Loughborough University

Certificate of Originality

Thesis Access Conditions and Deposit Agreement

Students should consult the guidance notes on the electronic thesis deposit and the access conditions in the University's Code of Practice on Research Degree Programmes

Author……Saleh Abdulaziz Altwaijri……..…………………………………………………………………..

Title…… Analysing Traffic Crashes in Riyadh City Using Statistical Models and Geographic Information Systems……..

I [Please insert name and address], "the Depositor",
would like to deposit [Please insert title of dissertation], hereafter referred to as the "Work", once it has successfully been examined in Loughborough University Institutional Repository

Status of access OPEN / RESTRICTED / CONFIDENTIAL

Moratorium Period…………………………………years,
ending…………../…………20……………………….

Status of access approved by (CAPITALS): Dr. MOHAMMED A QUDDUS
Supervisor (Signature)……………………………………………...…………………………………...

School of: Civil and Building Engineering

Author's Declaration *I confirm the following:*

CERTIFICATE OF ORIGINALITY
This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgements or in footnotes, and that neither the thesis nor the original work therein has been submitted to this or any other institution for a degree

user's declaration: for signature during any Moratorium period (Not Open work): ***I undertake to uphold the above conditions:***			
Date	Name (CAPITALS)	Signature	Address

Analysing Traffic Crashes in Riyadh City Using Statistical Models and Geographic Information Systems

By

Saleh Abdulaziz Altwaijri

A doctoral thesis submitted in partial fulfilment of the requirements for the award of the degree Doctor of Philosophy (PhD) from the Department of Civil and Building Engineering, at Loughborough University
[September 2012]

ABSTRACT

Road safety is a serious societal concern in Riyadh city, Kingdom of Saudi Arabia. Because of the negative impact of traffic crashes which cause losses in the form of deaths, injuries and property damage, in addition to the pain and social tragedy affecting families of the victims, it is important for transport policy makers to reduce their impact and increase safety standards by reducing the severity and frequency of crashes in the city of Riyadh. It is therefore important to fully understand the relationship between traffic crash severity and frequency and their contributing factors so to establish effective safety policies which can be implemented to enhance road safety in Riyadh city.

Data used in previous research have only consisted of basic information as there was unavailability of suitable and accurate data in Riyadh and there are very few studies that have undertaken as small area-wide crash analysis in Riyadh using appropriate statistical models. Therefore safety policies are not based on rigorous analyses to identify factors affecting both the severity and the frequency of traffic crashes. This research aims to explore the relationship between traffic crash severity and frequency and their contributing factors by using statistical models and a GIS approach. The analysis is based on the data obtained over a period of five years, namely AH 1425, 1426, 1427, 1428, and 1429 (roughly equivalent to 2004, 2005, 2006, 2007, and 2008). Injury crash severity data were classified into three categories: fatal, serious injury and slight injury.

A series of statistical models were employed to investigate the factors that affect both crash severity (i.e. ordered logit and mixed logit models) and area-wide crash frequency (i.e. classical Poisson and negative binomial models). Because of a severe underreporting problem on the slight injury crashes, binary and mixed binary logistic regression models were also estimated for two categories of severity: fatal and serious crashes. The mixed binary logit model and the negative binomial model are found to be the best models for crash severity and crash frequency analyses respectively. The model estimation results suggest that the statistically significant factors in crash severity are the age and nationality of the driver who is at fault, the time period from 16.00 to 19.59, excessive speed, road surface and lighting conditions, number of vehicles involved and number of casualties. Older drivers are associated with a higher probability of having a fatal crash, and, as expected, excessive speeds were consistently associated with fatal crashes in all models. In the area-level crash frequency models, population, percentage of illiterate people, income per capita and income per adult were found to be positively associated with the frequency of both fatal and serious injury crashes whereas all types of land use such as percentages of residential use, transport utilities, and educational use in all models were found to be negatively associated with the frequency of occurrence of crashes.

Results suggest that safety strategies aimed at reducing the severity and frequency of traffic crashes in Riyadh city should take into account the structure of the resident population and greater emphasis should be put on native residents and older age groups. Tougher enforcement should be introduced to tackle the issue of excessive speed.

This thesis contributes to knowledge in terms of examining and identifying a range of factors affecting traffic crash severity and frequency in Riyadh city.

KEY WORDS

Riyadh city, Crash severity, Crash frequency, ordered response models, nominal response models, GIS.

DEDICATIONS

First and foremost, I thank God Almighty for all His favours, mercies and guidance that propelled me successfully through this research. To almighty Allah, who has mercifully blessed me with everything which I have, and to whom all praises and all prays should be directed, and to his last prophet Muhammad (SAS).

My deep gratitude is reserved to my father and my mother for their love, endless support, dedication and prayers which has always been a source of inspiration. My special thanks to my wife and my children (Nader, Abdulaziz, Raghad and Faris) for their encouragement, understanding, and more importantly their patience throughout the entire long episode of this research. Not to forget thanking my brothers, sisters and all my family who didn't stop encouraging me throughout this study.

I hope that Allah accepts this work and reward me for my intention to improve my skills and enhance my abilities to add a great value and real contribution to the original body of knowledge.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors Professor Abigail Bristow and Dr Mohammed Quddus, for their guidance, constructive comments and critical reviews and exceptional support.

I would like to thank the High Commission for the Development of Riyadh (HCDR) for providing me with the data used in this study. A special thank is extended to Mr. Mohammed Al-Ahmari who was a member of staff at the (HCDR).

I am also grateful to the staff and friends at the Civil and Building Engineering Department, in particular Dr. Chao Wang, Dr. Nagendra R Velaga and Helen Newbold for their support.

Most importantly, I would like to express my deepest appreciation to my parents and family for their patience, continuous support, encouragement and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

Road traffic crashes are one of the most serious issues facing developed and developing countries. The World Health Organization has described traffic crashes as the "Epidemic of Civilised Societies" and anticipates that road crashes to be the third-leading cause of injury and disability globally by 2020. Road traffic crash deaths are projected to increase from 1.3 million in 2004 to 2.4 million in 2030 (WHO, 2009).

The costs of road traffic crashes to individuals, property, and society have become significant. For example, because of road crashes, more than 40,000 people die and over one million are injured every year in the European Union (CARE, 2008). According to IRTAD (2007), the UK is one of the safest countries in the world in which to drive, with an average of 6.4 people killed per 1 billion veh-km, which is considered to be low compared with other European countries, Japan and the US. According to the UK Department for Transport (DfT), there was a total of 203,950 road casualties in 2011, of whom 1,901 were killed and 23,122 were seriously injured (DfT, 2011). It has been also estimated that the economic costs of road traffic crashes for high income countries are about 2% of their Gross National Product (GNP) (IRTAD, 2007).

