)	p-value	No. of times in best 20 SM
Intercept			0,015	0,008	1,967	0,04934	
AADT_Heavy	0,682	0,017	0,000	0,000	39,491	0,00000	20
85th Percentile Speed (Ops)	0,032	0,016	0,000	0,000	1,981	0,04770	9
Lane Width	0,137	0,016	0,017	0,002	8,403	0,00000	20
Ground_SW	0,108	0,017	0,009	0,001	6,380	0,00000	10
Terrain_Vertical	0,112	0,017	0,016	0,002	6,746	0,00000	17
AADT_Light	Excluded	-	-	-	-	-	0
No_Lanes	Excluded	-	-	-	-	-	5
Surface_type	Excluded	-	-	-	-	-	0
Shoulder_type	Excluded	-	-	-	-	-	4
Surface_SW	Excluded	-	-	-	-	-	6
Horizontal (Curves/ length)	Excluded	-	-	-	-	-	3
Access_Density	Excluded	-	-	-	-	-	3
Pavement _Condition	Excluded	-	-	-	-	-	3
SSD	Excluded	-	-	-	-	-	1

The study results indicate that nine (9) covariates were excluded from the model due to their limited impact on crash rates. Using the standard regression coefficients (b*) generated by CPM 2 and presented in Table 4.30, the impact of the covariates included in the model on the crash rates explain that:

- Heavy vehicle annual average daily traffic (AADT_Heavy) on HORRs has the highest absolute influence (b*=0.682) on crash rates, with an increase in heavy vehicle traffic resulting in the crash rate increasing.
- An increase in the widths of the lanes (LW) on HORRs would result in an increase in the crash rate, with the LW covariate demonstrating the second highest influence (b*=0.137) on crash rate levels.
- The positive association (b*=0.112) between crash rate and the vertical terrain means that increasing the hilliness of the terrains in the vertical alignment would result in the crash rate on HORRs increasing.
- Increasing the ground shoulder width (GSW) would result in the HORR crash rate increasing, as shown by the positive association (b*=0.108) exhibited by the GSW covariate.

An increase in the 85th percentile operating speed on HORRs would result in the crash rate increasing, with the association demonstrated by the positive standardised coefficient (b* = 0.032).

4.5.3.4. CPM 3 (Robust MLR) tests and parameter estimates (Low Order Rural Roads)

Table 4.31 presents the results of the Breusch-Pagan test for the Low Order Rural Roads (LORRs) crash prediction model (CPM 3). The Breusch-Pagan test proves that statistically significant ($p=0.02<0.05$) differences exist between the error terms of the selected covariates in the final fitted LORR crash prediction model at 95 percent confidence interval. The assumption of homoskedasticity is thus negated and heteroskedasticity is assumed.

Table 4.31 CPM 3 Breusch-Pagan test

Breusch-Pagan Test for heteroskedasticity		
BP	df	p-value
13.13	5	0.02

Figure 4.35 shows the plot of the predicted LORR crash rate values vs the observed LORR crash rate values. This represents a visual description of the assumed statistically significant differences between the error terms of the LORR crash prediction model.

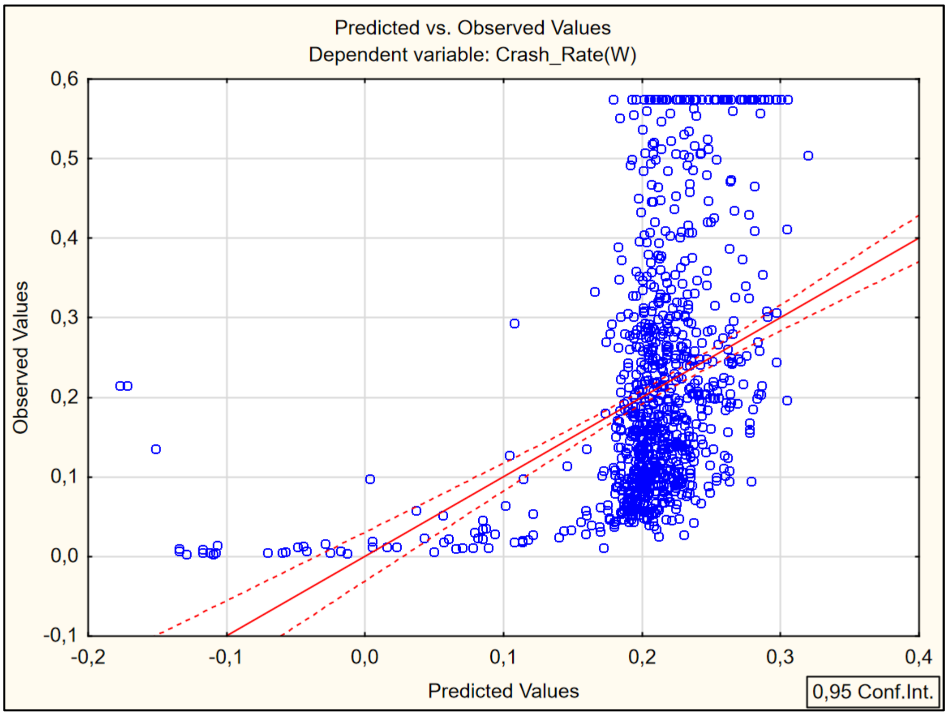


Figure 4.35 CPM 3 Predicted model values vs observed dataset values

Table 4.32 presents a summary of the best 20 sub models (SMs) generated while testing the novel LORR crash prediction model. The summary comprises R-square and standardised regression coefficient values for all the best performing covariates for the tested SMs.

Table 4.32 Summary of best subset models for CPM 3

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Winsorized) (Low Order Rural Roads) R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_Light	AADT_Heavy	85th Operating Speed (Ops)	Lane_Width	No_Lanes	Surface_type	Shoulder_type	Surface_SW	Ground_SW	Horizontal (C/L)	Terrain_Vertical	Access_Density	Pavement_Condition	Sight SD
1	**0,296**	**5**	**0,326**		**0,177**					**-0,173**	**-0,274**		**0,069**			
2	0,295	5	0,339		0,190					-0,169	-0,271	0,067				
3	0,295	5	0,280		0,166		-0,090			-0,154	-0,282					
4	0,293	5	0,359		0,208	-0,068				-0,190	-0,264					
5	0,292	5	0,323		0,184					-0,169	-0,276				0,038	
7	0,292	5	0,325		0,185					-0,172	-0,276			-0,032		
8	0,292	5	0,345		0,187				0,155		-0,262	0,068				
10	0,291	5	0,332		0,176				0,156		-0,264		0,064			
11	0,291	5	0,330		0,185					-0,173	-0,272					0,011
14	0,291	5	0,285		0,166		-0,091		0,137		-0,273					
15	0,288	5	0,330		0,182				0,155		-0,267				0,037	
16	0,288	5	0,362		0,202	-0,060			0,173		-0,255					
17	0,288	5	0,331		0,183				0,158		-0,267			-0,032		
18	0,288	5		-0,251	0,172		-0,124			-0,107	-0,270					
20	0,287	5	0,336		0,183				0,159		-0,262					0,013
23	0,284	5		-0,252	0,176		-0,128		0,080		-0,261					
24	0,284	5	0,309				-0,141			-0,192	-0,285		0,086			
25	0,284	5		-0,307	0,197					-0,123	-0,255		0,059			
26	0,283	5	0,259		0,207		-0,144				-0,250	0,077				
27	0,282	5		-0,250	0,189		-0,144				-0,246		0,059			

The study results indicate that subset (SM) 1 was found to be the best performing sub model for CPM 3 on low order rural roads, with a R square value of 0.296 at 95 percent confidence interval (CI). The best subset (SM 1) comprised five covariates with varying significance and correlation performance with the crash rates. These covariates are:

1. The ground shoulder width (GSW) on the LORR sections covariate was identified as one of the best predictors of the crash rate in the subset summary. The GSW covariate was selected as a predictor in all 20 sub models tested at 95 percent CI.
2. The operating speed (Ops) on the low order rural road was identified as one of the best contributors to the performance of sub models generated for CPM 3. The Ops covariate was selected in 19 of the best 20 sub models developed and tested for CPM 3.
3. The proportion of light vehicles in the annual average daily traffic (AADT_Light). The AADT_Light covariate was selected in 16 of the best 20 sub models, including the best performing SM 1 at 95 percent CI.
4. The width of the paved hard shoulders (SSW) covariate was identified as one of the best performing covariates in the sub models. The SSW covariate was selected in half (10) of all the sub models generated and presented in the subset summary.
5. The vertical terrain characteristics of the LORRs was identified as one of the best predictors of crash rates in the best performing sub model. The vertical terrain was only selected in 5 of the 20 sub models generated for CPM 3.

Using the MLR modelling technique, the five (5) selected covariates were investigated and the parameter estimates for the final fitted Low Order Rural Road crash prediction model (CPM 3) were generated. The LORR CPM 3 parameter estimates are presented in Table 4.33. The results show that four (4) of the 5 selected CPM 3 covariates exhibited statistically significant ($p<0.05$) correlations with crash rates at 95 percent CI. The adjusted R-square value of CPM 3 suggests that 15.9 percent of the variance in the LORR crash rates is accounted for by the covariates in the model. Of the five covariates, two covariates demonstrated a negative association to the crash rates These covariates are: (1) the surface shoulder width and (2) the ground shoulder width. The remaining three covariates; (1) the AADT_Light, (2) the operating speed and (3) the vertical terrain, exhibited positive associations to the crash rates. The continuous covariate summary for CPM 3 is presented in Table C.9 in Appendix C.

Table 4.33 CPM 3 Parameter Estimates

N=957	Regression Summary for Dependent Variable: Crash_Rate(W) (Low Order Rural Roads) R= 0.40419606; R^2= 0.16337445; Adjusted R^2= 0.15897579; CV-R^2=0.15 F (5,941) = 37.142; p<0.0000 Std. Error of estimate = 0.13791						
	b*	**Std. Err. of b***	**b**	**Std. Err. of b**	**t (2226)**	**p-value**	**No. of times in best 20 SM**
Intercept			0,325	0,017	19,588	0,00000	
AADT_Light	0,315	0,030	0,000	0,000	10,530	0,00000	16
85^{th} Percentile Speed (Ops)	0,049	0,030	0,000	0,000	1,646	0,10005	19
Surface_SW	-0,138	0,031	-0,153	0,034	-4,515	0,00001	10
Ground_SW	-0,205	0,030	-0,054	0,008	-6,739	0,00000	20
Terrain_Vertical	0,066	0,030	0,024	0,011	2,231	0,02588	5
AADT_Heavy	Excluded						4
Lane_Width	Excluded						2
No_Lanes	Excluded						7
Surface_type	Excluded						0
Shoulder_type	Excluded						8
Horizontal (Curves/ length)	Excluded						3
Access_Density	Excluded						2
Pavement _Condition	Excluded						2
SSD	Excluded						2

The MLR modelling technique excluded nine (9) of the 14 tested covariates in the final fitted novel LORR crash prediction model CPM 3. Using the standard regression coefficients b* generated by CMP 3 and presented in Table 4.33, the following is concluded about the impacts of the covariates selected for the crash prediction model on LORRs.

- The proportion of light vehicles in the annual average daily traffic (AADT_Light) exhibited the highest statistically significant ($p<0.05$) absolute influence on road crash rates. The AADT_Light coefficient estimate (b*=0.315) explains that an increase in the light vehicle volume on the low order roads would result in an increase in the crash rates.
- The second highest absolute influence on crash rates was demonstrated by the ground shoulder width (GSW) on the road sections. The GSW coefficient estimate (b*=-0.205) indicates that the road crash rates would decrease as a result of widened unpaved hard shoulders on the rural road sections.
- The width of the paved hard shoulders (SSW) covariate exhibited a statistically significant correlation to the LORR crash rates. The coefficient estimate of the SSW covariate describes that the crash rates would decrease on the road sections as a result of widening the paved hard shoulders.
- The hilliness of the vertical alignment covariate demonstrated statistically significant positive associations (b*=0.066) with the crash rates. The positive associations mean that an increase

in the hilliness of the vertical terrain would result in an increase in the crash rates on the LORRs.

- The speeds (Ops) selected by drivers on the LORRs did not exhibit any statistical significance ($p=0.100>0.05$) in predicting the crash rates. The positive coefficient estimate ($b^*=0.049$) of the Ops covariate explains that an increase in driver speed selections would result in an increase in the crash rates. Despite a lack of statistical significance, the Ops covariate was found to demonstrate some influence on the overall prediction of crash rates by CPM 3.

4.5.3.5. Evaluation of CPMs performance

The crash prediction models (CPMs) performances were assessed by evaluating the goodness-of-fit measures presented in Table 4.23. The study further examined and applied the generated standardized residual values, representing the difference between the observed and mean value predicted by the crash models developed (residuals). The study generated the residual values while fitting the CPMs to the crash dataset. These generated residuals, adopted for heteroskedasticity, were used to test the CPMs fitting performance, through determining whether the models were underestimating (positive residual value – suggests predicted value is less than observed value) or overestimating (a negative residual value- suggests predicted value is greater than observed value) the effects of design and traffic covariates, with reference to their difference from zero. For that purpose, the covariate effects (coefficient b estimates) from testing the residual values adopted for heteroskedasticity are compared to the best-fitting CPMs developed in Table 4.34. The result indicate that no marked difference exists between the covariate effects generated by the best-fitted CPMs and those generated using residuals adopted for heteroskedasticity. All the covariates exhibited the similar effects albeit slight differences in their extent. In summary, this implies satisfactory performance by the CPMs.

Table 4.34 Standardised residuals CPMs performance test

Parameter	Parameter Estimate (regression coefficient b*)					
	CPM 1: ARR	**CPM 1: adopted for heteroskedasticity**	**CPM 2: HORR**	**CPM 2: adopted for heteroskedasticity**	**CPM 3: LORR**	**CPM 3: adopted for heteroskedasticity**
Intercept	0,075639	0,075639	0,015033	0,015033	0,324966	0,324966
AADT_Heavy	0,000131	0,000131	0,000099	0,000099	-	-
85th Percentile Speed (Ops)	0,000056	0,000056	0,000033	0,000033	0,000172	0,000172
Lane Width	0,012227	0,012227	0,017089	0,017089	-	-
Surface_SW	-0,013159	-0,013159	-	-	-0,153058	-0,153058
Terrain_Vertical	0,022135	0,022135	0,016033	0,016033	0,024086	0,024086
AADT_Light	-	-	-	-	0,000040	0,000040
No_Lanes	-	-	-	-	-	-
Surface_type	-	-	-	-	-	-
Shoulder_type	-	-	-	-	-	-
Ground_SW	-	-	0,008970	0,008970	-0,054447	-0,054447
Horizontal (Curves/ length)	-	-	-	-	-	-
Access_Density	-	-	-	-	-	-
Pavement _Condition	-	-	-	-	-	-
SSD	-	-	-	-	-	-

4.5.3.6. Comparison of CPMs performance

The goodness-of-fit measures for the developed models presented in Table 4.23, evidently demonstrate that the General Regression Multivariate (MLR) model approach has the highest adjusted R-squared and significant F-test values. As a result, the MLR models were found to be the most suitable for the national rural road crash data in the study. A summarised comparison between the parameters of all crash prediction models is also provided in this section.

a) Best-fit CPMs performance: Standardised regression coefficient b*

The estimates (standardised regression coefficient b*) for all the General Multivariate Crash Prediction Models (MLR-CPMs) developed and applied in this study to fit the datasets used (All rural roads (ARR), high order rural roads (HORR) and low order rural roads (LORR)) are summarised in Table 4.35 The crash prediction models tested all the study covariates with the ARR and HORR MLR-CPMs eventually exhibiting the highest number (five covariates) of covariates showing statistically significant ($p<0.05$) relationships with crash rates. The CPM on LORR had four covariates that presented statistically significant associations with crash rates.

Table 4.35 Best-fit Road Crash Prediction Models (MLR-CPMs) performance

Parameter	Standardised Regression Coefficient (Coefficient b*)		
	CPM 1 All Rural Roads	CPM 2 High Order Rural Roads	CPM 3 Low Order Rural Roads
AADT_Heavy (AADTH)	0,464	0.682	-
85th Percentile Speed (Ops)	***0,028***	***0,032***	***0,049***
Lane Width (LW)	0,293	0,137	-
Surface_SW (SSW)	-0,069	-	-0,138
Terrain_Vertical (TV)	***0,082***	***0,112***	***0,066***
AADT_Light (AADTL)	-	-	0,315
No_Lanes (NL)	-	-	-
Surface_type (ST)	-	-	-
Shoulder_type (ShoT)	-	-	-
Ground_SW (GSW)	-	0,108	-0,205
Horizontal (Curves/ length) (Hor)	-	-	-
Access_Density (AD)	-	-	-
Pavement _Condition (PC)	-	-	-
SSD	-	-	-

Of the fourteen covariates tested in the models, seven (7) different covariates were found to exhibit statistically significant ($p<0.05$) effects in the CPMs. The following seven covariates tested in all the crash prediction models did not show any statistically significant affiliation with the crash rates on the rural roads: (1) the number of lanes available to traffic (NL); (2) the type of surface on the different road classifications (ST); (3) the proportion of hard shoulder surfaces (ShoT); (4) the number of horizontal curves per km length of rural road section (Hor); (5) the number of access points per km road section length (AD); (6) the condition of the pavement surface (PC); and (7) the stopping sight distance available on the rural road sections (SSD)

Two (2) covariates were shown to be influential ($p<0.05$) in all three of the CPMs, with varying effects on the crash rates. These covariates are: (1) the operating speed (Ops) on the road sections and (2) the vertical terrain – hilliness (TV) (all indicated in blue in Table 4.35). The 85th percentile operating speed exhibited positive association with crash rates in all the CPMs, with coefficient estimates ranging from 0.028 to 0.049. Similar to the operating speed, the hilliness of the vertical alignment exhibited a positive association to crash rates on all rural road classifications. The coefficient estimates generated by the CPMs for the hilliness covariate range from 0.066 to 0.122.

Two covariates exhibited statistically significant ($p<0.05$) parameter estimates in both CPM 1 and CPM 2. These covariates are: (1) the proportion of heavy vehicles in the annual average daily traffic (AADTH), and (2) the width of the rural road lanes (LW). Of the two covariates in CPM 1 and CPM 2, the proportion of heavy vehicles in the traffic stream (AADTH) covariate showed positive associations with the crash rates. The CPM 1 and CPM 2 standardised regression coefficient b* for the AADTH covariate are 0.464 to 0.682 respectively. In CPM 1, for all rural roads, the model results indicate that the lane widths (LW) are positively related (b*= 0.293) to the crash rates. In the same way, on higher order rural roads on the road network, the lane width (LW) covariate demonstrated a positive statistically significant correlation (b*= 0.137) to crash rates.

Two other covariates emerged with statistically significant coefficient b* estimates in two of the three developed crash prediction models: (1) the surfaced shoulder width (SSW) and (2) the ground shoulder width (GSW) on the road sections. The SSW was identified to have a negative relation with crash rates in CPM 1 (b*= -0.069) and CPM 3 (b*= -0.138). The GSW demonstrated significant relations to the crash rates in both CPM 2 and CPM 3. In CPM 2, the GSW covariate indicated a positive relation to crash rates, with a 0.108 coefficient b* estimate. In contrast, the GSW covariate in CPM 3 indicated a negative relation to crash rates, with a -0.205 coefficient b* estimate.

The proportion of light vehicles in the annual average daily traffic (AADTL) covariate was found to only exhibit statistically significant relations with crash rates in CPM 3. In CPM 3, the AADTL has a positive relation coefficient b* of 0.315 to rural road crash rates.

4.6 Impact of compliance with rural road design guidelines on developed Crash Prediction Models (CPMs)-Sensitivity Test

The Impact of compliance of the national rural roads to the Technical Recommendations for Highways 17 on the Geometric Design of Rural Roads (TRH17), the Technical Recommendations for Highways 20 on the Structural Design, Construction and Maintenance of Unpaved Roads (TRH 20) and Technical Recommendations for Highways 26 on Road Classification and Access Management (TRH 26) on the crash rates were tested to learn of the sensitivity of the parameter estimates to the road characteristic changes. For that reason, three additional models were included in the study analysis. One model was developed to test the sensitivity of the crash rates on all the rural roads (CPM 4). The other model was developed to test the sensitivity on high order rural roads (CPM 5) while the last model was developed to test the sensitivity of crash rates on low order rural roads (CPM 6). The sensitivity analysis intends to test possible mediating effects of road design variables

The three additional models were developed using the General Multivariate (MLR) crash predictive modelling approach with reference to the 16 covariates tested in CPMs 1, CPM 2 and CPM 3, to allow for a better basis for comparison of parameter estimates. All covariates in the MLR-CPMs were adjusted to meet the TRH 17, TRH 20 and TRH 26 minimum requirements. The coefficient b* estimates for the statistically significant ($p<0.05$) covariates are demonstrated and compared in Table 4.36 (ARR), Table 4.37 (HORR) and Table 4.38 (LORR) in the sections below.

4.6.1. Impact of compliance (CPM 4) on CPM 1 (All Rural Roads)

Table 4.36 presents the parameter estimate (coefficient b*) results of the road design compliance sensitivity analysis. Similar to CPM 1, CPM 4 also generated five (5) statistically significant ($p<0.05$) covariates. The results show a slightly improved adjusted R-square value (CPM $1_{adj\ R\text{-}sq.}= 0.443$; CPM $4_{adj\ R\text{-}sq.}= 0.476$) in the models due to compliance. As a result of compliant covariates, one covariate demonstrated a change in effect on the outcome variable in the sensitivity test results in CPM 4. This covariate is:

- The proportion of heavy vehicles in AADT ($b^*_{MLR\text{-}CPM\ 1}= 0.464$ to $b^*_{MLR\text{-}CPM\ 4}= -0.380$);

An increase in the contribution to the outcome variable with reference to the magnitude of coefficient b* is only apparent in one significant covariate This variable is:

- The operating speed on the rural road sections ($b^*_{MLR\text{-}CPM\ 1}= 0.028$ to $b^*_{MLR\text{-}CPM\ 4}= 0.036$);

In the same way, one covariate has a decreasing influence on the ARR crash rates when the road characteristics are compliant with TRH 17 and TRH 26 guidelines. This covariate is:

- The vertical terrain on ARRs ($b^*_{MLR\text{-}CPM\ 1}$= 0.082 to $b^*_{MLR\text{-}CPM\ 4}$= 0.076);

It is evident in the model results that the highest positive association between the outcome variable and a covariate was demonstrated by the proportion of heavy vehicles in the AADT ($b^*_{MLR\text{-}CPM\ 1}$= 0.464) covariate in CPM 1 (existing road characteristics), while in CPM 4 (compliant with TRH 17 and TRH 26), the proportion of paved shoulders ($b^*_{MLR\text{-}CPM\ 1}$= 0.378) generated a significantly higher coefficient b* estimate on ARRs. In contrast, the magnitude of the speed selection covariates contribution to the outcome variable is the lowest in both ARR CPMs ($b^*_{MLR\text{-}CPM\ 1}$= 0.028; $b^*_{MLR\text{-}CPM\ 4}$= 0.036), with a lower positive parameter estimate value in CPM 1.

The lane width ($b^*_{MLR\text{-}CPM\ 1}$= 0.293) and the width of the paved hard shoulders ($b^*_{MLR\text{-}CPM\ 1}$= -0.069) covariates which reflected statistically significant positive and negative associations with the outcome variable in CPM 1 respectively, failed to influence the crash rates in CPM 4 as a result of road compliance with TRH 17 and TRH 26.

In a different way, the ground shoulder width ($B_{GP\text{-}CPM\ 4}$= 0.369) and the proportion of paved shoulders ($b^*_{MLR\text{-}CPM\ 1}$= 0.378) reflect a significant association with the crash rates when compliance is tested in CPM 4. However, the significant association exhibited by the ground shoulder width and proportion of paved shoulders covariates in CPM 4 is absent in CPM 1, when existing road characteristics are tested. For comparison purposes, the full performance tests and parameter estimates outputs for CPM 4 are provided in Appendix C-3.

Table 4.36 Sensitivity test on parameter estimates to road design guidelines (Comparing CPM 1-CPM 4)

Parameter	Parameter Estimate (Coefficient b*)	
	CPM 1 All Rural Roads (Existing Road Characteristics)	CPM 4 All Rural Roads (TRH 17 & TRH 26 Compliant Road Characteristics)
AADT_Heavy (AADTH)	0,464	-0.380
85^{th} Percentile Speed (Ops)	0,028	0.036
Lane Width (LW)	0,293	-
Surface_SW (SSW)	-0,069	-
Terrain_Vertical (TV)	0,082	0.076
AADT_Light (AADTL)	-	-
No_Lanes (NL)	-	-
Surface_type (ST)	-	-
Shoulder_type (ShoT)	-	0.378
Ground_SW (GSW)	-	-0.078
Horizontal (Curves/ length) (Hor)	-	-
Access_Density (AD)	-	-
Pavement _Condition (PC)	-	-
SSD	-	-

4.6.2. Impact of compliance (CPM 5) on CPM 2 (High Order Rural Roads)

The estimates for the road design guidelines (TRH 17 and TRH 26) compliance sensitivity analysis of the crash prediction model (CPM) on high order rural roads (HORR) are presented in Table 4.37. The crash prediction model tested with the compliant road design characteristics on HORRs (CPM 5) generated five (5) covariates with significant effects on the outcome variable. In the same way, the same number (5) of covariates demonstrated significant effects on crash rates on the CPM developed with existing road characteristics on HORRs (CPM 2). The CPM developed with compliant design characteristics (CPM $5_{adj\ R\text{-}sq.}$= 0.445) demonstrated an improved adjusted R-square value compared to the CPM with existing rural road characteristics (CPM $2_{adj\ R\text{-}sq.}$= 0.421).

In the crash prediction model developed with existing road characteristics, the proportion of heavy vehicles in the AADT on the road section ($b^*_{MLR\text{-}CPM\ 2}$= 0.682) reflected the highest absolute value of coefficient b*. The same covariate (heavy vehicles in AADT) demonstrated the highest absolute coefficient b* value in CPM 5 ($b^*_{MLR\text{-}CPM\ 5}$= -0.594). In contrast to the association with the crash rates demonstrated in CPM 2, the heavy traffic AADT covariate showed an opposite signed effect on HORR crash rates in CPM 5.

Two (2) of the covariates reflected an increased effect on the output variable after the design guideline compliance test, with reference to the estimate value of coefficient b*. These covariates are:

- The operating speed on HORRs ($b^*_{MLR\text{-}CPM\ 2}$= 0.032 to $b^*_{MLR\text{-}CPM\ 5}$= 0.041); and.
- The vertical terrain on the HORRs ($b^*_{MLR\text{-}CPM\ 2}$= 0.112 to $b^*_{MLR\text{-}CPM\ 5}$= 0.120).

The sensitivity analysis results presented in Table 4.37, indicate that two covariates that are statistically significant in influencing the crash rates on HORRs with existing road characteristics, do not influence the outcome variable in the model developed with road design compliant road characteristics. These covariates are:

- The lane width on high order rural roads ($b^*_{MLR\text{-}CPM\ 2}$= 0.137); and
- The ground shoulder width on high order rural roads ($b^*_{MLR\text{-}CPM\ 2}$= 0.108).

As a result of compliant road design characteristics, the proportion of paved shoulders ($b^*_{MLR\text{-}CPM\ 5}$= 0.234) and the number of horizontal curves per rural road length ($b^*_{MLR\text{-}CPM\ 5}$= -0.033) covariates demonstrated statistically significant effects on the crash rates on HORRs. This significant association is however absent in the model (CPM 2) tested using existing rural road characteristics on high order roads. For comparison purposes, the full performance tests and parameter estimates for CPM 5 are provided in Appendix C-3.

Table 4.37 Sensitivity test on parameter estimates to road design guidelines (Comparing CPM 2-CPM 5)

Parameter	Parameter Estimate (Coefficient b*)	
	CPM 2 High Order Rural Roads (Existing Road Characteristics)	CPM 5 High Order Rural Roads (TRH 17 & TRH 26 Compliant Road Characteristics)
AADT_Heavy (AADTH)	0.682	-0.594
85^{th} Percentile Speed (Ops)	0.032	0.041
Lane Width (LW)	0.137	-
Surface_SW (SSW)	-	-
Terrain_Vertical (TV)	0.112	0.120
AADT_Light (AADTL)	-	-
No_Lanes (NL)	-	-
Surface_type (ST)	-	-
Shoulder_type (ShoT)	-	0.234
Ground_SW (GSW)	0.108	-
Horizontal (Curves/ length) (Hor)	-	-0.033
Access_Density (AD)	-	-
Pavement _Condition (PC)	-	-
SSD	-	-

4.6.3. Impact of compliance (CPM 6) on CPM 3 (Low Order Rural Roads)

Table 4.38 presents the sensitivity analysis of the crash prediction model (CPM) parameter estimates on low order rural roads (LORR) to changes in compliance with TRH 17 and TRH 26 design guidelines. The crash prediction model for LORRs using design compliant parameters (CMP $6_{adj\ R\text{-}sq.}$=0.386) showed a markedly high improvement due to compliance compared to the model with existing road characteristics (CPM $3_{adj\ R\text{-}sq.}$=0.159), as indicated by the adjusted R-square values of the respective models. In response to changes in design compliance, the crash prediction model developed for LORRs generated three (3) statistically significant covariates (CPM 6), compared to the four (4) significant covariates generated by the developed CPM 3 using the existing road characteristics.

The model results indicate that the proportion of light vehicles in the AADT ($b^*_{MLR\text{-}CPM\ 3}$= 0.315) demonstrated the highest absolute influence (coefficient b* estimate) on the outcome variable in the LORR CPM 3. As a result of road characteristic compliance, the light vehicle AADT $b^*_{MLR\text{-}CPM\ 6}$= -0.204) covariate exhibited a reduced and opposite signed association to crash rates in CPM 6. On the other hand, the model results indicate that the ground shoulder width ($b^*_{MLR\text{-}CPM\ 3}$= -0.205; $b^*_{MLR\text{-}CPM\ 6}$= -0.412) covariate showed and increased coefficient b* estimate and exhibited the highest absolute influence on crash rates in CPM 6. In the same way, an increased influence on crash rates, though not statistically significant (p>0.05), is demonstrated by the vertical terrain ($b^*_{MLR\text{-}CPM\ 3}$= 0.062; $b^*_{MLR\text{-}CPM\ 6}$= 0.086) as a result of compliance to road design guidelines.

The model results also indicate that the operating speed ($b^*_{MLR\text{-}CPM\ 3}$= 0.049), which did not demonstrate a statistically significant coefficient estimate, and the surface shoulder width ($b^*_{MLR\text{-}CPM\ 3}$= -0.138) covariates lose their statistical significance in influencing the crash rates, as a consequence of compliant road characteristics to guidelines. In contrast, as a result of compliance to guidelines, the proportion of paved shoulder ($b^*_{MLR\text{-}CPM\ 6}$= 0.241) and stopping sight distance ($b^*_{MLR\text{-}CPM\ 6}$= 0.081) on LORRs demonstrated positive associations to crash rates. The paved shoulder significant association to crash rates was not recognised by the CPM developed for existing road characteristics on LORRs. Despite the stopping sight distance exhibiting some influence on the crash rates in CPM 6, it was however found to be statistically insignificant (p>0.05). The full performance tests and parameter estimates for CPM 6 are presented in Appendix C-3.

Table 4.38 Sensitivity test on parameter estimates to road design guidelines (Comparing CPM 3-CPM 6)

Parameter	Parameter Estimate (Coefficient b*)	
	CPM 3 Low Order Rural Roads (Existing Road Characteristics)	CPM 6 Low Order Rural Roads (TRH 17 & TRH 26 Compliant Road Characteristics)
AADT_Heavy (AADTH)	-	-
85^{th} Percentile Speed (Ops)	0.049	-
Lane Width (LW)	-	-
Surface_SW (SSW)	-0.138	-
Terrain_Vertical (TV)	0.066	0.086
AADT_Light (AADTL)	0.315	-0.204
No_Lanes (NL)	-	-
Surface_type (ST)	-	-
Shoulder_type (ShoT)	-	0.241
Ground_SW (GSW)	-0.205	-0.412
Horizontal (Curves/ length) (Hor)	-	-
Access_Density (AD)	-	-
Pavement _Condition (PC)	-	-
SSD	-	0.081

4.7 Driver characteristics and risk factors – roadway condition analysis models (TSC Model): The synergy

This section presents a novel undertaking to investigate the interaction between the most probable combination of risk factors (see Table 4.18) and the demographic, temporal (see Section 4.2.1), and roadway and environmental elements (see Section 4.5.2), which are also applied in the development of crash prediction models in the study. The study applied the Two-Step Cluster (TSC) analysis method to develop a model that identifies covariate combinations (clustered) with an impact on the types and distribution of risk factors across the national rural road network. The covariates used for the TSC analysis are presented in Table 4.39. This section signifies the importance that all these covariates together determine how the national rural road environment is perceived and what driver behaviour is elicited in response. Therefore, it supports the importance of considering the road environment as a whole when investigating the its effect on road safety, reinforced by demographics and temporal data, without isolating single design and traffic covariates.

Table 4.39 Descriptive statistics variables included in the models

Variable Name	Mean (Standard Deviation)
Risk factors Combination Variables -Estimated variables	
Risk factor combination (*See Table 4.2*): ***Recognition error = 1*** ***Decision error = 2*** ***Performance error = 3*** ***Intentional error = 4*** ***Physiological error = 5*** ***Roadway and environmental = 6*** ***Vehicle factor = 7***	-
Demographic and Temporal Explanatory Variables	
Night/ Unlit Indicator (1 if true, 0 otherwise)	0.33(0.469)
Dawn/dusk indicator (1 if true, 0 otherwise)	0.08(0.266)
Young driver indicator (1 if driver ≤ 25 years old, 0 otherwise)	0.10(0.301)
Male driver indicator (1 if male, 0 otherwise)	0.85(0.355)
Weekday indicator (1 if Monday-Thursday, 0 otherwise)	0.46(0.498)
Weekend indicator (1 if Friday-Sunday, 0 otherwise)	0.54(0.498)
Road & Traffic Explanatory Variables	
Narrow lane width indicator (1 if LW<3.2 m, 0 otherwise) – Paved (1 if LW <8 m, 0 otherwise) – Unpaved	0.08(0.277)
Wider lane width indicator (1 if LW>3.5 m, 0 otherwise) – Paved (1 if LW> 10 m, 0 otherwise) - Unpaved	0.17(0.373)
Narrow shoulder width indicator (1 if SW< 1.5 m, 0 otherwise)	0.95(0.212)
Wider shoulder width indicator (1 if SW>2.1 m, 0 otherwise)	0.03(0.172)
No overtaking/ crossing line indicator (1 if true, 0 otherwise)	0.42(0.493)

Unpaved road indicator (1 if road unpaved, 0 otherwise)	0.31(0.463)
Unpaved shoulder indicator (1 if shoulder unpaved, 0 otherwise)	0.82(0.383)
Poor road surface condition indicator (1 if poor, 0 otherwise)	0.29(0.452)
Poor sight distance Indicator (1 if insufficient, 0 otherwise)	0.23(0.420)
Two road lane indicator (1 if true, 0 otherwise)	0.68(0.467)
High AADT Indicator (1 if AADT > 2 000, 0 otherwise)	0.39(0.489)
Low density of horizontal curves (HC/km) indicator (1 if HC < 0.35 HCs/km, 0 otherwise)	0.89(0.313)
Flat terrain indicator (1 if true, 0 otherwise)	0.81(0.389)
High Access Density (AD/km) Indicator (1 if AD >0.21, 0 otherwise)	0.13(0.337)
High operating speed (OS) indicator (1 if OS > 85th percentile speed, 0 otherwise)	0.01(0.110)

A list of the coded crash risk factor combinations is presented in Table D.1 in Appendix D. Out of a possible 343 possible risk factor combinations, the study identified a total of 93 crash risk factor combinations from the crash dataset for the development and analysis of Two-Step Cluster Models. The identified risk factor combinations are presented in Table D.2 in Appendix D, with five frequently occurring risk factor combinations highlighted. These risk factor combinations are:

1. The combination of recognition, decision and intentional risk factors – code 2 (7 percent)
2. The combination roadway and environmental, and recognition risk factors – code 90 (6 percent)
3. The combinations of a recognition and decision risk factors – code 33 (5.6 percent)
4. The combination of a recognition, decision and roadway and environmental risk factor – code 4 (5.2 percent), and
5. The combination of two intentional risk factor errors and a recognition error – code 78 (4.1 percent).

4.7.1 The Two-Step Cluster (TSC) Combination Model

In the development of the TSC combination model, the study sought to investigate how demographic, temporal and road and traffic characteristics impact the combination of risk factors on the national rural roads. The study tested 21 covariates (see Table 4.39) in the initial model (TSC-1) development attempt. Only six (6) variables generated a predictor importance value above the model threshold of 0.4. The 6 covariates were used to develop a final recalibrated model (TSC-2), with 3 recognised cluster groups, for better performance in identifying the factors that affect crash risk factor combinations. The auto-clustering for TSC-2 presented in Table 4.40, summarises the process by which the number of clusters were generated and chosen by the Two-Step Clustering model.

Table 4.40 TSC-2 Auto-Clustering Parameters

TSC- 2 Model Auto-Clustering				
Number of Clusters	Akaike's Information Criterion (AIC)	AIC Change[a]	Ratio of AIC Changes[b]	Ratio of Distance Measures[c]
1	3 011.761			
2	1 099.495	-3 512.267	1.000	1.821
3 (TSC-2)	568.073	-1 225.600	0.349	2.264
4	671.600	-1 005.895	0.286	1.364

a. The changes are from the previous number of clusters in the table.

b. The ratios of changes are relative to the change for the two-cluster solution.

c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

The Akaike's Information Criterion was computed for each of the number of clusters. Smaller AIC values indicate the better cluster model. Furthermore, the Ratio of AIC Changes (RAICC) and Ratio of Distance Measures (RDM) are evaluated to determine the best cluster solution. The "best" cluster solution will have a reasonably large RAICC and a large RDM. As presented in Table 4.40, the three (3) cluster solution exhibited the smallest AIC value of 568.073. Also indicative of the good solution provided by the TSC in the study, the three-cluster solution exhibited the largest ratio of distance measure (2.264) and a reasonably large ratio of change (0.349) with respect to the change at the two clusters, applying cluster two solution as the base cluster.

The quality of TSC-2 is further illustrated in Figure 4.36, in which a comparison between TSC-1 and TSC-2 is presented. The cluster quality for TSC-1 fell within the "fair" value of the Silhouette measure (SM). After the recalibration and removal of covariates with a threshold value less than 0.4, the new model (TSC-2) showed an improved cluster quality, with a "good" Silhouette quality measure.

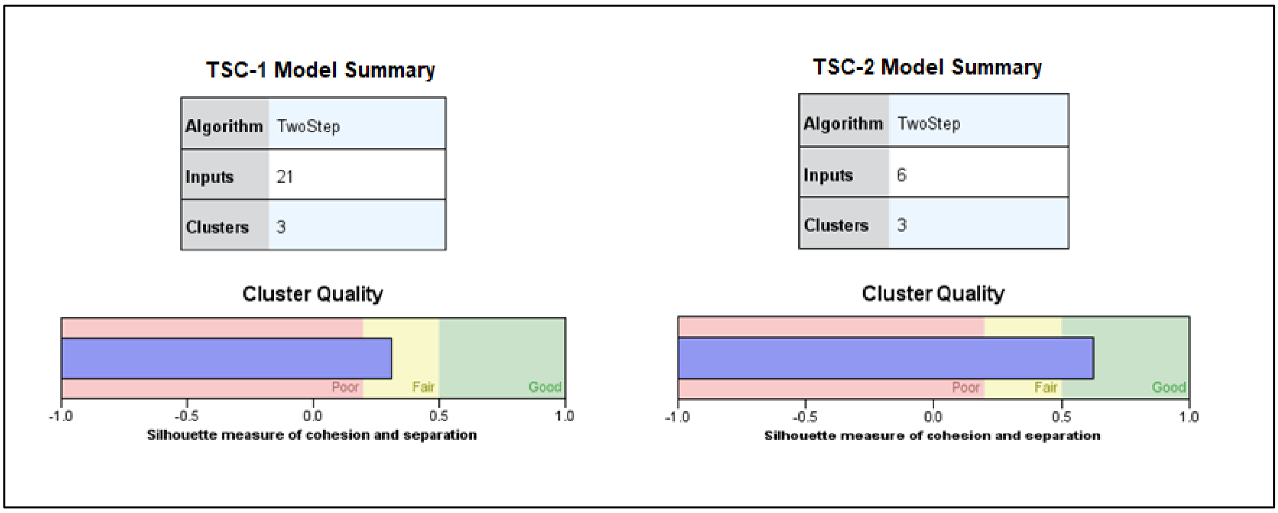


Figure 4.36 Cluster quality of the TSC models

The results indicate that three cluster sizes were identified by the TSC-2 model. The ratio of the largest cluster group to the smallest cluster group is 1.21 (see Figure 4.37), which lies between 1 to 3. This is indicative of good cluster groupings. The largest cluster grouping represented 36.7 percent of all the crash records analysed, with the smallest cluster grouping representing a slightly less 30.3 percent of the crash records.

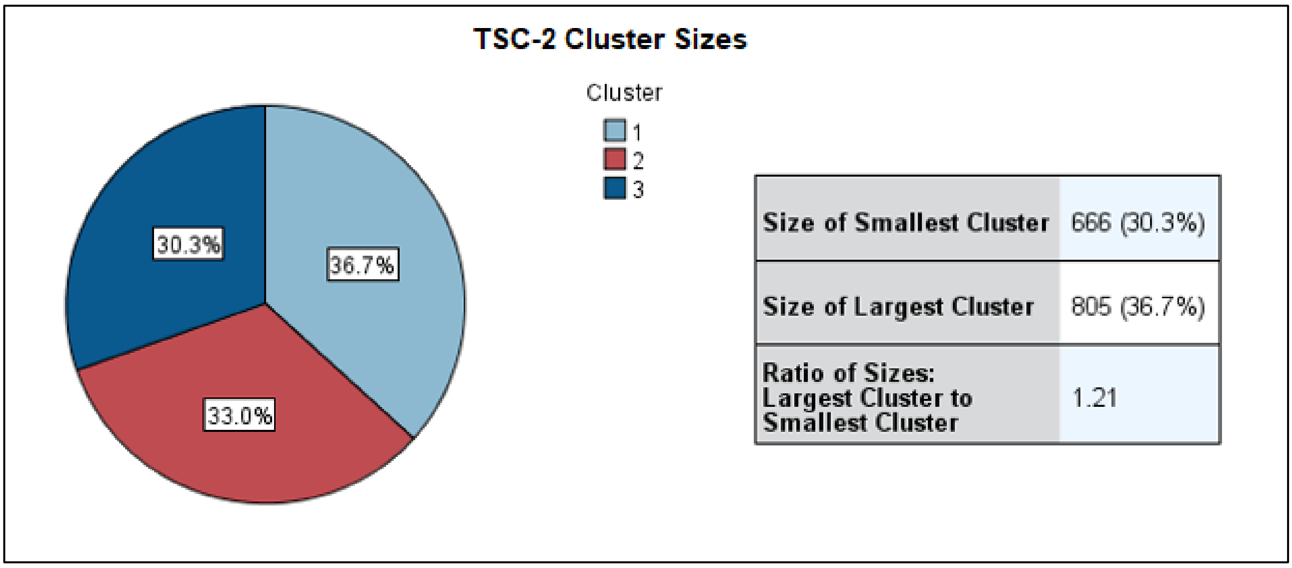


Figure 4.37 TSC-2 Cluster sizes

The TSC-2 model developed identified the following 6 significant (SM>0.4) covariates with an impact on the frequency of various crash risk factor combinations. The following four (4) covariates were found to have an importance equal to one (1):

- Two lane road indicator (SM = 1)
- Unpaved road indicator (SM = 1)
- Weekend indicator, and (SM = 1)
- Weekday indicator (SM = 1)

The following covariates exhibited SM measures above 0.5 mark, as shown in Figure 4.38.

- Poor pavement condition indicator (SM = 0.54), and
- No overtaking/ crossing road mark indicator (SM = 0.53)

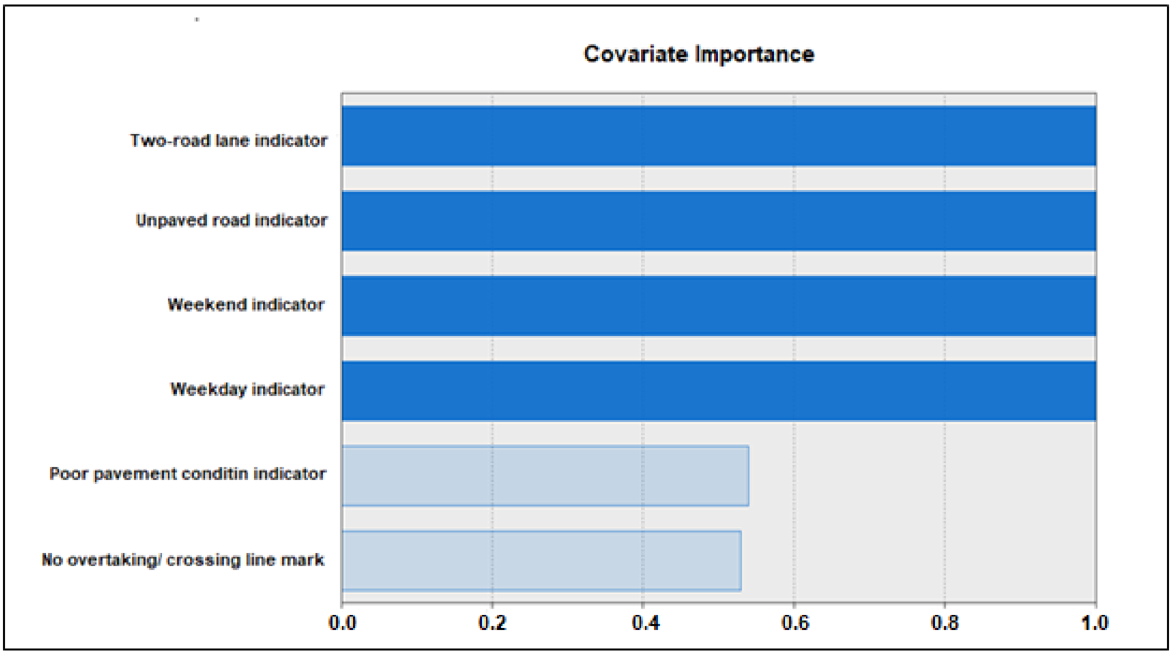


Figure 4.38 Covariate importance in TSC-2 Model

The study further investigated the importance of the significant covariates in the different cluster groupings determined by the TSC-2 model. The distribution and importance of the covariates are presented in Figure 4.39.

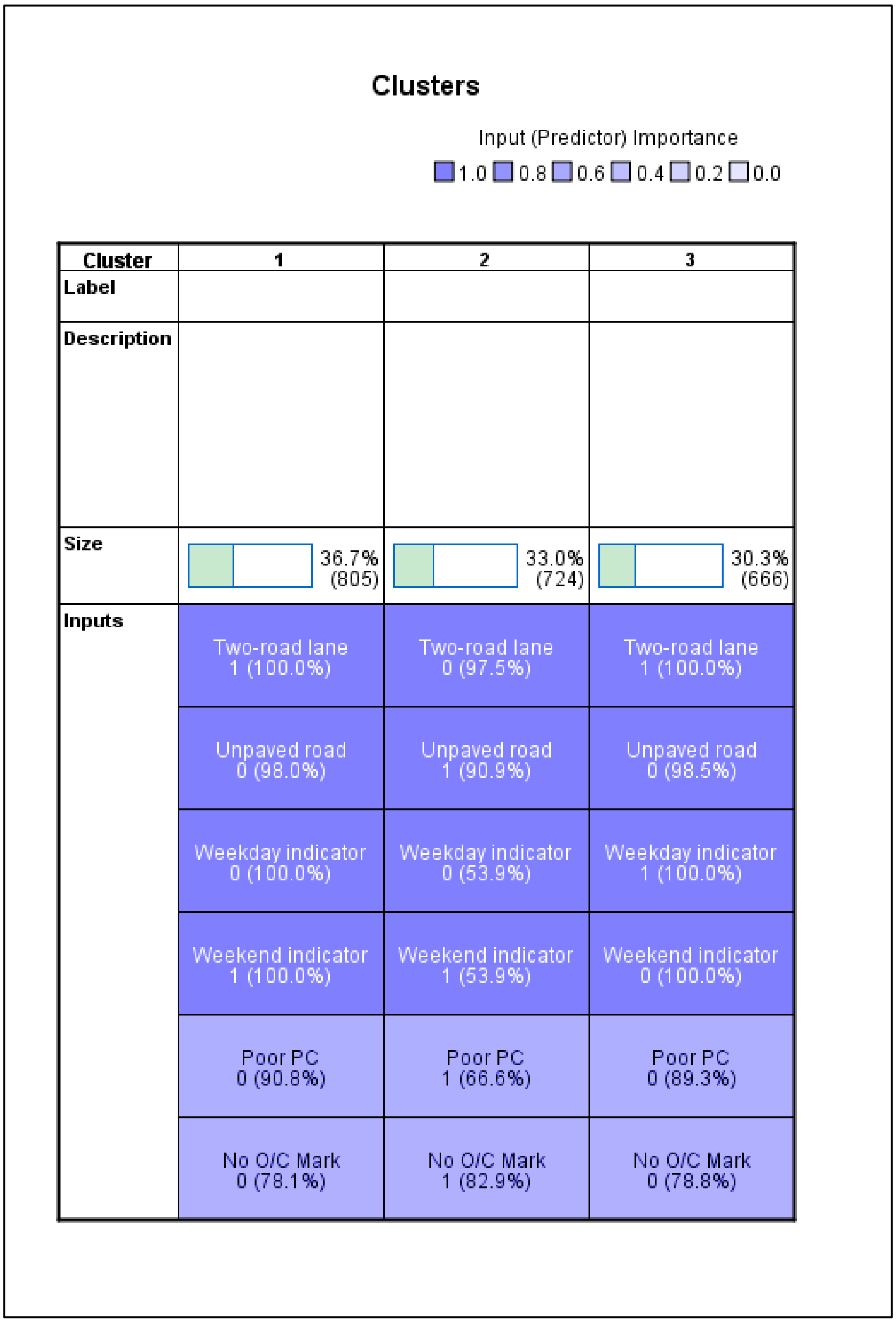


Figure 4.39 Covariate effects in the cluster groups

The following sections discuss the distribution of the covariates across the three determined cluster groups, underpinned by the distribution presented in Figure 4.39.

4.7.1.1 TSC-2 Cluster 1

The crash risk factor covariates in cluster 1, shown in Figure 4.39 and also illustrated in Figure 4.40, were found to influence the combination of national rural road risk factors. The distribution of these covariates is given below:

- Two-road lanes – All (100 percent) the rural road crashes in cluster 1 were found to have occurred on roads with two lanes (single carriageway)
- Paved roads – the study found that 98 percent of the crash records in cluster 1 occurred on paved roads.
- The weekdays were not found to have any impact on the crash records in cluster 1. This is shown by the 0 (100 percent) "otherwise" indication.
- All (100 percent) the crash records in crash cluster 1 were found to have occurred during weekends.
- The indicator on the pavement condition indicated that only 9.2 percent of the crash records in cluster 1 were in any way affected by poor pavement conditions.
- The study found that 78.1 percent of the crashes in cluster 1 occurred on road sections with visible overtaking/ crossing road markings.

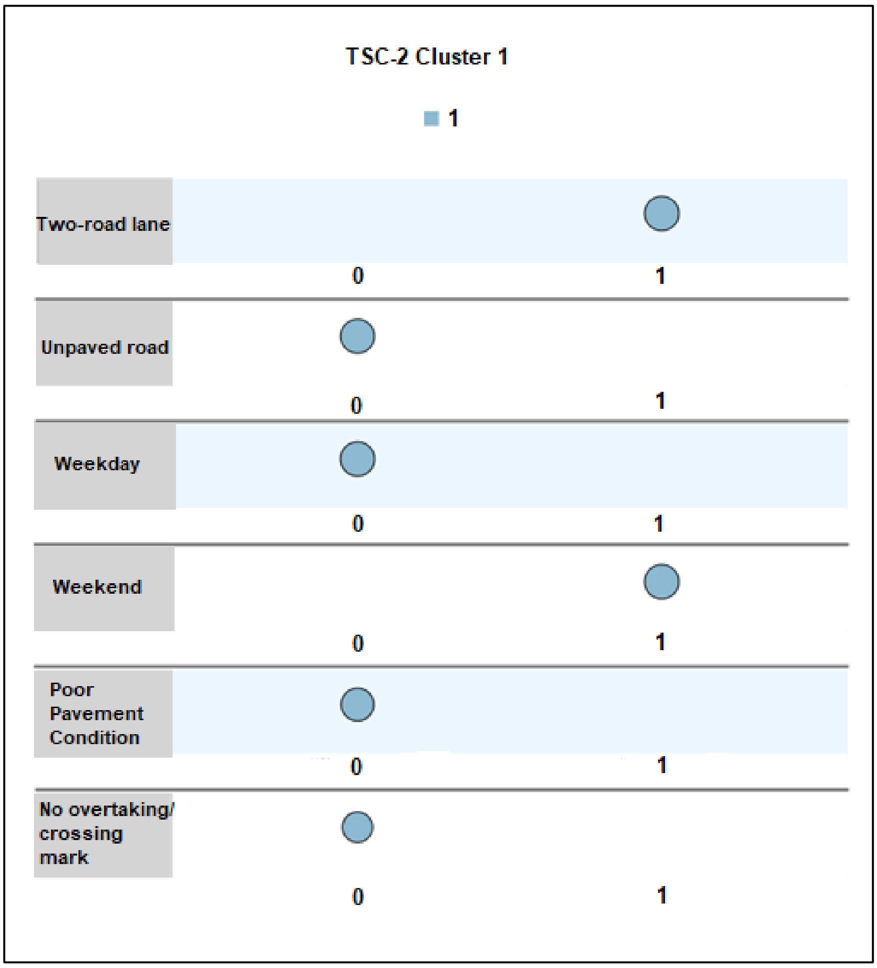


Figure 4.40 Covariates distribution in Cluster 1

The study further identified the extent to which the foremost (The five highest occurring crash risk factor groupings) risk factor combinations occurred in the model cluster groups (presented in Section 4.7) generated by TSC-2. The TSC-2 model identified 805 crash records within cluster 1. From the results, the distribution of the highest occurring risk factor combinations was determined as follows:

- The combination of recognition, decision and intentional risk factors in cluster 1 (code 2), represents the highest (6.46 percent) combination of all the risk factor combinations influenced by the covariate combination determined in cluster 1.
- The risk factor combination – recognition and decision risk factors (code 33), represents the second highest (6.09 percent) combination of all combinations influenced by cluster 1 covariates.
- The third highest combination of recognition and roadway and environmental risk factors (code 90) due to cluster 1 covariate combinations, represents 5.59 percent of the risk factor combinations.
- The combination of recognition, decision, and roadway and environmental risk factors (code 4) represents the fourth highest (4.72 percent) combination among the foremost model risk factors impacted significantly influenced by the combination of covariates in cluster 1.
- The combination of intentional risk factors and a recognition risk factor (code 78) in the dataset is shown to represent the fifth highest (4.22 percent) of all the combinations that constitute cluster 1.

4.7.1.2 TSC-2 Cluster 2

Figure 4.39 and Figure 4.41 describe and illustrate the distribution of risk combination factors due to covariates in cluster 2. The distribution of the covariates as determined by TSC-2 is discussed below:

- A high majority (97.5 percent) of road crashes grouped in cluster group two occurred on roads other than two-lane roads, this being 1 lane roads (mostly gravel roads) and dual carriageways.
- Most (90.9 percent) of the crash records in cluster 2 were reported to have occurred on roads with unpaved surfaces.
- The study results indicated that the weekends (Friday-Sunday) had a slightly higher impact on road crashes compared to weekdays. This impact is owing to the majority (53.9 percent) of crashes recorded over the weekends.
- A majority (66.6 percent) of road crashes in cluster 2 occurred on rural road sections with poor road conditions.
- A significantly high majority (82.9 percent) of the road crashes in cluster 2 were recorded on roads with no overtaking/ crossing road markings. This is expected as most of the unpaved roads are gravel in the study area.

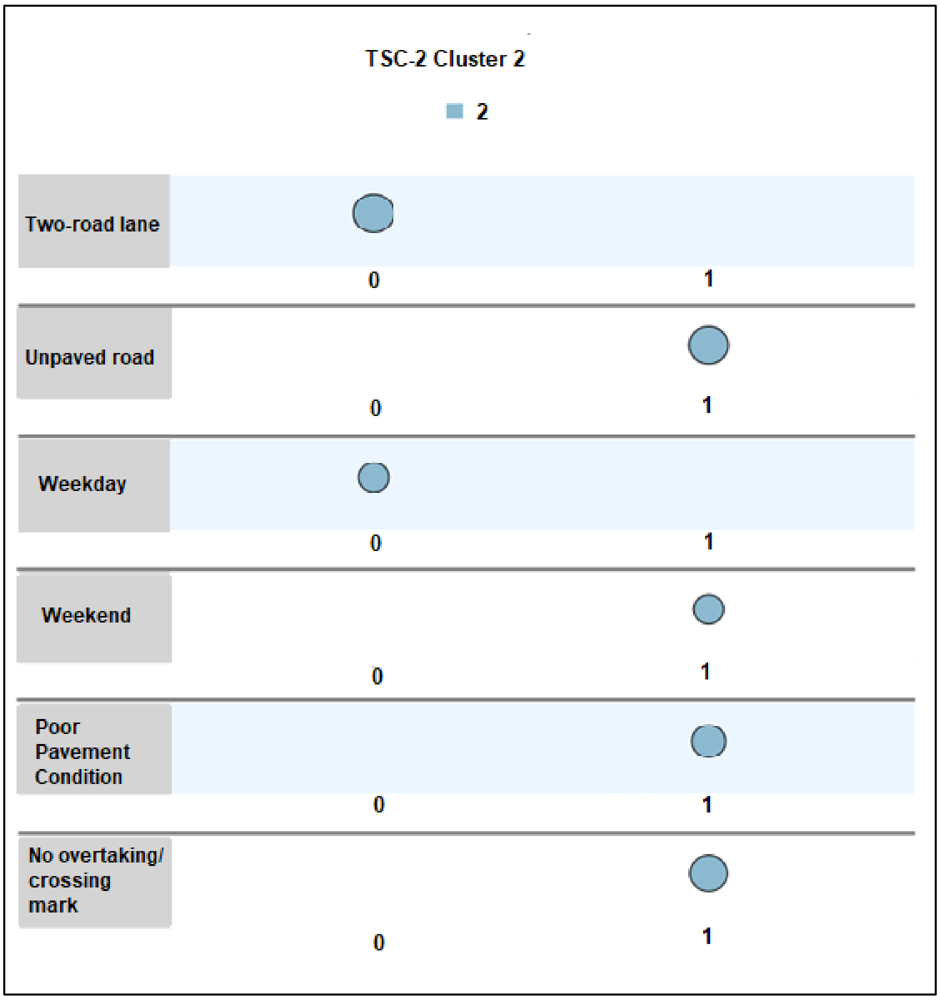


Figure 4.41 Covariates distribution in Cluster 2

The TSC-2 model identified 724 crash records influenced by the covariate combination in cluster 2. The study therefore interrogated the distribution of the foremost risk factor combinations identified in the crash dataset. The distribution of the risk factor combinations in cluster 2 is given below in descending order:

- The combination of recognition, decision and intentional risk factors (code 2) represents 7.46 percent of all the crash risk combinations influenced by the cluster 2 covariates
- The study results showed that the combination of recognition, decision, and roadway and environmental risk factors (code 4) represents 6.22 percent of the risk factor combinations identified in cluster 2.
- Similar to risk factor combination in code 4, the risk factor combinations in code 90 – the recognition and roadway and environmental risk factors, represent 6.22 percent of all risk factor combinations in cluster 2.
- The combination of recognition and decision risk factors (code 33) represents approximately 5.80 percent of all risk factor combinations owing to covariates combinations in cluster 2.
- The results shown that the combination of intentional risk factors and recognition risk factor, represented by code 78, account for 3.59 percent of all risk factor combinations in cluster 2.

4.7.1.3 TSC-2 Cluster 3

The results of the TSC-2 model presented in Figure 4.39 and Figure 4.42 show the combinational impact of the various covariates tested on the risk factor combinations found in the crash dataset. The distribution of the covariates in cluster 3 is discussed below, according to Figure 4.39:

- The study results indicated that all (100 percent) the crash records grouped into cluster 3 by the TSC-2 model occurred on roads with two lanes – single carriageways.
- A markedly low (1.5 percent) proportion of the crash records in cluster 3 occurred on rural roads with unpaved surfaces. This is indicative of the high percentage (98.5 percent) of crash records occurring on paved surfaces in cluster 3.
- All (100 percent) the crash records in cluster 3 were found to have occurred during the weekdays (Monday- Thursday)
- The results showed that a majority (89.3 percent) of the crash records in cluster 3 occurred on roads with good pavement conditions. This is also illustrated by the dummy variable shown in Figure 4.42.
- A majority (78.8 percent) of the crash records in cluster 3 were also found to have occurred on roads with visible overtaking/ crossing road markings.

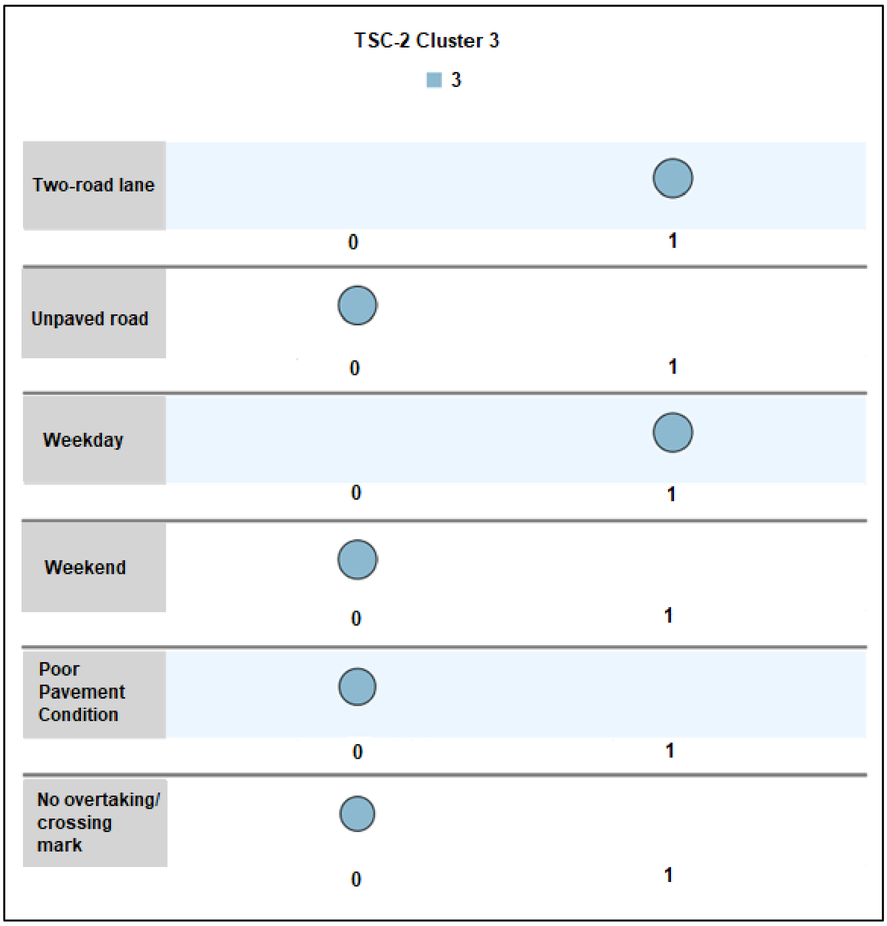


Figure 4.42 Covariates distribution in Cluster 3

The results show that the TSC-2 model identified 666 road crashes that occurred as a result of the covariate combinations in cluster 3. The study further interrogated the distribution of the foremost risk factor combinations due to the covariates identified in cluster 3. The interrogation is presented here, in descending order (contribution to total risk factor combinations in cluster):

- The results indicated that the combination of recognition, decision and intentional risk factors (code 2) represents the highest (7.21 percent) risk factor combination as a result of covariates in cluster 3.
- The risk factor combination coded 90, comprising recognition and roadway and environmental risk factors, represent 6.31 percent of all the risk factor combinations identified in cluster 3.
- The risk factor combination represented by code 33 – combination of recognition and decision risk factors, accounts for 4.95 percent of all risk factor combinations influenced by the covariate combination determined in cluster 3.
- The study results indicated the combination of recognition, decision, and roadway and environmental risk factors (code 4) in cluster 3 account for 4.64 percent of all risk factor combinations identified in the cluster.
- Of the foremost risk factor combinations, the combination of intentional and recognition risk factors (code 78) represents approximately 4.35 percent of all the risk factor combinations identified by the TSC-2 model in cluster 3.

4.7.1.4 Comparison of covariate combinations in TSC-2 model clusters

The TSC-2 model tested the interactive relationship between the numerous road crash risk factor combinations and covariates that were found to exhibit a Silhouette Measure (SM) greater than 0.4. The results found six (6) significant covariates with different dummy variable combinations across three (3) generated road crash cluster groups. The dummy variable combination of the covariates across the 3 cluster groups is presented in Table 4.41.

Table 4.41 Dummy variable combinations in TSC-2 model cluster groups

Covariate	TSC-2 Cluster 1	TSC-2 Cluster 2	TSC-2 Cluster 3
Two-lane road indicator	1	0	1
Unpaved road indicator	0	1	0
Weekday indicator	0	0	1
Weekend indicator	1	1	0
Poor pavement indicator	0	1	0
No overtaking/ crossing mark indicator	0	1	0

The TSC-2 model generated various covariate combinations with possible impacts on the risk factor combinations identified in the three cluster groups (see Table 4.41). The two-lane road indicator was found to have an impact on risk factor combinations in cluster 1 and 3. In cluster 3, the risk factor combinations were found to be influenced by roads with either one or more than two-lanes. As expected, the unpaved surface nature of the roads was only found to have an impact on the combination of risk factors in cluster 2.

The TSC-2 model found that weekdays (Monday to Thursday) were having a possible influence on the combination of risk factors in cluster 3. In comparison, the TSC-2 model found that the risk factor combinations in cluster 1 and cluster 2 were influenced by the weekends (Friday to Sunday). The results of the TSC-2 model indicated that poor pavement conditions only had an impact on risk factor combinations found in cluster 2. In the same way, road with no overtaking, crossing markings were found affect the risk factor combinations in cluster 2 as well. This is expected as the roads in cluster 2 are mostly gravel roads.

4.7.1.5 Comparison of risk factor distribution levels across the cluster groups

The TSC-2 model in the study interrogated the distribution of the foremost risk factor combinations in the various cluster groups. This presents an opportunity to identify and investigate the impact of the covariate combinations in the various clusters on the most common risk factor combinations in the crash dataset. The distribution of the five highest occurring risk factor combinations in the clusters is presented in Table 4.42.

Table 4.42 Risk factor combination distribution across TSC-2 cluster groups

Top five risk factor combinations (Code)	**TSC-2 Cluster 1**	**TSC-2 Cluster 2**	**TSC-2 Cluster 3**
Recognition, Decision and Intentional risk (**Code 2**)	6,46%	7,46%	7,21%
Recognition, Decision, and Roadway and Environmental risk (**Code 4**)	4,72%	6,22%	4,64%
Recognition and decision risk (**Code 33**)	6,09%	5,80%	4,95%
Intentional and Recognition risk (**Code 78**)	4,22%	3,59%	4,35%
Recognition and roadway and environmental risk (**Code 90**)	5,59%	6,22%	6,31%

The results of the risk factor distribution in the cluster groups found that risk factors in ***code 2*** (recognition, decision and intentional risk factors) represent the account for the highest combinations in all three clusters ($C_{1,2}$ = 6.46 percent; $C_{2,2}$ = 7.46 percent; $C_{3,2}$ = 7.21 percent). In cluster 1 (4.72 percent) and cluster 3 (4.64 percent), the combination of recognition, decision, roadway and environmental risk factors (***code 4***) represent the fourth highest occurring combinations due to the covariate cluster combinations. In comparison, the code 4 combination represents the second highest occurring combination in cluster 2 (6.22 percent). In the same way, the combination of recognition and, roadway and environmental risk factor (***code 90***) in cluster 2 (6.22 percent) and cluster 3 (6.31 percent), represents the second highest occurring combination in the clusters. However, code 90 risk factor combinations represent the third highest occurring combination in cluster 1 (5.59 percent). The study results showed that the combination of recognition and decision risk factors (***code 33***) accounts for the second highest occurring combination in cluster 1 (6.09 percent). Code 33 combinations however account for the third highest occurring risk factor combination in cluster 2 (5.80 percent) and cluster 3 (4.95 percent)). The combination of intentional and recognition risk factors (***code 78***) were found to be the least occurring risk factor combination of all the foremost combinations identified by TSC-2 across all cluster groups ($C_{1,78}$ = 4.22 percent; $C_{2,78}$ = 3.59 percent; $C_{3,78}$ = 4.35 percent).

4.8 Summary of key results

A mix analysis method was applied in the study to develop crash predictive models for national rural roads and examine the relationship between road characteristics and road crashes. The study applied several analysis methods including descriptive, inferential, spatial and statistical modelling techniques on the crash dataset. A summary of the key results from the analyses carried out are presented in this section.

4.8.1. Univariate and bivariate crash analyses

4.8.1.1 National rural road crash frequencies

- On average, 638 fatal and serious injury crashes were recorded annually between 2012 and 2016 on the national rural road network in Namibia.
- A road crash rate of 21.3 fatal and serious injury road crashes per 100 000 population was computed for the period from 2012 to 2016.
- The frequency of road crashes was found to peak over holiday months (May-April, August and December)
- The highest mean weekly crash counts were observed over the third quarter of the calendar year (12.94±4.419). The lowest road crash counts were recorded over the first quarter (10.98±4.414).
- Statistically significant ($p=0.041<0.05$) interactions were found between the mean values of the quarterly weekly crash counts of the first quarter of the year and the third quarter.
- Weekly road crash occurrences were found to be consistent ($p>0.05$) over the second and fourth quarters of the calendar year.
- The highest frequency of road crashes was observed over the second week after pay week (12.60±4.260), followed by weeks other than the first two weeks of the month (12.35±4.990). The lowest frequency of road crashes over the study period was observed during the pay week (11.78±4.244).
- As revealed by the Post-hoc test results, statistically significant ($p=0.002<0.05$) interaction was found between the individual mean values of the pay week and those of the second week after the pay week.
- The highest week day road crash frequencies were observed over the days of the weekend, with a peak on Saturdays (128±31.757). Sunday (107±16.956) and Friday (103.6±11.803) had the next highest crash frequencies recorded over the study period. The lowest crash frequencies were observed over Holidays (31±5.612), Wednesday (58.4±8.989) and Tuesdays (59±9.028).

- Statistically significant ($p<0.05$) relationships were identified between the weekend and all the week days (Monday, Tuesday, Wednesday and Thursday). Also, mean values of crashes over holidays were found to significantly interact with weekend days (Friday, Saturday and Sunday).
- The study observed that the highest frequency of road crashes occurred in the late afternoon (17h00 to 18h00). Lower peaks were observed during the mid (11h00 12h00) to and early hours (07h00 to 08h00) of the day respectively. The lowest crash frequencies were observed during the early morning hours (03h00 to 04h00) of the day.
- An overrepresentation of male drivers was observed in the crash dataset. Male drivers were more likely (M: F= 5.86) to be involved in road crashes on national rural roads than female drivers.
- Male drivers were found to be at a higher crash risk across the whole day (24 hours). The highest crash risk for males (M: F=24) was found to be in the early morning hours (02h00 to 03h00) despite the lowest crash frequency observed during that stretch of time (02hh00 to 05h00). The lowest crash risk for males (M: F= 4) was found to be during the afternoon (13h00 to 14h00).
- The mean driver age for the crash dataset was found to be 28.16 years (S.D 14.33). The oldest driver on the national rural roads was recorded as being 85 years while the youngest was 11 years.
- The road crashes were disproportionally distributed across the driver age groups. The highest crash frequencies emerged in the driver age group of 31 to 35 years. The frequency of road crashes rose drastically from the age group 21 to 25 years, with the high frequency stretch maintained until the 41 to 45 years age group.
- The highest male to female driver crash risk ratio emerged in young adults (21 to 25 years) and teenagers (16 to 20 years), with male drivers more than ten times (M: F>10) likely to get involved in a road crash.

4.8.1.2 National rural road casualties

- The fatal and serious injuries casualty dataset comprised 6 712 cases. More male road users (M: F=2.25) were likely to be FSI casualties than female road users.
- The distribution of FSI casualties across the time of day was found to be disproportionate. The highest FSI casualty frequency occurred in the late afternoon (16h00 to 17h00)
- The crash risk ratio emerged higher for male road users (M: F= 3.36) across the whole day, with a peak occurring in the early hours of the morning (01h00 to 02h00). Other higher risk casualty ratios emerged for male road users in the morning (06h00 to 07h00) and late evening hours (22h to 23h00).
- FSI casualty frequency emerged highest over the weekend (Fridays, Saturdays and Sundays), with a peak observed on Saturdays (268.80±86.085).

- FSI casualties were found to be consistent (p>0.05) between Mondays and all the days of the week. In the same way, between Saturdays and all the days of the week.
- Fatal injury casualty counts were found to be fairly consistent across all the months of the year, with a slight peak emerging in December.
- Serious injury casualties emerged with a marked peak in December and slightly lower peaks emerging in May and August.

4.8.1.3 Driver risk factors and behavioural characteristics

Driver-gender based crash risk analysis

- Inadequate surveillance of the road environment emerged highly among both driver genders as a primary risk factor – more dominant in female drivers than male drivers.
- Inattention also emerged significantly among both driver genders, with female drivers more prone to this risk factor than male drivers.
- Both driver genders showed similar degrees of traffic violations
- Dangerous manoeuvres and following too closely were identifiable in both driver gender- more in female than male drivers.
- Misjudgement of gaps between vehicles was notable in both driver gender as a primary crash risk factor- more in males than female drivers
- Encounters with animals on national rural roads were high for both driver genders. Animals emerged as the highest primary contributing factor for bother genders.
- Poor visibility emerged as a primary contributing factor for both gender- similar impact extent in both driver genders.
- Speed differential (congestion) was identifiable as a primary risk factor for both driver genders- emerging more for females than male drivers.

Driver-age based crash risk analysis results

- Human-related errors (86 percent) emerged strongly in driver in the adolescent age group (less than 18 years). Of the human errors in this age group, delay in response to traffic situations, inadequate surveillance and driving too fast for curves were most notable.
- Majority of primary factors in road crashes when young adults (18 to 25 years) were driving were human-related errors (75 percent). Animals (19 percent) and traffic violations (11 percent) were identifiable primary crash factors for young adults.
- Human-related errors (64 percent) were found to be the main crash risk factor for the age group 26 to 35 years. The most notable risk factors in this age group were animals (17 percent), inattention (10 percent), inadequate surveillance of the surrounding environment (8 percent) and dangerous manoeuvres on the roads (8 percent).

- As with other age groups, human-related errors group was the main errors leading to crashes in adults age group (35 to 65 years). Unexpectedly, the condition of the road surface (10 percent) was the leading primary factor (level 1) in road crash occurrences for the adult age group.
- For the elderly (greater than 65 years), a significantly high number of road crashes were primarily influenced by human errors (95 percent). The most common primary factors among the elder were a false assumption of other road users' actions (11 percent) and panicking/ freezing in complex traffic situations (11 percent).

Relationship between driver risk factors

- The highest possible level 2 and level 3 crash risk factors contribution (49%) was observed in road crashes were intentional risks were the leading primary factors (28%).
- As expected, roadway and environmental risk factors were the second highest (27 percent) contributing factor towards road crashes in both the level 1 and level 2 analyses.
- Animals were identified as the highest individual primary and level 2 and 3 possible risk factors in crash occurrences on the national rural roads
- Dangerous road manoeuvres (15 percent), misjudgement of gaps or other driver actions (14 percent) and traffic violations (12 percent) were also identifiable as individual risk factors in crash occurrences.

4.8.2. Road crash geospatial analyses

Using the Kernel Density Estimation (KDE) in QGIS, the study carried out a geospatial analysis and developed raster maps to investigate the distribution of fatal and serious injury crashes and identify hazardous sections- crash densities on the different classifications of the national rural roadway. The key results of this method are summarised in this Section.

Distribution of FSI crashes on All Rural Roads (ARR- R1 to R6 classifications)

- The highest FSI crash densities were observed on the Northern part of the national rural road network- trunk, main and district roads. The national rural road network within the area connecting the following towns was identified to be hazardous (extreme clustering of road crashes):
  - Ongwediva
  - Oshakati
  - Oshikuku
  - Oniipa
  - Eenhana
  - Helao Nafidi
- The national rural roads leading to and from the following towns/ cities on the Northern and Central road network were identified to have higher road crash intensities:
  - Okahandja
  - Windhoek
  - Rehoboth
  - Omuthiya
- On the Western part of the national rural road network, higher crash densities were observed on the rural road between the following towns:
  - Arandis
  - Swakopmund
  - Walvis Bay
- Moderate crash densities were identifiable on the road sections around and between the following towns:
  - Usakos and Karibib
  - Otjiwarongo and Okahandja
  - Otjiwarongo and Otavi
  - Nkurenkuru
  - Rundu
  - Okahao
  - Outapi

- Lower degrees of crash densities were observed on national roads in the Southern regions (Hardap and !Karas) of Namibia and towards the Eastern (Omaheke) and North- Eastern (Kavango East and Zambezi) parts of the road network.

Distribution of FSI crashes on High Order Rural Roads (HORR- R1 to R3)

- The FSI road crashes on high order rural roads- classified R1 to R3, were visualised at a bandwidth of 1 000 m using the KDE tool. Extreme crash densities were observed on rural roads around and between the following localities:
    - Oniipa
    - Ongwediva
    - Eenhana
    - Okahandja
    - Windhoek
    - Rehoboth
- Higher crash densities were observed on high order rural roads between and around the following localities:
    - Walvis Bay
    - Arandis
    - Usakos
    - Karibib
    - Oniipa
    - Omuthiya and Tsumeb
    - Nkurenkuru
- Rural roads in the central part of Namibia displayed moderate crash densities during the period between 2012 to 2016. These included national roads between:
    - Otjiwarongo and Otavi
    - Otavi and Tsumeb
    - Otjiwarongo towards Okahandja
- Lower crash densities were observed on high order rural roads traversing in the following regions:
    - Hardap
    - !Karas
    - Kunene
    - Kavango East
    - Zambezi

Distribution of FSI crashes on Low Order Rural Roads (LORR- R4 to R6)

- The geospatial analysis of FSI crashes that occurred on national low order rural roads only from 2012 to 2016 allowed for a more focused identification of hazardous low order roads on the national road network.
- Extreme FSI crash densities on LORRs were observed on the roads surrounding the following areas/ in regions:
    - Oshikuku
    - Outapi
    - Okahao
    - Ongwediva
    - Oniipa
    - North of Gobabis
- Higher to moderate crash densities were mostly observed on rural roads around the following regions:
    - Kunene
    - Kavango East
    - Kavango West
    - Erongo
- Lower crash densities were mostly observed in the following regions:
    - Khomas
    - Otjozondjupa
    - !Karas
    - Zambezi
    - Oshikoto

4.8.3. Road design and traffic characteristics compliance summary

4.8.3.1. Compliance summary results

- In all three datasets examined - all rural roads, high order and low order rural roads, more than half (50 percent) of the unpaved rural roads lane widths were non-compliant with the design guidelines.
- Less than a quarter (25 percent) of the surfaced shoulder widths in all datasets were found to be compliant with design guidelines.
- The proportion of roads recommended to have paved shoulders and complying with the design guidelines on the national road network were less than a quarter (25 percent) of the sample size examined.
- The extent of compliance of ground shoulder widths on paved roads was found be significantly lower than half (50 percent) of all paved roads investigated.
- More than three-fifths (60 percent) of paved roads in all the datasets complied with the design guidelines set for lane widths
- More than four-fifths (80 percent) of the unpaved roads studied on the national rural road network complied with the ground shoulder width requirements set in the TRH 17.
- Significantly higher levels (above 90 percent) of compliance for stopping sight distances were observed on paved roads
- Stopping sight distance compliance levels on unpaved roads was found to be less than half (39.5 percent) in the low order roads dataset and slightly above nine-tenths (9.24 percent) in the high order roads dataset.
- The conditions of the pavements on paved and unpaved roads, guided by Table 2.20 were found to be ranging between 69 percent and 80 percent.