In the Kingdom of Saudi Arabia, rates of road crashes on the road network have increased dramatically. It has become a phenomenon affecting negatively all sectors of the society. It causes human suffering to individuals, and huge financial losses to the public as a result of deaths, injuries and disabilities. A conservative estimate of costs associated with the road traffic crashes is 14 billion Saudi Riyals (£2.2 billion) in 2008 and expected to be 24 billion Saudi Riyals (£3.7 billion) in 2018 (Al-Ahmadi, 2011).

The Kingdom of Saudi Arabia (KSA) is the largest country on the Arabian Peninsula and is geographically divided into 13 regions (see Figure 1.1).

KSA has some characteristics of a developed country and some of a developing nation; it falls into a middle ground. However, it has the road safety record of a developing nation. In relation to Human Development Index (HDI) developing countries are in general countries that have not achieved a significant degree of industrialisation relative to their populations. HDI which combines an economic measure, national income, with other measures, indices for life expectancy (ILE), standard living index (SLI), and education index (EI) has become prominent. Furthermore three other indices are added as supplementary indices which are: the human poverty index (HPI), gender-related development index (GDI) and gender empowerment measure (GEM).

The GEM leads us to the effect of male driving dominance in KSA especially as it may increase the driving aggressiveness and reduce the road safety level, if women were allowed to drive this might be more balanced as will be discussed in chapter 2 where evidence suggests that women are less associated with severity and frequency of road crashes (which means women have fewer and less severe crashes). Based on United Nations data the HDI for KSA is low as is the Road Safety Development Index (RSDI), which ranks road safety progress in different countries/ regions across the globe. KSA may therefore be viewed as a developing country even though its income and the economic growth are high, as the country is not sufficiently industrialised, has a poor road safety record and a low HDI.

Like other developing countries, the KSA is suffering from both human losses and economic losses as a result of high numbers of traffic crashes (Al-Ghamdi, 1993). In Saudi Arabia, twenty years ago, one in five victims of road traffic crashes was a child, and unrestrained child passengers were at great risk of serious injuries or death (Shawan et al., 1993). Such risks were feared to increase as the KSA has experienced a tremendous growth in the transport sector, including changes in socio-economic activities, lifestyle, and mobility (Al-Zahrani and Jadaan, 1995).

Figure 1-1Regions of KSA (Source: http://www.saudi-us-relations.org/maps/saudi-arabia-province-map.html)

The Kingdom of Saudi Arabia (KSA), the largest country of the Arabian Peninsula, has a serious road safety problem that results in both human and economic losses. A study in 1999 (Jacobs et al, 2000) found that although countries in the Middle East and North Africa (i.e. MENA countries) had only 2% of the world's motor vehicles and 4% of the world's population, the region experienced 6% of global road fatalities in 1996 (Jacobs et al, 2000). In KSA in 1994 there were 21 deaths per 100,000 people (Jacobs et al, 2000) and by 2007 as shown in Figure 1.2, this had risen to 29 deaths per 100,000 people (WHO, 2009).

It is necessary to clarify the term 'fatal road crash': in the Kingdom of Saudi Arabia it is defined as a road crash where death has occurred at the scene or within 24 hours of the crash. However, most countries (certainly those of Western Europe and North America) define a fatal road crash as one where death occurs within 30 days of the event.

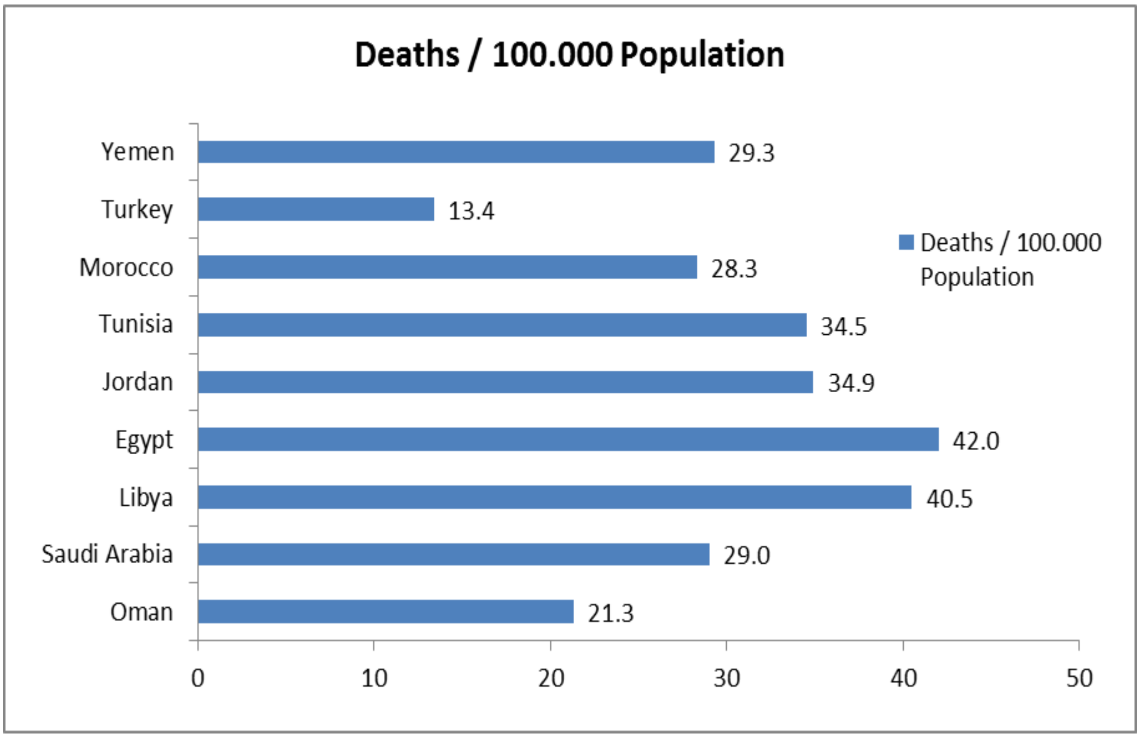

Figure 1-2 Middle East and North Africa fatality rates per 100,000 population

"World Health Organization, 2009": http://whqlibdoc.who.int/publications/924156220X.pdf.
http://en.wikipedia.org/wiki/List_of_countries_by_traffic-related_death_rate

Riyadh, the capital of the Kingdom of Saudi Arabia, has also been suffering human and economic losses as a result of a high rate of road traffic crashes. This is due to the fact that Riyadh has experienced tremendous growth in the transport sector and changes in socio-economic activities, lifestyle and mobility since 1990 (Al-Ghamdi, 1993; 1996a). Riyadh, the second most heavily populated region in the KSA, is one of the fastest growing cities in the Middle East, as its population has increased from 150,000 in the 1960s to over 4.5 million in 2005, and is expected to reach 10.5 million by 2020, according to the Saudi General Directorate of Statistics (SDGS, 2005) (see Figure 1.3).