4.8.3.2. Distribution of road crashes by non-compliance

Lane width non-compliance spatial distribution results

- The spatial distribution of road crashes by the non-compliance of lane widths LW was evident on both high order and low order rural roads on the Northern part of the national road network. With extreme crash densities around the following localities:
    - Oniipa
    - Ongwediva
    - Eenhana
    - Oshakati

- Higher crash densities due to LW non-compliance were observed on national roads close to the following localities:
  - Outapi
  - Oshikuku
  - Okahao
- The national rural road north of the Gobabis locality showed moderate FSI crash densities due to LW non-compliance.
- National rural roads with across the rest of the country exhibited FSI lower crash densities due to LW non-compliance levels. These roads are in the following regions:
  - Kunene
  - Kavango East
  - Kavango West
  - Zambezi
  - Erongo
  - Khomas
  - !Karas
  - Hardap

Surfaced shoulder width non-compliance spatial distribution results

- The non-compliance of surfaced shoulder widths (SSW) was found to cause an extreme density of FSI road crashes on the far Northern part of the national rural road network. The extreme densities were identified on roads around the following localities:
  - Ongwediva
  - Eenhana
  - Helao Nafidi
- Higher FSI crash densities were identifiable on the road networks between and around the following localities:
  - Walvis Bay
  - Arandis
  - Windhoek
  - Nkurenkuru
- Moderate FSI crash densities due to SSW non-compliance levels were prominent on roads in the following localities:
  - Karibib
  - Rundu
  - Katima Mulilo
  - Oshikuku
  - Outapi

        - Oniipa
  - Lower crash densities were identifiable on national rural roads in the following regions:
    - Hardap
    - Otjozondjupa
    - Kunene
    - Oshikoto
  - The extent of compliance of the SSW parameter was not found to significantly influence the occurrence of FSI crashes on national roads in the !Karas region.

Ground shoulder width non-compliance spatial distribution results

- The ground shoulder width extent of non-compliance was identified to cause extreme densities of FSI crashes on the national rural roads between and around the following localities:
  - Oniipa
  - Ongwediva
  - Oshakati
- Higher crash densities are identifiable on national roads between and around the following areas:
  - Eenhana
  - Helao Nafidi
  - Oshikuku
  - Outapi
  - Omuthiya towards Oniipa
  - Okahao
  - Windhoek
- Moderate FSI crash densities were found to be prominent on national roads around and between localities on the Western, Central and North-Eastern parts of the network: These areas are:
  - Walvis Bay
  - Arandis
  - Okahandja
  - Nkurenkuru
  - Rundu
- Lower crash densities were identifiable on national rural roads around and between the following localities:
  - Gobabis
  - Katima Mulilo
  - Tsumeb

- Grootfontein
- Keetmanshoop to Mariental
- Mariental to Rehoboth
- Rehoboth towards Windhoek
- Okahandja to Otjiwarongo
- Otjiwarongo towards Otavi
- Rundu towards Katima Mulilo

Shoulder type (proportion of paved shoulders) non-compliance spatial distribution results

- The spatial distribution of FSI road crashes due to the extent of shoulder type non-compliance is mostly concentrated along the high order rural roads across the national road network.
- Extreme FSI crash densities are prominent on the Northern part of the national road network. The densities are identifiable around the following localities:
  - Oniipa
  - Ondangwa
  - Ongwediva
  - Eenhana
  - Helao Nafidi
- Higher crash densities were identifiable around and between the following localities:
  - Omuthiya towards Oniipa
  - Okahao
  - Oshikuku
  - Outapi
  - Between Windhoek and Okahandja
  - Between Walvis Bay and Arandis
- Moderate to lower crash intensities were mostly prominent around the Central, Southern and North Eastern parts of the national road network. These densities are around and between the following areas:
  - Between Otjiwarongo and Okahandja
  - Between Otjiwarongo and Otavi
  - Between Karibib and Usakos
  - Tsumeb
  - Otavi
  - Nkurenkuru
  - Rundu
  - Katima Mulilo
  - Rehoboth towards Mariental and Keetmanshoop

Stopping sight distance (SSD) non-compliance spatial distribution results

- The non-compliance of stopping sight distance was found to have a pronounced impact on crash occurrence on the central part of the national rural road network. Localities with extreme densities include:
    - National roads around Windhoek.
    - National roads around Okahandja.
    - Between Windhoek and Rehoboth.
- Higher crash densities were marked on the following roads:
    - Okahandja towards Otjiwarongo.
    - Otjiwarongo towards Otavi.
    - National roads around Grootfontein.
- Moderate to lower FSI crash densities were identifiable on the following roads:
    - National rural roads in the Kunene region - around Opuwo town.
    - Road between Walvis bay and Arandis.
    - Roads in the Southern regions - Hardap and !Karas region.
    - Roads around Nkurenkuru and Rundu towns in Kavango West.
    - Roads in the Omaheke region.
- SSD non-compliance levels was found not to have an impact on crash occurrence on roads in the Northern part of the network, which are significantly affected by the non-compliance of other parameters.

Pavement condition (PC) non-compliance spatial distribution results

- The overall poor condition of the ride surface – non-compliance thereof, was found to have a marked impact on the occurrence of FSI road crashes on the central and northern parts of the national rural road network. The “extreme” crash densities were identifiable around the following localities:
    - Oniipa
    - Ongwediva
    - Oshakati
    - Windhoek towards Okahandja
- Moderate crash densities were identifiable on the national road network through and around the following localities:
    - Okahandja to Otjiwarongo
    - Walvis Bay to Omuthiya through the following areas: Arandis, Usakos, Karibib, Omaruru, Otjiwarongo, Otavi and Tsumeb.
    - Rehoboth towards Mariental and Keetmanshoop

- Lower crash densities were identifiable on the national network around Katima Mulilo and Rundu.

4.8.4. Road crash prediction model development results

- The General Multivariate Regression (MLR) models developed demonstrated better crash prediction performance – higher and statistically significant adjusted R-square and F-test values, compared with the Base Mean Multivariate (BMM) models at predicting the rate of fatal and serious injury (FSI) crash occurrences.
- The crash prediction model 1 (CPM 1), fitted to FSI crashes on all national rural roads comprised five (5) statistically significant ($p<0.05$) covariates with effects of various magnitudes.
  - The following covariates demonstrated statistically significant positive associations (coefficient b* estimate) with FSI road crashes on the rural roads, in descending order:
    i. The proportion of heavy vehicles in the annual average daily traffic – AADT_H (0.464)
    ii. The lane width (0.0.293)
    iii. The vertical terrain – hilliness (0.082)
    iv. The operating speed (0.028)
  - The following covariate was shown to be negatively associated with the occurrence of FSI crashes on rural roads:
    i. The surfaced shoulder width (-0.069)
- The CPM 2 was fitted to road crashes that occurred on high order rural roads (HORRs) and comprised five statistically significant covariates.
  - All five covariates demonstrated positive associations (coefficient b* estimates) with the occurrence of FSI road crashes on rural roads. These covariates are shown below in descending order:
    i. The proportion of light vehicles in the AADT (0.682)
    ii. The lane width (0.137)
    iii. The vertical terrain (0.112)
    iv. The ground shoulder width (0.108)
    v. The operating speed on road sections (0.032)
- The FSI road crashes that occurred on low order rural roads were used to develop CPM 3. The best performing CPM 3 comprised four (4) statistically significant covariates with various effects on road crash occurrence. One (1) covariate exhibited influence on the crash rates but had no statistical significance.
  - The following two covariates demonstrated positive associations (coefficient b* estimates) with the occurrence of FSI road crashes on low order roads (in descending order):

I. The proportion of light vehicles in the AADT (0.315).
II. The vertical terrain – hilliness (0.066).

- The operating speed (0.049) is the only covariate in CPM 3 that demonstrated positive associations with the occurrence of road crashes on the roads classified as low order, with no statistical significance demonstrated.
- The following two covariates were shown to be negatively associated with the occurrence of FSI crashes on rural roads, in descending order:

I. The surfaced shoulder width (-0.138)
II. The ground shoulder width (-0.205)

4.8.5. Impact of compliance of crash predictive models

The best-performing crash prediction models (General Multivariate (MLR) crash prediction models (CPMs)), fitted to the crash datasets and existing road conditions, had a sensitivity to design compliance test performed to examine the extent to which design compliance affects the outcome variables (covariate effects).

Sensitivity test results of CPM 1 – CPM 4

- The sensitivity test on MLR-CPM 1 - fitted to all the FSI crashes on national rural roads, had a greater influence on the following covariates:
  - The proportion of heavy vehicles in AADT (opposite effect on crash rates)
  - The ground shoulder width (gained statistical significance)
  - The shoulder type – proportion of paved shoulder on paved roads (gained statistical significance).
- The following covariate demonstrated an increased influence (coefficient b* value) on the outcome variable when design conditions are considered “ideal”:
  - The operating speed
- The following covariates exhibited reduced influence on the occurrence of FSI road crashes on all national rural roads:
  - The vertical terrain on the road sections – hilliness.
- The following variables did not demonstrate statistically significant effects in CPM 5 after the sensitivity to road design compliance test:
  - The lane width on the road sections.
  - The width of the paved hard shoulders on the road sections.

Sensitivity test results of CPM 2 – CPM 5

- The sensitivity test results of MLR-CPM 2 for crashes on high order rural roads indicate that the following covariates had no statistically significant influence on the occurrence of road crashes when "ideal" road characteristics are considered:
  - The lane width
  - The ground shoulder width
- The following variables showed an increased influence (coefficient b* estimate) on the occurrence of road crashes on high order rural roads:
  - The operating speed
  - The vertical terrain on the road sections
- The compliance of design characteristics had a major influence on the following covariates:
  - The proportion of heavy vehicles in the AADT (exhibiting a change of effect on crash rates).
  - The proportion of paved shoulder on the high order roads (exhibiting statistically significant effects on crash rates)
  - The number of horizontal curves per length high order rural road (exhibiting statistically significant effects on crash rates).

Sensitivity test results of CPM 3 – CPM 6

- As a result of design compliance, the sensitivity test had a greater influence on the following variables on low order rural roads:
  - The operating speed on the road sections (influence on crash rates lost)
  - The surfaced shoulder width (loses influence and statistical significance)
  - The proportion of light vehicle in the AADT (change in effect on crash rates)
  - The shoulder type on the road sections (demonstrates statistically significant effects).
  - The stopping sight distance (shows influence on crash rates but no statistical significance).
- The following variable demonstrated an increased absolute effect on the frequency of road crashes:
  - The ground shoulder width.
  - The vertical terrain – hilliness (increase influence but loses statistical significance).

4.8.6. Driver characteristics and risk factors – roadway condition analysis models (TSC Model)

- The risk factor combination analysis in this section identified 93 combinations of a possible 343 risk factor combinations in the dataset.
- The Two-Step Cluster (TSC-2) Model exhibited the lowest AIC value (568.073) and the largest ratio of AIC changes (0.349) and ratio of distance measures (2.264) with respect to the base cluster.
- The Two-Step Cluster (TSC-2) analysis generated three (3) cluster groups for the crash dataset, in which the following indicators (six (6) of twenty-one (21) indicators) were found to have a Silhouette Measure (SM) greater than the threshold of 0.4, which is indicative of the importance of the predictor.
  - Two-lane road indicator (SM = 1.0)
  - Unpaved road indicator (SM = 1.0)
  - Weekday indicator (SM = 1.0)
  - Weekend indicator (SM = 1.0)
  - Poor pavement condition indicator (SM = 0.54), and
  - No overtaking/ crossing line mark indicator (SM = 0.53)
- In cluster one (1), the following covariate indicators were found to exhibit a dummy variable of one (1), indicating a validation of the indicators impact.
  - Two-lane road indicator
  - Weekend indicator
- The following covariate indicator exhibited a validation dummy variable in cluster two (2).
  - Unpaved road indicator
  - Weekend indicator
  - Poor pavement condition indicator
  - No overtaking/ crossing line mark indicator
- In the same way to cluster 1, cluster three (3) also had two covariates exhibiting validating indicator dummy variables
  - Two-lane road indicator
  - Weekday indicator
- The following risk factor combinations, in descending order, were identified as the foremost occurring combinations of all the risk factor combinations identified in the crash dataset.
  - Recognition, Decisions and Intentional risk factors – Code 2 (7 percent)
  - Recognition, and Roadway and Environmental risk factors – Code 90 (6 percent)
  - Recognition and Decision risk factors – Code 33 (5.6 percent)
  - Recognition, Decisions, and Roadway and Environmental risk factors – Code 4 (5.2 percent), and

    - Recognition and Intentional risk factors – Code 78 (4.1 percent)
- A further examination of the risk factor combinations distribution across the TSC cluster groups, found that the combination of the recognition, decision and intentional risk factors was the highest occurring combination across all the TSC model cluster groups.
- The following individual risk factors were found to feature the most (in descending order) among the risk factor combinations identified in the dataset:
    - Recognition risk factor (100 percent)
    - Decision risk factor (60 percent)
    - Intentional (40 percent), and Roadway and Environmental risk factors (40 percent).

Chapter 5: Discussion of results

Road crashes are a complex event influenced by a multiplicity of interacting factors – human, road environment and vehicle related. Human related factors are globally affirmed as the leading crash factors. However, it is not easy to directly control and predict human related factors on roads. One way to directly impact human related factors is through investigation into the roadway environment (Gaudry and Vernier, 2002; Farahmand and Boroujerdian, 2018; Islam *et al.*, 2019). The study took a mixed approach to understand and examine the crash dataset from 2012 to 2016, to determine the various factors affecting the occurrence of road crashes in Namibia and to develop road crash prediction models driven by road and traffic conditions on national rural roads. A univariate and bivariate approach was taken to examine the crash frequencies over the study period, investigating the temporal and demographic variations of crashes on driver risk factors, therefore creating a basis to understand how a change in the crash prediction models developed will affect human related factors in the future. The crash prediction models are novel in the context that they investigate multiple interactive road environmental factors (geometric and traffic characteristics) on national rural roads, different from the usual approach of investigating the impact of a single road characteristic on road crashes, when in fact road elements work in tandem to create a road environment understandable by road users. In an effort to explore how all the study findings impact the driver risk factor combinations preceding a crash occurrence, the study explored how several covariates (demographic, temporal, roadway and environmental) influenced the combination of several identified driver risk factors by using the Two-Step Cluster analysis method. This approach is new in Namibia and to an extent in Sub-Saharan Africa - where literature on the impact of the road environment on risk factors and actions preceding road crashes are almost non-existent. On the whole, the mixed approach applied in the study is novel in Sub-Saharan Africa and globally and contributes to the attempt by researchers to understand the impact of the road environment on road crashes holistically.

The results of the study are discussed in the sections below:

5.1 Discussion of univariate and bivariate analyses results

5.1.1. Univariate and bivariate analyses of crash datasets

Between the years 2012 to 2016, the study results show that an average of 638 fatal and serious injury (FSI) crashes were recorded annually on the national rural road network in Namibia. This represents an annual FSI crash rate of 21.3 FSI crashes per 100 000 population. The study found an overrepresentation of male drivers involved in FSI road crashes and FSI casualty counts in the crash datasets. This is however not a novel finding, as previous studies have also shown a high proportion of male road users compared to females (NRSC, 2012; Namibia Statistics Agency, 2015; Nteziyaremye, 2018; World Health Organisation, 2018). As a result, male road users were found to be at a higher crash risk than their female counterparts across all age groups. Findings from previous studies were also telling of the high risk-taking behaviour among male drivers, which can be attributed to the innate risk-taking nature of male road users and young drivers (Schulze and Koßmann, 2010; Berhanu Bezabeh, 2013; World Health Organisation, 2018; Jones *et al.*, 2019).

An analysis of road crash frequencies by driver age group found that road crash frequencies rose drastically from the 21 to 25 age group. The high crash frequency remained steady in this economically active cluster until the 41 to 45 age group. The holistic peak in the high crash frequencies stretch emerged in the 31 to 35 age group. A more detailed examination of the distribution of road crashes across the age groups showed that the highest crash frequencies for female drivers were in the age group 31 to 35. For male drivers, the highest frequencies are observed in the 26 to 30 age group. A report by the African Development Bank (AfDB) on Road Safety in Africa also reported that casualties among male road users are highest in the 15 to 29 age groups in Sub-Saharan Africa (Berhanu Bezabeh, 2013). Driver crash frequencies dipped drastically from the 41 to 45 age group, with a steady decrease in crashes as the age group years increased. The reduction in crash frequencies in both advanced driver age groups could possibly be attributed to the reduced exposure that older drivers get on the national roads. Literature on distribution of road crashes on national roads in Namibia and the factors influencing it are non-existent, as a result, the reasons on the distribution of road crashes across the age groups cannot be validated.

The study results found three distinct high road crash frequency peaks over a virtual day. The study revealed that the safest time to be on the road was the early morning hours (00h00 to 06h00). From the early morning onwards, road crashes increased steadily and peaked in the late afternoon. Lower peaks were observed during the middle of the day (11h00 to 12h00) and in the morning hours (07h00 to 08h00). This is in line with findings reported in previous studies in which crash frequencies were observed to be highest during the peak traffic hours of the day (Botha, 2005; NRSC, 2012; Carey and Sarma, 2017). The leading primary risk factor was found to be animals on the national road, with road crashes involving animals mostly observed to have occurred in the late afternoon and evening hours, during which the highest peak crash frequencies are noted to have occurred. This finding corroborates findings from previous studies on road crashes in Namibia (Eggleston *et al.*, 2016; Nghishihange, 2018).

An examination of road crashes over the days of the week revealed that the highest road crash incidents occurred over the weekends – Friday, Saturday and Sunday. The lowest crash frequencies were observed on Tuesdays and Wednesdays. This finding is in line with findings from other studies (NRSC, 2012; Nghishihange, 2018). Previous studies have also observed high crash frequencies over the weekends, which are exacerbated by the high traffic volumes on national roads as most people travel to and from their regions of origin to visit families and friends. The relative riskiness and specific psychological conditions of young drivers operating long-distance public transport services are well documented (Sinclair, 2013; Amweelo, 2016). Other factors such as the monotonous road environment and road design also play a crucial role in this regard as they play a crucial role on drivers' mental workload (Farahmand and Boroujerdian, 2018).

The study examined the occurrence of FSI road crashes by the week of the month. The study results showed that in the Namibian context, the highest occurrence of road crashes occurred during the second week after the pay week and the remaining weeks after that. It was expected for road user activity and risk factors to be more pronounced during the pay week as traffic activities are highest due to increased road use because of increased public and individual economic activities among communities. No studies exist in Namibia that attempt to explain the reasons why high crash frequencies are observed in the weeks after pay week, despite pay week usually being the most active week.

The temporal variation analysis of road crashes across the months of the year showed the highest frequencies of FSI crashes occurred around December, August and April to May respectively over the years examined. These high crash frequencies and casualties are mostly observed during August and December when festivities take place in Namibia. Many Namibians travel from the coastal and central regions towards the northern regions during the festive season (NRSC, 2012; Eggleston *et al.*, 2016; Nghishihange, 2018). As a result, high traffic volumes are observed on the national rural roads during this time, causing increasing likelihood of road crashes. The high crash frequencies observed in this analysis corroborate findings from previous studies that April to May, August and December are high risk months for road users on national rural roads (NRSC, 2012; Nambahu, 2018).

5.1.2. Driver risk factors and behaviour analyses

An examination of driver risk factors found that inattention, inadequate surveillance and dangerous manoeuvres were some of the predominant driver risk factors. However, these risk factors were found to be more prevalent among female drivers on rural roads than in male drivers. These risk factors are a possible indication of driver fatigue and can be attributed to the monotonous road conditions that exist on Namibian national roads, where long straight sections with few geometric changes exist between major towns. These monotonous road environments cause a highly predictable and dull driving experience, which often leads to boredom and trigger hazardous risk-taking behaviours by drivers. Researchers have also noted that such conditions generate physiological states that can worsen driver fatigue (Karlaftis and Golias, 2009; Gastaldi *et al.*, 2014; Farahmand and Boroujerdian, 2018).

Poor visibility and animals also emerged as leading primary contributing risk factors on the national rural roads. This finding was validated by results from previous studies which noted that animals on the national roads are one of the leading causes of road crashes (NRSC, 2012; Eggleston *et al.*, 2016; Nghishihange, 2018). Road crashes caused by animals are significant in the study because most of the national roads traverse through communal and commercial farm lands, with little to no barriers separating the roads from the animals. The animal-related crashes were observed to have occurred during the night hours (18h00 to 06h00). Visibility at night time is constrained, for that reason, the risk of getting involved in an animal-related crash is higher for drivers on these national roads.

Human-related errors strongly emerged in all the age groups. Even though this was self-evident, it served to corroborate the risk assessment and gave assurance that the application of risk was applied correctly. Traffic violations and dangerous manoeuvres were notable in the 18 to 35 age groups. There are common assertions as to why young drivers are more likely to be involved in road crashes. Prominent among these assertions is that young drivers are prone to crashes due to life-stage perceptions evident in other youth behaviours. As a result, human-centred factors that contribute to crashes are often pronounced in younger drivers (Blockey and Hartley, 1995; Stevenson *et al.*, 2001; Johnson and Jones, 2011; Adanu *et al.*, 2018). The elderly (greater than 65 years) were found to be prone to panicking or freezing in complex traffic environments and falsely assuming other road users' actions. The inability of older drivers to process complex road environments may be attributed to their reduced physical and physiological capabilities, which can cause slower processing and reactions to traffic situations that require drivers to act promptly (Hakkert and Braimaister, 2002; Huvarinen *et al.*, 2017; Cox *et al.*, 2017).

The combination of risk factors on crashes has not been studied extensively in Namibia nor internationally. The study analysis found that in FSI road crashes where intentional risk was the primary risk crash factor, a significantly high proportion of risk was contributed by level two and three risk factors. The results also evidently showed that other contributing factors played a significant role in road crashes where the primary crash risk factor was found to be roadway and environment related. The study results confirmed that road crashes are a result of a combination of several interrelated factors and are mostly not caused by a singular risk factor (Persia *et al.*, 2016; Ouni and Belloumi, 2019; Islam *et al.*, 2019).

5.2 Discussion of geospatial analyses and design compliance results

The study carried out a geospatial analysis to locate where road crashes occurred and to assess the specific patterns of distribution through heat map visualisations. The study applied the Kernel Density Estimation technique to analyse the distribution of FSI road traffic crashes on national rural roads. The study applied a bandwidth value of 1000 m to achieve the best hotspot visualisation on a grid size of 30 m by 30 m.

The results of the geospatial analysis on all the rural roads identified that the highest (extreme crash densities) occurrence of FSI crashes occurred on rural roads leading to or close to localities in central and northern Namibia. These localities form part of the socio-economic hub of the country, with high volumes of commercial vehicles observed on the rural roads daily. In addition, high traffic volumes are observed on the rural roads between central and northern Namibia over the holiday seasons. This predisposes the road users to higher crash risks due to the lengthy time spent on the road and the high traffic peaks.

On high order rural roads, higher crash densities were observed on the rural roads on the western part (coastal area) of the road network. The western part of the road network connects the coastal towns, which harbours the national ports and other high value commercial activities, to the rest of the country. These rural roads form an integral part of the Trans-Kalahari Corridor and the Walvis Bay-Ndola-Lubumbashi Development Road, formerly the Trans-Caprivi Corridor. The aforementioned rural roads accommodate a high volume mix of commercial and passenger vehicles traveling to and from the western part, to other parts of the Namibia and land locked countries. The high combination of commuters between these coastal localities, holiday-makers and long-distance drivers who are mostly predisposed to fatigue, creates an undesirable safety hazard for all users and can be attributed to the “higher” crash densities recorded on that part of the high order rural road network.

As expected, moderate crash densities were found on rural roads leading to and from the north-central parts of the high order rural roads. The localities on this part of the road network mostly serve the purpose of rest-stops for passenger and commercial vehicles en-route to their final destinations. Drivers driving from these towns are usually rested due to the pit-stops and are thus more aware of the complex traffic conditions that may occur on the road, to which they can appropriately react to.

Lower crash intensities were observed on rural roads in the southern, north-western and north-eastern parts of the high order road network. In the historical context of Namibia, only one high order rural road network was developed, running from the central part of Namibia to the southern borders. This road ran through sparsely populated areas and was primarily for the purpose of transporting commercial goods and resources from Namibia to South Africa. However, this status quo largely still remains. The high order rural roads on the southern part of the network serves mostly commercial vehicles, with little interaction with passenger vehicles. Despite the high volumes of commercial vehicles, the lower crash densities can possibly be attributed to the wider roads on these sections, which provide the larger commercial vehicles with more space to travel compared to other parts of the high order rural roads in other parts of Namibia. The north-western part of the road network, which is one of the largest regions in Namibia, remains one of the most under-developed (mostly low order roads) when it comes to the extent of high order roads in the regions. The low traffic volumes on roads in the north-western and eastern parts of the high order network could also contribute to the lower crash intensities recorded on these parts of the network.

On low order rural roads, extreme and higher crash densities were observed on rural roads mostly around localities in the north and north-eastern part of the country. This part of the country is densely populated with rural communities using the connector and low order unpaved roads to commute to urban areas, which house most basic services and markets (Starkey et al., 2017). The low safety conditions of these low order roads caused by high traffic volumes, especially during the rainy months over which most of the crashes on these roads were recorded, can be attributed to the extreme crash densities. The conditions on these roads are further exacerbated by the low visibility and poor road conditions that accompany the rains and the high number of domestic animals on the roads (Nghishihange, 2018; Jones et al., 2019).

The compliance assessment of the roads showed that the majority of the unpaved rural roads on the road network are not compliant with lane width design guidelines. The road lanes on unpaved roads were found to be narrower than the recommendations set out in the TRH 20. Similarly, the majority of unpaved shoulder widths on national rural paved roads that were examined were not compliant with the design guidelines. More worrisome, the compliance assessment also showed that only less than a quarter of the paved roads authorised to have paved hard shoulders were compliant, with a quarter of those having the appropriate width as recommended by the TRH 17. The sight distances on all the roads were found to be highly compliant with the set-out design guidelines in TRH 17. The SSD compliance is affirmed by the favourable effects that the SSD has shown in the crash predictive models developed and discussed in Section 5.3.

5.3 Crash predictive models (CPM) results

5.3.1. CPM results

The results of the crash prediction models developed provide a platform to further link and examine the impact of road and traffic characteristics on driver behavioural traits and their distribution across the national rural road network. Three crash models were developed to investigate the impact of road and traffic characteristics on FSI crashes. These models focused on road crashes on all the national rural roads (CPM1), high order rural roads (CPM2) and low order rural roads (CPM3) on the road network.

The combinational influence of road and traffic characteristics on the safety of road users has not been investigated extensively locally or internationally. The study developed a novel crash prediction model for all the national rural roads classifications. The study found several positive associations between road characteristics and fatal and serious injury crashes in all the CPMs developed. The vertical terrain was found to demonstrate a positive association to road crash occurrence, with an increase in the degree of hilliness causing an increase in crash occurrences on the rural roads. Several studies have found no significant correlation between hilliness as a single variable on the occurrence of road crashes (Bester and Makunje, 1998; Taylor *et al.*, 2002; Gitelman *et al.*, 2016). However, high correlations between hilliness and bendiness have been reported by researchers, where the combination has been found to lead to an increase in the frequency of road crashes. This increase has been attributed to poor coordination between the horizontal and vertical alignment, leading to poor driver perceptions and driving errors (Bester and Makunje, 1998; Walmsley *et al.*, 1998; Hanno, 2004; Laird *et al.*, 2010). This agrees with the study finding that hilliness is a significant contributor to crash risk in combination with other road design and traffic parameters examined.

The results from CPM 1 and CPM 2 found that an increase in the width of the travel lanes increased the occurrence of FSI crashes. On high order rural roads where lane widths were found to be mostly wide (LW > 3.5 m) with extremely narrow (SSW < 1.5 m) or no surfaced shoulder widths, drivers tend to select high operating speeds, with high levels of lateral lane deviations mostly observed. This has led to a higher same direction road crash frequency (National Road Safety Council, 2012; Nghishihange, 2018). The study findings can be attributed to the dangerous behaviour by drivers attempting to move to the narrow shoulder of the road to make way for faster drivers to overtake.

Another important finding of the study is the influence of the proportion of heavy vehicles (HV) and light vehicles (LV) in the traffic stream on the safety of road users on national rural roads. The novel models demonstrated that an increase in the proportion of both vehicle types increased the occurrence of FSI crashes. The effect contributed by this modal split can possibly be attributed to the speed differences between the different types of vehicles on the highways – speed variations. HVs are mostly always moving at lower speeds than the modus speed and this inadvertently exposes the entire traffic stream to higher crash risks, as the patience of other drivers may dwindle and possibly lead to dangerous road manoeuvres. Another possible contributing factor is that HV drivers usually have to drive long distances, as they deliver products from commercial hubs locally and regionally. These drivers are mostly predisposed to poor sleep quality and fatigue, due to the long hours they spend on national roads. In general, sleep plays an important role in physical and mental well-being (Bener *et al.*, 2017; Kwon *et al.*, 2019). Lack of quality sleep and severe fatigue are significantly associated with more frequent human errors (Aworemi *et al.*, 2010; Gastaldi *et al.*, 2014). The risk of road crashes involving both vehicle types may also be exacerbated by other factors including twilight and night time driving, weather and the high presence of animals on the national highways. As per the researcher’s best knowledge, no study exists investigating the impact of heavy vehicles on road safety locally.

The operating speed is fundamental to the development of any roadway facility through determining the appropriate design speed and subsequently the development of geometric design elements. The study found that operating speed demonstrated a statistically significant positive association to the occurrence of FSI crashes in all the CPMs. This finding suggests that a higher crash frequency is associated with higher operating speeds. The positive relationship can be attributed to the wider road lanes available on the national roads. The wider roads can give the driver the perception that they have enough space to correct their driving errors, therefore increasing the driver’s appetite for risk. Another possible factor that can lead to high speed selections is the monotonous road environment and long straight sections on the road network, which predispose the driver to risk-taking perceptions, as they perceive an adequate stopping sight distance from the road environment should any dangerous situation occur. The study findings corroborated previous studies on the impact of operating speeds on Namibian national roads (Ambunda, 2018).

The study found that the ground shoulder width (GSW) had dissimilar statistically significant influences on high order and low order national rural roads. On high order rural roads, the model demonstrated that ground shoulder widths have a positive association with FSI crash frequencies. This finding suggests that despite increasing the width of the ground shoulder on the road section, an increase in crash frequencies will be observed. This finding corroborates the design compliance findings, that there is a high presence of wrong shoulder types (unpaved) on high order roads. As a consequence, the high crash frequencies cannot be addressed by increasing the shoulder widths available to drivers, but rather by making available the correct shoulder type. On low order rural roads – mostly low volume paved roads or one lane gravel roads – the ground shoulder width demonstrated a negative correlation to the occurrence of FSI road crashes. This novel finding in the local context suggested that an increase in the width of the ground shoulder results in the decrease of crash frequencies. This finding confirms the finding that ground shoulder widths on low order rural roads are mostly non-compliant (existing GSW < recommended TRH 17 GSW) with design guidelines, and thus increasing the GSW could reduce crash frequencies. A possible contributing factor is that driver speed selections tend to be lower on roads with gravel shoulders due to visual cues (colour difference between the paved roadway surface and the gravel surfaced shoulder) that give a perception of a narrower driving lane. The finding on the impact of GSW on low order rural roads corroborates results from several international researchers (Zegeer V *et al.*, 1987; Gitelman *et al.*, 2019).

The surfaced shoulder width (SSW) demonstrated a negative association to the frequency of FSI road crashes on all the rural roads (CPM1). This means that increasing the width of paved shoulders on road sections results in a decrease in road crashes. This finding goes hand in hand with the design compliance finding, where existing SSW were found to be significantly non-compliant (existing SSW < recommended TRH 17 SSW) with design guidelines. International studies explain that drivers tend to select lower speeds on narrow travel lanes (LW < 3.2 m) with narrow surfaced shoulders (SSW < 1.5 m) due to the perception of lower safety. However, in the existing local context, the combination of narrow shoulders and wider travel lanes (LW > 3.5 m) provides a situation where drivers select high speeds due to a false sense of security and perceived space to correct driving errors. These actions are also confirmed in the appetite shown by drivers to make dangerous manoeuvres shown in Section 4.7.1.3. Despite the wider lanes, the narrow shoulder could also inadvertently lead drivers to steer away from the left shoulder and drive closer to the centre of the rural road (Liu *et al.*, 2016; Ambunda and Sinclair, 2019). In this case, the likelihood of head-on crashes increases significantly. The high head-on crash likelihood is also confirmed by crash statistics from the Namibian National Road Safety Council (NRSC, 2012).

5.3.2. Compliance impact on CPMs

The study investigated the sensitivity of the models to the compliance of the design parameters to the TRH 17, TRH 20 and TRH 26. To the researcher's best knowledge, no local or international study exists examining the aspect of how road and traffic design fundamentals impact rural road safety. The novel findings from the sensitivity analysis are discussed in this section.

The sensitivity analysis on the model developed (CPM 4) for all rural roads found that the operating speed demonstrated an amplified influence, discussed in Section 5.3.1, on the occurrence of FSI crashes. It is important to note that speed will always play a key role in the functioning of a road. It is expected that should all design parameters ideally comply with design guidelines, parameters such as the operating speed and other road environment and land use factors will play a key role in the safety of the roadway. More interesting, the sensitivity analysis found that the proportion of heavy vehicles in the annual average daily traffic ($b^*_{MLR\text{-}CPM\ 1} = 0.464$ to $b^*_{MLR\text{-}CPM\ 4} = -0.380$) demonstrated completely opposite effects (coefficient b* estimates) on the occurrence of road crashes. The sensitivity analysis also proved that several direct design parameters – the lane widths and surfaced shoulder widths – will exhibit a reduced or no statistically significant influence on road crashes due to the combined effect of compliant parameters to design guidelines. The expected improved driver perceptions could also play a significant role in the reduction of road crashes as a result of “better communicating” parameters. On the other hand, as a result of compliant covariates, the proportion of paved shoulders and the ground shoulder widths were found to significantly influence crash occurrence. An increase in the ground shoulder widths was observed to lead to a decrease in crash rates. The opposite effect (increase in crash rates) was however observed as a result of increasing the proportion of paved shoulders.

A detailed analysis of the results on high order rural roads (CPM 5) shows that the sensitivity test amplifies the influence of the operating speed and the vertical terrain – hilliness on FSI crash rates. The operating speed parameter was also observed as showing an increased influence in CPM4. The proportion of heavy vehicles in the traffic stream ($b^*_{MLR\text{-}CPM2}$ = 0.682 to $b^*_{MLR\text{-}CPM5}$ = -0.594) demonstrated a change in effect on road crashes. Unsurprisingly, the sensitivity test caused several of the direct-design parameters to lose effect (statistical significance) on road crash frequencies. These parameters are: (1) the lane width and (2) the ground shoulder width. Similar to CPM4, an increase in the proportion of paved shoulders on the higher order roads was found to result in an increase in road crash rates, due to design compliance. This result is rather surprising as the compliance test on the high order roads indicates that the majority of these roads do not have the appropriate shoulder types to accommodate the observed high traffic volumes and expected high traffic speed selections by drivers. A number of factors could explain this correlation between shoulder types and crashes. An increase in the proportion of a paved shoulder combined with wider lane widths may result in perceived space to correct errors and thus higher speed selections and in-lane deviations. This however increases the risk for run-off crashes. Also, drivers may decide to use the hard-paved shoulder as an “extra” lane to give space to vehicles making overtaking manoeuvres in the traffic stream. This unacceptable practice can present dangerous situations for other drivers, especially when combined with factors such as night-time driving, non-compliant ground shoulder widths and high traffic speeds. Several studies have investigated the impact of present shoulder types on road sections, without delving into whether the appropriate shoulder type is provided (Stamatiadis *et al.*, 2009; Sisiopiku, 2011; Ambunda and Sinclair, 2019).

The sensitivity test results on higher order roads indicated that increasing the extent of bendiness resulted in a decrease in the frequency of road crashes. In the local context, conditions are such that long straight sections in monotonous road environments are prevalent on the road network (Adanu *et al.*, 2020; Ambunda and Johannes, 2020). These sections can predispose rural road drivers to fatigue-related crashes. Fatigue can affect driving skills by increasing the frequency, amplitude and variability of errors (Dagli, 2004; Bener *et al.*, 2017). Therefore, the model findings explain that increasing the bendiness, which indirectly leads to an increase in the level of driver engagement in the driving process, may reduce road crashes due to monotonous environment-related fatigue.

The sensitivity test on low order rural roads (CPM 6) had a significant impact on the effect that the surfaced shoulder width and the hilliness of the vertical alignment have on crash occurrence. Both parameters lost their statistically significant influence on crash frequency, due to the combined effects of design compliant parameters. The proportion of light vehicles demonstrated a change in effect while the ground shoulder width on the road sections demonstrated an increased absolute effect on crash frequency. Similar to CPM4 and CPM5, an ideal design environment on lower order roads resulted in the proportion of paved hard-shoulders demonstrating a statistically significant positive association to crash rates. The stopping sight distance (SSD) was found to exhibit “some” influence in the sensitivity test, though statistically insignificant. The statistical insignificance of the SSD is expected due to the road environment on the rural road network – mostly flat terrains and long road sections with high levels of forward visibility.

5.4 Two-Step Cluster analysis model results

The TSC analysis findings are novel in their nature as they inform on the impact of several predictors in the rural road environment on the nature and combination of risk factors preceding a crash occurrence. The results of the novel TSC analysis models further strongly add to the importance of investigating the impact of all potential risk factors that impact the occurrence of road crashes on national rural roads. The TSC model identified three (3) clusters categorising the impact of the predictors on the dependant variable (risk factors). The three-cluster solution was identified as the "best" solution. The "best" solution was a result of the TSC-2 model exhibiting the lowest AIC value and the largest ratio of AIC changes and ratio of distance measures. The cluster groups developed by the TSC analysis demonstrate that covariates have different impacts on the cluster crashes, depending on the level of rural road classification and to a certain extent, the purpose of the trip. Therefore, the TSC model presents a synergy between the crash analyses carried out on the demographic, temporal, road and traffic characteristics and their impact of crash causation risk factors identified in the crash dataset.

In cluster one (1), the TSC analysis found a relationship between the characteristics of the two-lane road during weekends (Friday to Sunday) and numerous combinations of crash causation risk factors. The crash records in cluster 1 represent drivers using high order roads (HORR) to travel long distances, in this case, possibly holiday-makers and long distance private and commercial drivers. The majority of drivers in cluster 1 were found to be prone to the combination of recognition, decision and intentional crash risk factors. A second class of drivers in cluster 1 was also found to be highly prone to the combination of only recognition and decision risk factors. These risk factor combinations confirm the high impact of the primary level contributing factors on crash occurrence, identified in Section 4.2.3. These primary level contributing factors, as part of the main risk factor groups, include inadequate surveillance, inattention, false assumption of other drivers' actions, fatigue and dangerous manoeuvres. The high number of crash records on HORRs found in cluster 1 is indicative of the impact and role played by features of the HORR environment (see Section 4.5.3) on crash occurrence and crash risk factor combinations identified in the TSC analysis.

A high number of crash records in cluster two (2) were found to have occurred on unpaved low order rural roads (LORRs). The poor surface conditions on the unpaved roads were identified to have a high impact on driver risk factor combinations during the weekends. The high number of trips made in rural communities on unpaved rural roads to commercial regional centres over the weekends are a key contributor to the high number of crashes recorded on these roads. The TSC analysis demonstrates that the interaction of these covariates contributes to several high-risk factor combinations. In the same way to road crashes in cluster 1, the combination of recognition, decision and intentional risk factors was the highest observed risk factor combination preceding crashes in the cluster. The TSC analysis also found that drivers in cluster 2 were also highly prone to risk factor combinations involving the roadway and environment, such as (1) the combination of recognition, decision, and roadway and environmental risk factors, and (2) the recognition and roadway and environmental risk factors. The findings of the TSC analysis on the impact of roadway and environment risk factors are reinforced by findings from several studies and reports on how animals, as part of roadway and environmental risk factors, are one of the leading causes of road crashes on the Namibian national rural roads (NRSC, 2012; Eggleston *et al.*, 2016).

The TSC analysis identified several road crashes that occurred on paved lower order rural roads (LORRs) over the weekdays (Monday to Thursday) – cluster three (3) records. The records in cluster 3 possibly represent drivers using the LORRs to commute or travel between several small towns every day. Similar to the two previous TSC clusters discussed, the combination of recognition, decision and intentional risk factors emerged as the highest combination of risk factors preceding crashes on paved LORRs. Drivers using paved LORRs were also found to be highly prone to recognition, roadway and environmental risk factor errors during the weekdays, stemming from the interactive covariate relations discussed and modelled in Section 4.5.3.

Chapter 6: Conclusions

6.1 Introduction

The study developed novel road crash predictive models and two-step analysis clusters that explored the interactive relationship between road characteristics on national rural roads, demographic and temporal factors, fatal and serious road injury (FSI) crashes and driver actions and risk factors preceding crashes. A road crash occurs when there is failure in the road traffic system at multiple levels, therefore, the study explored driver behaviour by creating an unprecedented analysis to inform on driver behaviour and risk characteristics on national rural roads with specific characteristics in the Namibian road and traffic environment. Human factors are globally affirmed as leading crash risk factors, they are however unpredictable and difficult to directly control. This study has created a basis on which driver behaviour on national rural roads can be directly influenced to some extent, through investigating the roadway characteristics.

The study objectives were five-fold. The first objective of the study was to examine road crash profiles and factors attributed to rural road crashes. The goal of this objective was to create a new basis to assess the relationship between road characteristics and driver risk factors preceding road crashes – a two step cluster analysis. This will serve as a basis for comparison for any future studies.

The second objective was to identify high risk traffic crash locations on the different national rural road classifications. The third objective was to assess how the spatial analysis varied in the distribution of FSI crashes across the national rural road network. The second and third objectives aided in the understanding of how population characteristics and road design guidelines compliance (fourth objective) influenced the distribution of FSI road crashes across the rural road network. The fourth objective was to investigate the compliance of the rural road design characteristics with road design guidelines. Recommendations on the suitability of the design standards are based on the first three and the fifth objectives of the study.

The fifth objective was to develop novel road crash predictive models in the context of the Namibian national rural road environments. This objective is underpinned by the other four objectives in examining the spatial distribution of the road crashes, the response of crash distribution to design compliance levels and the sensitivity of the novel CPMs to changes in design characteristics. The fifth objective also provides a basis to examine how the sensitivity of CPMs to design characteristics affects driver risk factors on the national rural roads.

Therefore, this chapter presents a summary of the study by talking to the points below:

I. Summary of key findings of the study.
II. Summary of original contributions and practical implications of the study.

III. Model transferability.
IV. Discussion of key challenges of the study.
V. Future research and developments.

6.2 Key findings of the study

The main findings of the study are summarised below with reference to the stated study objectives.

6.2.1 Study objective one

The study applied various analytical methods that demonstrated a multitude of relationships between the characteristics of the driver population and road crash incidences on national rural roads. This objective was focused on examining driver risk factors on the current road environment and finding out whether a measurable link exists between the driver population characteristics and the high severity road crash dataset used in the study.

1. The study examined how population characteristics (driver gender and age) and temporal variations are distributed in the fatal and serious injury road crash dataset. These interactive demographic factors are interrelated with the driver risk factors. The average age of the driver population was found to be 28 years, with the male driver population at a higher risk of being involved in FSI crashes than females. The male driver population group crash risk pointedly began to increase from the age group of 21 to 25 years (young adult) and peaked at the age group of 31 to 35 years (adult). The young adult driver population group demonstrated the highest ratio of male to female crash risk. Young male adults were more than ten times predisposed to risky situations than their female counterparts.

2. The study found an annual fatal and serious injury road crash rate of 21.3 road crashes per 100 000 population. The crash rate was found to be slightly lower than the road fatality crash rate reported by previous studies in Namibia.

3. The analysis on the temporal variation and distribution of road crashes found statistically significant t-test relationships between the FSI crashes and the following temporal distributions:
    i. Time of the day – higher crash occurrences during the peak traffic hours
    ii. Day of the week – higher crash occurrences over weekends and holidays
    iii. Week of the month – higher crash frequency over the 2nd week after pay week and all other weeks of the month
    iv. Month of the year – higher crash frequencies over the holiday months (April to May, August and December)

v. Yearly quarters – higher crash frequencies observed over the first and third quarters of the calendar year

4. The driver risk factor and behavioural characteristics analysis revealed that inattention, inadequate surveillance and dangerous manoeuvres were prominent risk factors among the driver population. These human-centred risk factors were found to be more pronounced in the young adults and adults, which encompassed drivers aged 18 to 35 years. The study applied the Two-Step Cluster analysis technique to explore the relationship between the combination of human-centred risk factors preceding the occurrence of a road crash and the demographic, temporal, and road and traffic environmental factors on Namibian national rural roads. Several risk factor combinations were identified as playing key roles in crash occurrences on high and low order road classifications. The study revealed that human-related factors play a key role in crash occurrences where road and environmental factors were the primary risk contributors. This is an indication of the inter-relationship between the crash risk factors and that no single factor is responsible for a road crash.

6.2.2 Study objective two and three

The study applied the planar Kernel Density Estimation (KDE) geospatial analysis technique to detect clusters of FSI road crashes across the national rural road network. The KDE was applied on three data sets predicated on the classification of rural roads (high and low order roads). The KDE with a bandwidth value of 1 000 m over grid sizes of 30 m by 30 m achieved the best hotspot visualisation of the FSI crashes. On high order rural roads, the KDE technique highlighted higher to extreme FSI crash clusters were observed on roads leading to and from localities in the northern, central and western parts of the network – these localities are situated primarily along some of the most active trade routes across the country and thus experience an abnormally high amount of diverse road users. On lower order rural roads, extreme to higher crash densities were observed on roads in the northern regions of the country, on rural roads around localities in rural and peri-urban areas. These parts of the country are densely populated and have a high dense network of connecting low order unpaved roads with high traffic volumes.

6.2.3 Study objective four

The compliance to road design guidelines (TRH 17, TRH 20 and TRH 26) assessment revealed that the majority of the unpaved rural low order roads are not compliant with lane width design standards – the combination of non-compliant lane width and high traffic volume on unpaved roads is best visualised in Figure 4.23, where higher to extreme crash clusters are observed on the northern zone of the national rural road network. In the same way, the majority of unpaved hard shoulders on paved rural roads assessed were not compliant with design standards set out in the TRH 17, to the detriment of road safety on these roads. A much less desirable revelation from the compliance assessment was that less than a quarter of the paved roads had the appropriate hard shoulder, with less than a quarter of those having the appropriate width for the traffic conditions on those roads.

6.2.4 Study objective five

The study developed and calibrated multiple novel crash predictive models (CPMs) as tools to examine the relationship between the road design and traffic environment and the frequency of FSI road crashes in the Namibian context. The model measures of goodness-of-fit indicated that the General Multivariate CPMs were the best performing models suitable for the FSI crash datasets on both high and low order rural roads. The developed CPMs were further used to carry out a sensitive analysis on the design parameters applied in the study. The sensitivity analysis indicated that despite applying design compliant parameters in the models, several model covariates demonstrated more pronounced effects detrimental to the safety of the road system – indicative of how important an assessment of the much deep-seated over-reliance on “international” design guidelines is needed, in an effort to localise guidelines to suit the road environment and behavioural characteristics of road users. Inferences on the suitability of the road design standards applied in Namibia are underpinned by the design sensitivity analysis using the CPMs.

6.3 Applicability of CPMs

The crash predictive models (CPMs) were calibrated and tested for the Namibian national rural road environment. The models provide a road crash risk assessment tool that relies on road characteristics and traffic information across the national road network. The CPMs allow for the identification of design parameters that may pose a hazard to the safety of road users. Furthermore, the CPMs present an opportunity to examine the impact of the road environment on human-centred crash risk factors by comparing the changes that may result from road characteristic improvements through, before and after studies. The replication and applicability of the models at an aggregated level in countries with similar rural road environments will need to be investigated further, as the models are predicated on road design and traffic data, which requires an extensive amount of time to collect and a comprehensive road management system.

6.4 Summary of contributions and practical implications

6.4.1 Key contributions

The study has developed a novel tool for road safety assessment in Sub-Saharan Africa and beyond, underpinned by design and traffic data on the Namibian rural road environment. The study is novel in the following ways:

i. The type of study is undoubtedly rare globally and novel in the Namibian context in that the safety analysis has included a large number of design and traffic related parameters describing the rural road environment in Namibia. This approach is underpinned by the understanding that road elements work in tandem and as such it is important to consider the impact of the road environment on road safety as a whole.
ii. The study has developed a methodology that comprehensively improved the quality of the historical road crash data and allowed for the aggregation of crash data with road characteristic data at a macro road network level for safety analysis. In the same way, the study developed a novel approach to identify driver crash risk factors at multiple levels linked to road crashes on rural roads.
iii. The study has primarily understood that road elements are designed taking into consideration average road user behaviour and traffic conditions and that driver behaviour is a direct result of how drivers perceive the road environment. The study has thus informed on driver crash risk factors and behaviour during the study period, by applying the Two-Step Cluster analysis technique to explore the relationship. This has allowed for a novel link to be developed between several combinations of human-centred factors, the road environment, temporal and demographic factors. Human factors are very difficult to predict and directly address. As a result, one way to impact these factors is through examining the roadway environment. This study has thus formed the basis on which future comparisons of the impact of several key factors including demographics, temporal and road characteristics on driver behaviour and risk factors can be built.

The study developed FSI road crash predictive models that will be useful in forecasting future road crash occurrences using comprehensive design and traffic parameters datasets. These CPMs also represent a tool that is significantly able to explore the nature and magnitude of the relationship between the rural road environment and FSI road crashes at a macro road network level.

6.4.2 Practical implications of the study on road safety

The insights from the study will have a long-standing significant impact on rural road safety in Sub-Saharan Africa (SSA) and beyond. The study is one of the few compositions of literature that has greatly contributed to the knowledge gap that exists on road safety studies on rural roads and has significantly improved the understanding of the combinatorial effects of road design and traffic attributes on rural road FSI crashes. The study explored the role of driver crash risk factors in the rural road safety system and has built a foundation on which the sensitivity of crash risk factors can be tested against changes in road parameters. The study has highlighted multiple areas in the rural road safety system that urgently need to be addressed to provide a safer environment for road users on the network. As Namibia prepares the new Decade of Action (DoA) Strategic Plan for the year 2021 to 2030, the insights from the study provide a backbone on which rural road safety can be addressed in the DoA, with an approach that is aimed at reducing and eliminating so-called latent gaps in the components of a safe road system.

The new DoA strategic plan for the period 2021 to 2030 is developing its strategies (performance indicators and targets), based on the principles of the five (5) road safety pillars designed to guide strategic planning and action and anchored on the twelve (12) United Nations (UN) global road safety performance targets (Peden *et al.*, 2017; Olivier, 2020). The new DoA plan is also in line with the eight (8) African Union (AU) guidelines / cross-cutting issues (CCIs) guided by the Sub-Saharan Africa Transport Policy Program (SSATPP) (World Bank, 2012). The CCIs focus on road safety in rural areas. A study by the African Union in 2018 had found that a majority of African countries, including Namibia, have taken only very minimal steps to implement the recommended rural road safety activities – states should undertake rural road safety audits, ensure that safety audits are taken into consideration in the design and construction of rural roads, improve rural transport safety through mixed transport measures and sensitise road users using national rural roads on road safety issues (African Union, 2018).

I. Global road safety performance pillars and cross-cutting issues

The five global road safety pillars and AU guidelines provide a good understanding of the areas where insights from the study will enhance road safety performance. The road safety pillars and cross-cutting issues are presented in Figure 6.1 and Figure 6.2 respectively. The target areas directly impacted by the study insights are highlighted in red in both figures and discussed thereafter.

The 8 AU Cross Cutting Issues (CCI)
CCI 1: Improved awareness of road safety
CCI 2: Strengthened capacity of road safety stakeholders at national and regional level
CCI 3: Crowding - in private sector participation in road safety
CCI 4: Digitisation of road safety management
CCI 5: Effective monitoring and evaluation of road safety
CCI 6: Rural road safety
CCI 7: Urban road safety
CCI 8: Safety of regional corridors

Figure 6.1 The 8 AU cross cutting issues

The 12 UN Global Road Safety Targets
Pillar 1: Road Safety Management
Target 1: By 2020, all countries establish a comprehensive multisectoral national road safety action plan with time-bound targets.
Target 2: By 2030, all countries accede to one or more of the core road safety related UN legal instruments.
Pillar 2: Safer Roads & Mobility
Target 3: By 2030, all new roads achieve technical standards for all road users that consider road safety, or meet a three-star rating or better.
Target 4: By 2030, more than 75 percent of travel on existing roads is on roads that meet technical standards for all road users that take into account road safety.
Pillar 3: Safe Vehicles
Target 5: By 2030, 100 percent of new (defined as produced, sold or imported) and used vehicles meet high quality safety standards, such as the recommended priority UN Regulations, Global Technical Regulations, or equivalent recognized national performance requirements.
Pillar 4: Safe Road Users
Target 6: By 2030, halve the proportion of vehicles travelling over the posted speed limit and achieve a reduction in speed-related injuries and fatalities.
Target 7: By 2030, increase the proportion of motorcycle riders correctly using standard helmets to close to 100 percent.
Target 8: By 2030, increase the proportion of motor vehicle occupants using safety belts or standard child restraint systems to close to 100 percent.
Target 9: By 2030, halve the number of road traffic injuries and fatalities related to drivers using alcohol, and/or achieve a reduction in those related to other psychoactive substances.
Target 10: By 2030, all countries have national laws to restrict or prohibit the use of mobile phones while driving.
Target 11: By 2030, all countries to enact regulation for driving time and rest periods for professional drivers, and/or accede to international/regional regulation in this area.
Pillar 5: Post-crash Response
Target 12: By 2030, all countries establish and achieve national targets in order to minimize the time interval between road traffic crash and the provision of first professional emergency care.

Figure 6.2 The 12 UN Global road safety performance targets

The study will assist in the success of the following AU cross-cutting issues (CCIs) 1, 2, 3, 4 ,5 ,6 and 8 highlighted in Figure 6.1 and the global road safety pillars – Pillar 1, Pillar 2 and Pillar 4, highlighted in Figure 6.2. The pillars and CCIs are discussed below:

Pillar 1 and CCIs 2,3, 4 and 5: Road safety management, strengthened stakeholder capacity, private sector participation, digitisation and effective monitoring and evaluation of road safety

Road safety management serves as the key pillar on which the other four global road safety pillars are based. Therefore, a comprehensive road crash data collection system through which data is collected regularly, disseminated and used to improve the effectiveness of road safety measures. The study provided a challenge in that three road crash databases exist in Namibia – MVA crash data, NRSC data and police reported data. This provided a challenge in addressing the deficiencies from each individual database and aggregating the data while avoiding duplications. This challenge was indicative of the need for strengthened and effective stakeholder collaborations and partnerships. Other limitations such as missing records, crash locations – or rather a lack thereof - and inaccurate crash records also presented a daunting challenge to the application of the databases in the study. The study developed a method to address the deficiencies in the databases and significantly improved the quality and management of the data. These challenges highlighted the importance of having a centralised and well managed high-quality crash data centre geared towards driving data centred road safety decisions and actions. The method applied in the study could be used as reference on how important high-quality and comprehensive databases are in encouraging innovative and value-adding road safety investigations. The novel crash prediction models (CPMs) and two step analysis clusters (TSC) developed in the study will help in the formation of pro-active safety management systems that are geared towards identifying potential deficiencies in the safety of road users. This modern approach will lead to a more digitised and effective road safety monitoring system. This digital approach will bring about a better synergy between the development and implementation of road safety policies, as decisions are more "data proof".

In order to create and enable an environment in which insights from this study will be best applied and effective, the Law Reform and Development Commission (LRDC) of Namibia undertook the review of existing road safety laws[24] and development of a new Road Safety Management Bill (RSMB) in 2018 to bring about uniformity and effective cooperation among road safety stakeholders (LRDC, 2018). The RSMB will culminate in the development of a New Decade of Action (DoA) road safety strategy for the period 2021 to 2030 in Namibia.

[24] Road safety act No. 29 of 1992

Pillar 2 and CCI 6 and 8: Safer roads and mobility, rural road safety and safety of regional corridors

The second pillar focusses on safer roads and road environments (new and existing roads), for all road users, based on high technical / design standards with reference to road safety and incorporating safe system principles. Pillar 2 is also in line with the AU cross cutting issues 6 and 8. The road design and traffic characteristics play a crucial role in the behaviour of drivers on national rural roads. Design principles are thus crucial to achieve a road environment that is cognisant of road safety and expands on the idea of the interaction between humans (drivers) and road factors.

The study carried out a crash hot spot analysis for the difference classification of national rural roads – high order[25] and low order roads[26]. The crash hotspot analysis is crucial in that it allows for stakeholders to develop targeted remedial measures and prioritise road safety funding. To further expand on the identification of hazardous and potential hazardous road sections on rural roads, the study carried out a road design standards compliance assessment. The design compliance assessment identified areas on the national roads with non-compliant design elements. The distribution of the FSI crashes across the rural road network in reference to the non-compliance of design elements carried out, expanded on the relationship between road design principles and road safety – determining potential defects in the perception of the road and surrounding environment by a driver, which may lead to erroneous actions and running the risk of a crash incident. The design compliance assessment identified numerous design deficiencies, which the design non-compliance geospatial analysis found to be detrimental to the safety of national road users. These design shortcomings in the system will need to be urgently addressed to address the high FSI rate on Namibian national rural roads.

The study developed novel crash prediction models (CPMs) for the various road classification. Using the existing rural road design data as the key cog in the CPMs, the models developed are intended to supplement and potentially replace road safety traditional tools, as their application and insights will further expand the stakeholder's ability to determine road sections with potential crash risk and eliminate the risk for road users. These CPMs are novel in that they determine and quantify the operational characteristics of the roads and identify elements which do not comply with the function of the roads and therefore disorient the drivers, causing a breach in the smoothness of psychological perception of the road and creating an element of surprise and ambiguity on the road.

Inherently, roads designed according to suitable design principles should absorb the potential risk that other road users could pose by adhering to sustainable safety principles. Design principles enable road characteristics to play a clear role in guiding drivers of all categories as to the type and

---

[25] High order rural roads classification includes R1 to R3 TRH 26 classified roads

[26] Low order rural roads classification includes R4 to R6 TRH 26 classified roads

function of the road, as well as inform on the level of risk that they should prepare for. Road designs need to create the right impressions to solicit expectations from all drivers. Design and planning authorities should therefore consider spatial knowledge, the skills and awareness of road users that develop over time and facilitate the development of skills, hazard and risk perception, inter alia, manoeuvring in relation to the road characteristics, estimation of vehicle speeds and the ability to judge and accept gaps.

With this in mind, the novel CPMs developed were used to carry out a sensitivity analysis using the design standards that were applied in the design of most of the national rural roads, to test how the model parameters would react to potential remedial design measures and indirectly test the level of safety incorporated into the design principles. The insights from the sensitivity analysis were unnerving and pointed towards the application of remedial measures on the rural roads and revision of some of the design principles used. Also, it is important to note that the capacity of some of the roads has been far exceeded over the course of the years. This emphasises the urgency to audit some of these roads and apply findings from the study towards developing a safe system for current and future road users. In summary, the novel CPMs provide a crucial opportunity and step towards building a crash risk control system that embraces all crash risk factors throughout the life cycle stages of the roads.

Underpinned by the understanding that road environment, temporal and demographic factors do not influence crash occurrences in silos, the study explored how the combination of human-centred risk factors preceding road crashes are influenced by multiple factors, including the road and traffic environment. This provides valuable new information on how the safety of rural roads depends on addressing multiple perceptive and behavioural issues triggered by the environment in which drivers find themselves.

Pillar 4 and CCI 1 and CCI 6: Safe road users, improved awareness of road safety and rural road safety

In Namibia, and most countries, the cause of a road crash with combined crash risk factors of human related errors and the road environment, is mostly blamed on the driver's inability to control the vehicle. This solely puts the fault of the crash on the driver whether they consciously or subconsciously committed an error that led to the crash. It is however important to recognise that road crashes occur as a result of a combination of factors, among which the driver's ability plays a fundamental part.

In light of this, the study carried out and reported on driver crash risk factors and behaviours. The study applied a multi-level approach to the identification of crash risk factors. This was done with the understanding that road crashes are potentially caused by a multitude of interrelated factors. The insights from the primary analysis showed that human-related errors were the most predominant risk factors in FSI crashes. The main categories of human failures revolved around driver's ignorance towards safe driver behaviour norms (intentional risks) and their inability to recognize complex traffic and road situations, which could be exacerbated by the "tricky" road environment (several high design non-compliance levels). The findings also largely identified high levels of manoeuvring failures coupled with high impatience levels among drivers on the rural road network – a combination of decision and performance related errors.

The second level analysis was geared towards identifying factors leading to a crash and allowing for a more relational assessment of the crash risk factors. The study identified with the second level analysis that the highest risk to FSI crashes was posed by significant levels of ignorance to principles of safe driver behaviour. The overall results indicated that crash risk factors related to the roadway ranked the second highest in the primary analysis and the more nuanced secondary level crash risk factor analysis. In crashes where roadway factors where the primary causation factor, the human-related responses were remarkably high. This was indicative of the significant relation between these factors and how addressing roadway deficiencies could have a possible significant impact on the more unpredictable human-related issues. The other factors that proved to have a considerable impact on crash incidences were decision errors and performance errors.

The study findings are clear on the significant role played by the human-related and roadway crash risk factors on FSI crashes in Namibia. The adolescent (less than 18 years) and young adult (18 to 25 years) age groups were particularly prone to performance and intentional crash risk factors. This finding raises the importance of basic driver traffic safety training during the early phases of licensing to ensure that drivers are well versed in safe driving and behavioural practices. Inattention and inadequate surveillance (recognition errors) were found to be more prevalent among the adult driver population – 25 to 35 years. These type of errors could be attributed to the thought process of young adults who believe they have enough driving experience and feel that they are not prone to driving errors compared to most age groups.

The study then developed Two-step Cluster models to discern how multiple factors, explored in the study, impact on the interaction between the different level driver risk factors prior to a crash occurring. This analysis sought to enhance the understanding of underlying factors that influence the safety of road users and by extension, serve as an illustrative example for an analysis of similar data within the context of countries with similar road and traffic conditions. The insights from the study also raise the importance of using educational initiatives to constantly inform and increase awareness among all driver categories of the risks associated with poor driving behaviour.

It is recommended that the crash reporting system applied in Namibia should fall in line with international best practices. This allows for the recording of specific crash risk factors associated with a crash. This will present an opportunity for the development of new, improved and safer road user practices, introduced and enforced to sensitise and improve road safety on rural roads, using quality data as a backbone.

II. Harmonisation of the study insights, pillars and guidelines in a safety strategy

The strategy of the study, aided by international experience and best practices is focused on making Namibian rural roads safer and inspire safer road users through positive attitudes and behaviours towards good safety practices. This strategy is aimed towards reducing the frequency and level of severity of crashes on national rural roads. Pillar 2 and Pillar 4 are particularly aimed at reducing driver crash risks and preventing crashes. These two core pillars, together with other pillars (Pillar 3 and Pillar 5) have to be supported by effective and efficient road safety management. This foundation should be built on Pillar 1, which is geared towards making sure institutional arrangements are in place to provide an enabling environment for road safety programmes to take off. An enabling environment will involve a system designed to house road safety responsive legislation, ensure sustainable funding, promote good stakeholder collaborations and partnerships, and an effective integrated road safety management system. The strategy is presented in Figure 6.3.

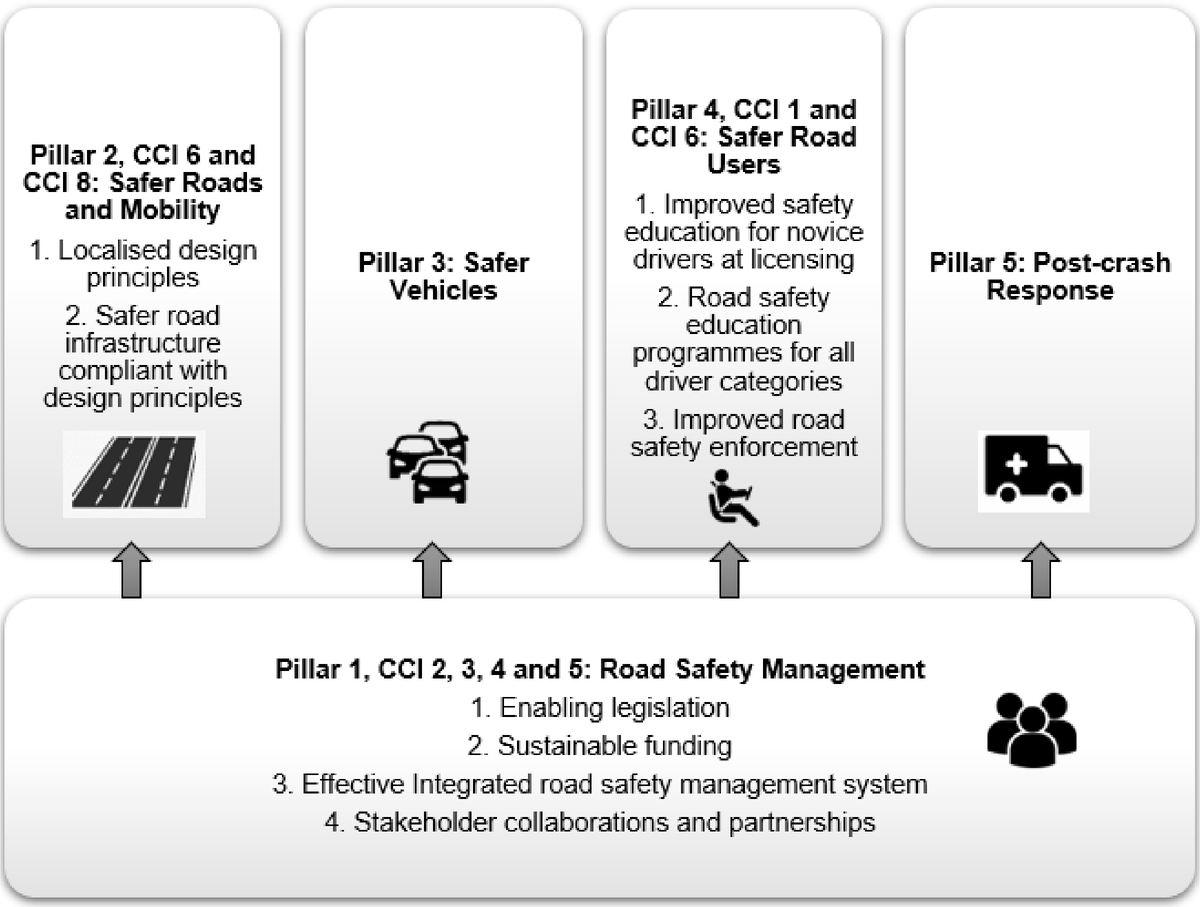


Figure 6.3 Road safety strategy map

6.5 Discussion of challenges

Several data limitations were observed in the study:

1. **Missing crash information and incorrect crash data information**: Some of the crash records in the database were missing critical information on the description of the road crashes, inter alia, the type and number of casualties involved. Crash records with missing data needed for the study were removed for the analysed database due to their quality deficiencies. This hindered the level and quality of analysis that could be carried out to determine driver crash risk factors and behavioural traits.
2. **Missing crash location information**: Several crash records did not have a description of the crash location and the database was not georeferenced. For crashes where the location could not be determined from the description of the location in the records, the record had to be removed. The study observed that 21 percent of all the crash records could not be located due to the missing location descriptors. This represented a high number of crash records that would help to improve the level of detail and quality of the geospatial and CPM analysis.

6.6 Future research

The study has developed novel crash prediction models in an attempt to address the knowledge gap that exists in the investigation of combinational road elements on road safety in Sub-Saharan Africa and globally. The study has primarily built a foundation for investigating the impact of road characteristics on FSI road crashes and has determined the driver crash risk factors linked to these characteristics. With that in mind, the future research and development are suggested below:

1. As the CPMs were developed, they were referenced on the Namibian crash data set. It will be imperative to test the transferability of the models to countries with similar road conditions to the Namibian national rural road network, where long straight sections exist between towns.
2. A before and after study on the impact of road design changes on driver risk factors on national rural roads will help to garner more insights into the relationship between the road traffic elements in the CPMs and driver crash risk factors. Such a study will improve the understanding of stakeholders on the effectiveness of the remedial measures applied and will improve decision making in the formulating of road safety guidelines and policies.
3. In an effort to improve proactive road safety measures and improve road safety management, real time monitoring of the road environment through the application of CPMs built on historical crash data could be used to identify potential hazardous road sections and proactively move towards the potential reduction and removal of crash risks for road users in the road system. Real-time crash prediction will present a huge opportunity to test the application of the novel CPMs using real-time road design and traffic data in crash prevention.

4. Applying immersive technologies, virtual and augmented reality, quantifying the perceptive impact that the road and traffic environment has on drivers through controlling environments, and further investigating and presenting innovative applications for crash causation risk factors, prevention and education among drivers on different levels of national rural road classifications. Applying immersive technologies to explore the risk factors on different road classifications may prove beneficial to addressing crash causation knowledge gaps that exist in road safety.

References

AASHTO (2010) *Highway Safety Manual*. 1st edn. Washington DC: AASHTO.

AASHTO (2011) *A Policy on Geometric Design of Highways and Streets*. Fouth. Washignton D.C: AASHTO.

Abdulhafedh, A. (2017) 'Road Traffic Crash Data: An Overview on Sources, Problems, and Collection Methods', *Journal of Transportation Technologies*, 07(02), pp. 206–219. doi: 10.4236/jtts.2017.72015.

Abele, L. and Møller, M. (2011) 'The Relationship between Road Design and Driving Behavior', *RSS 2011: Road Safety and Simulation 2011 Conference*, 2012, pp. 1–16. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:THE+RELATIONSHIP+BETWEEN+ROAD+DESIGN+AND+DRIVING+BEHAVIOR#1%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Relationship+between+Road+Design+and+Driving+Behavior%231.

Achcar, J. A., Achcar, A. and Martinez, E. Z. (2013) 'Robust Linear Regression Models: Use of a Stable Distribution for the Response Data', *Open Journal of Statistics*, 03(06), pp. 409–416. doi: 10.4236/ojs.2013.36048.

Adanu, E. K., Penmetsa, P., Jones, S. and Smith, R. (2018) 'Gendered Analysis of Fatal Crashes among Young Drivers in Alabama, USA', *Safety*, 4(3), p. 29. doi: 10.3390/safety4030029.

Adanu, E. K., Penmetsa, P., Wood, D. and Jones, S. (2019) 'Incorporating systems thinking approach in a multilevel framework for human-centered crash analysis', *Transportation Research Interdisciplinary Perspectives*, 2. doi: 10.1016/j.trip.2019.100031.

Adanu, E. K., Jones, S. and Odero, K. (2020) 'Identification of factors associated with road crashes among functionally classified transport modes in Namibia', *Scientific African*. Elsevier B.V., 7, p. e00312. doi: 10.1016/j.sciaf.2020.e00312.

Afghari, A. P. (2018) *Detecting Motor Vehicle Crash Blackspots based on their Underlyinh Behavioural, Engineering and Spatial Causes*. University of Queensland.

African Union (2018) *EVALUATION OF THE IMPLEMENTATION OF THE AFRICAN ROAD SAFETY ACTION PLAN, THE AFRICAN CHARTER ON ROAD SAFETY AND THE INTERGOVERNMENTAL CONVENTION ON NORMS AND STANDARDS OF TRANS-AFRICAN HIGHWAYS (TAH)*. Addis Ababa.

Agerholm, N., Knudsen, D. and Variyeswaran, K. (2017) 'Speed-calming measures and their effect on driving speed – Test of a new technique measuring speeds based on GNSS data', *Transportation*

Research Part F: Traffic Psychology and Behaviour. Elsevier Ltd, 46, pp. 263–270. doi: 10.1016/j.trf.2016.06.022.

Aghajani, M. A., Dezfoulian, R. S., Arjroody, A. R. and Rezaei, M. (2017) ‘Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics (case study: Ilam Province, Iran)’, *Transportation Research Procedia*. Elsevier B.V., 25, pp. 2126–2138. doi: 10.1016/j.trpro.2017.05.409.

Aghayari, M., Pahlavani, P. and Bigdeli, B. (2017) ‘A geographic weighted regression for rural highways crashes modelling using the Gaussian and Tricube kernels: A case study of USA rural highways’, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W4), pp. 305–309. doi: 10.5194/isprs-archives-XLII-4-W4-305-2017.

Ahmed, I. (2013) ‘Road Infrastructure and Road Safety’, *Transport and Communications Bulletien for Asia and the Pacific*, (83), pp. 19–25.

Ajayi, O. V. (2017) ‘Primary sources of data and secondary sources of data’, *Distinguish between primary sources of data and secondary sources of data*, pp. 1–5.

Al-Matawah, J. A. (2009) *An Investigation of Driver Attitudes towards Road Safety in Kuwait*, *School of Civil and Environment Engineering*. University of Southampton. Available at: http://eprints.soton.ac.uk/158357/.

Alexopoulos, E. (2010) ‘Introduction to Multivariate Regression Analysis’, *Hippokratia*, 14(Suppl 1), pp. 23–28.

Ali, Z. and Bhaskar, S. B. (2016) ‘Basic statistical tools in research and data analysis’, *Indian Journal of Anaesthesia*, 60(9), pp. 662–669. doi: 10.4103/0019-5049.190623.

Alsubeai, A. M. (2017) *Evaluating the Interactive Effects of Traffic Volumes and Access Density on Crash Frequency*. South Dakota State University.

Amarasingha, N. and Dissanayake, S. (2015) ‘Factors Associated with Rural Run-Off-Road and Urban Run-off-road Crashes : A Study in the United States’, *Journal of Society for Transportation and Traffic Studies*, 5(4), pp. 39–51.

Ambros, J. (2013) *Self-explaining roads*. Lisenska.

Ambros, J., Jurewicz, C., Turner, S. and Kieć, M. (2018) *An international review of challenges and opportunities in development and use of crash prediction models*, *European Transport Research Review*. Iowa State University. doi: 10.1186/s12544-018-0307-7.

Ambunda, R. and Johannes, P. (2020) 'Assessment of the Impact of Namibian Rural Road Characteristics on Crash Incidence', *Athens Journal of Technology and Engineering*, 7(1), pp. 15–30.

Ambunda, R. (2018) *Analysis of the Relationship between Fatal Rural Road Crashes and Road Design Elements : Namibian Case Study*. Stellenbosch University.

Ambunda, R. and Sinclair, M. (2019) 'Effect of Two-Lane Two-Way Rural Roadway Design Elements on Road Safety', *International Journal of Innovative Technology and Exploring Engineering*, 8(4), pp. 632–637.

American Association of State and Transportation Officials (2001) *A policy on the geometric design of highways and streets*. 4th edn. Washington DC.

Amweelo, M. (2016) 'Invited Article The Road Safety in Namibia : Focus on Road Traffic Accidents', pp. 103–119.

Arani, M. G., Pahlavani, P., Effati, M. and Alamooti, F. N. (2017) 'The crash intensity evaluation using general centrality criterions and a geographically weighted regression', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W4), pp. 367–371. doi: 10.5194/isprs-archives-XLII-4-W4-367-2017.

Archer, J. and Vogel, K. (1999) 'The Traffic safety Problem in Urban Areas', pp. 1–30.

Arria, A. M., Caldeira, K. M., Vincent, K. B., Garnier-Dykstra, L. M. and O'Grady, K. E. (2011) 'Substance-related traffic-risk behaviors among college students', *Drug and Alcohol Dependence*. Elsevier Ireland Ltd, 118(2–3), pp. 306–312. doi: 10.1016/j.drugalcdep.2011.04.012.

Arthur, O. (1995) 'Local Spatial Autocorrelation Statistics: Distributional Issues and an Application', 27(4).

Asbridge, M., Hayden, J. A. and Cartwright, J. L. (2012) 'Acute cannabis consumption and motor vehicle collision risk: Systematic review of observational studies and meta-analysis', *BMJ (Online)*, 344(7846), pp. 1–9. doi: 10.1136/bmj.e536.

ASTM Standard E867-06 (2017) *Standard Terminology relating to Vehicle-Pavement Systems*. West Conshohocken.

Aworemi, R. J., Abdul-Azeez, A. I., Oyedokun, J. A. and Adewoye, O. J. (2010) 'Efficacy of Drivers' Fatigue on Road Accident in Selected Southwestern States of Nigeria', *International Business Research*, 3(3), pp. 225–233.

Ayati, E. and Abbasi, E. (2014) ‘Modeling Accidents on Mashhad Urban Highways’, *Open Journal of Safety Science and Technology*, 04(01), pp. 22–35. doi: 10.4236/ojsst.2014.41004.

Bacher, J. (2000) ‘A Probabilistic Clustering Model for Variables of Mixed type’, *Quality and Quanitity*, 34, pp. 223–235.

Bacher, J., Wenzig, K. and Vogler, M. (2004) ‘SPSS Two-Step Cluster - A First Evaluation’, *Lehrstuhl fur Soziologie*, 2(2).

Bagha, A. and Madisetti, V. (2019) *Big Data Science & Analytics: A Hands-On Approach*.

Baguley, C. J., Sexton, B. and Done, S. (2006) *Road Accident Modelling for Highway Development and Management in Developing Countries : Main Report : Trials in India and Tanzania*.

Bamdad-Mehrabani, B. and Mirbaha, B. (2018) ‘Evaluating the Relationship between Operating Speed and Collision Frequency of Rural Multilane Highways Based on Geometric and Roadside Features’, *Civil Engineering Journal*, 4(3), p. 609. doi: 10.28991/cej-0309120.

Barrel, J., Fletcher, J., Charman, S. and Sterling, T. (2014) *Existing Roads : Proactive Approaches*, *Road Safety Manuals for Africa*. Tunis.

Barua, S., El-Basyouny, K. and Islam, M. T. (2016) ‘Multivariate random parameters collision count data models with spatial heterogeneity’, *Analytic Methods in Accident Research*. Elsevier, 9, pp. 1–15. doi: 10.1016/j.amar.2015.11.002.

Bassan, S. (2016) ‘Sight distance restriction on highways’ horizontal curves: insights and sensitivity analysis’, *European Transport Research Review*. European Transport Research Review, 8(3). doi: 10.1007/s12544-016-0208-6.

Basu, S. and Saha, P. (2017) ‘Regression Models of Highway Traffic Crashes: A Review of Recent Research and Future Research Needs’, *Procedia Engineering*. The Author(s), 187, pp. 59–66. doi: 10.1016/j.proeng.2017.04.350.

Batrakova, A. G. and Gredasova, O. (2016) ‘Influence of Road Conditions on Traffic Safety’, *Procedia Engineering*. Elsevier B.V., 134, pp. 196–204. doi: 10.1016/j.proeng.2016.01.060.

Bauer, K. and Harwood, D. (2014) ‘Safety Effects of Horizontal Curve and Grade Combinations on Rural Two-Lane Highways’, *Transportation Research Record: Journal of the Transportation Research Board*, 2398(January), pp. 37–49. doi: 10.3141/2398-05.

Bax, C., Leroy, P. and Hagenzieker, M. P. (2014) ‘Road safety knowledge and policy: A historical institutional analysis of the Netherlands’, *Transportation Research Part F: Traffic Psychology and*

Behaviour. Elsevier Ltd, 25(PART B), pp. 127–136. doi: 10.1016/j.trf.2013.12.024.

Bayliss, D. (2009) *Accident Trends by Road Type*.

Bella, F. (2005) ‘Driver Perception of Crest and Sag Combinations at the Driving Simulator : effects on driver ’ s behaviour Driver Perception of Crest and Sag Combinations at the Driving Simulator : effects on driver ’ s behaviour’, *3rd International SIIV Congress*.