66% of residents are Saudi and 34% non-Saudi. Previous studies of the traffic safety problem have highlighted that traffic safety is a serious issue for Riyadh and there is an urgent need to develop safety policies aimed at reducing both traffic crashes and their severity (Lee, 1986; Al-Ghamdi, 1996a, 1996b and 1999). They have all agreed that this problem is serious and needs a comprehensive solution.

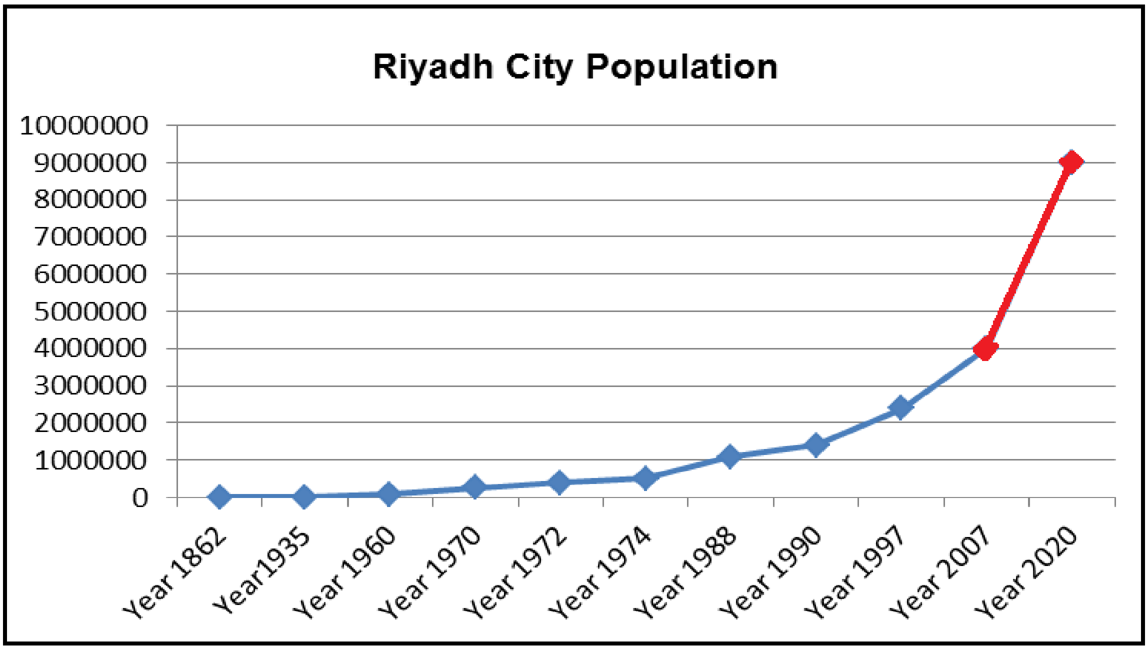

Figure 1-3 Population in Riyadh city from 1862 to 2020 Source: Saudi General Directorate of Statistics (SGDS, 2005).

The major problems behind the high number of crashes in Riyadh are identified as:

- High car ownership: there were 1.2 million registered vehicles in 2005 resulting in an average car ownership of approximately 1.72 vehicles per household for Riyadh city (SMOT, 2007).
- Migration of people to Riyadh city for 'work', 'study' or 'business': 34% of the city's population are non-Saudi (1.6 million in 2005) (Al-Gabbani, 2009).
- Daily trips made by vehicles inside the city reached about 6 million compared to about 1 million twenty years ago; the majority of those trips were made by private cars (about 85%), followed by private buses (8%), taxis (5%) and public transport (2%). This increase in trips is primarily due to increases in population, vehicle ownership, and income. Therefore, good transport planning and implementation are essential to increase safety standards and reduce the severity of crashes in the city of Riyadh (SMOT, 2007).
- High rates of income growth
- Low cost of petrol
- Drivers from different nationalities from 190 different countries (primarily from India, Egypt, Pakistan, Yemen and Bangladesh). They are primarily unskilled or semi-skilled with low incomes (average monthly salary 1,800 Saudi Riyals).

It is thought that the high proportion of the foreign population may have profound effects on Saudi society (Al-Gabbani, 2009).

- High proportion of young drivers and many other reasons resulting in more fatalities, severe injuries, disabilities, and property damage.
- T Existing regulations covering irresponsible driving, speeding, crossing red lights and use of mobile phones while driving which are some of the main reasons for road crashes are not adequately enforced. Moreover there may be a lack of discipline as people are not respecting the rules and the police are not firmly applying the law. As an example of the gap between regulations and application by population a survey on 80% out of 297 students at Abha University in KSA where 70% and 72% had cars and driving licenses respectively. More than 50% of the whole students had been involved in road traffic crashes, 22% of these had been injured in these crashes and 13% admitted to hospital for an average of nine days. High speed was found to be the main cause of their crashes. More than 75% of the participants mentioned that they had problems with the use of seat belts, the most common of which were forgetfulness and anxiety.

In terms of transport, the city of Riyadh has a modern hierarchical roadway system with a total road length of 11,372 miles (i.e. 11% of the KSA’s total road length). It has a modern highway system. The main Eastern Ring Road connects the city's south and north regions, while the Northern Ring Road connects the east and the west of the city. King Fahd Road runs through the centre of the city from north to south and is 30 km long. Mecca Road runs from east to west across the city centre and is 40 km long.

In 2005, a total of 47,341 traffic crashes occurred on the Riyadh roads, which represents 19% per cent of all KSA traffic crashes. According to the Riyadh General Department of Traffic (RGDT, 2007), although the number of traffic crashes declines over time in many developed countries, this is not true for the case of the KSA, especially in Riyadh city where 52% of the road fatalities were attributed to young people aged 16 to 30.

1.2 Research gap

As discussed earlier, road crashes in Riyadh city are a burden on society; it is therefore important to reduce the impacts of road traffic crashes in the city. Because of the negative impact of traffic crashes, which cause losses in the form of deaths, injuries and property damage, in addition to the pain and social tragedy affecting families of the victims, it is important to carry out a careful investigation to understand the relationship between traffic crash severity and frequency and their contributing factors, in order to establish effective safety policies which can be implemented to reduce the severity and frequency of road crashes in Riyadh city.

Dimitriou (1990) identified two sets of transport problems in Third World countries. The first group is problems such as a drastic increase in car ownership, poor management of traffic, non-application of traffic laws, inadequate transport facilities, high growth in population, and regulation of land use. The second group is the occasional problems such as traffic congestion and a high rate of road crashes. Therefore, the need for an effective and safe transport system becomes a big challenge.

Jacobs et al. (2000) identified a range of factors (e.g. speeding, high rate of car ownership, traffic congestion and land-use patterns) that are associated with traffic crashes in Riyadh. Their study also highlighted that there is a lack of appropriate safety planning and policies aimed at improving road traffic safety. There are very few studies that analyse road traffic crashes in Riyadh city. This is primarily due to the lack of reliable data on both road traffic crashes and their contributory factors.