Ben-Bassat, T. and Shinar, D. (2011) ‘Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior’, *Accident Analysis and Prevention*. Elsevier Ltd, 43(6), pp. 2142–2152. doi: 10.1016/j.aap.2011.06.004.

Benedek, J., Ciobanu, S. M. and Man, T. C. (2016) ‘Hotspots and social background of urban traffic crashes: A case study in Cluj-Napoca (Romania)’, *Accident Analysis & Prevention*, 87(February), pp. 117–126.

Bener, A., Yildirim, E. and Lajunen, T. (2017) ‘Driver sleepiness , fatigue , careless behavior and risk of motor vehicle crash and injury : Population based case and control study’, *Journal for Traffic and Transportation Engineering*, 4(5), pp. 496–502. doi: 10.1016/j.jtte.2017.07.005.

Berhanu Bezabeh, G. (2013) ‘Road Safety in Africa Assessment of Progresses and Challenges in Road Safety Management System’, (December), p. 54. Available at: https://www.afdb.org/fileadmin/uploads/afdb/Documents/Events/ATFforum/Road_Safety_in_Africa_-_Assesement_of_Progresses_and_Challenges_in_Road_Safety_-_AfDB.pdf.

Bester, C. J. and Makunje, J. A. (1998) ‘The effect of rural road geometry on safety in Southern Africa’, *International Symposium on Highway Geometric Design Practices*, pp. 1–10.

Bidulka, S., Sayed, T. and Hassan, Y. (2002) *Influence of vertical alignment on horizontal curve perception: Phase 1 Examining the Hypothesis*. Washignton D.C.

Blockey, P. and Hartley, L. (1995) ‘Aberrant driving behaviour: errors and violation’, *Ergonomics*, 38(9), pp. 1759–1771. doi: 10.1080/00140139508925225.

Botha, G. (2005) ‘Contributory factors to road traffic crashes’, *24th Annual Southern African Transport Conference, SATC 2005: Transport Challenges for 2010*, (July), pp. 242–254.

Botha, G (2005) ‘Measuring Road Traffic Safety Performance’, (July), pp. 515–529.

Brookhuis, K. A. (2014) ‘The role of traffic psychology in psychopharmacological research’, *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 25(PART B), pp. 120–126. doi: 10.1016/j.trf.2013.10.011.

Bruce, P. and Bruce, A. (2017) *Practical Statistics for Data Scientists: 50 Essential Concepts*, *O'Reilly Media Inc*. O'Reilly Media.

Butchart, A. and Mikton, C. (2014) *Global Status Report on Violence Prevention 2014*. Geneva, Switzerland.

Cairney, P. and Bennet, P. (2008) 'Relationship between surface characteristics on selected rural roads in Victoria rural roads', in *ARRB Conference*. ARRB Group.

Câmara, G., Monteiro, A. M., Fucks, S. D. and Sá, M. (2002) 'Spatial Analysis and GIS : A Primer'.

Cameron, C. and Trivedi, P. (1998) 'Regression Analysis of Count Data (Econometric Society Monographs)', (30), pp. 96–137. doi: 10.1017/CCOL0521632013.

Carey, R. N. and Sarma, K. M. (2017) 'Impact of daylight saving time on road traffic collision risk: A systematic review', *BMJ Open*, 7(6). doi: 10.1136/bmjopen-2016-014319.

Carfora, A., Campobasso, C. P., Cassandro, P., Petrella, R. and Borriello, R. (2018) 'Alcohol and drugs use among drivers injured in road accidents in Campania (Italy): A 8-years retrospective analysis', *Forensic Science International*. Elsevier Ireland Ltd, 288(June), pp. 291–296. doi: 10.1016/j.forsciint.2018.05.003.

Cenek D, P., Davies B, R. and Henderson J, R. (2012) *Crash risk relationships for improved road safety management*. Available at: http://www.nzta.govt.nz/resources/research/reports/488/%5Cnhttps://trid.trb.org/view/1148289.

Chambers, R. L., Kokic, P., Smith, P. and Cruddas, M. (2000) 'Winsorization for identifying and treating outliers in business surveys', *Proceedings of the Second International Conference on Establishment Surveys*, (January), pp. 717–726.

Chan, Y., Huang, B., Yan, X. and Richards, S. (2008) *Effects of Asphalt Pavement Conditions on Traffic Accidents in Tennessee Utilizing Pavement Management System (PMS)*. Knoxville.

Chen, F., Chen, S. and Ma, X. (2016) 'Crash Frequency Modeling Using Real-Time Environmental and Traffic Data and Unbalanced Panel Data Models', pp. 1–16. doi: 10.3390/ijerph13060609.

Chen, S., Rakotonirainy, A., Loke, S. W. and Krishnaswamy, S. (2007) 'A Crash Risk Assessment Model for Road Curves', *Proceedings 20th International Technical Conference on the Enhanced Safety of Vehicles*, 2007, pp. 1-8.

Cheng, Z., Zu, Z. and Lu, J. (2018) 'Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections', *Sustainability (Switzerland)*, 11(1). doi:

10.3390/su11010160.

Chiu, T., Fang, D., Chen, J. and Wang, Y. (2001) ‘A robust and scalable clustering algorithm for mixed type attributes in large database environment’, in *7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 263–268.

Choi, Y. Y., Kim, D. K., Kho, S. Y. and Park, P. Y. (2019) ‘Safety evaluation of four safety countermeasures on freeways in South Korea’, *Journal of Transportation Safety and Security*, (February). doi: 10.1080/19439962.2018.1564945.

Choudhary, J., Ohri, A. and Kumar, B. (2015) ‘Spatial and statistical analysis of road accidents hot spots using GIS’, (December).

Choudhary, P., Imprialou, M., Velaga, N. R. and Choudhary, A. (2018) ‘Impacts of speed variations on freeway crashes by severity and vehicle type’, *Accident Analysis and Prevention*. Elsevier, 121(September), pp. 213–222. doi: 10.1016/j.aap.2018.09.015.

Čičković, M. (2016) ‘Influence of Human Behaviour on Geometric Road Design’, *Transportation Research Procedia*, 14(0), pp. 4364–4373. doi: 10.1016/j.trpro.2016.05.358.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioural Sciences*. 2nd edn. New York: Lawrence Erlbaum Associates.

Cohen, J. (1992) ‘Statistical power analysis’, *Current Directions in Psychological Science*, 1(3), pp. 98–101.

Comittee for State Road Authorities (1990) *The structural design, construction and maintenance of unpaved roads*, *Technical Recomendations for Highways*. Pretoria, South Africa: Department of Transport.

Committee of State Road Authorities (1988) ‘Geometric Design of Rural Roads’, in *Technical Recommendations for Highways, TRH 17*. Pretoria, South Africa: Department of Transport.

Committee of State Road Authorities (1988) ‘Road Classification and Access Management’, in *Technical Recommendations for Highways, TRH 26*. Pretoria, South Africa: Department of Transport.

Compton, R. P., Blomberg, R. D., Moskowitz, H., Burns, M., Peck, R. C. and Fiorentino, D. (2002) ‘Crash Risk of Alcohol Impaired Driving’, pp. 39–44. Available at: https://trid.trb.org/view/699712.

Compton, R. P. and Berning, A. (2015) ‘Drug and Alcohol Crash Risk *’, *Journal of Drug Addiction, Education, and Eradication*, 11(1), pp. 29–46. Available at:

http://ezproxy.fiu.edu/login?url=https://search-proquest-com.ezproxy.fiu.edu/docview/1841978960?accountid=10901.

Connor, J. (2002) 'Driver sleepiness and risk of serious injury to car occupants: population based case control study', *British Medical Journal*, 324(7346), pp. 1125–1125. doi: 10.1136/bmj.324.7346.1125.

Cox, J. A., Beanland, V. and Filtness, A. J. (2017) 'Risk and safety perception on urban and rural roads : E ff ects of environmental features , driver age and risk sensitivity', *Traffic Injury Prevention*. Taylor & Francis, 18(7), pp. 703–710. doi: 10.1080/15389588.2017.1296956.

Dagli, R. (2004) 'Driver fatigue and Road safety - Implcation in an indian context', *International journal of advances in Engineering and Tehnology*, 9(4), pp. 1–5.

Daniel, D. (2016) 'Applied Univariate, Bivariate, and Multivariate Statistics', *Journal of Statistical Software*, 72(Book Review 2), pp. 1–4. doi: 10.18637/jss.v072.b02.

Davidović, J., Pešić, D. and Antić, B. (2018) 'Professional drivers' fatigue as a problem of the modern era', *Transportation Research Part F: Traffic Psychology and Behaviour*, 55, pp. 199–209. doi: 10.1016/j.trf.2018.03.010.

Dehuri, A. N. (2013) 'Impacts of Roadway Condition, Traffic and Manmade Features on Road Safety', *Asian Journal of Civil and Road Engineering*, 12(3), pp, 1031-1039.

Deller, J. (2013) 'The influence of road design speed, posted speed limits and lane widths on speed selection: a literature synthesis', in *Australasian Transport Research Forum*, p. 14p. Available at: http://www.atrf.info/papers/2013/index.aspx.

Demasi, F., Loprencipe, G. and Moretti, L. (2018) 'Road Safety Analysis of Urban Roads : Case Study of an Italian Municipality', *Safety*, 3(58). doi: 10.3390/safety4040058.

Demissie, M. (2017) *Risk Factors associated with Serious and Fatal Road Traffic Accidents in Manzini city , Swaziland*. University of the Western Cape.

Denis, D. (2021) 'Linear Regression'. doi: 10.1002/9781119583004.ch7.

Dereli, M. A. and Erdogan, S. (2017) 'A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods', *Transportation Research Part A: Policy and Practice*. Elsevier Ltd, 103, pp. 106–117. doi: 10.1016/j.tra.2017.05.031.

Deublein, M., Schubert, M., Adey, B. T., Köhler, J. and Faber, M. H. (2013) 'Prediction of road accidents: A Bayesian hierarchical approach', *Accident Analysis and Prevention*. Elsevier Ltd, 51, pp. 274–291. doi: 10.1016/j.aap.2012.11.019.

Dong, C., Nambisan, S. S., Richards, S. H. and Ma, Z. (2015) 'Assessment of the effects of highway geometric design features on the frequency of truck involved crashes using bivariate regression', *Transportation Research Part A: Policy and Practice*, 75. doi: 10.1016/j.tra.2015.03.007.

Duivenvoorden, K. (2010) *The Relationship between Traffic Volume and Road Safety on the Secondary Rural Road Network*. Leidschendam. doi: D-2010-2.

Dutch Ministry of Transport (2005) *Sustainable safe road design*.

Dwikat, M. G. (2014) *Modeling Relationship between Geometric Design Consistency and Road Safety for Two-Lane Rural Highways in the West Bank*. An- Najah National University.

Easa, S. M. (2003) 'Improved speed-profile model for two-lane rural highways', *Canadian Journal of Civil Engineering*, 30(6), pp. 1055–1065. doi: 10.1139/l03-021.

Easa, S. M., Dong, H. and Li, J. (2007) 'Use of Satellite Imagery for Establishing Road Horizontal Alignments', *Journal of Surveying Engineering*, 133(1), pp. 29–35. doi: 10.1061/(ASCE)0733-9453(2007)133:1(29).

Edquist, J., Rudin-Brown, C. and Lenne, M. (2009) *Road Design Factors and their Interactions with Speed and Speed Limits*. Victoria.

Eenink, R., Reurings, M., Elvik, R., Cardoso, J., Wichert, S. and Stefan, C. (2005) 'Accident Prediction Models and Road Safety Impact Assessment: Recommendations for using these tools', *RiPCORD iSEREST*, 506184(January 2005), pp. 1–20.

Eggleston, Z., Hansen, P. and Carrera, F. (2016) 'Improvement of Road Safety : Route - based Assessment on Major Roads in Namibia'.

El-basyouny, K. and Sayed, T. (2009) 'Collision prediction models using multivariate Poisson-lognormal regression', 41, pp. 820–828. doi: 10.1016/j.aap.2009.04.005.

Elvik, R., Christensen, P. and Amundsen, A. (2004) 'Speed and road accidents: An evaluation of the Power Model', *TOI report*, 740(December), p. 134.

European Transport Safety Council (2001) *The Role of Driver Fatigue in Commercial road Transport Crashes*. Brussels.

Farahmand, B. and Boroujerdian, A. M. (2018) 'Effect of road geometry on driver fatigue in monotonous environments: A simulator study', *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 58, pp. 640–651. doi: 10.1016/j.trf.2018.06.021.

Farkas, D., Hilton, B., Pick, J., Ramakrishna, H., Sarkar, A. and Shin, N. (2016) 'A tutorial on geographic information systems: A ten-year update', *Communications of the Association for Information Systems*, 38(1), pp. 190–234. doi: 10.17705/1CAIS.03809.

Federal Highway Administration (1989) *Highway Functional Classification*.

Federal Highway Administration (2014) *Highway Performance Monitoring System (HPMS) Field Manual*. Available at: https://www.fhwa.dot.gov/policyinformation/hpms/fieldmanual/.

Feuillet, M., Egido, À. and Lerbet-Sereni, F. (2015) 'Speed and road accidents: risk perception, knowledge and attitude towards penalties for speeding', *Psychofenia: Ricerca Ed Analisi Psicologica*, 18(31), pp. 63–76.

Field, A. (2013) *Discovering Statistics using IBM SPSS Statistics*. 4th edn, *British Library Cataloguing*. 4th edn. London: SAGE Publications Inc.

Foss, R. and Goodwin, A. (2003) 'Enhancing the effectiveness of graduated driver licensing legislation', *Journal of Safety Research*, 34(1), pp. 79–84. doi: 10.1016/S0022-4375(02)00083-X.

Garber, N. J. and Hoel, L. A. (2009) *Traffic and Highway Engineering*. Fourth, *Cengage Learning*. Fourth. Edited by H. Gowans. Toronto, Canada: Nelson Education Ltd. doi: 10.1061/(ASCE)HY.1943-7900.0000746.

Garcia, R. and Abreu, L. (2016) 'Road safety in rural roads of two lanes', *Revista Ingenieria De Construccion*, 31(1), pp. 54–60. Available at: %3CGo%0Ato.

Gargoum, S. A. and El-Basyouny, K. (2016) 'Exploring the association between speed and safety: A path analysis approach', *Accident Analysis and Prevention*. Elsevier Ltd, 93, pp. 32–40. doi: 10.1016/j.aap.2016.04.029.

Gastaldi, M., Rossi, R. and Gecchele, G. (2014) 'Effects of driver task-related fatigue on driving performance', *Procedia - Social and Behavioral Sciences*. Elsevier B.V., 111, pp. 955–964. doi: 10.1016/j.sbspro.2014.01.130.

Gastwirth, J. L., Gel, Y. R. and Miao, W. (2010) 'The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice', *Statistical Science*, 24(3), pp. 343–360. doi: 10.1214/09-sts301.

Gaudry, M. and Vernier, K. (2002) 'Effects of road geometry and surface on speed and safety', *Accident Analysis & Prevention 34.3*, 34(November 1998), pp. 357–365.

Getis, A. and Ord, J. K. (2010) 'The analysis of spatial association by use of distance statistics', *Advances in Spatial Science*, 61(October), pp. 127–145. doi: 10.1007/978-3-642-01976-0_10.

Ghadi, M. and Török, Á. (2017) 'Comparison Different Black Spot Identification Methods', *Transportation Research Procedia*. Elsevier B.V., 27, pp. 1105–1112. doi: 10.1016/j.trpro.2017.12.104.

Ghanbari, A. (2017) *Impact of Pavement Condition Indicators on Road Safety in New Brunswick Arash Bagher Ghanbari Master of Science in Engineering*. University of New Brunswick.

Gichaga, F. J. (2017) 'The impact of road improvements on road safety and related characteristics', *IATSS Research*. Elsevier Ltd, 40(2), pp. 72–75. doi: 10.1016/j.iatssr.2016.05.002.

Gitelman, V., Pesahov, F., Carmel, R. and Bekhor, S. (2016) 'The Identification of Infrastructure Characteristics Influencing Travel Speeds on Single-carriageway Roads to Promote Self-explaining Roads', *Transportation Research Procedia*. Elsevier B.V., 14(June), pp. 4160–4169. doi: 10.1016/j.trpro.2016.05.387.

Gitelman, V., Doveh, E., Carmel, R. and Hakkert, S. (2019) 'The influence of shoulder characteristics on the safety level of two-lane roads: A case-study', *Accident Analysis and Prevention*. Elsevier, 122(February 2018), pp. 108–118. doi: 10.1016/j.aap.2018.10.003.

Glavić, D., Mladenović, M., Stevanovic, A., Tubić, V., Milenković, M. and Vidas, M. (2016) 'Contribution to Accident Prediction Models Development for Rural Two-Lane Roads in Serbia Mark', *Promet-Traffic & Transportation*, 28(4), pp. 415–424. doi: 10.7307/ptt.v28i4.1908.

Godavarthy, R. P. and Russell, E. R. (2016) 'High-Risk Rural Road Safety Study and Determining the Crash-Reduction Factors for High-Risk Rural-Road Usage', *Journal of Transportation Technologies*, 06(01), pp. 1–8. doi: 10.4236/jtts.2016.61001.

Godley, S., Triggs, T. and Fildes, B. (2004) 'Perceptual lane width, wide perceptual road centre marking and driving speeds', *Ergonomics*, 3(47), pp. 237–256.

Gogtay, N. (2010) 'Principles of sample size calculation', *Indian Journal of Ophthalmology*, 58(6), p. 517. doi: 10.4103/0301-4738.71692.

Golob, T. F. and Recker, W. W. (2003) 'Relationships Among Urban Freeway Accidents, Traffic Flow, Weather, and Lighting Conditions', *Journal of Transportation Engineering*, 129(4), pp. 342–353. doi: 10.1061/(ASCE)0733-947X(2003)129:4(342).

Golob, T. F., Recker, W. W. and Alvarez, V. M. (2004) 'Freeway safety as a function of traffic flow', *Accident Analysis and Prevention*, 36(6), pp. 933–946. doi: 10.1016/j.aap.2003.09.006.

Gomes, M. J., Cunto, F. and Silva, A. R. (2017) 'Geographically weighted negative binomial regression applied to zonal level safety performance models', *Accident Analysis and Prevention*. Elsevier, 106(June), pp. 254–261. doi: 10.1016/j.aap.2017.06.011.

Gomes, M. M., Pirdavani, A., Brijs, T. and Souza-Pitombo, C. (2019) 'Assessing the impacts of enriched information on crash prediction performance', *Accident Analysis and Prevention*. Elsevier, 122(February 2018), pp. 162–171. doi: 10.1016/j.aap.2018.10.004.

Gooch, J. P., Gayah, V. V. and Donnell, E. T. (2016) 'Quantifying the safety effects of horizontal curves on two-way, two-lane rural roads', *Accident Analysis and Prevention*, 92. doi: 10.1016/j.aap.2016.03.024.

Government of Namibia (2020) *Yearly Calendar*. Available at: https://www.gov.na/.

Gower, J., Lubbe, S. and Le Roux, N. (2011) *Understanding Biplots*, *Understanding Biplots*. doi: 10.1002/9780470973196.ch10.

Grandjean, E. (1979) 'Fatigue in industry.', *British journal of industrial medicine*, 36(3), pp. 175–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/40999%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1008561.

Grant, E., Salmon, P. M., Stevens, N. J., Goode, N. and Read, G. J. (2018) 'Back to the future: What do accident causation models tell us about accident prediction?', *Safety Science*. Elsevier, 104(December 2017), pp. 99–109. doi: 10.1016/j.ssci.2017.12.018.

Gupta, H. (2017) 'Development of Crash Prediction Model Using Multiple Regression Analysis', *International Journal of Current Engineering and Scientific Research*, 4(6), pp. 82–86.

Haddon, W. (1972) 'A Logistic Framework for Catergorizing Highway Safety Phenomena and Activity', *The Journal of Trauma*, 12, pp. 193–207.

Hagenzieker, M. P., Commandeur, J. J. and Bijleveld, F. D. (2014) 'The history of road safety research: A quantitative approach', *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 25(PART B), pp. 150–162. doi: 10.1016/j.trf.2013.10.004.

Hakkert, A. S., Gitelman, V. and Vis, M. A. (2007) 'Institutional Repository Road Safety Performance Indicators: Theory Deliverables D3.6 of the EU FP6 project SafetyNet'. Available at: https://dspace.lboro.ac.uk/.

Hakkert, S. and Braimaister, L. (2002) *The uses of exposure and risk in road safety studies*. Leidschendam.

Hamzeie, R., Savolainen, P. T. and Gates, T. J. (2017) 'Driver speed selection and crash risk: Insights from the naturalistic driving study', *Journal of Safety Research*. National Safety Council and Elsevier Ltd, 63, pp. 187–194. doi: 10.1016/j.jsr.2017.10.007.

Hanno, D. (2004) *Effects of the Combination of Horizontal and Vertical Alignments on Road Safety*. The University of British Columbia.

Hartley, L. (1998) *Managing Fatigue In Transportation*, *Proceedings of the 3rd Fatigue in Transportation Conference*. doi: 10.1016/j.smrv.2005.03.002.

Hartley, L. and Arnold, P. (1996) 'Recommendations on Fatigue in Transportation', in *Second international Conference on Fatigue in Transportation*. Perth.

Hartman, R. L. and Huestis, M. A. (2013) 'Cannabis effects on driving skills', *Clinical Chemistry*, 59(3), pp. 478–492. doi: 10.1373/clinchem.2012.194381.

Hashimoto, S., Yoshiki, S., Saeki, R., Mimura, Y., Ando, R. and Nanba, S. (2016) 'Development and application of traffic accident density estimation models using kernel density estimation', *Journal of Traffic and Transportation Engineering (English Edition)*. Elsevier Ltd, 3(3), pp. 262–270. doi: 10.1016/j.jtte.2016.01.005.

Hassan, Y. and Easa, S. M. (2003) 'Effect of Vertical Alignment on Driver Perception of Horizontal Curves', *Journal of Transportation Engineering*, 129(4), pp. 399–407. doi: 10.1061/(ASCE)0733-947X(2003)129:4(399).

Hauer, E. (2014) 'Statistical Road Safety Modeling', (January 2004). doi: 10.3141/1897-11.

Haynes, R., Lake, I. R., Kingham, S., Sabel, C. E., Pearce, J. and Barnett, R. (2008) 'The influence of road curvature on fatal crashes in New Zealand', *Accident Analysis and Prevention*, 40(3), pp. 843–850. doi: 10.1016/j.aap.2007.09.013.

Herrstedt, L. (2015) *The Basis of " The Self-Explaining Road " The Basis of ” The Self - Explaining Road ”*. Copenhagen.

Hicks, S. and Fetter, M. (1991) 'an Evaluation of Robust Estimation Techniques'.

Ho, R. (2006) *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*, *Taylor & Francis Group*. doi: 10.1198/000313008X332287.

Housley, S. (2015) 'Design Criteria', in *Highway Design Manual*. 86th edn.

Huanghui, Z. (2012) *Estimation of safety effectiveness of composite shoulders on rural two-lane highways*. University of Kansas.

Hughes, B. P., Newstead, S., Anund, A., Shu, C. C. and Falkmer, T. (2015) ‘A review of models relevant to road safety’, *Accident Analysis and Prevention*. Elsevier Ltd, 74, pp. 250–270. doi: 10.1016/j.aap.2014.06.003.

Huvarinen, Y., Svatkova, E., Oleshchenko, E. and Pushchina, S. (2017) ‘Road Safety Audit’, *Transportation Research Procedia*. Elsevier B.V., 20(September 2016), pp. 236–241. doi: 10.1016/j.trpro.2017.01.061.

Hyder, A. A., Paichadze, N., Toroyan, T. and Peden, M. M. (2017) ‘Monitoring the Decade of Action for Global Road Safety 2011–2020: An update’, *Global Public Health*, 12(12), pp. 1492–1505. doi: 10.1080/17441692.2016.1169306.

Hyldekær, K. and Giacomo, C. (2016) ‘Statistical modelling of the frequency and severity of road accidents Statistical modelling of the frequency and severity of road’.

Imprialou, M. I., Quddus, M., Pitfield, D. E. and Lord, D. (2016) ‘Re-visiting crash-speed relationships: A new perspective in crash modelling’, *Accident Analysis and Prevention*. Elsevier Ltd, 86, pp. 173–185. doi: 10.1016/j.aap.2015.10.001.

Islam, M. H., Hua, L., Hamid, H. and Azarkerdar, A. (2019) ‘Relationship of Accident Rates and Road Geometric Design’, *IOP Conference Series: Earth and Environmental Science*, 357(1). doi: 10.1088/1755-1315/357/1/012040.

Iyinam, A., Iyinam, S. and Ergun, M. (1997) ‘Analysis of relationship between highway safety and road geometric design elements: Turkish case’, *Faculty of Civil Engineering, Turkey*. Available at: http://www.trafficforum.ethz.ch/vwt_2003/beitraege/vwt19proceedings_contribution_91.1-91.8.pdf.

Jaiswal, A. and Bhatore, A. (2016) ‘Effect on Road Safety by Roadway condition, traffic and manmade features’, *International Journal of Latest Trends in Engineering and Technology*, 7(3), pp. 41–48. doi: 10.21172/1.73.006.

Jinghui, Y. and Xuesong, W. (2018) ‘Modeling the Safety Effect of Access and Signal Density on Suburban Arterials : Using Macro Level Analysis Method’.

Joanne, M. (2013) *RAP Road Risk Mapping Manual : Design Specification*.

Johnson, S. B. and Jones, V. C. (2011) ‘Adolescent development and risk of injury: using developmental science to improve interventions’, *Injury Prevention*, 17(1), pp. 50 LP – 54. doi: 10.1136/ip.2010.028126.

Jones, I. S. (1976) *The Effect of Vehicle Characteristics on Road Accidents*. Pergamon Press.

Jones, R., Shinar, D. and Walsh, J. (2003) *State of knowledge of drug-imapired driving*. DOT HS 809 642. Washignton D.C.

Jones, S., Odero, K. and Adanu, E. K. (2019) 'Road crashes in Namibia: Challenges and opportunities for sustainable development', *Development Southern Africa*. doi: 10.1080/0376835X.2019.1659131.

Karlaftis, M. G. and Golias, I. (2002) 'Effects of Road Geometry and Traffic Volumes on Rural Roadway Accident Rates', *Accident Analysis and Prevention*, 34(3), pp. 357–365. doi: 10.1016/S0001-4575(01)00033-1.

Karlaftis, M. G. and Golias, I. (2009) 'Impacts of Roadway Condition , Traffic and Manmade Features on Road Safety', *International Journal of Civil Engineering & Technology*, 4(July), pp. 19–25. doi: 10.1061/(ASCE)HY.1943-7900.0000746.

Kassu, A. and Anderson, M. (2018) 'DETERMINANTS OF SEVERE INJURY AND FATAL TRAFFIC ACCIDENTS ON URBAN AND RURAL HIGHWAYS', *International Journal for Traffic and Transport Engineering*, 8(3), pp. 294–308.

Kelly, E., Darke, S. and Ross, J. (2004) 'A review of drug use and driving: epidemiology, impairment, risk factors and risk perceptions.', *Drug and alcohol review*, 23(3), pp. 319–44. doi: 10.1080/09595230412331289482.

Khan, G., Bill, A., Chitturi, M. and Noyce, D. (2014) 'Safety Evaluation of Horizontal Curves on Rural Undivided Roads', *Transportation Research Record: Journal of the Transportation Research Board*, 2386(1), pp. 147–157. doi: 10.3141/2386-17.

Kilawa, S. and Nyongole, O. V (2015) 'Injury prevention; Motorcyclists' Responses and Practices on the Use of Helmets in Mwanza, Tanzania.', *East & Central African Journal of Surgery*, 20(2), pp. 49–55. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=111494196&site=ehost-live.

King, B. A. (2014) 'The Effect of Road Roughness on Traffic Speed and Road Safety', (November), pp. 1–146.

Kiran, B. N., Kumaraswamy, N. and Sashidhar, C. (2017) 'A review of road crash prediction models for developed countries', *American journal of traffic and transportation engineering*, 2(2), pp. e10–e10. doi: 10.11648/j.ajtte.20170202.11.

Kockelman, K. (2006) 'Crash Frequency and Severity Modeling Using Clustered Data from Washington State', pp. 1621–1626.

Kockelman, K. M., Murray, W. J. and Fellow, J. (2008) ‘A Multivariate Poisson-Lognormal Regression Model for Prediction of Crash Counts by Severity using Bayesian Methods’, *Accident Analysis and Prevention*, 40(3), pp. 964–975.

Kononov, J., Bailey, B. and Allery, B. K. (2008) ‘Relationships between Safety and Both Congestion and Number of Lanes on Urban Freeways’, *Transportation Research Record: Journal of the Transportation Research Board*, 2083(1), pp. 26–39. doi: 10.3141/2083-04.

Kopits, E. and Cropper, M. (2005) ‘Traffic fatalities and economic growth’, *Accident Analysis and Prevention*, 37(1), pp. 169–178. doi: 10.1016/j.aap.2004.04.006.

Krug, E. and Sharma, G. (2009) ‘Risk factors for road traffic injuries’, *Asociatia Victimelor Accidentelor de Circulatie din Romania*, 7(2), pp. 23–39.

Kundakci, E. (2014) ‘Identification of Urban Traffic Accident Hot Spots and Their Characteristics By Using Gis’, (January).

Kwon, S., Kim, H., Kim, G. S. and Cho, E. (2019) ‘Fatigue and poor sleep are associated with driving risk among Korean occupational drivers’, *Journal of Transport and Health*. Elsevier Ltd, 14(June), p. 100572. doi: 10.1016/j.jth.2019.100572.

Laird, J., Harris, R. and Shen, S. (2010) *Accidents on rural roads- For better or worse*. Leeds.

Law Reform and Development Commission of Namibia (2018) *Report on the Road Safety Management Bill*. Windhoek.

Legal Assistance Centre (2016) *Proposals for Enhancing Road Safety in Namibia*. Windhoek.

Lewis-Evans, B. and Charlton, S. G. (2006) ‘Explicit and implicit processes in behavioural adaptation to road width’, *Accident Analysis & Prevention*, 38(November), pp. 610–617. doi: 10.1016/j.aap.2005.12.005.

Li, G. and Baker, S. P. (2012) ‘Epidemiologic Methods’, in Li, G. and Baker, S. P. (eds) *Injury Research: Theories, Methods, and Approaches*. Boston, MA: Springer US, pp. 203–220. doi: 10.1007/978-1-4614-1599-2_9.

Li, G., Brady, J. E. and Chen, Q. (2013) ‘Drug use and fatal motor vehicle crashes: A case-control study’, *Accident Analysis and Prevention*. Elsevier Ltd, 60, pp. 205–210. doi: 10.1016/j.aap.2013.09.001.

Li, Y. and Huang, J. (2015) ‘Safety Impact of Pavement Conditions’, *Transportation Research Record: Journal of the Transportation Research Board*, 2455(1), pp. 77–88. doi: 10.3141/2455-09.

Li, Y., Liu, C. and Ding, L. (2013) 'Impact of pavement conditions on crash severity', *Accident Analysis and Prevention*. Elsevier Ltd, 59, pp. 399–406. doi: 10.1016/j.aap.2013.06.028.

Li, Z., Wang, W., Liu, P., Bigham, J. M. and Ragland, D. R. (2013) 'Using Geographically Weighted Poisson Regression for county-level crash modeling in California', *Safety Science*. Elsevier Ltd, 58, pp. 89–97. doi: 10.1016/j.ssci.2013.04.005.

Liu, C. and Subramanian, R. (2009) *Factors Related to Fatal Single-Vehicle Run-Off-Road Crashes*. Washington DC.

Liu, S., Wang, J. and Fu, T. (2016) 'Effects of lane width, lane position and edge shoulder width on driving behavior in underground urban expressways: A driving simulator study', *International Journal of Environmental Research and Public Health*, 13(10), pp. 1–14. doi: 10.3390/ijerph13101010.

Lloyd, C. (2010) 'Spatial Data Analysis: An introduction for GIS Users'.

Loo, B. P., Cheung, W. S. and Yao, S. (2011) 'The Rural-Urban Divide in Road Safety : The Case of China', *The Open Transportation Journal*, 5(October 2015), pp. 9–20. doi: 10.2174/1874447801105010009.

Lord, D. and Ivan, J. N. (2006) 'Further notes on the Application of Zero Inflated Models in Highway Safety', (860).

Mackie, H. W., Charlton, S. G., Baas, P. H. and Villasenor, P. C. (2013) 'Road user behaviour changes following a self-explaining roads intervention', *Accident Analysis and Prevention*. Elsevier Ltd, 50(June), pp. 742–750. doi: 10.1016/j.aap.2012.06.026.

Maji, A., Singh, D., Agrawal, N. and Zaman, M. (2018) 'Operating speed prediction models for tangent sections of two-lane rural highways in Oklahoma State', *Transportation Letters*. Taylor & Francis, 00(00), pp. 1–8. doi: 10.1080/19427867.2018.1536424.

Mannering, F. L. and Bhat, C. R. (2014) 'Analytic methods in accident research: Methodological frontier and future directions', *Analytic Methods in Accident Research*. Elsevier, 1, pp. 1–22. doi: 10.1016/j.amar.2013.09.001.

Maqbool, Y. (2019) 'Road safety and Road Accidents : An Insight Road safety and Road Accidents : An Insight', *International Journal of Information and Computing Science*, 6(4), pp. 93–105.

Marchesini, P. and Weijermars, W. (2010) 'The relationship between road safety and congestion on motorways', *SWOV Institute for Road Safety Research*, p. 28. doi: R-2010-12.

Massie, D. L. and Campbell, K. L. (1993) *Analysis of accident rates by age, gender, and time of day* *based on the 1980 Nationwide Personal Transportation Survey*. Michigan. Available at: http://libaccess.mcmaster.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=gnh&AN=59833&site=ehost-live&scope=site.

Maxwell, L. K. and David, C. H. (1995) 'The application of the durbin-watson test to the dynamic regression model under normal and non-normal errors', *Econometric Reviews*, 14(4), pp. 487–510. doi: 10.1080/07474939508800333.

May, A. D. (1990) *Traffic Flow Fundamentals*. Available at: http://books.google.com/books?id=JYJPAAAAMAAJ&pgis=1.

Maydeu-Olivares, A. and Garcia-Forero, C. (2010) 'Goodness-of-Fit Testing', *International Encyclopedia of Education*, 7, pp. 190–196.

Meng, L., Kees, W. and Evangelos, B. (2006) 'Traffic safety principles and physical road infrastructure measures', in *19th International Cooperation on Theories and Concepts in Traffic Safety*. Lund.

Miaou, S. P. and Lord, D. (2007) 'Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods', *Transportation Research Record: Journal of the Transportation Research Board*, 1840(1), pp. 31–40. doi: 10.3141/1840-04.

Miranda-Moreno, L., Fu, L., Saccomanno, F. and Labbe, A. (2007) 'Alternative Risk Models for Ranking Locations for Safety Improvement', *Transportation Research Record: Journal of the Transportation Research Board*, 1908(January), pp. 1–8. doi: 10.3141/1908-01.

Mitchel, A. (2005) *The ESRI Guide to GIS Analysis, Volume 2: Spatial measurements and statistics*. Redlands, CA, USA: ESRI Press.

Mitra, S., Haque, M. and King, M. J. (2017) 'Effects of access, geometric design, and heterogeneous traffic on safety performance of divided multilane highways in India', *Journal of Transportation Safety and Security*, 9. doi: 10.1080/19439962.2016.1237600.

Mohammed, A., Umar, S. Y., Samson, D. and Ahmad, T. Y. (2017) 'The Effect of Pavement Condition on Traffic Safety: A Case Study of Some Federal Roads in Bauchi State', *IOSR Journal of Mechanical and Civil Engineering*, 12(03), pp. 139–146. doi: 10.9790/1684-120301139146.

Mohammed, A. A., Bin Ambak, K., Mosa, A. M. and Syamsunur, D. (2018) 'Classification of Traffic Accident Prediction Models: a Review Paper', *International Journal of Advances in Science Engineering and Technology*, ISSN(6), pp. 2321–9009. Available at: http://iraj.in.

Mohammed, H. (2013) 'The Influence of Road Geometric Design Elements on Highway Safety', *International Journal of Civil Engineering & Technology*, 4(4), pp. 146–162.

Mohanty, M. and Gupta, A. (2015) 'Factors affecting road crash modeling', *Journal of Transport Literature*, 9(2), pp. 15–19. doi: 10.1590/2238-1031.jtl.v9n2a3.

Mollel, E., Msumba, E. and Ngowi, N. (2011) *Road Geometric Design Manual*. 2011th edn. Dar-es-Salaam: Tanzania Ministry of Works.

Montella, A., Andreassen, D., Tarko, A. P., Turner, S., Mauriello, F., Imbriani, L. L., Romero, M. A. and Singh, R. (2012) 'Critical Review of the International Crash Databases and Proposals for Improvement of the Italian National Database', *Procedia - Social and Behavioral Sciences*, 53, pp. 49–61. doi: 10.1016/j.sbspro.2012.09.859.

Montgomery, D. C. and Runger, G. C. (2014) *Applied Statistics and Probalisty for Engineers*.

Moodley, S. and Allopi, D. (2008) 'An Analytical Study of Vehicle Defects and Their Contribution To Road Accidents', *27th Southern African Transport Conference*, (July), pp. 469–479.

Moran, P. A. (1948) 'THe interpretation of statistical maps', *Journal of the Royal Statistical Society: Series B (Methodological)*, 10(2), pp. 243–251.

Munteanu, P. L., Rosu, M., Panaitescu, V. and Pungă, A. (2014) 'Human and environmental factors contributing to fatal road accidents in a Romanian population', *Romanian Journal of Legal Medicine*, 22(2), pp. 97–100. doi: 10.4323/rjlm.2014.97.

Murthy, N. and Rao, R. S. (2015) 'Development of model for road accidents based on intersection parameters using regression models', 5(1), pp. 1–8.

Nambahu, B. B. (2018) *Modeling of Road Traffic Fatalities in Namibia: A Generalized Linear Model Approach*. University of Namibia. doi: 10.1590/s1809-98232013000400007.

Namibia National Road Safety Council (2012) *Road Accidents in Namibia Statistical Report 2012*. Windhoek. doi: 10.1017/CBO9781107415324.004.

Namibia Statistics Agency (2011) *Namibia Population and Housing Census 2011 - Metadata*. Windhoek.

Namibia Statistics Agency (2015) *Namibia Social Statistics 2010 - 2015*. Windhoek.

National Safety Transportion Board (1995) *Factors that affect fatigue in heavy truck accidents*. Washignton D.C.

National Safety Transportion Board (1999) *Evaluation of U.S. Department of Transportation Efforts in the 1990s to Address Operator Fatigue*. Washignton D.C.

Neill, M. O. (2006) 'Levene' s mean-based test : exact and approximate distributions Technical report'.

Nghishihange, S. (2018) *Factors contributing to motor vahicleaccidents and severity of related injuries in Oshana region, Namibia*. University of Namibia. doi: 10.22201/fq.18708404e.2004.3.66178.

Niewiadomska-Bugaj, M. and Bartoszynski, R. (2021) 'Testing Statistical Hypotheses', pp. 373–428. doi: 10.1002/9781119243830.ch12.

Nilsson, G. (2004) *Traffic safety dimensions and the Power model to describe the effect of speed on safety*, *National Road and Transport Research Institute*. Lund.

Nilsson, G. (2005) 'Road Safety in Sweden and the Effect of Speed on Safety', pp. 1–11.

Noce, F., Tufk, S. and De Mello, M. T. (2008) 'Professional drivers and working time: Journey span, rest, and accidents', *Sleep Science*, 1(1), pp. 20–26.

Noland, R. B. and Oh, L. (2004) 'The effect of infrastructure and demographic change on traffic-related fatalities and crashes: A case study of Illinois county-level data', *Accident Analysis and Prevention*, 36(4), pp. 525–532. doi: 10.1016/S0001-4575(03)00058-7.

Nteziyaremye, P. (2018) *Investigating the link between the built environment and the incidence of pedestrian crashes in Cape Town , South Africa*. Stellenbosch University.

Nusholtz, G. (2011) *Mass Change , Complexity and Fleet Impact Response*.

Ogden, K. W. (1994) 'Traffic Engineering Road Safety : a Practitioner ' S Guide', *Design Issues*.

Olivier, A. (2020) *Decade of Aaction Strategy for Road Safety (2021-2030)*. Windhoek.

Oppong, R. A. (2012) 'Statistical Analysis of Road Accidents Fatality in Ghana Using Poisson Regression', *Thesis submitted to the Department of Mathematics , Kwame Nkrumah University of Science and Technology*, p. 80.

Organisation for Economic Co-operation and Development (2006) *Speed Management: A road safety manual for decision-makers and practitioners*, *GRSP Publication*. Geneva: World Health Organisation. doi: 10.1787/9789282103784-en.

Organisation for Economic Co-Operation and Development (2003) *Road Safety: Impact of New*

Technologies. Paris: OECD Publishing.

Othman, S. and Thomson, R. (2007) 'Influence of Road Characteristics on Traffic Safety', in *ESV 20th Conference*. Lyon, pp. 1–10.

Othman, S., Thomson, R. and Lannér, G. (2009) 'Identifying critical road geometry parameters affecting crash rate and crash type.', *Annals of advances in automotive medicine / Annual Scientific Conference ... Association for the Advancement of Automotive Medicine. Association for the Advancement of Automotive Medicine. Scientific Conference*.

Ouni, F. and Belloumi, M. (2018) 'Spatio-temporal pattern of vulnerable road user's collisions hot spots and related risk factors for injury severity in Tunisia', *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 56(June), pp. 477–495. doi: 10.1016/j.trf.2018.05.003.

Ouni, F. and Belloumi, M. (2019) 'Pattern of road traffic crash hot zones versus probable hot zones in Tunisia: A geospatial analysis', *Accident Analysis and Prevention*. Elsevier, 128(May), pp. 185–196. doi: 10.1016/j.aap.2019.04.008.

Padmanaban, J., Rajaraman, R., Stadter, G., Narayan, S. and Ramesh, B. (2010) 'Analysis of in-depth crash data on Indian national highways and impact of road design on crashes and injury severity', *Berichte Der Bundesanstalt Fuer Strassenwesen. Unterreihe Fahrzeugtechnik*, (77).

Pais, J. C., Amorim, S. I. and Minhoto, M. J. (2013) 'Impact of Traffic Overload on Road Pavement Performance', *Journal of Transportation Engineering*, 139(9), pp. 873–879. doi: 10.1061/(ASCE)TE.1943-5436.0000571.

Papadimitriou, E., Machata, K., Bauer, R., Stadlbauer, S., Soteropoulos, A., Daniels, S., Elvik, R., Ziakopoulos, A., Theofilatos, A. and Yannis, G. (2018) 'Safety effects of infrastructure road safety measures', in *7th Transport Research Arena TRA 2018*. Vienna.

Parizel, P. M. and Phillips, C. D. (2004) 'Neuroradiological Diagnosis of Craniocerebral and Spinal Trauma: Current Concepts in World report on road traffic injury prevention: summary 2004', *Diseases of the Brain, Head and Neck, Spine*, pp. 60–72. doi: 10.1007/978-88-470-2131-0_11.

Park, B. J., Fitzpatrick, K. and Lord, D. (2010) 'Evaluating the Effects of Freeway Design Elements on Safety', *Transportation Research Record: Journal of the Transportation Research Board*, 2195(2195), pp. 58–69. doi: 10.3141/2195-07.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A., Jarawan, E. and Mathers, C. (2017) *World Report on Road Traffic Injury Prevention*, *World Health Organisation*. Geneva. doi: 10.1016/j.puhe.2005.09.003.

Persia, L., Usami, D., De Simone, F., Beaumelle, V. F., Yannis, G., Laiou, A., Han, S., Machata, K., Pennisi, L., Marchesini, P. and Salathè, M. (2016) 'Management of Road Infrastructure Safety', *Transportation Research Procedia*. Elsevier B.V., 14, pp. 3436–3445. doi: 10.1016/j.trpro.2016.05.303.

Pinard, M. I., Ellis, S. D. and Eriksson, C.-H. (2003) *SADC Guideline on Low- Volume Sealed Roads*. Goldfields Press (Pty) Ltd.

Pirdavani, A., Bellemans, T., Brijs, T. and Wets, G. (2014) 'Application of Geographically Weighted Regression Technique in Spatial Analysis of Fatal and Injury Crashes', *Journal of Transportation Engineering*, 140(8), p. 04014032. doi: 10.1061/(asce)te.1943-5436.0000680.

Pljakić, M., Jovanović, D., Matović, B. and Mićić, S. (2019) 'Identification of Accident Hotspot Locations Using Network Kernel Density', (September), pp. 7–13.

Porter, R., Donnell, E. and Mason, J. (2012) 'Geometric Design, Speed, and Safety', *Transportation Research Record: Journal of the Transportation Research Board*, 2309, pp. 39–47. doi: 10.3141/2309-05.

Prestor, J., Klemen, B. and Zotlar, S. (2014) 'Self-Explaining Roads: Concept Analysis and a Proposal for the Establishment in Slovenia', in *XII International Symposium*.

Qian, G. and Künsch, H. (1996) *On Model Selection in Robust Linear Regression*. Zurich. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.26.7185&rep=rep1&type=pdf.

Rakha, H., Arafeh, M., Abdel-Salam, A. G., Guo, F. and Flintsch, A. M. (2010) *Linear Regression Crash Prediction Models: Issues and Proposed Solutions*. Virginia.

Ramaekers, J. G., Kuypers, K. P., Bosker, W. M., Brookhuis, K. A., Veldstra, J. A., Simons, R., Martens, M., Hjälmdahl, M., Forsman, Å. and Knoche, A. (2012) 'Effects of stimulant drugs on actual and simulated driving: perspectives from four experimental studies conducted as part of the DRUID research consortium', *Psychopharmacology*, 222(3), pp. 413–418. doi: 10.1007/s00213-012-2766-1.

Reason, J. (1990) *Human Error*. Cambridge University Press.

Reifman, A. and Keyton, K. (2010) 'Winsorize', *Encyclopedia of Research Design*, pp. 1636–1638.

Richter, T., Ruhl, S., Ortlepp, J. and Bakaba, E. (2016) 'Prevention of Overtaking Accidents on Two-lane Rural Roads', in *Transportation Research Procedia*. doi: 10.1016/j.trpro.2016.05.385.

Ridout, M., Demetrio, C. G. . and Hinde, J. (1998) 'Models for count data with many zeros', *International Biometric Conference*, (December).

Robinson, R. (1984) *Problems in the urban environment : traffic congestion and its effects*.

Rogers, M. (2003) *Highway Engineering*, *Blackwell Science*. Oxford: Blackwell Publishing Ltd. doi: 10.1016/0016-0032(55)91076-0.

Rohe, K. and Zeng, M. (2020) 'Vintage Factor Analysis with Varimax Performs Statistical Inference', pp. 1–82. Available at: http://arxiv.org/abs/2004.05387.

Rose, D., Castro, D., Jordaan, G., Judd, D. and Paige-Green, P. (2014) *South African Pavement Engineering Manual*, *South African Pavement Engineering Manual*.

Runji, J. (2003) 'Road Sector Reform: the Case of Namibia', pp. 1–14.

Saffari, S. E. and Adnan, R. (2011) 'Zero-Inflated Negative Binomial Regression Model with Right Censoring Count Data', *Journal of Materials Science and Engineering B*, 1(9). doi: 10.17265/2161-6221/2011.09.020.

Saffari, S. E., Adnan, R. and Greene, W. (2012) 'Hurdle negative binomial regression model with right censored count data', *Sort*, 36(2), pp. 181–194.

Saha, P. and Ksaibati, K. (2016) 'An optimization model for improving highway safety', *Journal of Traffic and Transportation Engineering (English Edition)*. Elsevier Ltd, 3(6), pp. 549–558. doi: 10.1016/j.jtte.2016.01.004.

Salmon, P. and Johston, I. (2013) 'A systems perspective on road user error in Australia: swiss cheese and the road transport system', in *57th Annals of Advances in Automotive Medicine (AAAM Annual) Conference*, pp. 1–11.

Satria, R. and Castro, M. (2016) 'GIS Tools for Analyzing Accidents and Road Design: A Review', *Transportation Research Procedia*. The Author(s), 18(June), pp. 242–247. doi: 10.1016/j.trpro.2016.12.033.

Saxena, K. M., Johnson, R. A. and Bhattacharyya, G. K. (2006) 'Statistics Principles and Methods', *The American Statistician*, 48(1), p. 59. doi: 10.2307/2685091.

Schmidt, K., Carriquiry, A. and Kaiser, M. (2012) 'Modeling crash frequency data'. *Graduate Theses and Dissertations*, 12911. Available at: https://lib.dr.iastate.edu/etd/12911.

Schulze, H. and Koßmann, I. (2010) 'The role of safety research in road safety management', *Safety Science*. Elsevier Ltd, 48(9), pp. 1160–1166. doi: 10.1016/j.ssci.2009.12.009.

Van Scoor, O., Van Niekerk, J. and Grobbelaar, B. (2001) ‘Mechanical failures as a contributing cause to motor vehicle accidents- South Africa’, *Accident Analysis & Prevention*, (33), pp. 713–721.

Semar, H. (2003) *Design Manual for Roads and Bridges*. 6th edn, *Roads Directorate*. 6th edn. Floriana: Malta Transport Authority. doi: 10.1034/j.1600-0846.2003.00019.x.

Shafabakhsh, G. A., Famili, A. and Bahadori, M. S. (2017) ‘GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran’, *Journal of Traffic and Transportation Engineering (English Edition)*. Elsevier Ltd, 4(3), pp. 290–299. doi: 10.1016/j.jtte.2017.05.005.

Shalom-Hakkert, A. and Gitelman, V. (2014) ‘Thinking about the history of road safety research: Past achievements and future challenges’, *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 25(Part B), pp. 137–149. doi: 10.1016/j.trf.2014.02.005.

Shibani, A. (2016) ‘Analysis of Traffic Accident Severity on Great Britain Roadways and Junctions’.

Shinar, D. (2007) ‘Crash Countermeasures and Design of Safety’, in *Traffic Safety and Human Behaviour*. Emerald Inc.

Shinar, D. (2017) *Traffic safety and Human Behaviour*. doi: 10.1108/9781786352217.

Swedish Institute for Transport and Communication Analysis-SIKA (2000) *Road Traffic Injuries 2000, Official Statistics of Sweden*. Stockholm.

Sinclair, M. (2013) ‘Attitudes, norms and driving behaviour: A comparison of young drivers in South Africa and Sweden’, *Transportation Research Part F: Traffic Psychology and Behaviour*, 20, pp. 170–181.

Singh, G., Sachdeva, S. N. and Pal, M. (2004) ‘Analysis of Causal Factors of Accidents on Highways in Haryana’, *National Conference on Advances in Engineering, Technology & Management (AETM’15)”*, pp. 7–12.

Singh, K. (2017) *Road Accidents in India – 2017*.

Singh, S. K. (2017) ‘Road Traffic Accidents in India: Issues and Challenges’, *Transportation Research Procedia*. Elsevier B.V., 25(December), pp. 4708–4719. doi: 10.1016/j.trpro.2017.05.484.

Sisiopiku, V. P. (2011) ‘Crash prediction on rural roads’, pp. 1–14.

Sjogren, L., Anund, A., Charman, S., Ahern, A., Pumberger, A. and Cocu, X. (2012) *Speed Adaption Control by Self-Explaining Roads (SPACE)*.

Slop, M. (1994) ‘Safety effects of road design standards’, *International conference ‘Road safety in*

Europe and Strategic Highway Research Program', (September), pp. 26–28.

Smith, M., Goodchild, M. and Longley, P. (2009) *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, *Annals of the Association of American Geographers*. doi: 10.1080/00045600802708705.

Sommer, S. and Wade, T. (2006) 'A to Z GIS: an illustrated dictionary of geographic information systems', *Choice Reviews Online*. doi: 10.5860/choice.44-5991.

Songchitruksa, P. and Zeng, X. (2010) 'Getis-ord spatial statistics to identify hot spots by using incident management data', *Transportation Research Record*, (2165), pp. 42–51. doi: 10.3141/2165-05.

Songpatanasilp, P., Yamada, H., Horanont, T. and Shibasaki, R. (2015) 'Traffic accidents risk analysis based on road and land use factors using GLMs and zero-inflated models', *CUPUM 2015 - 14th International Conference on Computers in Urban Planning and Urban Management*. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85026379960&partnerID=40&md5=5b38b5ff5f0f3fb6b6c4e0a6fcd950c0.

South African National Road Agency Limited (2003) *Geometric Design Guidelines*. Pretoria: South African National Roads Agency Limited.

Stamatiadis, N., Pigman, J.G., Lord, D., Sacksteder, J. and Ruff, W. (2009) *Impact of Shoulder Width and Median Width on Safety*, *Impact of Shoulder Width and Median Width on Safety*. doi: 10.17226/14252.

Stellenbosch University (2013) *Policy for Responsible Research Conduct at Stellenbosch University*. Available at: http://www0.sun.ac.za/research/assets/files/Policy_Documents/POLICY (Accessed: 17 May 2019).

Stephan, K. and Newstead, S. (2017) 'Measuring the Influence of the Road and Roadside in the Safe System', pp. 1–11.

Stevenson, M. R., Palamara, P., Morrison, D. and Ryan, G. A. (2001) 'Behavioural Factors as Predictors of Motor Vehicle Crashes in Young Drivers', *Journal of Crash Prevention and Injury Control*, 2(4), pp. 247–254. doi: 10.1080/10286580108902569.

Swart, E. and Sinclair, M. (2015) 'Fatigue Management – Lessons From International Legislation and Practice', in *34th Southern African Transport Conference*, pp. 88–104.

SWOV (2010) *Road Functionality and Homogeneity*. Leidschendam. Available at: https://www.swov.nl/feiten-cijfers.

Taha, A. (2016) ‘Knowledge Discovery In GIS Data’, (January 2016).

Tamayo, A. (2009) ‘Occurrence of Traffic Accidents in the Philippines: An Application of Poisson Regression Analysis’, *Ssrn*, 63(082), pp. 1–16. doi: 10.2139/ssrn.1438478.

Taylor, M. C., Baruya, A. and Kennedy, J. V (2002) *The relationship between speed and accidents on rural single-carriageway roads Prepared for Road Safety Division, Department for Transport, Local Government and the Regions*, *TRL Report TRL511*. Berkshire. Available at: http://www.safespeed.org.uk/TRL511.pdf.

Taylor, M. C., Lynam, D. A. and Baruya, A. (2000) ‘The effect of drivers’ speed on the frequency of road accidents’, *Transportation Research Laboratory*, p. 56. doi: 10.1051/ matecconf 2016 00 / 47 3 04.

Tehrani, S. S. and Falls, L. C. (2015) ‘A Comparison of IRI and Collision Rates in the Province of Alberta’, *2015 CSCE Conference*, (November), pp. 1–9. doi: 10.13140/RG.2.1.1599.1129.

Thakali, L., Kwon, T. J. and Fu, L. (2015) ‘Identification of crash hotspots using kernel density estimation and kriging methods: a comparison’, *Journal of Modern Transportation*. Springer Berlin Heidelberg, 23(2), pp. 93–106. doi: 10.1007/s40534-015-0068-0.

Theeuwes, J. and Godthelp, H. (1995) ‘Self-explaining roads’, *Safety Science*, 19, pp. 217–225.

Thomas, L., Srinivasan, R., Lan, B., Hunter, W., Martell, C. and Rodgman, E. (2013) *Speed and Safety in North Carolina*. Available at: https://connect.ncdot.gov/projects/research/RNAProjDocs/2011-08finalreport.pdf%0Ahttps://trid.trb.org/view/1277053.

Titi, H. H., Coley, N. J. and Latifi, V. (2018) ‘Evaluation of Pavement Performance due to Overload Single-Trip Permit Truck Traffic in Wisconsin’, *Advances in Civil Engineering*, 2018, pp. 1–11. doi: 10.1155/2018/1070653.

Tolouei, R., Maher, M. and Titheridge, H. (2012) *Vehicle Mass and Injury Risk in Two-Car Crashes: A Novel Methodology*. London.

Toran, A. and Moridpour, S. (2015) ‘Identifying Crash Black Spots in Melbourne Road Network Using Kernel Density Estimation in GIS’, *Road Safety and Simulation*.

Transportation Research Board (2000) *Highway capacity manual*, *National Academy Press*. Washignton D.C. doi: 10.1061/(ASCE)HY.1943-7900.0000746.

Transportation Research Board (2003) *NCHRP Report 500: A Guide for addressing Head-On*

Collisions. Los Angeles.

Truong, L. T. and Somenahalli, S. V. (2011) 'Using GIS to identify pedestrian- vehicle crash hot spots and unsafe bus stops', *Journal of Public Transportation*, 14(1), pp. 99–114. doi: 10.5038/2375-0901.14.1.6.

Turner, B. (2005) *Austroads Road Safety Engineering Risk Assessment*. Brisbane.

Turner, B., Woolley, J. and Cairney, P. (2015) 'An analysis of driver behaviour through rural curves: Exploratory results on driver speed', *Australasian Road Safety Conference*, 10, pp. 1–10.

Ul, A., Kayani, H. and King, M. J. (2014) 'A qualitative investigation of drug use among Pakistani road users', (August 2013), pp. 1–6.

United Nations (2011) *Global Plan for the Decade of Action for Road Safety 2011-2020*. Available at: http://www.who.int/roadsafety/publications/global_launch.pdf. (Accessed: 9 November 2017).

Vadeby, A., Turner, B., Machata, K., Hollo, P. and Shelton, T. (2018) *Speed and crash risk*. Paris. doi: 10.1007/s11219-006-6001-3.

Vayalamkuzhi, P. and Amirthalingam, V. (2016) 'Influence of geometric design characteristics on safety under heterogeneous traffic flow', *Journal of Traffic and Transportation Engineering (English Edition)*. Elsevier Ltd, 3(6), pp. 559–570. doi: 10.1016/j.jtte.2016.05.006.

Venter, K. (2019) *South African Novice Drivers : Exploring Hazard Perception Development during the Learner Driver Training Phase*. Stellenbosch University. Available at: https://scholar.sun.ac.za/handle/10019.1/764.

Verstraete, A. G., Legrand, S. A., Vandam, L. and Hughes, B. G. (2014) *Drug Use, Impaired Driving and Traffic Accidents*. 2nd edn. Lisbon: European Monitoring Centre of Drugs and Drug Addiction. Available at: http://www.emcdda.europa.eu/attachements.cfm/att_229259_EN_TDXD14016ENN.pdf.

Vinayakamurthy, M., Mamlouk, M., Underwood, S. and Kaloush, K. (2017) 'Effect of pavement condition on accident rate', (June), p. 59.

Walker, J. and Maddan, S. (2009) 'Factor analysis, path analysis, and structural equation modeling', *Statistics in Criminology and Criminal Justice: Analysis and Interpretation*, pp. 325–350.

Walmsley, D. A., Summersgill, I. and Binch, C. (1998) *Accidents on modern rural single-carriageway trunk roads. Prepared for the Highways Agency, Traffic, Safety and Environment Division*. Berkshire. Available at:

http://swov.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwbZ3LSsQwFIaD6EZ04aDieIHzANPb9JJmOY4dxJXoILgaYi4ijim0VXHvg3tOqhTUbektITl_cvJ_CWPpNIyDXzHBplZKrVBOs8JylEFR5NMsV4nJEi4lLfBW58XVfXZ3w5cDI_aTeGvf67fQraPe6x9ZYvsjhVWIoYZWUgWqHkF8Zc79BCyekzELH9oZQuBij2.

Wang, C., Quddus, M. A. and Ison, S. G. (2013) 'The effect of traffic and road characteristics on road safety: A review and future research direction', *Safety Science*, 57, pp. 264–275. doi: 10.1016/j.ssci.2013.02.012.

Wang, C., Quddus, M. and Ison, S. (2009) 'The effects of area-wide road speed and curvature on traffic casualties in England', *Journal of Transport Geography*, 17(5), pp. 385–395. doi: 10.1016/j.jtrangeo.2008.06.003.

Wedajo, T., Quezon, E. T. and Mohammed, M. (2017) 'Analysis of Road Traffic Accident Related of Geometric Design Parameters in Alamata-Mehoni- Hewane Section', (June).

Wegman, F. (2017) 'The future of road safety: A worldwide perspective', *IATSS Research*. doi: 10.1016/j.iatssr.2016.05.003.

Wegman, F. and Elsenaar, P. (1997) 'Sustainable solutions to improve road safety in The Netherlands', p. 28. Available at: http://www.swov.nl/rapport/D-97-08.pdf.

Weller, G., Schlag, B., Friedel, T. and Rammin, C. (2008) 'Behaviourally relevant road categorisation : A step towards self-explaining rural roads', *Accident Analysis & Prevention*, 40(April), pp. 1581–1588. doi: 10.1016/j.aap.2008.04.009.

Woolley, J. E., Zito, R., Dyson, C. B., Stazic, B. and Taylor, M. A. (2002) 'Impacts of Lower Speed Limits in South Australia', *IATSS Research*. International Association of Traffic and Safety Sciences, 26(2), pp. 6–17. doi: 10.1016/S0386-1112(14)60038-8.

World Bank (2012) 'African Transport Policy Program', in *Long-Term Strategy for Moving SSATP Forward*. Addis Ababa, Ethipia.

World Health Organisation (2015) *Global Status Report on Road Safety*. Available at: http://ra.ocls.ca/ra/login.aspx?url=http://search.ebscohost.com/login.aspx?direct=true&db=enr&AN=56097888&site=ehost-live.

World Health Organisation (2018) *Global status report on road safety 2018*. Geneva, Switzerland.

Wretstrand, A., Holmberg, B. and Berntman, M. (2014) 'Safety as a key performance indicator: Creating a safety culture for enhanced passenger safety, comfort, and accessibility', *Research in Transportation Economics*. Elsevier Ltd, 48, pp. 109–115. doi: 10.1016/j.retrec.2014.09.008.

Yalcin, G. (2013) 'Non-spatial Analysis for the Road Traffic Accidents', *Procedia - Social and Behavioral Sciences*, 92, pp. 1033–1038. doi: 10.1016/j.sbspro.2013.08.795.

Yan, X., Wang, B., An, M. and Zhang, C. (2012) 'Distinguishing between Rural and Urban Road Segment Traffic Safety Based on Zero-Inflated Negative Binomial Regression Models', 2012. doi: 10.1155/2012/789140.

Yang, L., Li, X., Guan, W. and Zhang, H. M. (2017) 'Effect of traffic density on lane change and overtaking manuvers: A driving simulator based study', *97th Annual Meeting of the Transportation Research Board*, 9588(May), pp. 1–14. doi: 10.1080/15389588.2018.1471470.

Yannis, G., Dragomanovits, A., Laiou, A., Richter, T., Ruhl, S., La Torre, F., Domenichini, L., Graham, D., Karathodorou, N. and Li, H. (2016) 'Use of Accident Prediction Models in Road Safety Management - An International Inquiry', in *6th Transport Research Arena*. Warsaw. doi: 10.1016/j.trpro.2016.05.397.

Yingxue, Z. (2009) 'Analysis of the Relation between Highway Horizontal Curve and Traffic Safety', *2009 International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2009*, 3, pp. 479–481. doi: 10.1109/ICMTMA.2009.511.

Zegeer, C. V, Deen, R. C. and Mayes, J. G. (1981) 'Effect of Lane and Shoulder Widths on Accident Reduction on Rural, Two-Lane Roads.', *Transportation Research Record*, p. pp 33-43. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-0019686903&partnerID=40&md5=676c7ceb963bc4a9b5e7a856996051af%5Cnhttp://scholar.google.com/scholar_lookup?title=EFFECT+OF+LANE+AND+SHOULDER+WIDTHS+ON+ACCIDENT+REDUCTION+ON+RURAL,+TWO-LANE+ROADS&author=C.+Z.

Zegeer, V. C., Hummer, J., Reinfurt, D., Herf, L. and Hunter, W. (1987) *Safety Effects of Cross-Section Design for Two-Lane Roads - Volume I - Final Report*. Available at: http://ntl.bts.gov/lib/31000/31000/31037/FHWA-RD-87-008.pdf%5Cnhttps://trid.trb.org/view/274947%5Cnhttp://ntl.bts.gov/lib/30000/30800/30875/FHWA_RD_87_009.pdf%5Cnhttps://trid.trb.org/view/274948.

Zhang, H., Wu, C., Yan, X. and Qiu, T. Z. (2016) 'The effect of fatigue driving on car following behavior', *Transportation Research Part F: Traffic Psychology and Behaviour*. Pergamon, 43, pp. 80–89. doi: 10.1016/J.TRF.2016.06.017.

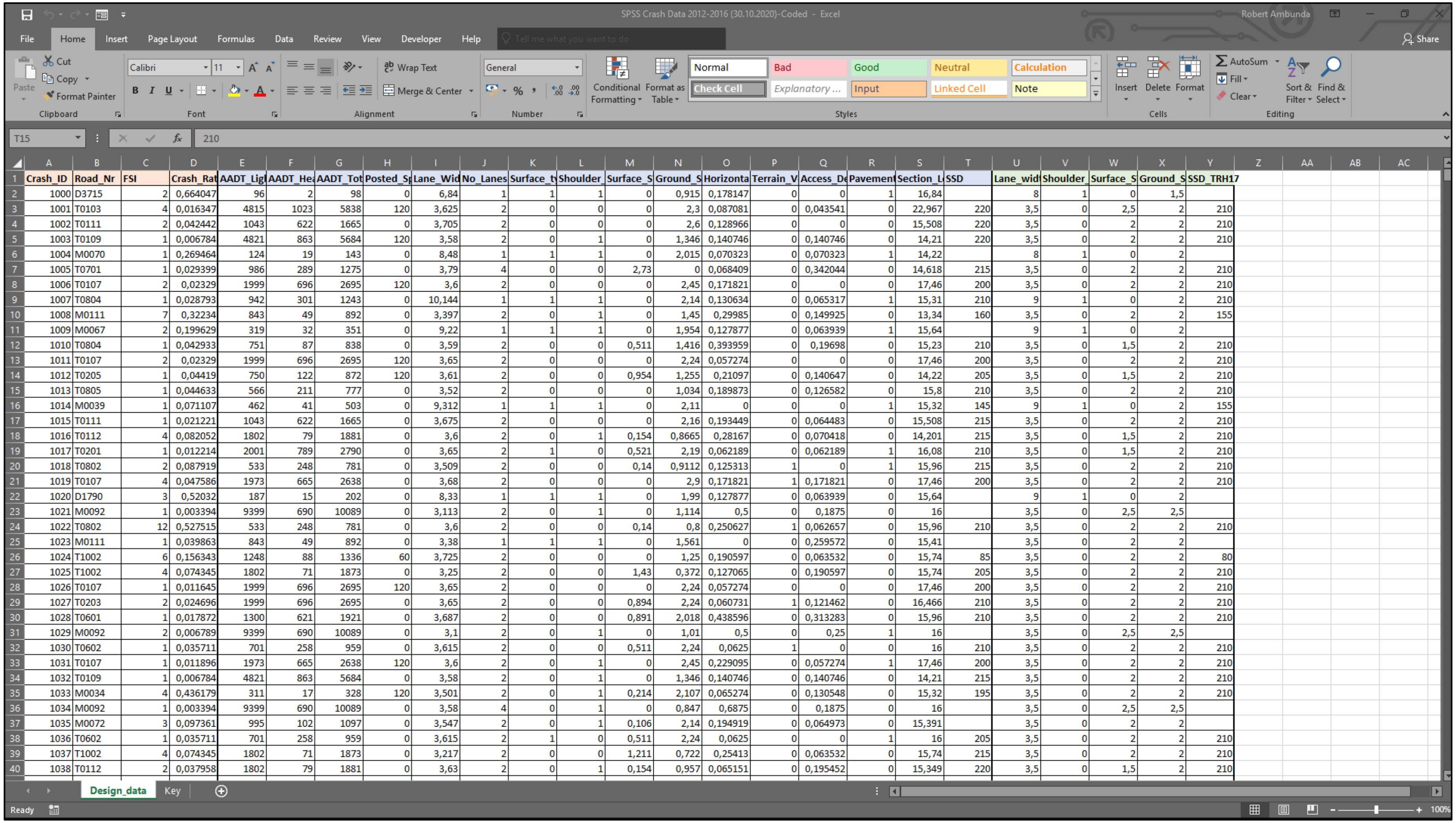
Zhang, Y., Liu, T., Bai, Q., Shao, W. and Wang, Q. (2018) 'New systems-based method to conduct analysis of road traffic accidents', *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 54, pp. 96–109. doi: 10.1016/j.trf.2018.01.019.

Zheng, L., Michael, R. R., Khattak, A. and Wang, X. (2011) 'All Accidents are Not Equal: Using Geographically Weighted Regressions Models to Assess and Forecast Accident Impacts', pp. 1–13.

Zimmerman, K., Mzige, A. A., Kibatala, P. L., Museru, L. M. and Guerrero, A. (2012) 'Road traffic injury incidence and crash characteristics in Dar es Salaam: A population based study', *Accident Analysis and Prevention*, 45, pp. 204–210. doi: 10.1016/j.aap.2011.06.018.

Appendices

A. Appendix A: Design Compliance



	Crash_ID	Road_Nr	FSI	Crash_Rat	AADT_Ligl	AADT_He	AADT_Tot	Posted_Sp	Lane_Wid	No_Lanes	Surface_ty	Shoulder_	Surface_S	Ground_S	Horizonta	Terrain_V	Access_De	Pavement	Section_L	SSD	Lane_widt	Shoulder_	Surface_S	Ground_S	SSD_TRH17
2	1000	D3715	2	0,664047	96	2	98	0	6,84	1	1	1	0	0,915	0,178147	0	0	1	16,84		8	1	0	1,5	
3	1001	T0103	4	0,016347	4815	1023	5838	120	3,625	2	0	0	0	2,3	0,087081	0	0,043541	0	22,967	220	3,5	0	2,5	2	210
4	1002	T0111	2	0,042442	1043	622	1665	0	3,705	2	0	0	0	2,6	0,128966	0	0	0	15,508	220	3,5	0	2	2	210
5	1003	T0109	1	0,006784	4821	863	5684	120	3,58	2	0	1	0	1,346	0,140746	0	0,140746	0	14,21	220	3,5	0	2	2	210
6	1004	M0070	1	0,269464	124	19	143	0	8,48	1	1	1	0	2,015	0,070323	0	0,070323	1	14,22		8	1	0	2	
7	1005	T0701	1	0,029399	986	289	1275	0	3,79	4	0	0	2,73	0	0,068409	0	0,342044	0	14,618	215	3,5	0	2	2	210
8	1006	T0107	2	0,02329	1999	696	2695	120	3,6	2	0	0	0	2,45	0,171821	0	0	0	17,46	200	3,5	0	2	2	210
9	1007	T0804	1	0,028793	942	301	1243	0	10,144	1	1	1	0	2,14	0,130634	0	0,065317	1	15,31	210	9	1	0	2	210
10	1008	M0111	7	0,32234	843	49	892	0	3,397	2	0	1	0	1,45	0,29985	0	0,149925	0	13,34	160	3,5	0	2	2	155
11	1009	M0067	2	0,199629	319	32	351	0	9,22	1	1	1	0	1,954	0,127877	0	0,063939	1	15,64		9	1	0	2	
12	1010	T0804	1	0,042933	751	87	838	0	3,59	2	0	0	0,511	1,416	0,393959	0	0,19698	0	15,23	210	3,5	0	1,5	2	210
13	1011	T0107	2	0,02329	1999	696	2695	120	3,65	2	0	0	0	2,24	0,057274	0	0	0	17,46	200	3,5	0	2	2	210
14	1012	T0205	1	0,04419	750	122	872	120	3,61	2	0	0	0,954	1,255	0,21097	0	0,140647	0	14,22	205	3,5	0	1,5	2	210
15	1013	T0805	1	0,044633	566	211	777	0	3,52	2	0	0	0	1,034	0,189873	0	0,126582	0	15,8	210	3,5	0	2	2	210
16	1014	M0039	1	0,071107	462	41	503	0	9,312	1	1	1	0	2,11	0	0	0	1	15,32	145	9	1	0	2	155
17	1015	T0111	1	0,021221	1043	622	1665	0	3,675	2	0	0	0	2,16	0,193449	0	0,064483	0	15,508	215	3,5	0	2	2	210
18	1016	T0112	4	0,082052	1802	79	1881	0	3,6	2	0	1	0,154	0,8665	0,28167	0	0,070418	0	14,201	215	3,5	0	1,5	2	210
19	1017	T0201	1	0,012214	2001	789	2790	0	3,65	2	1	0	0,521	2,19	0,062189	0	0,062189	1	16,08	210	3,5	0	1,5	2	210
20	1018	T0802	2	0,087919	533	248	781	0	3,509	2	0	0	0,14	0,9112	0,125313	1	0	1	15,96	215	3,5	0	2	2	210
21	1019	T0107	4	0,047586	1973	665	2638	0	3,68	2	0	0	0	2,9	0,171821	1	0,171821	0	17,46	200	3,5	0	2	2	210
22	1020	D1790	3	0,52032	187	15	202	0	8,33	1	1	1	0	1,99	0,127877	0	0,063939	0	15,64		9	1	0	2	
23	1021	M0092	1	0,003394	9399	690	10089	0	3,113	2	0	1	0	1,114	0,5	0	0,1875	0	16		3,5	0	2,5	2,5	
24	1022	T0802	12	0,527515	533	248	781	0	3,6	2	0	0	0,14	0,8	0,250627	1	0,062657	0	15,96	210	3,5	0	2	2	210
25	1023	M0111	1	0,039863	843	49	892	0	3,38	1	1	1	0	1,561	0	0	0,259572	0	15,41		3,5	0	2	2	
26	1024	T1002	6	0,156343	1248	88	1336	60	3,725	2	0	0	0	1,25	0,190597	0	0,063532	0	15,74	85	3,5	0	2	2	80
27	1025	T1002	4	0,074345	1802	71	1873	0	3,25	2	0	0	1,43	0,372	0,127065	0	0,190597	0	15,74	205	3,5	0	2	2	210
28	1026	T0107	1	0,011645	1999	696	2695	120	3,65	2	0	0	0	2,24	0,057274	0	0	0	17,46	200	3,5	0	2	2	210
29	1027	T0203	2	0,024696	1999	696	2695	0	3,65	2	0	0	0,894	2,24	0,060731	1	0,121462	0	16,466	210	3,5	0	2	2	210
30	1028	T0601	1	0,017872	1300	621	1921	0	3,687	2	0	0	0,891	2,018	0,438596	0	0,313283	0	15,96	210	3,5	0	2	2	210
31	1029	M0092	2	0,006789	9399	690	10089	0	3,1	2	0	1	0	1,01	0,5	0	0,25	1	16		3,5	0	2,5	2,5	
32	1030	T0602	1	0,035711	701	258	959	0	3,615	2	0	0	0,511	2,24	0,0625	1	0	0	16	210	3,5	0	2	2	210
33	1031	T0107	1	0,011896	1973	665	2638	120	3,6	2	0	1	0	2,45	0,229095	0	0,057274	1	17,46	200	3,5	0	2	2	210
34	1032	T0109	1	0,006784	4821	863	5684	0	3,58	2	0	1	0	1,346	0,140746	0	0,140746	0	14,21	215	3,5	0	2	2	210
35	1033	M0034	4	0,436179	311	17	328	120	3,501	2	0	1	0,214	2,107	0,065274	0	0,130548	0	15,32	195	3,5	0	2	2	210
36	1034	M0092	1	0,003394	9399	690	10089	0	3,58	4	0	1	0	0,847	0,6875	0	0,1875	0	16		3,5	0	2,5	2,5	
37	1035	M0072	3	0,097361	995	102	1097	0	3,547	2	0	1	0,106	2,14	0,194919	0	0,064973	0	15,391		3,5	0	2	2	
38	1036	T0602	1	0,035711	701	258	959	0	3,615	2	1	0	0,511	2,24	0,0625	0	0	1	16	205	3,5	0	2	2	210
39	1037	T1002	4	0,074345	1802	71	1873	0	3,217	2	0	0	1,211	0,722	0,25413	0	0,063532	0	15,74	215	3,5	0	2	2	210
40	1038	T0112	2	0,037958	1802	79	1881	0	3,63	2	0	1	0,154	0,957	0,065151	0	0,195452	0	15,349	220	3,5	0	1,5	2	210

Figure A.1 Compliance assessment of national rural roads (existing – purple; design standards – green)

B. Appendix B: Crash data analysis: Best-fit models (MLR)

This section presents extra information on the factor selection and analysis carried out for the development of the novel best-fit crash predictive models (MLR-CPMs) in the study. The information for CPM 2 was used to explain the method in Chapter 3, therefore not included here.