The existing studies in road crashes in the Kingdom of Saudi Arabia (KSA) in general and in Riyadh city in particular have focused only on the following:

- driver characteristics (such as age and singularity of gender) (Koushki, 1988; Shanks et al., 1994)
- examination of seat-belt law and helmet usage since the enactment of the law in 5 December 2000 (Ansari et al., 2000; Bendak, 2005)
- weather effects (Nofal and Saeed, 1997)

- Nationality impacts on road crashes due to the high proportion of people from different nationalities in Riyadh city who have come to work and for business (Al-Ghamdi, 2000; Al-Ghamdi, 2002).

Because the data used in previous research consist of basic information that is relatively easy to obtain (i.e. age, gender, nationality and weather), other factors affecting road traffic crashes have not been explored because there is a lack of suitable data in the KSA. The unavailability of suitable data for crash analysis is a common problem, not only in the KSA but also in most developing countries (Jacobs and Sayer, 1983; Mekky, 1985; Al-Saif, 1997). Researchers who attempted to estimate the factors influencing traffic crashes in Riyadh city used inaccurate, imprecise, and unreliable data (Al-Ghamdi, 2002; Al-Ghamdi, 2003). In contrast, data to be used in this research are recent, and only started to be collected from 2004 onwards, especially crash data, which were recorded by police officers at the time of the crash using advanced instruments; the officers reported all the crash details needed for the analysis, which was carried out using the most appropriate statistical prediction models.

Only a few studies have undertaken an area-wide crash analysis in the KSA using advanced statistical models to analyse crash data and to estimate the effect of statistically significant factors such as population, land use, and car ownership at an aggregate level.

In other words, very few attempts used appropriate crash prediction models to investigate the severity and counts of crashes including their contributing factors. The existing studies shown in Chapter 2, section 2.3.2.3 do not provide solid evidence on that issue. Therefore, the research gap which has been identified is that this research will gather suitable data from different sources and will be using an appropriate statistical model to analyse area-wide crash data and develop a relationship between crashes and their contributing factors at area level, then use advanced statistical models to provide more robust empirical evidence, develop safety policies to reduce crashes, and identify the most hazardous locations by mapping the crashes in Riyadh city, KSA. In addition, GIS integration will be used to integrate data from different sources and to map the locations of crashes on the Riyadh city road network.

The research gaps can be summarised as follows:

- a lack of suitable and accurate data
- very few studies on the identification of factors affecting both severity and frequency of traffic crashes in Riyadh city
- very few attempts to develop and apply crash prediction models to provide solid empirical evidence using advanced econometric models
- a need for analysis of area-wide crash data
- a lack of safety policies based on rigorous analyses.

In conclusion, there is a lack of skilled and experienced manpower and credible research in most developing countries. Applying the results obtained from developed countries to developing countries would not be practical because of the differences in the nature of the problem, society, culture and human behaviour. Furthermore, not enough studies have been undertaken in the KSA carrying out an area-wide analysis using advanced statistical models to analyse crash data and estimate the effect of the statistically significant factors at an aggregate level. Due to the unavailability of suitable data for crash analysis in the KSA and most developing countries (Jacobs and Sayer, 1983; Mekky, 1985; Al-Saif, 1997), and the use of inaccurate, imprecise, and unreliable data for studies performed on traffic crashes in Riyadh city (Al-Ghamdi, 2002; Al-Ghamdi, 2003). While, this study uses detailed and recent crash data recorded by police officers at the time of the crash using advanced instruments and most appropriate statistical prediction models. Therefore, this thesis attempts to fill the above gap in the literature by analysing a reliable crash dataset maintained by the HCDR since 2004. This study also obtains data from other sources which will be integrated using GIS so as to conduct an area-wide analysis of traffic crashes in Riyadh city. This suggests that there is a clear gap in the literature on traffic crash analysis in the context of developing countries, especially in the KSA.

1.3 Research aim and objectives

With regard to the research problem described above, the aim of this thesis is to **explore factors affecting the severity and frequency of traffic crashes in Riyadh city using appropriate econometric models with the aid of GIS tools.**

To achieve the above aim, the following objectives are formulated:

- to understand theories of road safety
- to conduct an in-depth literature review on the factors affecting traffic crashes
- to integrate data from different sources using GIS
- to develop models for the identification of contributory factors affecting traffic crashes
- to identify, explore and interpret factors affecting the severity and frequency of road injury crashes in Riyadh city
- to develop safety policies for reducing traffic crashes in the city of Riyadh.

This thesis will provide empirical evidence on the factors affecting road traffic crashes by using appropriate statistical models with the aid of Geographic Information System (GIS) tools. This thesis also seeks to investigate the factors affecting the severity and frequency of road crashes in Riyadh city using statistical models which will assist transport policy makers in transport and safety planning, to enable them to develop different countermeasures to improve road safety in Riyadh city.

1.4 Outline of the thesis

This thesis is organised into nine chapters. This chapter has described the research background, research gap, aim and objectives, and the structure of the thesis; this section provides an overview of the following chapters.

Chapter 2 provides a literature review of various theories in road safety, and discusses various factors influencing road traffic crashes in developed and developing countries. The factors considered include traffic, driver, road and land-use characteristics, and socio-economic and environmental factors.

Chapter 3 reviews a range of econometric methods used in crash modelling. It includes econometric models used in modelling crash severity, and models of crash frequency, together with non-spatial models based on road segments and intersections.

Chapter 4 introduces the methodology used in this thesis, including details of the econometric models for both crash severity and crash frequency analysis. For crash severity analysis, ordered and nominal response models are employed. For crash frequency analysis, classical count outcome models are employed.

Chapter 5 explores the data used in this thesis. It starts by presenting descriptions of data which include crash, population, land-use, and road network data. This is followed by data validation and finally a description of variables used for both crash severity and crash frequency analysis.

Chapter 6 presents the results from the crash severity models, including the ordered response models and nominal response models which are detailed in Chapter 4, using the data described in Chapter 5 to explore the various factors affecting crash severity.

Chapter 7 presents the results from the crash frequency models, including the classical count outcome models which are detailed in Chapter 4, using the data described in Chapter 5 to explore the various factors affecting crash frequency.

Chapter 8 presents a further discussion on the factors affecting both crash severity and crash frequency, using the results from Chapters 6 and 7. Based on the findings, several potential policy implementations are proposed aiming to improve road safety in Riyadh city.

Chapter 9 summarises the research contribution in the area of road traffic safety in Riyadh city. It concludes this thesis with research limitations and directions for future improvement which can remedy the limitations of this work.