1. CPM 1: All Rural Roads

a) Factor analysis

Table B.1 Principle factor components from factor loadings-Varimax normalised for All Rural Roads

Variable	Factor Loadings (Varimax normalized) (Low Order Rural Roads Extraction: Principal components (Marked loadings are >.48)				
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
AADT_Heavy	-0,155	0,092	0,883	0,033	0,111
AADT_Total	-0,291	0,100	0,834	0,083	-0,183
Ops	0,028	0,165	-0,079	0,701	0,352
Lane_Width	0,860	0,139	-0,217	0,080	-0,151
No_Lanes	-0,472	-0,364	0,538	0,061	0,205
Shoulder_type	0,885	0,094	-0,183	-0,030	-0,155
Surface_SW	0,133	0,882	0,024	-0,043	0,155
Ground_SW	-0,112	-0,889	0,002	0,079	-0,010
Horizontal_(Curves/Length)	0,052	0,687	0,148	0,153	-0,390
Terrain_Vertical	0,194	0,073	0,379	-0,397	0,366
Access_Density	0,095	-0,138	0,206	0,700	-0,141
Pavement_Condition	-0,104	-0,054	0,061	0,044	0,735
SSD	0,731	0,058	-0,015	-0,005	0,072
Expl.Var	-0,511	-0,049	0,157	-0,093	-0,366
Prp.Totl	2,743	2,278	2,088	1,199	1,268

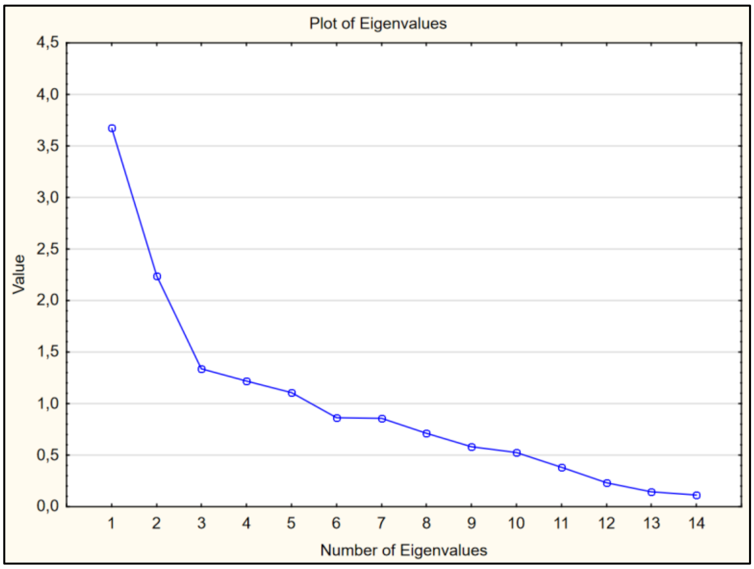


Figure B.1 Scree plot for All Rural Roads

b) Durbin-Watson test

Table B.2 Durbin-Watson Test for All Rural Roads CPM

Durbin-Watson d (CR Model and Serial Correlation of Residual)		
	Durbin-Watson d	Serial Corr.
Estimate	1.904595	0.045799

c) Outlier analysis

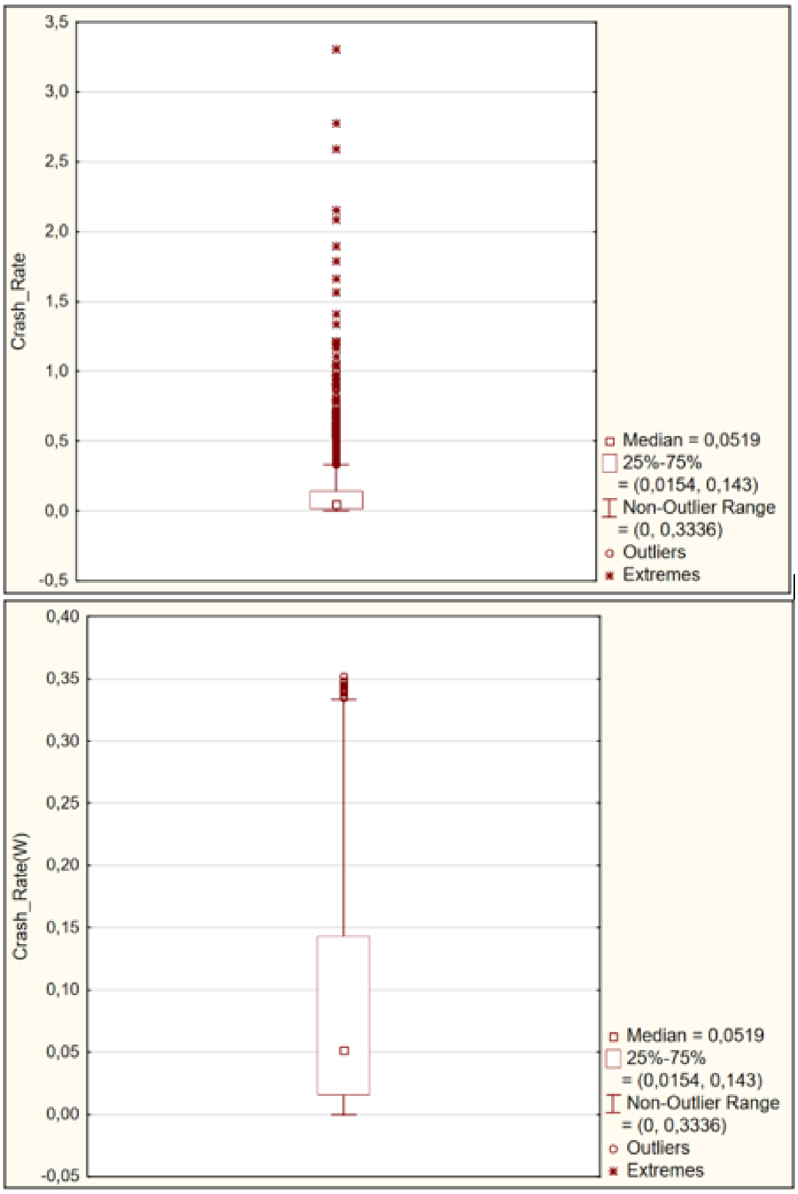


Figure B.2 2D Box Plots of the crash rate distribution before and after Winsorization: All Rural Roads

2. CPM 3: Low Order Rural Roads

a) Factor analysis

Table B.3 Principle factor components from factor loadings-Varimax normalised for Low Order Rural Roads

Variable	Factor Loadings (Varimax normalized) (Low Order Rural Roads Extraction: Principal components (Marked loadings are >.49)				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
AADT_Heavy	-0,924	-0,009	-0,055	0,033	0,009
AADT_Light	-0,924	0,120	0,067	0,035	0,008
Ops	0,586	0,387	-0,103	0,094	0,320
Lane_Width	0,787	0,431	-0,085	-0,136	-0,013
No_Lanes	-0,766	-0,275	0,168	0,164	0,002
Shoulder_type	0,190	0,903	0,073	0,002	-0,004
Surface_SW	-0,187	-0,905	-0,061	-0,001	0,034
Ground_SW	-0,090	0,513	-0,210	-0,362	0,018
Horizontal_(Curves/Length)	-0,182	-0,011	0,739	-0,200	-0,013
Terrain_Vertical	0,082	0,016	0,033	-0,167	0,851
Access_Density	0,009	0,019	0,782	0,209	0,031
Pavement_Condition	0,092	0,005	0,003	-0,800	0,193
SSD	-0,347	-0,197	-0,030	0,494	0,428
Expl.Var	3,504	2,363	1,266	1,183	1,050
Prp.Totl	0,270	0,182	0,097	0,091	0,081

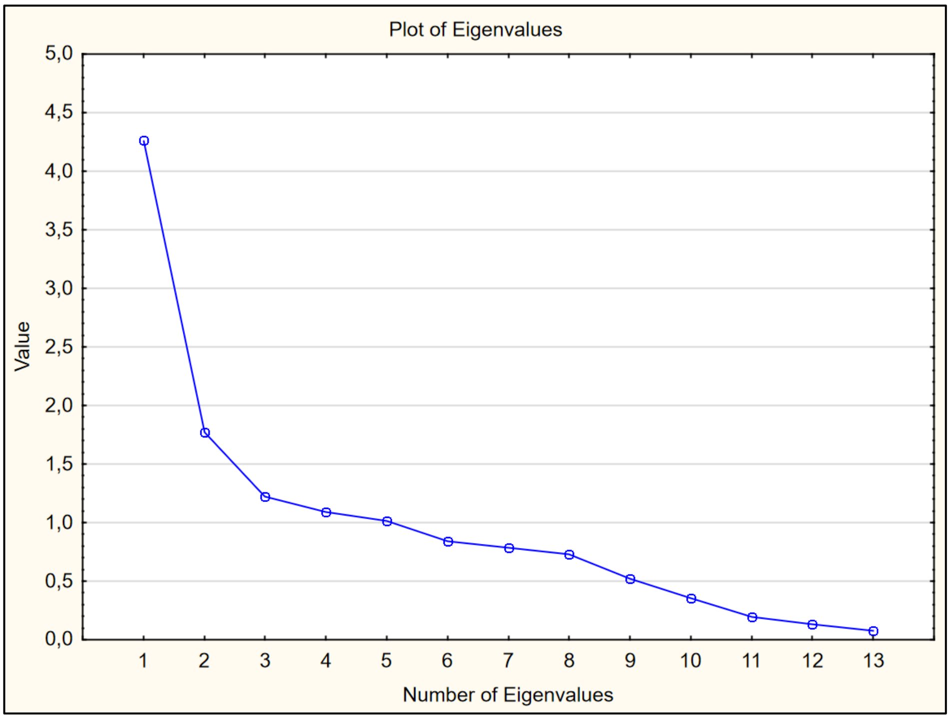


Figure B.3 Scree plot for Low Order Rural Roads

b) Durbin-Watson test

Table B.4 Durbin-Watson Test for Low Order Rural Roads CPM

Durbin-Watson d (CR Model and Serial Correlation of Residual)		
	Durbin-Watson d	Serial Corr.
Estimate	1.922284	0.035597

c) Outlier analysis

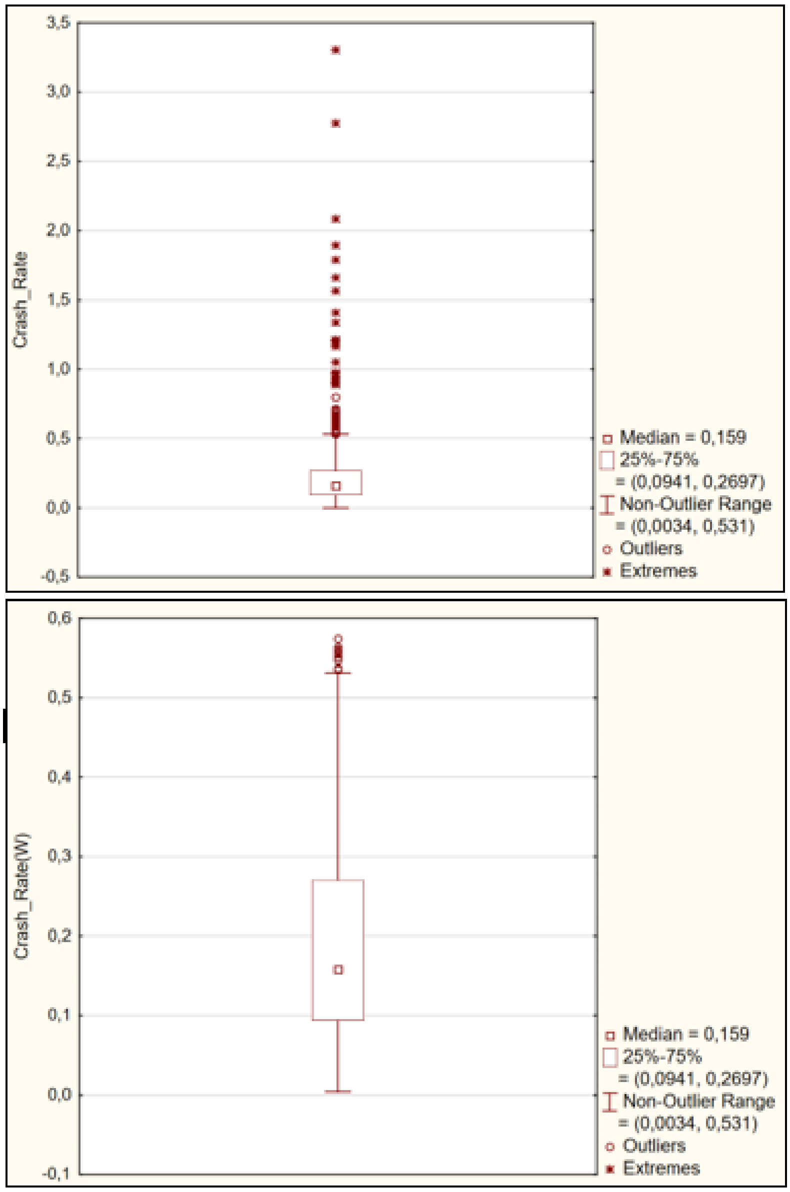


Figure B.4 2D Box Plots of the crash rate distribution before and after Winsorization: Low Order Rural Roads

C. Appendix C: Road Crash Prediction Models

This section presents extra model information on the novel crash predictive models (CPMs) developed in the study – both the best-fit models (MLR) and the base test models (BMM).

1. BMM: Developed test and parameter estimates

a) CPM 1: All Rural Roads

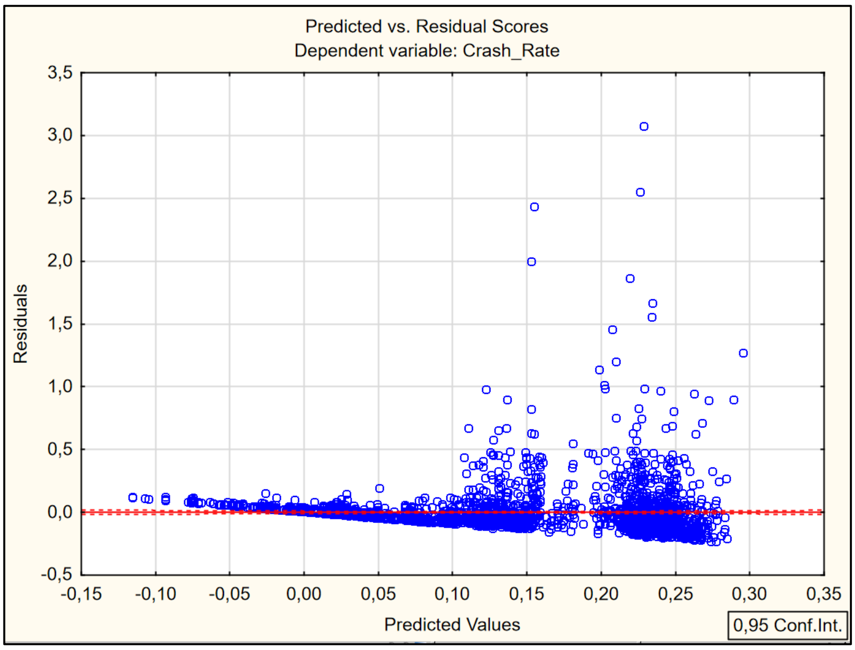


Figure C.1 BMM CPM 1 Predicted model values vs residual dataset values

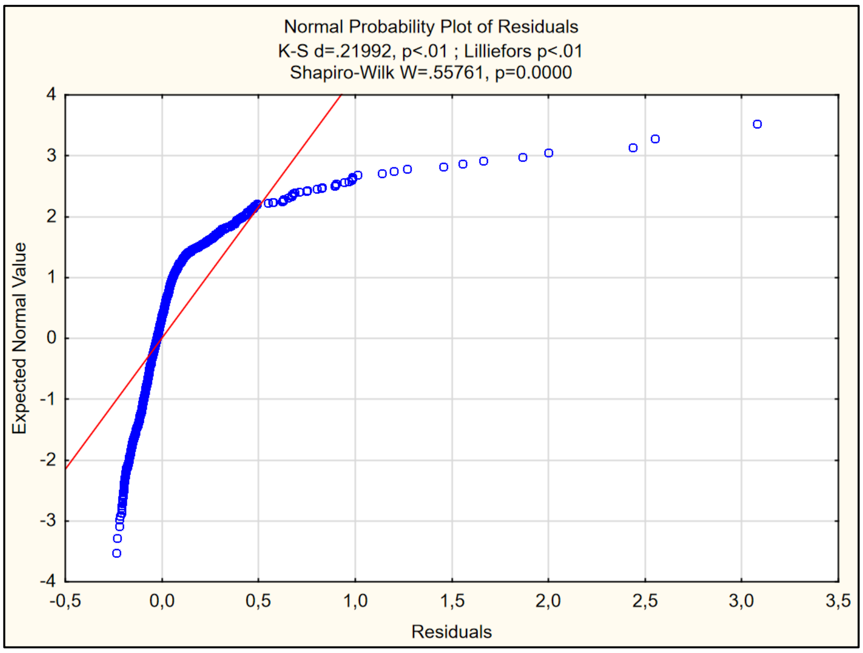


Figure C.2 Normal probability plot of residuals BMM CPM 1

Table C.1 Summary of best subset models for BMM CPM 1

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Base Mean) (All Rural Roads) R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_ Light	AADT_ Heavy	85th Operating Speed (Ops)	Lane_ Width	No_L anes	Surface _type	Shoulde r_type	Surfac e_SW	Groun d_SW	Horizont al (C/L)	Terrain_ Vertical	Access_ Density	Pavement_C ondition	Sight SD
4	0,201	**5**	**-**	0,308	0,070	0,207				-0,056			0,043			
5	0,201	5	-	0,287	0,079	0,221				-0,073	-0,050					
6	0,201	5	-	0,302	0,079	0,206				-0,050				-0,042		
12	0,200	5	-	0,307	0,080	0,221				-0,053						0,032
13	0,200	5	-	0,302	0,075	0,214							0,039	-0,044		
16	0,200	5	-	0,299	0,077	0,209			0,036					-0,045		
18	0,199	5	-	0,303	0,077	0,223				-0,053					-0,024	
21	0,199	5	-	0,304	0,068	0,212			0,039				0,043			
25	0,199	5	-	0,312	0,089			0,200		-0,060			0,048			
26	0,199	5	-	0,301	0,077	0,212				-0,051		0,002				
27	0,199	5	-	0,301	0,077	0,213	0,002			-0,052						
30	0,199	5	-	0,287	0,082	0,225					-0,027			-0,048		
31	0,199	5	-	0,300	0,083	0,226								-0,041		0,026
32	0,199	5	-	0,305	0,098			0,198		-0,053				-0,045		
37	0,199	5	-	0,285	0,075	0,224			0,050		-0,039					
40	0,199	5	-	0,297	0,081	0,226								-0,042	-0,018	
42	0,199	5	-	0,305	0,075	0,230							0,037			0,028
43	0,198	5	-	0,297	0,081	0,218						0,011		-0,044		
44	0,198	5	-	0,293	0,081	0,211	-0,014							-0,042		
45	0,198	5	-	0,303	0,078	0,226			0,035							0,032

Table C.2 BMM CPM 1 Parameter Estimates

N=3189	Regression Summary for Dependent Variable: Crash_Rate (All Rural Roads) R= 0.46665730; R²= 0.21776903; Adjusted R²= 0.21654027; CV-R^2=0.21 F (5,3183) = 238.39; p<0.0000 Std. Error of estimate = 0.05811						
	b*	Std. Err. of b*	b	Std. Err. of b	t (3183)	p-value	No. of times in best 20 SM
Intercept			0,084	0,011	7,777	0,000	
AADT_Heavy	0,323	0,018	0,000	0,000	17,797	0,000	20
85^{th} Percentile Speed (Ops)	0,050	0,016	0,000	0,000	3,096	0,002	20
Lane Width	0,211	0,019	0,016	0,001	11,051	0,000	18
Surface_SW	-0,057	0,016	-0,020	0,006	-3,454	0,001	9
Terrain_Vertical	0,045	0,016	0,022	0,008	2,827	0,005	5
AADT_Light	Excluded						0
No_Lanes	Excluded						2
Surface_type	Excluded						2
Shoulder_type	Excluded						4
Ground_SW	Excluded						3
Horizontal (Curves/ length)	Excluded						2
Access_Density	Excluded						9
Pavement _Condition	Excluded						2
SSD	Excluded						4

b) CPM 2: High Order Rural Roads

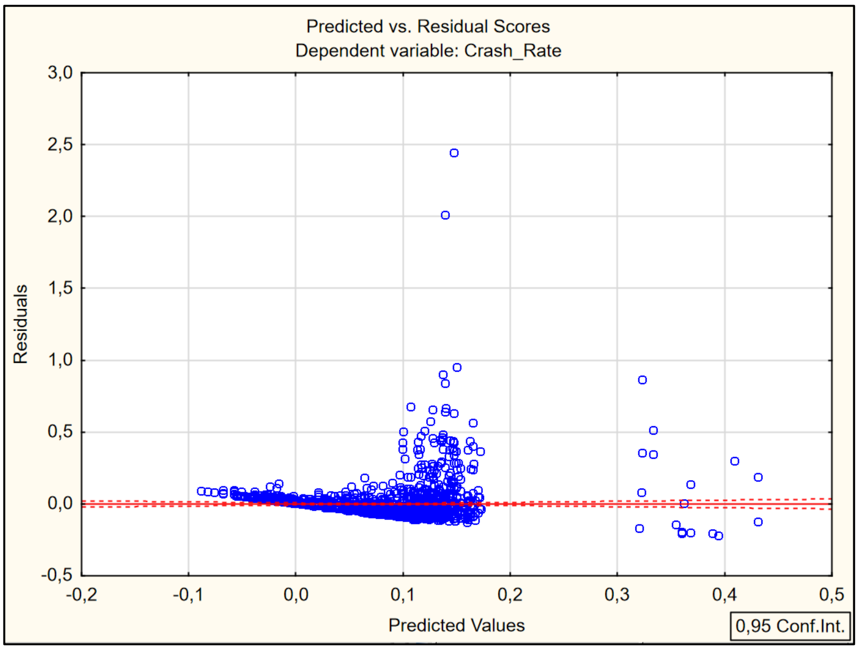


Figure C.3 BMM CPM 2 Predicted model values vs residual dataset values

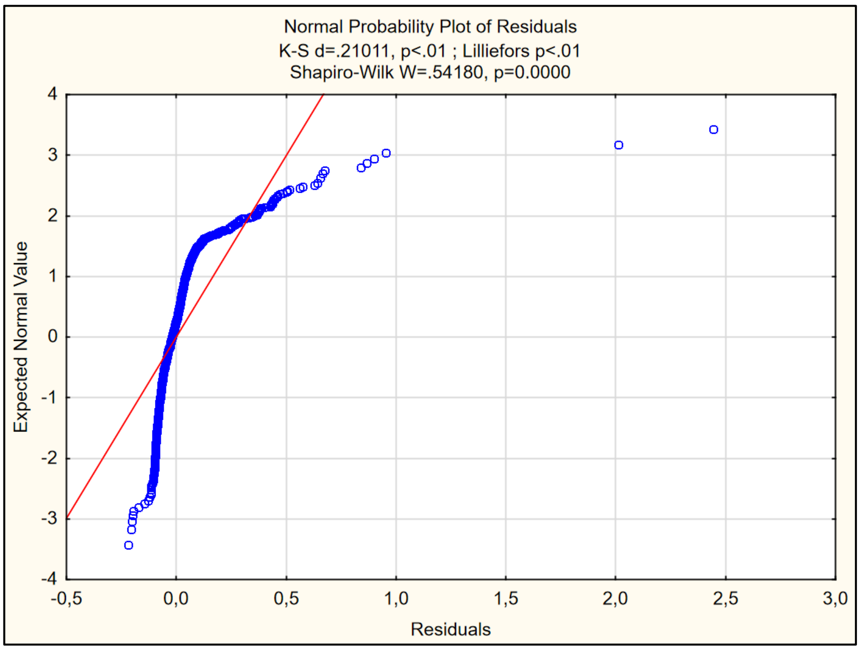


Figure C.4 Normal probability plot of residuals BMM CPM 2

Table C.3 Summary of best subset models for BMM CPM 2

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Base Mean) (High Order Rural Roads) R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_Light	AADT_Heavy	85th Operating Speed (Ops)	Lane_Width	No_Lanes	Surface_type	Shoulder_type	Surface_SW	Ground_SW	Horizontal (C/L)	Terrain_Vertical	Access_Density	Pavement_Condition	Sight SD
3	0,177	**5**	**-**	0,455	0,086	0,116		**-**			0,090		0,044			
4	0,176	5	-	0,449	0,094	0,125		-		-0,040	0,070					
6	0,176	5	-	0,459	0,091	0,123	0,029	-			0,098					
7	0,175	5	-	0,449	0,095	0,121		-			0,085			-0,020		
8	0,175	5	-	0,438	0,089	0,122		-		-0,076			0,050			
9	0,175	5	-	0,450	0,092	0,125		-	0,016		0,082					
10	0,175	5	-	0,451	0,094	0,122		-			0,090				-0,007	
11	0,175	5	-	0,450	0,093	0,123		-			0,090					0,002
12	0,175	5	-	0,450	0,093	0,123		-			0,090	-0,001				
14	0,174	5	-	0,432	0,100	0,126		-		-0,070				-0,035		
15	0,174	5	-	0,440	0,096	0,130	0,032	-		-0,082						
16	0,173	5	-	0,430	0,097	0,127		-		-0,073		-0,012				
17	0,173	5	-	0,433	0,099	0,128		-		-0,072						0,010
18	0,173	5	-	0,432	0,098	0,129		-		-0,072					-0,009	
21	0,173	5	-	0,434	0,086	0,125		-	0,059				0,050			
25	0,172	5	-	0,428	0,097	0,129		-	0,054					-0,041		
30	0,171	5	-	0,456		0,106		-		-0,046	0,074		0,062			
31	0,171	5	-	0,427	0,094	0,118		-					0,044	-0,040		
32	0,171	5	-	0,432	0,094	0,133	0,017	-	0,056							
33	0,171	5	-	0,428	0,096	0,131		-	0,054						-0,011	

Table C.4 BMM CPM 2 Parameter Estimates

N=2232	Regression Summary for Dependent Variable: Crash_Rate (All Rural Roads) R= 0.45963999; R^2= 0.21126892; Adjusted R^2= 0.20949729; CV-R^2=0.20 F (5,2226) = 141.88; p<0.0000 Std. Error of estimate = 0.04659						
	b*	Std. Err. of b*	b	Std. Err. of b	t (2226)	p-value	No. of times in best 20 SM
Intercept			-0,092	0,022	-4,149	0,000	
AADT_Heavy	0,466	0,020	0,000	0,000	23,077	0,000	20
85^{th} Percentile Speed (Ops)	0,039	0,019	0,000	0,000	2,076	0,038	19
Lane Width	0,176	0,019	0,055	0,006	9,226	0,000	20
Surface_SW	Excluded	-	-	-	-	-	8
Terrain_Vertical	0,057	0,019	0,020	0,007	2,943	0,003	5
AADT_Light	Excluded	-	-	-	-	-	0
No_Lanes	Excluded	-	-	-	-	-	3
Surface_type	Excluded	-	-	-	-	-	0
Shoulder_type	Excluded	-	-	-	-	-	5
Ground_SW	0,100	0,020	0,021	0,004	5,019	0,000	9
Horizontal (Curves/ length)	Excluded	-	-	-	-	-	2
Access_Density	Excluded	-	-	-	-	-	4
Pavement _Condition	Excluded	-	-	-	-	-	3
SSD	Excluded	-	-	-	-	-	2

c) CPM 3: Low Order Rural Roads

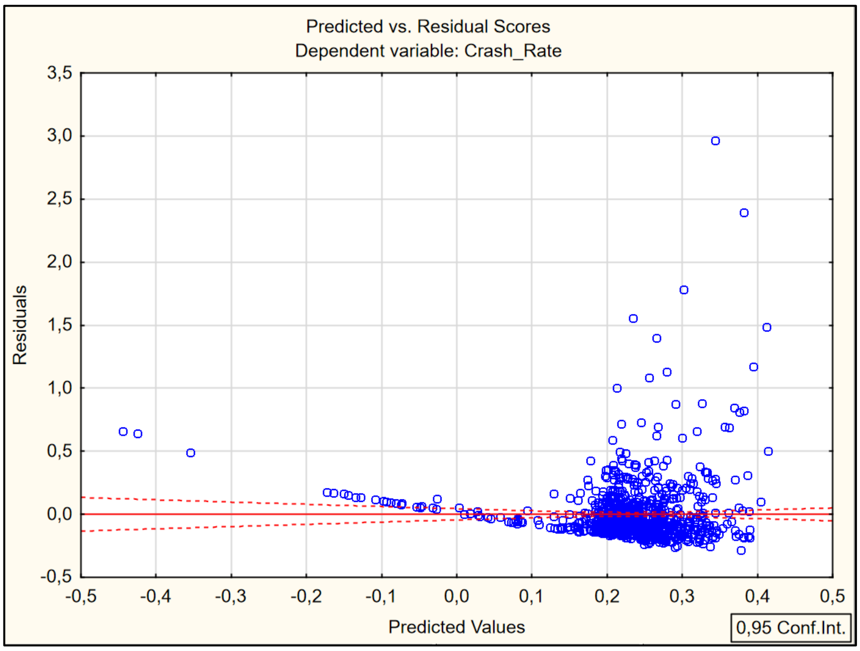


Figure C.5 BMM CPM 3 Predicted model values vs residual dataset values

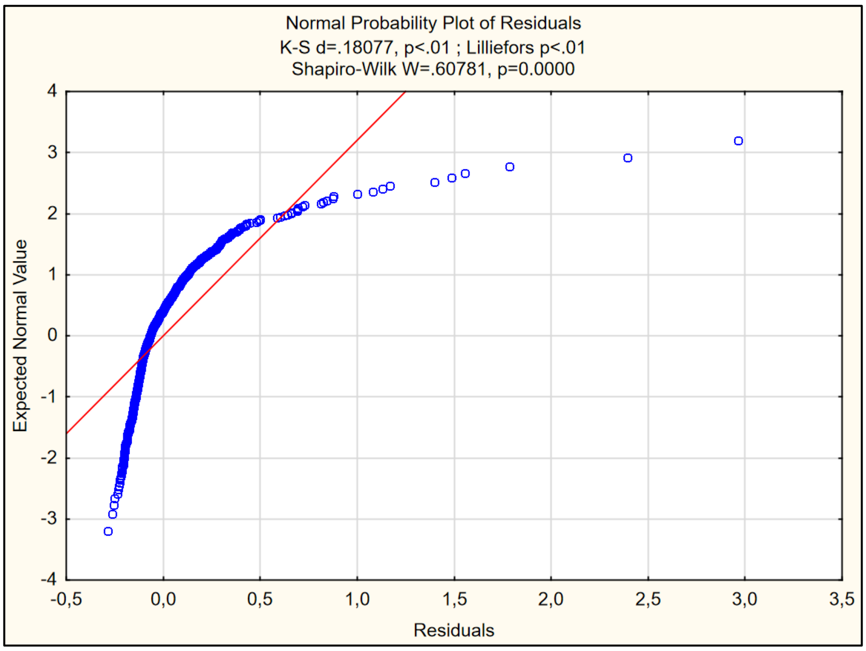


Figure C.6 Normal probability plot of residuals BMM CPM 3

Table C.5 Summary of best subset models for BMM CPM 3

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Base Mean) (Low Order Rural Roads) R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_Light	AADT_Heavy	85^{th} Operating Speed (Ops)	Lane_Width	No_Lanes	Surface_type	Shoulder_type	Surface_SW	Ground_SW	Horizontal (C/L)	Terrain_Vertical	Access_Density	Pavement_Condition	Sight SD
1	0,165	**5**	0,190		0,161					-0,128	-0,275	0,070				
2	0,165	5	0,174		0,154					-0,134	-0,284			-0,069		
3	0,163	5	0,220		0,186	-0,091				-0,155	-0,265					
4	0,162	5	0,194		0,161				0,113		-0,268	0,071				
5	0,162	5	0,179		0,153				0,120		-0,276			-0,069		
6	0,162	5	0,180		0,161					-0,129	-0,272				-0,041	
7	0,162	5	0,146		0,144		-0,059			-0,118	-0,283					
8	0,161	5	0,177		0,154					-0,129	-0,277		0,024			
9	0,161	5	0,182		0,155					-0,133	-0,275					0,018
14	0,160	5	0,221		0,182	-0,083			0,138		-0,257					
15	0,159	5		-0,154	0,167					-0,109	-0,276			-0,074		
16	0,159	5	0,185		0,160				0,115		-0,264				-0,042	
17	0,159	5	0,149		0,145		-0,060		0,102		-0,275					
18	0,158	5	0,184		0,203						-0,250	0,089		-0,079		
19	0,158	5	0,181		0,155				0,114		-0,269		0,020			
20	0,158	5	0,187		0,155				0,119		-0,267					0,018
21	0,158	5		-0,113	0,150		-0,086			-0,094	-0,279					
25	0,157	5	0,133		0,176		-0,102				-0,258	0,078				
26	0,156	5		-0,157	0,178					-0,102	-0,267	0,049				
27	0,156	5		-0,157	0,171				0,087		-0,267			-0,073		

Table C.6 BMM CPM 3 Parameter Estimates

N=957	Regression Summary for Dependent Variable: Crash_Rate (Low Order Rural Roads) R= 0.32908278; R^2= 0.10829547; Adjusted R^2= 0.10360723; CV-R^2=0.08 F (5,951) = 23.028; p<0.0000 Std. Error of estimate = 0.13083						
	b*	**Std. Err. of b***	**b**	**Std. Err. of b**	**t (951)**	**p-value**	**No. of times in best 20 SM**
Intercept			0,414	0,031	13,564	0,000	
AADT_Heavy	Excluded	-	-	-	-	-	4
85^{th} Percentile Speed (Ops)	0,098	0,031	0,001	0,000	3,181	0,002	20
Lane Width	Excluded	-	-	-	-	-	2
Surface_SW	-0,108	0,032	-0,204	0,060	-3,423	0,001	10
Terrain_Vertical	Excluded	-	-	-		-	2
AADT_Light	0,223	0,031	0,000	0,000	7,151	0,000	16
No_Lanes	Excluded	-	-	-	-	-	4
Surface_type	Excluded	-	-	-	-	-	0
Shoulder_type	Excluded	-	-	-	-	-	8
Ground_SW	-0,207	0,031	-0,094	0,014	-6,615	0,000	20
Horizontal (Curves/ length)	0,033	0,031	0,054	0,050	1,066	0,287	5
Access_Density	Excluded	-	-	-	-	-	5
Pavement _Condition	Excluded	-	-	-	-	-	2
SSD	Excluded	-	-	-	-	-	2

2. Continuous variable summary for best-fit MLR crash prediction models

Table C.7 CPM 1 (MLR) Continuous Variable Summary (All Rural Roads)

Parameter	All Rural Roads Continuous Summary									
	Valid N	Mean	Grubbs Test Stat.	P-value	Median	Min	Max	Low. Quartile	Upp. Quartile	Std. Dev.
Crash_Rate	3191	0,117	16,340	0,000	0,052	0,000	3,307	0,015	0,143	0,195
Crash_Rate(W)	3191	0,097	2,392	1,000	0,052	0,000	0,352	0,015	0,143	0,107
AADT_Light	3191	2328,440	3,997	0,200	952,000	85,000	14005,000	358,000	3111,000	2921,117
AADT_Heavy	3191	345,294	2,798	1,000	93,000	2,000	1400,000	41,000	690,000	376,970
Ops	3189	44,017	1,433	1,000	0,000	0,000	120,000	0,000	100,000	53,010
Lane_Width	3191	5,156	2,858	1,000	3,655	2,940	12,450	3,515	8,184	2,552
No_Lanes	3191	1,788	6,169	0,000	2,000	1,000	6,000	1,000	2,000	0,683
Surface_SW	3191	0,255	5,198	0,001	0,000	0,000	3,175	0,000	0,154	0,562
Ground_SW	3191	1,713	11,165	0,000	1,915	0,000	8,990	1,245	2,110	0,652
Horizontal_(Curves/Length)	3184	0,176	3,742	0,573	0,136	0,000	0,709	0,068	0,250	0,143
Access_Density	3191	0,121	3,337	1,000	0,127	0,000	0,409	0,064	0,188	0,086
SSD	1814	178,655	3,988	0,117	200,000	15,000	225,000	155,000	210,000	41,037

Table C.8 CPM 2 (MLR) Continuous Variable Summary (High Order Rural Roads)

Parameter	High Order Rural Roads Continuous Summary									
	Valid N	Mean	Grubbs Test Stat.	P-value	Median	Min	Max	Low. Quartile	Upp. Quartile	Std. Dev.
Crash_Rate	2234	0,070	18,272	0,000	0,026	0,000	2,591	0,011	0,068	0,138
Crash_Rate(W)	2234	0,051	2,341	1,000	0,026	0,000	0,180	0,011	0,068	0,055
AADT_Light	2234	461,074	2,446	1,000	402,000	3,000	1400,000	71,000	696,000	383,851
AADT_Heavy	2234	3551,247	3,447	1,000	2695,000	125,000	15362,000	838,000	5684,000	3426,048
Ops	2232	52,135	1,236	1,000	0,000	0,000	120,000	0,000	120,000	54,921
Lane_Width	2234	3,616	11,198	0,000	3,599	2,940	8,593	3,480	3,682	0,444
No_Lanes	2234	2,097	6,951	0,000	2,000	1,000	6,000	2,000	2,000	0,561
Surface_SW	2234	0,356	4,408	0,022	0,101	0,000	3,175	0,000	0,314	0,640
Ground_SW	2234	1,634	2,440	1,000	1,879	0,000	2,900	1,047	2,102	0,670
Horizontal_(Curves/Length)	2227	0,173	3,795	0,321	0,136	0,000	0,688	0,068	0,229	0,136
Access_Density	2234	0,128	3,105	1,000	0,130	0,000	0,387	0,064	0,188	0,083
SSD	1499	184,873	3,526	0,617	205,000	55,000	225,000	155,000	210,000	36,835

Table C.9 CPM 3 (MLR) Continuous Variable Summary (Low Order Rural Roads)

Parameter	Low Order Rural Roads Continuous Summary									
	Valid N	Mean	Grubbs Test Stat.	P-value	Median	Min	Max	Low. Quartile	Upp. Quartile	Std. Dev.
Crash_Rate	957	0,226	12,026	0,000	0,159	0,003	3,307	0,094	0,270	0,256
Crash_Rate(W)	957	0,203	2,468	1,000	0,159	0,003	0,575	0,094	0,270	0,150
AADT_Light	957	75,018	6,735	0,000	32,000	2,000	1152,000	18,000	63,000	159,920
AADT_Heavy	957	625,286	7,903	0,000	363,000	91,000	10089,000	252,000	517,000	1197,450
Ops	957	25,084	2,226	1,000	0,000	0,000	120,000	0,000	60,000	42,648
Lane_Width	957	8,749	3,429	0,558	9,116	3,000	12,450	8,660	9,450	1,676
No_Lanes	957	1,066	10,778	0,000	1,000	1,000	4,000	1,000	1,000	0,272
Surface_SW	957	0,020	13,088	0,000	0,000	0,000	1,794	0,000	0,000	0,136
Ground_SW	957	1,898	12,521	0,000	2,004	0,000	8,990	1,820	2,110	0,566
Horizontal_(Curves/Length)	957	0,182	3,344	0,764	0,140	0,000	0,709	0,065	0,267	0,158
Access_Density	957	0,105	3,354	0,738	0,072	0,000	0,409	0,000	0,144	0,091
SSD	315	149,063	2,866	1,000	145,000	15,000	220,000	115,000	195,000	46,770

3. CPMs developed with Road Design Guidelines (TRH 17 and TRH 26) performance tests and parameter estimates

a) CPM 4: All Rural Roads (MLR tests and parameter estimates)

Table C.10 CPM 4 Breusch-Pagan test

Breusch-Pagan Test for heteroskedasticity		
BP	df	p-value
402.43	5	0.000

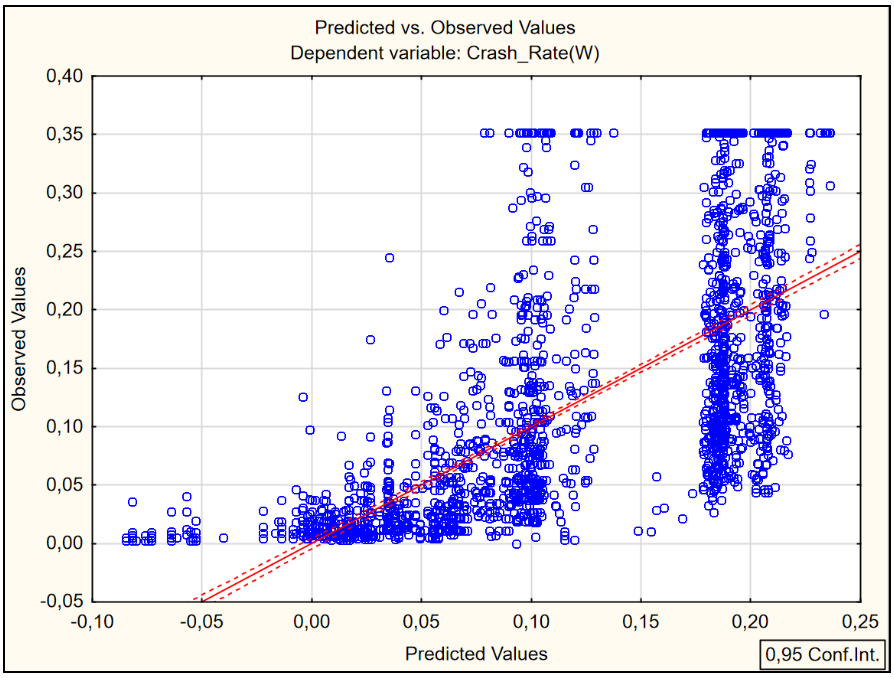


Figure C.7 CPM 4 Predicted model values vs observed dataset values

Table C.11 CPM 4 Principal Component summary

Principal Component	Eigenvalues (All rural Roads) Extraction: Principal components			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative (%)
1	5,280	37,717	5,280	37,717
2	1,812	12,943	7,092	50,660
3	1,173	8,380	8,266	59,040
4	1,111	7,935	9,377	66,976
5	0,913	6,523	10,290	73,499
6	0,863	6,168	11,153	79,666
7	0,781	5,578	11,934	85,244
8	0,643	4,590	12,577	89,834
9	0,607	4,337	13,184	94,172
10	0,378	2,700	13,562	96,871