2 LITERATURE REVIEW ON FACTORS AFFECTING ROAD CRASHES

2.1 Introduction

The purpose of this thesis is to investigate factors affecting road traffic crashes in Riyadh city; it is therefore important to understand the theories of road safety and the relationship between traffic crashes and their contributing factors in order to establish effective safety policies which can be implemented to reduce road crashes in Riyadh city.

The objective of this chapter is to provide a review of current literature relating to the various factors affecting road crashes. The review starts by looking into theories of road safety, which include risk compensation theory, engineering theory, economic theory, public health theory and physiological theory. The second section of this review involves identifying and evaluating factors affecting road safety in the literature, which consist of the following elements: factors influencing road traffic crashes in developed countries and factors influencing road traffic crashes in developing countries

2.2 Review of various theories on road safety

Many researchers have investigated road safety and its contributing factors from different aspects, such as economics, engineering and public health, including the design of vehicles and roads, care and treatment of victims in crashes, and enforcement of traffic safety legislation. In order to reduce crashes we need to understand road safety theories. These theories are: risk compensation, engineering, economic, public health, and physiological theories.

Over the last few decades, a range of road safety theories have been proposed by researchers to explain the number, trend and causes of road traffic crashes. The most common theories of road safety are associated with risk perception and compensation, vehicle and road design (i.e. engineering), GDP (i.e. economics), medical care and technology improvement (i.e. public health), and driver fatigue and drowsiness. A review of such theories is essential to an understanding of the complex nature of traffic crash occurrence. This section briefly describes theories of road safety.

2.2.1 Engineering Theory

The engineering theory of road safety is associated with the design of road infrastructure and vehicles. More specifically, it is speculated that ‘safer roads’ and ‘safer vehicles’ may reduce road traffic crashes.

Traffic safety specialists can influence traffic safety either by introducing road safety policies, law enforcement, and education or by improving roadway geometry, road signs and road markings, such as lane and road width, gradient, curvature (horizontal and vertical) and road layout. Many studies have indicated that improvements to road design could produce significant reductions in the number of crashes.

Milton and Mannering (1998) confirmed that road infrastructure designs do affect road safety when they looked at the annual crash frequency on sections of principal arterials in Washington State, they found that short sections are less likely to experience crashes than longer sections; narrow lanes (less than 3.5 m) and sharp horizontal curves tend to decrease crash frequency in Eastern Washington.

A further study undertaken by Jørgensen et al. (1999) studied the optimal use of roadway warning signs along the roads. Their analysis showed that the optimal number of signs was dependent on the road authorities’ objectives for road traffic and on how drivers form their risk perceptions. Simulations indicated that the safety and economic benefits of warning sign installations were not very high. When considering the whole road system, warning signs seem to have a greater positive impact on total driving costs than on crash costs.

Navin et al. (2000) studied road safety engineering in British Columbia, Canada and found that rear end collisions can be effectively reduced by using recognised engineering crash countermeasures such as enhanced signal visibility or through complex intersection geometric upgrades. They suggested that improvements in road safety engineering and better vehicle design can reduce the severity of whiplash injuries when crashes occur.

Wouters and Bos (2000) investigated whether drivers' crash rates would be reduced if drivers were monitored using in-car data-recording devices. Seven experimental vehicle fleets were involved; these varied in terms of which type of transport sector they are from, the type of vehicles used, and the traffic circumstances in which the vehicles were operated. Crash and exposure data were collected for 840 vehicles of which 270 were equipped with a recorder. During an observation period representing a total of about 3,100 vehicle years, these vehicles were involved in 1,836 road crashes. Analysis of the effects due to the use of data recorders in these fleets resulted in an average estimated crash reduction of 20%. Statistically significant crash reductions were found for some of the fleets, while the estimated risk ratios showed relatively wide confidence intervals. The lack of precision in this study is partly due to the small sample sizes available in some fleets, the use in other than the commercial fleet and the type of recorder used.

Carson and Mannering (2001) studied the effectiveness of ice warning signs in reducing both crash frequency and crash severity using three years of ice-crash data from Washington State using appropriate statistical models. They found that the presence of ice warning signs was not a significant factor in reducing ice-crash frequency or ice-crash severity.

Elvik and Greibe (2005) presented the road safety effect of porous asphalt in a systematic review evaluating the effects of porous road surfaces on crashes. They showed that the effects on road safety of porous asphalt were to a large extent unknown. Porous asphalt affects some risk factors associated with crash occurrence favourably, but road users adapt their behaviour to these changes, in particular by driving faster.

Ratrout (2005) studied tyre condition and drivers' practice in maintaining tyres in Saudi Arabia. A random sample of vehicles was stopped to check the condition of their tyres and at the same time to interview their drivers on tyre-related issues because tyre blowouts and tread separation are a very hot safety issue in the Kingdom of Saudi Arabia. The percentage of under-inflated tyres found in the study area was large. The results indicated that drivers need proper education on how to select, use, and maintain tyres. This will reduce the percentage of under-inflated tyres and minimise incorrect practices such as using tubes in tubeless tyres following punctures, and thus reduce traffic crashes resulting from tyre failure.

Chang (2006) investigated the association between major injuries and seat locations in a motor coach rollover crash in Taiwan. The results indicated that the majority of injuries occurred to passengers in the seats on the upper side who were thrown away from their seats and who crushed the neighbouring passengers in the lower seats.

Newstead et al. (2008) showed the relationship between vehicle crashworthiness in both the year of manufacture and the year of first registration from 1964 to 2006. Crashworthiness was measured by the number of injured drivers and drivers involved in the crash. The analysis illustrated a significant improvement in the crashworthiness of New Zealand light passenger vehicles over the years of manufacture; the risk of death or serious injury to drivers was reduced by 58% because there was an increase in the use of improved vehicles and safety regulations by the government.

Chen et al (2012) investigated the safety countermeasures and crash reduction in New York City, the effectiveness of 13 safety countermeasures and street designs installed in New York City between 1990 and 2008 were evaluated. They showed that signal related countermeasures that were designed to reduce conflicts including: split phase timing, signal installations, all pedestrian phase, and increasing pedestrian crossing time, have reduced crashes.

Many studies have indicated that improvements at the level of engineering design could produce significant reductions in the number of crashes and fatalities. These could include: training and education campaigns, vehicle design, road design and medical technology.

2.2.2 Economic Theory

The economic theory of road safety simply explains how the gross domestic product (GDP) of a country or a region may affect the level of traffic crashes.

The costs of road traffic crashes have considerable effects on people and property. Estimating the costs of road crashes and injuries helps to explain the significance of the problem and the benefits of preventing it.

It has been estimated that the economic cost of road crashes and injuries is 1% of the Gross National Product (GNP) in low income countries, 1.5% in middle income countries and 2% in high income countries (*World Report On Road Traffic Injury Prevention, 2004*). Between 1970 and 2004, there has been a 55% reduction in road fatalities per capita in countries of the Organization for Economic Cooperation and Development (OECD).