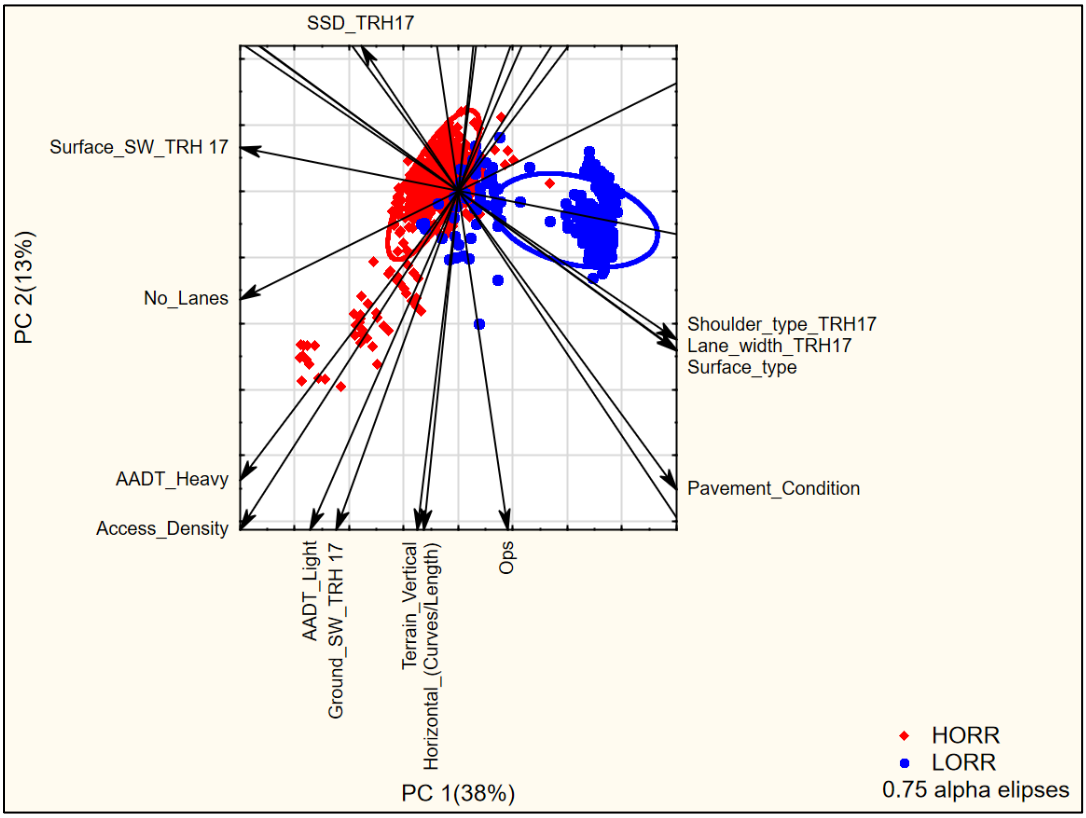


Figure C.8 CPM 4 Principal Component biplot

Table C.12 Summary of best subset models for CPM 4

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Winsorized) (All Rural Roads) R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_ Light	AADT_ Heavy	85^{th} Operating Speed (Ops)	Lane_ Width	No_L anes	Surface _type	Shoulde r_type	Surfac e_SW	Groun d_SW	Horizont al (C/L)	Terrain_ Vertical	Access_ Density	Pavement_C ondition	Sight SD
3	0,460	**5**	**-**	-0,422	0,077				0,335		-0,064		0,076			
15	0,458	5	-	-0,447	0,076				0,337				0,077	-0,037		
16	0,457	5	-	-0,450	0,075				0,352				0,075			0,032
17	0,457	5	-	-0,447	0,074				0,355				0,075		-0,025	
18	0,457	5	-	-0,444	0,073				0,344			-0,017	0,075			
23	0,456	5	-	-0,445	0,073		-0,002		0,342				0,077			
34	0,455	5	-	-0,413	0,088		0,043		0,361		-0,079					
35	0,455	5	-	-0,412	0,090				0,337		-0,062			-0,028		
37	0,455	5	-	-0,415	0,089				0,350		-0,061					0,028
39	0,455	5	-	-0,412	0,089				0,352		-0,063				-0,023	
41	0,454	5	-	-0,428					0,339		-0,055		0,088	-0,026		
42	0,454	5	-	-0,410	0,088				0,342		-0,063	-0,014				
45	0,454	5	-	-0,428			0,029		0,356		-0,068		0,083			
49	0,454	5	-	-0,430					0,348		-0,055		0,086			0,020
53	0,454	5	-	-0,427					0,350		-0,057		0,086		-0,017	
54	0,454	5	-	-0,426					0,343		-0,057	-0,012	0,086			
66	0,453	5	-	-0,440	0,088				0,354					-0,032		0,034
69	0,453	5	-	-0,452					0,352				0,087	-0,030		0,025
70	0,452	5	-	-0,437	0,087				0,356					-0,033	-0,027	
71	0,452	5	-	-0,450					0,354				0,087	-0,031	-0,021	

Table C.13 CPM 4 Parameter Estimates

N=3189	Regression Summary for Dependent Variable: Crash_Rate(W) (All Rural Roads) R= 0.69039066; R^2= 0.47663927; Adjusted R^2= 0.47581715; CV-R^2=0.47 F (5,3183) = 579.77; p<0.0000 Std. Error of estimate = 0.07719						
	b*	**Std. Err. of b***	**b**	**Std. Err. of b**	**t (3183)**	**p-value**	**No. of times in best 20 SM**
Intercept			0,176	0,014	12,779	0,000	
AADT_Heavy	-0,380	0,016	-0,000	0,000	-23,930	0,000	20
85^{th} Percentile Speed (Ops)	0,036	0,013	0,000	0,000	2,721	0,007	13
Lane Width	Excluded						0
Surface_SW	Excluded						0
Terrain_Vertical	0,076	0,013	0,020	0,004	5,808	0,000	13
AADT_Light	Excluded						0
No_Lanes	Excluded						3
Surface_type	Excluded						0
Shoulder_type	0,378	0,015	0,088	0,004	24,525	0,000	20
Ground_SW	-0,078	0,014	-0,037	0,007	-5,496	0,000	11
Horizontal (Curves/ length)	Excluded						3
Access_Density	Excluded						7
Pavement _Condition	Excluded						5
SSD	Excluded						5

b) CPM 5: High Order Rural Roads (MLR tests and parameter estimates)

Table C.14 CPM 5 Breusch-Pagan test

Breusch-Pagan Test for heteroskedasticity		
BP	df	p-value
354.69	5	0.000

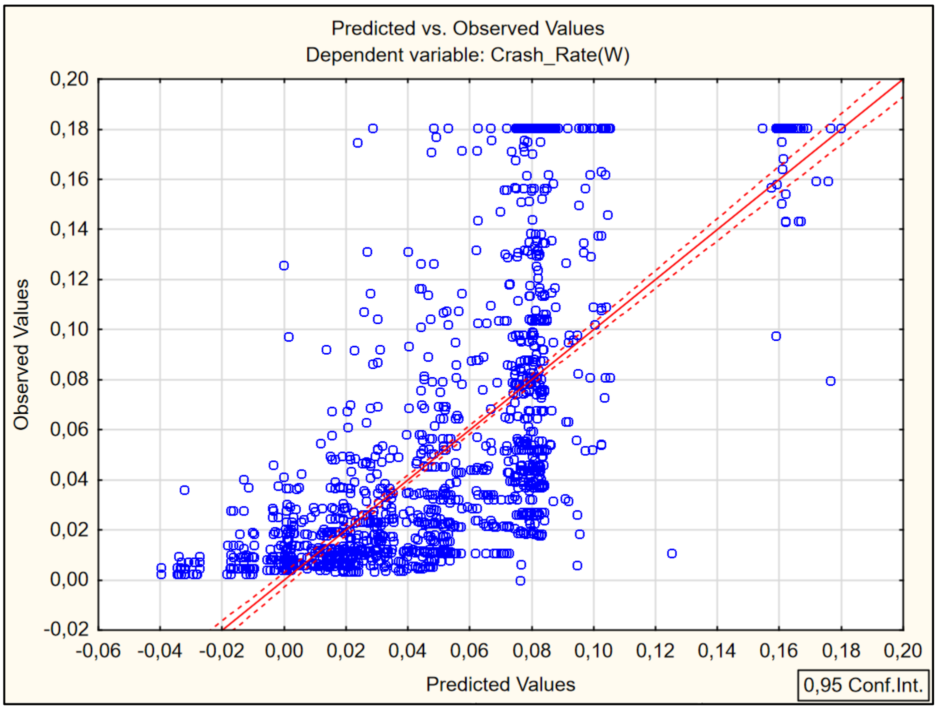


Figure C.9 CPM 5 Predicted model values vs observed dataset values

Table C.15 Summary of best subset models for CPM 5

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Winsorized) (High Order Rural Roads) -CPM 5 R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_ Light	AADT_ Heavy	85th Operating Speed (Ops)	Lane_ Width	No_L anes	Surface _type	Shoulde r_type	Surfac e_SW	Groun d_SW	Horizont al (C/L)	Terrain_ Vertical	Access_ Density	Pavement_C ondition	Sight SD
59	0,409	**5**	**-**	-0,606	0,076			**-**	0,135			-0,048	0,124			
61	0,409	5	-	-0,612	0,078			-	0,135				0,129	-0,046		
70	0,409	5	-	-0,586	0,081			-	0,127	-0,047			0,128			
75	0,408	5	-	-0,612	0,079			-	0,136				0,124		-0,034	
79	0,408	5	-	-0,609	0,075	0,025		-	0,129				0,128			
80	0,407	5	-	-0,617	0,074			-	0,138		0,018		0,128			
81	0,407	5	-	-0,612	0,077			-	0,136				0,127			0,015
82	0,407	5	-	-0,609	0,075		-0,005	-	0,137				0,128			
90	0,405	5	-	-0,610				-	0,135			-0,045	0,137	-0,039		
94	0,405	5	-	-0,620				-	0,137		0,032	-0,052	0,135			
96	0,404	5	-	-0,610				-	0,136			-0,047	0,133		-0,026	
97	0,404	5	-	-0,594				-	0,130	-0,029		-0,045	0,136			
98	0,404	5	-	-0,608		0,025		-	0,128			-0,047	0,136			
100	0,404	5	-	-0,626				-	0,138		0,031		0,141	-0,045		
101	0,404	5	-	-0,614		0,029		-	0,128				0,141	-0,043		
103	0,404	5	-	-0,599				-	0,131	-0,030			0,141	-0,039		
104	0,404	5	-	-0,609			0,004	-	0,136			-0,048	0,134			
105	0,404	5	-	-0,609				-	0,136			-0,047	0,135			0,004
108	0,404	5	-	-0,615				-	0,137				0,139	-0,039	-0,022	
113	0,403	5	-	-0,614				-	0,137				0,140	-0,041		0,002

Table C.16 CPM 5 Parameter Estimates

N=2225	Regression Summary for Dependent Variable: Crash_Rate(W) (All Rural Roads) – CPM 5 R= 0.66784454; R²= 0.44601633; Adjusted R²= 0.44476805; CV-R^2=0.44 F (5,2219) = 357.31; p<0.0000 Std. Error of estimate = 0.04121						
	b*	Std. Err. of b*	b	Std. Err. of b	t (2219)	p-value	No. of times in best 20 SM
Intercept			0,085	0,002	43,624	0,000	
AADT_Heavy	-0,594	0,016	-0,000	0,000	-36,212	0,000	20
85th Percentile Speed (Ops)	0,041	0,016	0,000	0,000	2,556	0,011	8
Lane Width	Excluded						3
Surface_SW	Excluded						3
Terrain_Vertical	0,120	0,016	0,017	0,002	7,421	0,000	20
AADT_Light	Excluded						0
No_Lanes	Excluded						2
Surface_type	Excluded	-	-	-	-	-	0
Shoulder_type	0,234	0,016	0,079	0,005	14,572	0,000	20
Ground_SW	Excluded						3
Horizontal (Curves/ length)	-0,033	0,016	-0,013	0,007	-2,064	0,039	8
Access_Density	Excluded						7
Pavement _Condition	Excluded						3
SSD	Excluded						3

c) CPM 6: Low Order Rural Roads (MLR tests and parameter estimates)

Table C.17 CPM 5 Breusch-Pagan test

Breusch-Pagan Test for heteroskedasticity		
BP	df	p-value
22.28	5	0.000

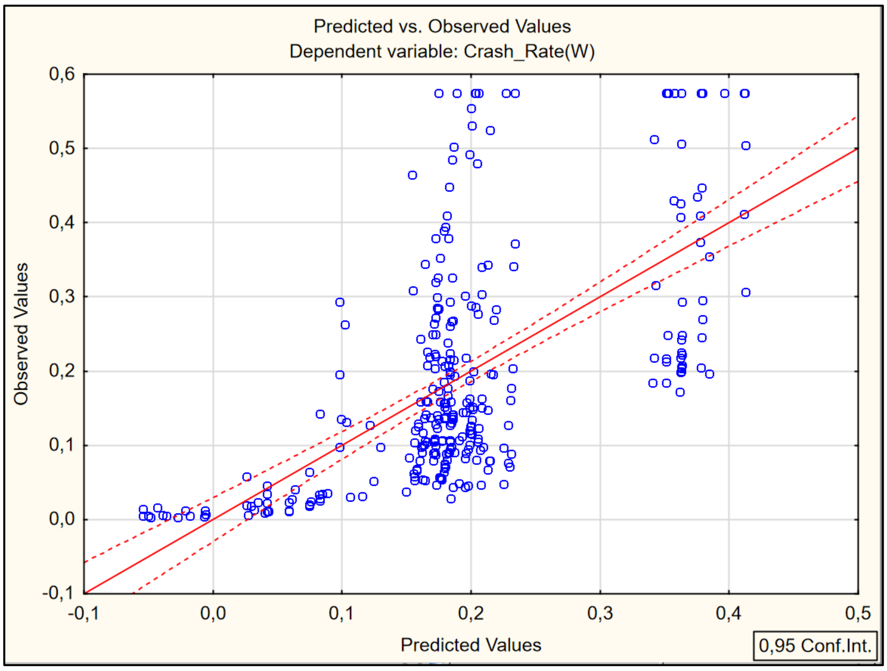


Figure C.10 CPM 6 Predicted model values vs observed dataset values

Table C.18 Summary of best subset models for CPM 6

Subset No.	Summary of best subsets; variable(s): Crash_Rate (Winsorized) (Low Order Rural Roads) -CPM 6 R square and standardized regression coefficients for each sub model															
	R Square	No. of Effects	AADT_Light	AADT_Heavy	85^{th} Operating Speed (Ops)	Lane_Width	No_Lanes	Surface_type	Shoulder_type	Surface_SW	Ground_SW	Horizontal (C/L)	Terrain_Vertical	Access_Density	Pavement_Condition	Sight SD
3	0,396	**5**	-0,204					**-**	0,241	**-**	-0,412		0,086			0,081
6	0,394	5	-0,222					-	0,200	-	-0,420	0,059	0,090			
10	0,393	5	-0,205		0,067			-	0,161	-	-0,416		0,084			
14	0,392	5	-0,206					-	0,201	-	-0,423		0,090	-0,037		
17	0,391	5	-0,220					-	0,246	-	-0,409	0,058				0,085
23	0,391	5		-0,252	0,115		-0,065	-	-	-	-0,418		0,072			
24	0,390	5		-0,283	0,132			-	-	-	-0,422		0,071	-0,046		
26	0,390	5	-0,208					-	0,202	-	-0,420		0,090		0,000	
30	0,390	5	-0,222		0,086			-	0,151	-	-0,412	0,063				
31	0,390	5	-0,204					-	0,247	-	-0,412			-0,038		0,086
34	0,390	5		-0,284	0,139			-	-	-	-0,419	0,035	0,070			
35	0,389	5	-0,205		0,048			-	0,211	-	-0,408					0,070
39	0,389	5		-0,282	0,134			-	-	-	-0,418		0,069		0,015	
40	0,388	5	-0,206					-	0,246	-	-0,409				0,017	0,087
41	0,388	5		-0,287	0,134			-	-	-	-0,418		0,070			0,013
48	0,388	5	-0,224					-	0,203	-	-0,421	0,070		-0,051		
52	0,388	5		-0,289	0,144			-	-	-	-0,419	0,048		-0,056		
57	0,387	5	-0,200		0,088	0,129		-	-	-	-0,440		0,082			
58	0,387	5		-0,255	0,126		-0,071	-	-	-	-0,414	0,044				
60	0,387	5	-0,208			0,188		-	-	-	-0,456	0,065	0,089			

Table C.19 CPM 6 Parameter Estimates

N=315	Regression Summary for Dependent Variable: Crash_Rate(W) (All Rural Roads) -CPM 6 R= 0.62892763; R²= 0.39554997; Adjusted R²= 0.38576922; CV-R^2=0.37 F (5,309) = 40.442; p<0.0000 Std. Error of estimate = 0.12416						
	b*	**Std. Err. of b***	**b**	**Std. Err. of b**	**t (309)**	**p-value**	**No. of times in best 20 SM**
Intercept			0,749	0,090	8,314	0,000	
AADT_Heavy	Excluded	-	-	-	-	-	7
85th Percentile Speed (Ops)	Excluded	-	-		-	-	11
Lane Width	Excluded	-	-	-	-	-	2
Surface_SW	Excluded	-	-	-	-	-	0
Terrain_Vertical	0,086	0,044	0,034	0,017	1,940	0,053	12
AADT_Light	-0,204	0,054	-0,000	0,000	-3,754	0,000	13
No_Lanes	Excluded	-	-	-	-	-	2
Surface_type	Excluded	-	-	-	-	-	0
Shoulder_type	0,241	0,059	0,101	0,025	4,074	0,000	11
Ground_SW	-0,412	0,046	-0,351	0,039	-9,020	0,000	20
Horizontal (Curves/ length)	Excluded	-	-	-	-	-	8
Access_Density	Excluded	-	-	-	-	-	5
Pavement _Condition	Excluded	-	-	-	-	-	3
SSD	0,081	0,049	0,000	0,000	1,635	0,103	6

D. Appendix D: Driver characteristics and risk factors – roadway condition analysis

1. Risk factor coding

Table D-1 presents coded risk factor combinations identified in the crash dataset used in the study.

Table D.1 Crash causation risk factor codes

No (Code)	Combinations		
1	1	2	3
2	1	2	4
3	1	2	5
4	1	2	6
5	1	2	7
6	2	3	4
7	2	3	5
8	2	3	6
9	2	3	7
10	3	4	1
11	3	4	5
12	3	4	6
13	3	4	7
14	4	5	1
15	4	5	2
16	4	5	6
17	4	5	7
18	5	6	1
19	5	6	2
20	5	6	3
21	5	6	7
22	6	7	1
23	6	7	2
24	6	7	3
25	6	7	4
26	1	1	
27	2	2	
28	3	3	
29	4	4	
30	5	5	
31	6	6	
32	7	7	
33	1	2	
34	1	3	
35	1	4	
36	1	5	
37	1	6	

38	1	7	
39	2	3	
40	2	4	
41	2	5	
42	2	6	
43	2	7	
44	3	4	
45	3	5	
46	3	6	
47	3	7	
48	4	5	
49	4	6	
50	4	7	
51	5	6	
52	5	7	
53	6	7	
54	1		
55	2		
56	3		
57	4		
58	5		
59	6		
60	1	1	2
61	1	1	3
62	1	1	4
63	1	1	5
64	1	1	6
65	1	1	7
66	2	2	1
67	2	2	3
68	2	2	4
69	2	2	5
70	2	2	6
71	2	2	7
72	3	3	1
73	3	3	2
74	3	3	4
75	3	3	5
76	3	3	6
77	3	3	7
78	4	4	1
79	4	4	2
80	4	4	3
81	4	4	5
82	4	4	6
83	4	4	7

84	5	5	1
85	5	5	2
86	5	5	3
87	5	5	4
88	5	5	6
89	5	5	7
90	6	6	1
91	6	6	2
92	6	6	3
93	6	6	4
94	6	6	5
95	6	6	7
96	7	7	1
97	7	7	2
98	7	7	3
99	7	7	4
100	7	7	5
101	7	7	6
102	1	3	4
103	1	3	5
104	1	3	6
105	1	3	7
106	1	4	2
107	1	4	5
108	1	4	6
109	1	4	7
110	1	5	2
111	1	5	3
112	1	5	6
113	1	5	7
114	1	6	2
115	1	7	3
116	1	7	4
117	2	4	5
118	2	4	6
119	2	4	7
120	6	6	6
121	7	5	3
122	7	5	2
123	1	1	1
124	2	2	2
125	3	3	3
126	4	4	4
127	7		

2. Frequency of risk factor combination

Table D.2 Crash causation risk factor frequency

Risk factor combinations				
Risk Factor Combination	Frequency (N)	Percent	Valid Percent	Cumulative Percent
1	51	2.3	2.3	2.3
2	153	7.0	7.0	9.3
3	46	2.1	2.1	11.4
4	114	5.2	5.2	16.6
6	36	1.6	1.6	18.2
7	8	0.4	0.4	18.6
8	13	0.6	0.6	19.2
9	2	0.1	0.1	19.3
11	41	1.9	1.9	21.1
12	31	1.4	1.4	22.6
13	4	0.2	0.2	22.7
14	3	0.1	0.1	22.9
15	29	1.3	1.3	24.2
16	7	0.3	0.3	24.5
18	4	0.2	0.2	24.7
19	1	0.0	0.0	24.7
23	8	0.4	0.4	25.1
24	7	0.3	0.3	25.4
25	13	0.6	0.6	26.0
26	12	0.5	0.5	26.6
27	8	0.4	0.4	26.9
28	5	0.2	0.2	27.2
29	17	0.8	0.8	27.9
31	16	0.7	0.7	28.7
32	1	0.0	0.0	28.7
33	124	5.6	5.6	34.4
34	61	2.8	2.8	37.1
35	41	1.9	1.9	39.0
36	8	0.4	0.4	39.4
37	66	3.0	3.0	42.4
39	25	1.1	1.1	43.5
40	46	2.1	2.1	45.6
41	2	0.1	0.1	45.7
42	22	1.0	1.0	46.7
43	4	0.2	0.2	46.9
44	22	1.0	1.0	47.9
45	6	0.3	0.3	48.2
46	5	0.2	0.2	48.4
47	19	0.9	0.9	49.2
48	7	0.3	0.3	49.6
49	16	0.7	0.7	50.3
50	9	0.4	0.4	50.7
51	1	0.0	0.0	50.8

53	15	0.7	0.7	51.4
54	48	2.2	2.2	53.6
55	14	0.6	0.6	54.3
56	29	1.3	1.3	55.6
57	28	1.3	1.3	56.9
58	1	0.0	0.0	56.9
59	63	2.9	2.9	59.8
60	21	1.0	1.0	60.7
61	18	0.8	0.8	61.5
62	57	2.6	2.6	64.1
63	1	0.0	0.0	64.2
64	1	0.0	0.0	64.2
66	36	1.6	1.6	65.9
67	1	0.0	0.0	65.9
68	13	0.6	0.6	66.5
70	8	0.4	0.4	66.9
72	6	0.3	0.3	67.2
73	1	0.0	0.0	67.2
74	1	0.0	0.0	67.2
75	1	0.0	0.0	67.3
76	1	0.0	0.0	67.3
77	4	0.2	0.2	67.5
78	89	4.1	4.1	71.6
79	28	1.3	1.3	72.8
80	14	0.6	0.6	73.5
81	47	2.1	2.1	75.6
82	39	1.8	1.8	77.4
90	132	6.0	6.0	83.4
91	47	2.1	2.1	85.6
92	5	0.2	0.2	85.8
93	22	1.0	1.0	86.8
94	2	0.1	0.1	86.9
95	5	0.2	0.2	87.1
102	79	3.6	3.6	90.7
103	12	0.5	0.5	91.3
104	29	1.3	1.3	92.6
106	3	0.1	0.1	92.7
107	12	0.5	0.5	93.3
108	44	2.0	2.0	95.3
110	1	0.0	0.0	95.3
111	1	0.0	0.0	95.4
112	5	0.2	0.2	95.6
115	1	0.0	0.0	95.6
117	5	0.2	0.2	95.9
118	55	2.5	2.5	98.4
119	3	0.1	0.1	98.5
120	2	0.1	0.1	98.6
121	3	0.1	0.1	98.7
122	2	0.1	0.1	98.8

123	2	0.1	0.1	98.9
124	3	0.1	0.1	99.0
126	20	0.9	0.9	100.0
Total	2195	100.0	100.0	

3. TSC-1 Model Information

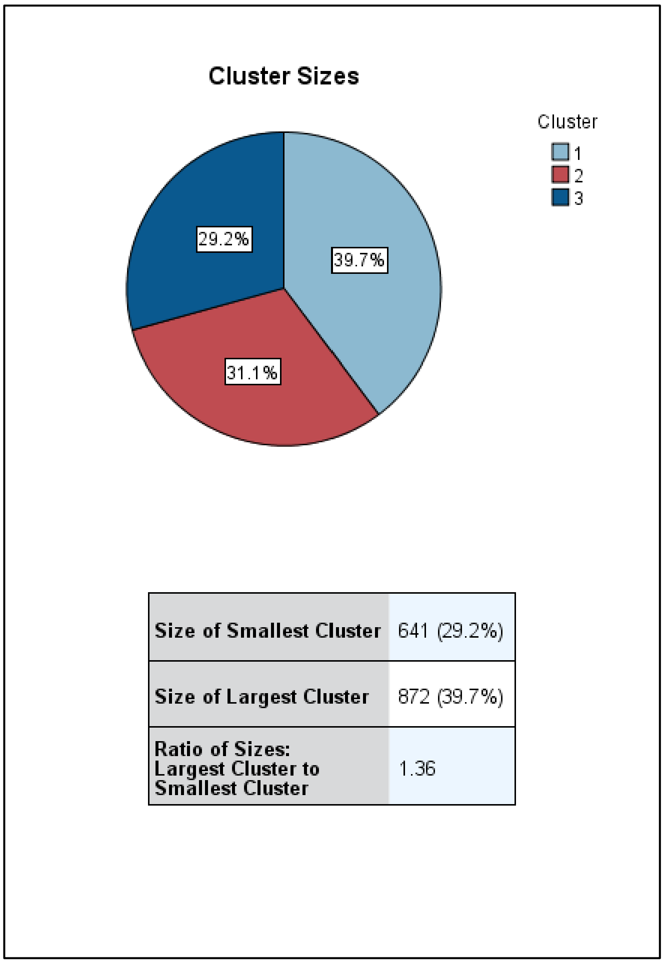


Figure D.1 TSC-1 Cluster sizes

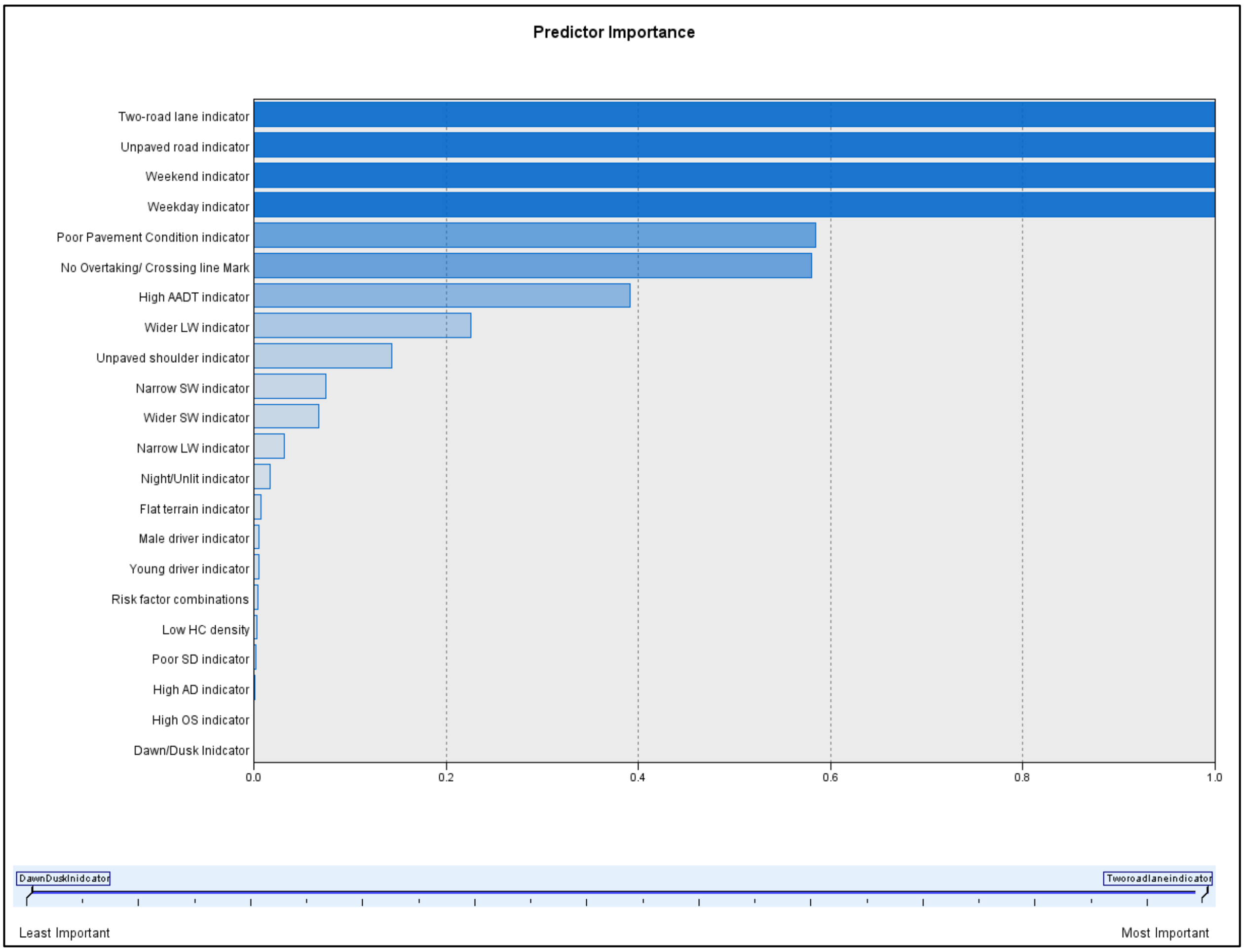


Figure D.2 Covariate importance in TSC-1 Model

Clusters

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	2	3
Label			
Description			
Size	39.7% (872)	31.1% (682)	29.2% (641)
Inputs	Two-road lane indicator	Two-road lane indicator	Two-road lane indicator
	Unpaved road indicator	Unpaved road indicator	Unpaved road indicator
	Weekday indicator 0 (97.1%)	Weekday indicator 1 (100.0%)	Weekday indicator 0 (54.3%)
	Weekend indicator 1 (97.1%)	Weekend indicator 0 (100.0%)	Weekend indicator 1 (54.3%)
	Poor Pavement Condition indicator	Poor Pavement Condition indicator	Poor Pavement Condition indicator
	No Overtaking/ Crossing line Mark	No Overtaking/ Crossing line Mark	No Overtaking/ Crossing line Mark
	High AADT indicator 1 (56.0%)	High AADT indicator 1 (54.4%)	High AADT indicator 0 (98.8%)
	Wider LW indicator 0 (92.5%)	Wider LW indicator 0 (92.4%)	Wider LW indicator 0 (61.2%)
	Unpaved shoulder indicator	Unpaved shoulder indicator	Unpaved shoulder indicator
	Narrow SW indicator 1 (89.7%)	Narrow SW indicator 1 (98.1%)	Narrow SW indicator 1 (100.0%)
	Wider SW indicator 0 (92.5%)	Wider SW indicator 0 (99.7%)	Wider SW indicator 0 (100.0%)
	Narrow LW indicator 0 (94.7%)	Narrow LW indicator 0 (93.4%)	Narrow LW indicator 0 (85.5%)
	Night/Unlit indicator 0 (62.5%)	Night/Unlit indicator 0 (67.0%)	Night/Unlit indicator 0 (74.3%)
	Flat terrain indicator 1 (81.4%)	Flat terrain indicator 1 (84.8%)	Flat terrain indicator 1 (77.8%)
	Male driver indicator 1 (82.8%)	Male driver indicator 1 (85.9%)	Male driver indicator 1 (87.8%)
	Young driver indicator	Young driver indicator	Young driver indicator
	Low HC density 1 (90.3%)	Low HC density 1 (89.4%)	Low HC density 1 (86.7%)
	Poor SD indicator 0 (75.9%)	Poor SD indicator 0 (77.0%)	Poor SD indicator 0 (79.3%)
	High AD indicator 0 (87.6%)	High AD indicator 0 (87.4%)	High AD indicator 0 (85.5%)
	High OS indicator 0 (98.9%)	High OS indicator 0 (98.5%)	High OS indicator 0 (98.9%)
	Dawn/Dusk Inidcator 0 (92.3%)	Dawn/Dusk Inidcator 0 (91.9%)	Dawn/Dusk Inidcator 0 (92.8%)

Figure D.3 Covariate effects in the cluster groups

a) TSC-1 Cluster 1

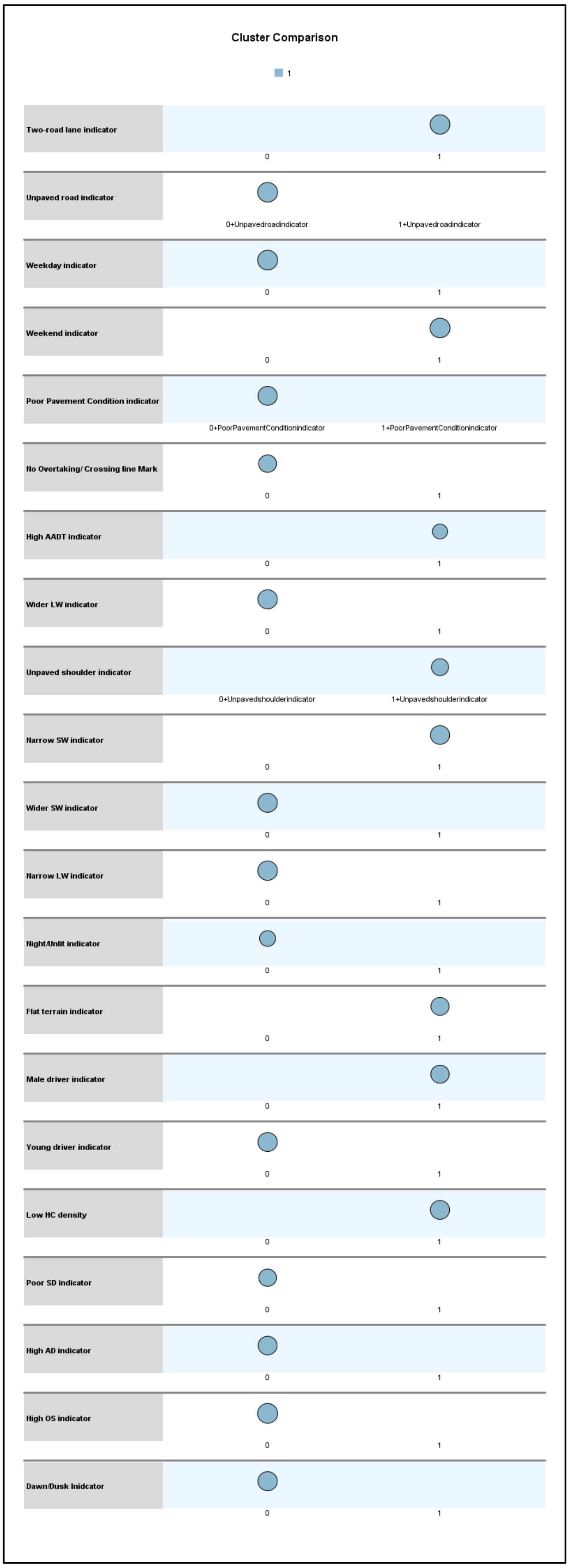


Figure D.4 Covariate distribution in TSC-1 cluster 1

b) TSC-1 Cluster 2

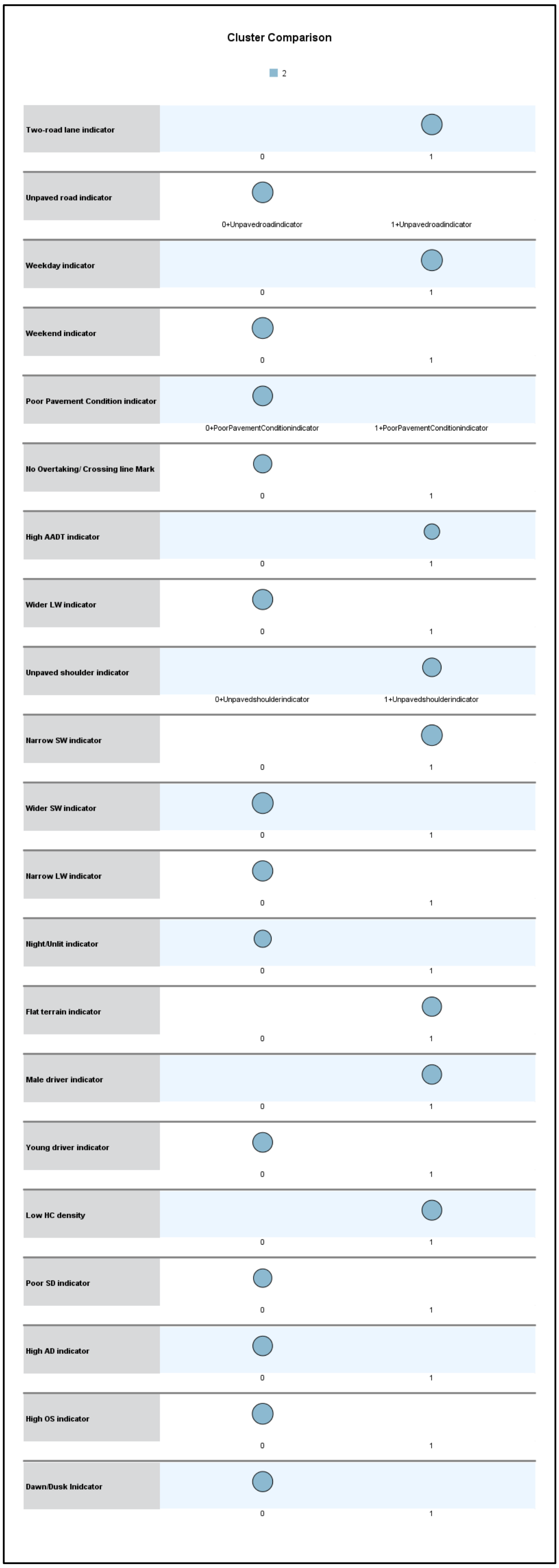


Figure D.5 Covariate distribution in TSC-1 cluster 2

c) TSC-1 Cluster 3

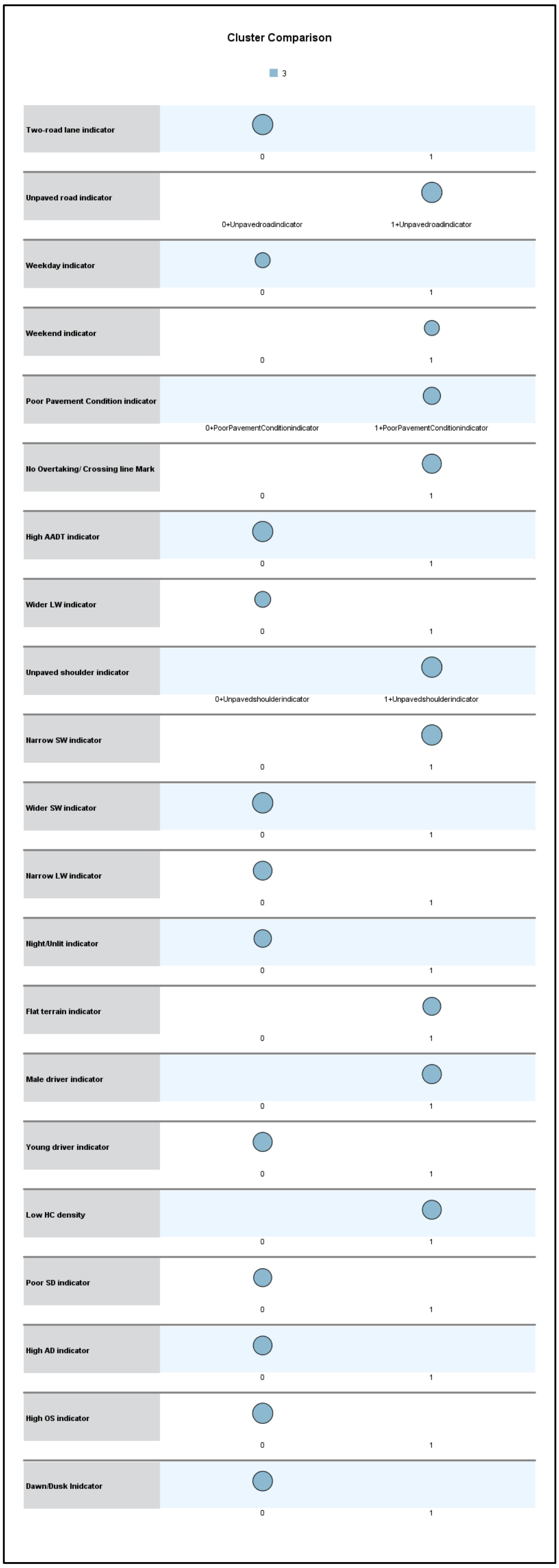


Figure D.6 Covariate distribution in TSC-1 cluster 3