Research undertaken by Kopits et al. (2005) examined the relationship between traffic fatality risk and per capita income.

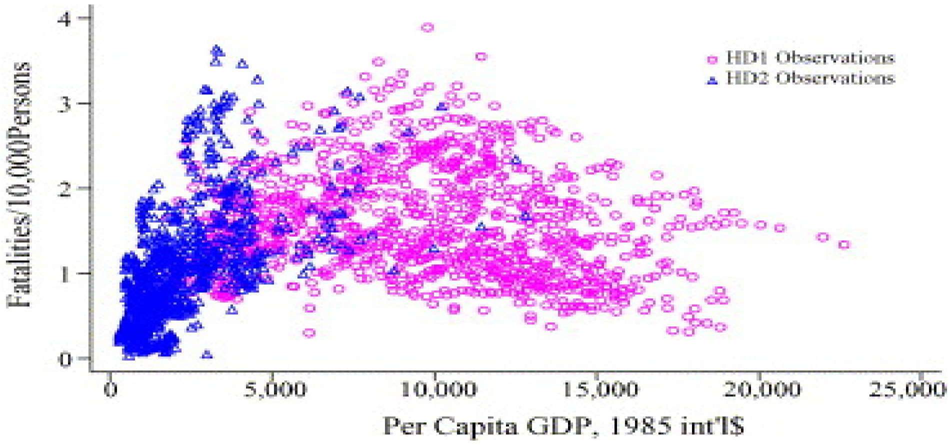

Figure 2-1 Traffic fatality risk vs income: all countries and years (Kopits et al. 2005)

It is noticeable from the results shown in Figure 2.1 (where Highly Developed countries (HD1) and Developing countries (HD2)) that as GDP increases the number of fatalities (per 10,000 persons) also increases, up to a certain level ($10,000 per capita), then the fatality rate tends to decline. They forecast traffic fatalities by geographic regions using panel data from 1963 to 1999 for 88 countries. Their results suggest that motorisation is strongly correlated with income as the rate of motorisation is higher in the developed countries with a Human Development Index (HDI) greater than 0.8 and with a comparatively high level of per capita GDP, whereas the road death rate appears to decline rapidly with income and the sharp decline in Fatalities/Vehicles (F/V). The income factor reflects that when income rises the number of travellers in vehicles will be higher than the number of pedestrians, which means that there is a lower probability of dying in the event of a crash. In addition, it might be that, if one uses safer cars, such as four wheel drive cars instead of two, Fatalities/Vehicles (F/V) will be reduced.

Kopits et al. concluded that it will take many years for developing countries to achieve the motor vehicle fatality risks of high income countries. The predictions and estimations of the income level at which traffic fatality risk begins to decline may be based on the policies.

Paulozzi et al. (2007) studied the economic development's effect on road transport-related mortality among different types of road users in 44 countries using death certificate data provided by the World Health Organization. Motor vehicle crash (MVC) mortality is expressed as deaths per 100,000 people and per 1000 motor vehicles. They found that overall MVC mortality peaked among low-income countries at about US$ 2000 gross national income (GNI) per capita and at about 100 motor vehicles per 1000 people. Overall mortality declined at higher national incomes up to about US$ 24,000. Most changes in MVC mortality associated with economic development were explained by changes in rates among non-motorised travellers, especially pedestrians.

Traynor (2008) estimated the cross-county correlation between per capita income and fatalities per vehicle mile travelled (VMT) in Ohio indicating that a significant interaction effect exists between per capita income and the percentage of highway VMT, indicating a nonlinear correlation between per capita income and fatality rates. Counties with very low shares of VMT on highways have a significant inverse correlation between per capita income and traffic fatality rates and counties with very high highway shares have a significant direct correlation between per capita income and traffic fatality rates, for most counties the correlation is not statistically significant.

Law et al. (2009) looked at the factors associated with the relationship between motorcycle deaths and economic growth by examining the Kuznets curve relationship for motorcycle deaths using data from the OECD countries for 25 countries covering the period 1970–1999. The results indicated that motorcycle deaths follow an inverted U-shape, or Kuznets curve, relationship with per capita income. Later on, Law et al. (2011) showed an empirical analysis of the Kuznets curve relationship between per capita income and road fatalities across 60 countries over the period 1972–2004. This relationship hypothesizes that the number of road fatalities increases with increasing motorization in the early stages of economic growth. Results indicated evidence of a

Kuznets curve relationship between per capita income and road fatalities for both highly developed and less developed countries.

Between 1970 and 2004, there has been a 55% reduction in road fatalities per capita in countries of the Organization for Economic Cooperation and Development (OECD). It has been found that crashes decreased in the countries with higher GDP.
Dickerson et al. (2000) investigated the external costs of crashes using road types and geographical areas and found that high traffic flows were related to strong negative crash externality.

Transport imposes direct costs on the national economy such as the building of the infrastructure of roads with its services, and it also imposes indirect costs such as noise, air pollution, traffic congestion, and traffic crashes which result in deaths and injuries.
Many economists have been involved in road safety research because crashes are considered as an external cost of transport.

2.2.3 Public Health Theory

Road traffic crashes are considered to be major public health problem worldwide. Public health theory of road safety implies how the improvement in emergency response, medical care and medical technology reduces traffic fatalities.
A few studies have focused on the role of improvements in medical care and technology in reducing traffic fatalities.

Antonio et al. (2001) and Avery et al. (2000) looked at the effect of the trauma care system on fatalities due to traffic crashes, and found that implementation of an organized trauma care system will considerably reduce traffic fatalities.

Research by Noland (2003a), Noland (2003b) and Noland and Quddus (2002) studied the impact of improvements in medical care and technology on traffic fatalities. Using various proxies for medical care such as the number of licensed physicians per capita, the average length of inpatient hospital stay, indicators of the number of people waiting for hospital treatment, and the number of average care days in hospital, these studies suggest that improvements in medical services have significantly reduced traffic fatalities in the U.S., Great Britain, and other industrialised countries. Results also suggest that medical technology improvements are more important than changes in medical care.

Many low income countries have weak emergency services available at the time of a crash, and a lack of qualified medical specialists, which causes a long delay between arrival at the hospital and the start of treatment. In most of the cases, injured people are evacuated and transported to hospitals by their relatives using their private vehicles.

For example, in Kenya, Nantulya et al. (2002) found that the police evacuate only 5.5% of crash survivors and that 2.9% of ambulance staff or police who arrive at the place of the crash before the emergency medical team are not trained or properly equipped to provide primary help to the injured. In addition, in low and middle income countries, people do not have a public health system or private health insurance (Hijar et al., 2003), and this leads to a high rate of deaths caused by traffic crashes due to a low level of medical services.

In 2006 the Cabinet Office Financial Year (COFY) of Japan investigated the economic losses of road traffic crashes and analysed them in terms of financial losses and non-financial losses in a one-year period. The total losses were 6,745 billion yen (53.31 billion GBP) which is about 1.4% of the gross domestic product (GDP). Accordingly the Japanese government improved the rescue and emergency medical system by a programme called the "doctor-helicopter programme" that provided better and faster air transport for medical patients.

Law et al. (2007) used three medical proxy variables (physicians per capita, average inpatient days in hospital, and infant mortality rate) to develop a relationship between motorcycle fatalities and the improvements in medical technology and care. Motorcycle crashes are more severe and therefore medical care or technology has little impact. However, such studies need wider data to be used to cover both developed and developing countries and also needs to have detailed medical information about the type of injuries.

2.2.4 Physiological Theory

This theory relates to how various physiological factors such as fatigue, drowsiness, and sleepiness affect driving performance which may lead to safety occurrence. People gradually perform more poorly in doing a specific task for a long period of time. One example is that of driving a vehicle for a long time without any breaks, especially at night. Many terms are used to indicate the cause of this poor performance; examples include drowsiness, sleepiness, fatigue and inattentiveness. Some have argued that drowsiness is a result of fatigue. Others have noticed that monotony also causes drowsiness. This theory relates poor driving performance to drowsiness and fatigue with the possibility of traffic crash occurrence.

Sagberg et al. (1997) suggested that the causes of the higher occurrence of self-reported drowsy driving in the US may be due to differences in road geometry, design and environment as well as exposure. He argued that the risk of falling asleep is higher on straight, boring roads in situations of low traffic. Fatigue during driving is a serious problem in transport systems and is believed to be a direct or contributing cause of road related crashes. Gander (1995) and Jiang et al. (2003) found that persons with higher levels of anxiety also reported higher levels of fatigue.

Horne and Reyner (1995) studied sleep related vehicle crashes in Southwest England and the midlands using police databases or on the spot interviews for drivers involved in 679 sleep related vehicle crashes. They found that of all vehicle crashes, sleep related vehicle crashes comprised 16% on major roads in Southwest of England, and over 20% on midlands motorways. They also found that during the 24 hour period there were three major peaks, at around 02:00, 06:00, and 16:00, where about half of these drivers were men under 30 years; and only few such crashes involved women.

Garder (1995) proposed that drowsy driving can be reduced by design efforts in highway engineering. This may be done by building roads with shorter tangents and regular alignments in short distances. Brown (1997) also suggested that tiredness is a contributing factor in 20% of crashes on roads. Verwey and Zaidel (2000) found that fatigue measured by deviations on the road during simulated driving was associated strongly with high scores on the extraversion-boredom. They also observed a poor association between physiological and more individual measures of fatigue.

In summary, it has been shown that various types of road safety theories have been introduced to explain the occurrence of traffic crashes. Each of these theories has rightly identified a number of factors that affect traffic crashes. For instance, risk compensation theory suggests that driver behaviour (e.g. attitudes towards in-vehicle safety devices) and socio-demographic factors (e.g. age, gender, nationality) could affect traffic crashes. The engineering theory of road safety implies that the geometry of roads and their infrastructure, the number and width of lanes, shoulder width, road curvature, and safety signs and road marking could affect traffic crashes. The economic condition of a nation is also associated with traffic crashes. Medical technology improvements such as better crash and emergency responses and services could reduce traffic-related fatalities.

2.2.5 Risk compensation theory

Risk compensation was first developed by Adams in 1981 who suggested that there was no correlation between the passing of seat belt legislation and the reductions in injuries or fatalities (Adams, 1981). Risk compensation is the name given to a theory which tries to understand the behaviour of people in potentially hazardous activities, such that an individual will accept a given level of risk in a certain activities. Another way of stating this is when individuals behave less cautiously in situations where they feel "safer" or more protected. This theory affects humans and is associated with the use of safety features such as car seat belts, bicycle helmets, safety equipment for children, and anti-lock systems. The theory has been built largely on road safety investigations (Wild, 1998, Assum et al, 1999 and Hedlund, 2000, Adams and Hillman, 2001).

There is evidence that such an effect is seen in humans, associated with the use of safety features such as seat belts, air bags, and bicycle helmets. The evidence is particularly convincing for the case of anti-lock braking systems. One of the best-known examples of risk compensation is when people who buy safe, modern cars with electronic stability control (ECS) and anti-lock brakes have a tendency to counter a greater deal of increased safety by driving faster and taking greater risks (Grant and Smiley, 1993 and Sagberg et al., 1997).

Peltzman (1975) showed that reductions in occupant fatalities were compensated by increases in pedestrian fatalities because drivers protected by the safety standards drove more riskily. According to Adams (1988), risk compensation theory (also called risk homoeostasis theory) suggests that safety measures will not reduce crash loss unless they lower 'target levels of risk'. The theory is likely has no index of crash loss, and the risk compensation effects are frequently underestimated**.**

Underwood et al (1993) explained the motives of road users' risk nomination towards the measures, and indicated when behavioural adaptation is likely to take place and its effects on road safety programmes. It also suggested that risk compensation, engineering safety measures alone are usually not sufficient, and, hopefully, carefully designed motivational safety measures lead to modify road users' behaviour toward safer traffic

Another explanation of this theory refers to the situation where drivers of vehicles equipped with (ABS) normally drive fast, closely following the car at the front and brake late when required (Grant and Smiley, 1993; Sagberg et al., 1997; Aschenbrenner and Bieh, 1994).

Wilde (1998) pointed out that the factors influencing the level of accepted risk are cultural, social, or psychological. The amount of risk that people can take depends on the accepted benefits of risky behaviour, e.g. increasing speed to save time, or the expected costs of safe behaviour such as the use of uncomfortable seat belts. He explained the mechanism of this theory by stating that a change in the degree of warning displayed in behaviour brings about a change in the injury rate, while changing the injury rate also leads to a change in behaviour. Drivers with cars equipped with airbags drive more aggressively, which increases the risk of death. Also, cars with anti-lock brakes systems (ABS) are driven faster and more carelessly. He also concluded that the crash prevention strategy that follows from this theory has been effective in many areas (Grant and Smiley, 1993; Sagberg et al., 1997).

Assum et al. (1999) studied risk compensation theory in the case of road lighting. They expected that drivers would not adjust their behaviour as a result of introducing road lighting. More specifically, drivers are not expected to increase their speed, reduce their concentration or travel more when road lighting is installed. They found, however, that drivers do compensate for road lighting in terms of increased speed and reduced concentration, which means that road lighting could have rather larger crash-reducing effects if compensation could be avoided. This means that risk compensation may occur even if the significant existing measure is proved to reduce crashes; the road users on the average do not seem to adapt their behaviour.

Wilde (2002) further explained in his later research that risk homeostasis theory predicts that, as safety features are added to vehicles and roads, drivers tend to increase their exposure to collision risk because they feel that they are better protected. He also illustrated the theory by referring to the Swedish experience when they changed from left- to right-hand driving (Figure 2.2): this was followed by a reduction in the traffic fatality rate, but it returned to its previous values after 18 months because drivers had responded to the increase in supposed danger by taking more care, which was later reduced.

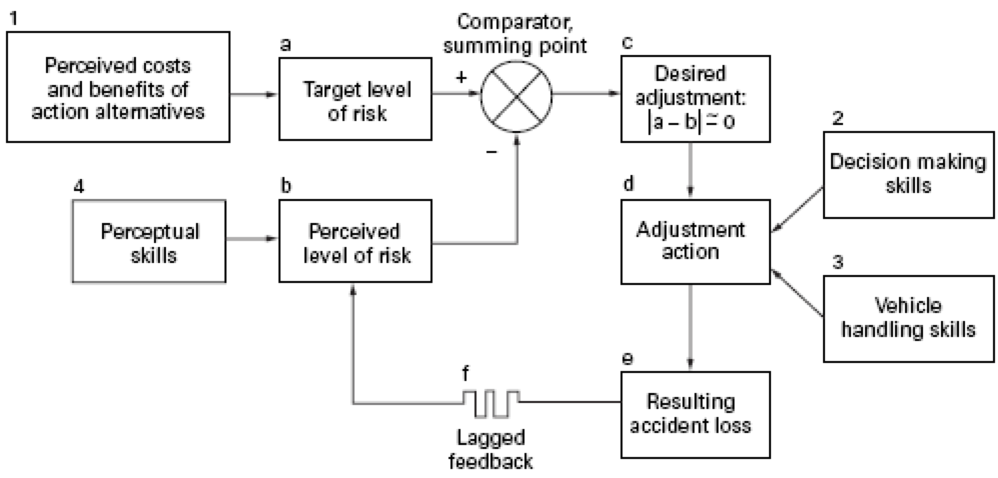

Figure 2-2 Compensation mechanisms (Wilde, 2002)

Morrongiello et al. (2007) investigated School-age children's self-reported risk compensation (greater risk-taking when wearing safety gear compared to when not doing so) using 6 common play situations. Results revealed greater risk taking scores under gear than no-gear conditions for every situation, indicating risk compensation operated for every activity. There was no significant variation in risk compensation with age or sex.

Fyhri et al. (2012) investigated whether the lack of effect of helmet wearing laws is due to risk compensation mechanisms or discouraging cyclists with the lowest crash risk, and thus increasing the overall average risk per cyclist. The cyclist population in Norway is divided into two sub-populations: one Speed-happy group that cycle fast and have lots of cycle equipment including helmets, and one traditional kind of cyclist without much equipment, cycling slowly. The results gave less support to a risk-compensation explanation because the speeding behaviour of the speed-happy group is more connected to other types of equipment than to bicycle helmets.

It is argued that the undetermined present state regarding the theory of risk homeostasis in traffic (RHT) needs to be resolved and that the following elements can serve as a basis for future concerted efforts directed towards the modelling of road users' decision making in traffic.

A further example is provided by Lardelli et al. (2003) who tried to obtain empirical data supporting or refuting the existence of the risk compensation mechanism with voluntary helmet use by Spanish cyclists. They suggested that a subgroup of cyclists with a higher risk of suffering a traffic crash are also those to whom the health consequences of the crash will be higher, but in this study careless cyclists who are more likely to be involved in a crash seem also to be among those for whom the health consequences of a crash are greater. Consequently, their results did not have a strong influence on risk compensation for voluntary helmet use in Spain.

If their perceived level of risk alters, their behaviour will compensate to place them back at their accepted level of risk. "Risk compensation is then seen as self-evident that individuals will tend to behave in a more cautious manner if their perception of risk or danger increases" (Stranks, 2007).

Morrongiello et al. (2007) studied risk compensation in children and why children show it in their reactions to wearing safety gear, by investigating six common play situations. Children responded to hypothetical scenarios by rating intended risk-taking when wearing safety gear and when not doing so, and by providing explanations for their behaviour. Results showed that greater risk-taking occurs when gear is worn than when no gear is worn in every situation, indicating that risk compensation operated for every activity. There was no significant variation in risk compensation with age or gender. Engaging in greater risk-taking when wearing safety gear revealed that the children believed wearing safety gear kept them safe from any degree of injury, protected them from serious injury, and resulted in them somehow being more competent to perform a higher-risk activity.

Generally, technology not always used for safety but also incentives for safety and health, expectations and hopes which make people more positive about their future and encourage them to look after themselves.

It can be concluded that the increase in the number of vehicles equipped with various safety devices may not be directly associated with the reduction in traffic crashes. In this case, the overall improvement in road safety depends on drivers' behaviour and how they perceive the risk of being involved in a traffic crash in relation to the introduction of safety devices in their vehicles.

2.2.6 Summary

It has been suggested that motorisation is strongly correlated with income as the rate of motorisation is higher in the developed countries with a comparatively high level of per capita GDP. It has been found that crashes decreased in the countries with higher GDP.

On the other hand, implementation of an organised trauma care system will considerably reduce traffic fatalities. Driver fatigue and drowsiness are also related to traffic occurrence, as indicated by the physiological theory of road safety, whereas, technology is not always used for safety but also used as incentives for safety and health.

It can also be concluded that the increase in the number of vehicles equipped with various safety devices may not be directly associated with the reduction in traffic crashes. In this case, the overall improvement in road safety depends on drivers' behaviour and how they perceive the risk of being involved in a traffic crash in relation to the introduction of safety devices in their vehicles.

Based on these road safety theories, literature identified a range of risk factors that influence traffic crashes. The following section discusses these theories in some detail.

2.3 Factors affecting road safety

An in-depth literature review on road safety suggests that factors affecting road traffic crashes in developed countries are different from those in developing countries (e.g. Smeed, 1949; Haight, 1980; Preston, 1980; Andreassen, 1985; Jacobs and Cutting, 1986; Navin et al., 1994).

In the case of developed countries, it is relatively easy to obtain all necessary data to conduct an analysis on the relationship between crashes and their contributing factors such as traffic flow, speed, and road geometry and weather conditions. This is not always true for the case of developing countries, in which it is more difficult to obtain data both on crashes themselves and their contributing factors. Moreover, there are reliability and quality issues concerning the data which are available. In order to see the differences and similarities in crash analyses between developed and developing countries, this section is divided into two parts: (1) factors influencing road safety in developed countries, and (2) factors affecting road safety in developing countries.

2.3.1 Factors influencing road traffic crashes in developed countries

A range of factors that affect road traffic crashes, and in order to reduce crashes it is essential to understand and analyse factors affecting road safety. These factors can be categorised as: (1) traffic characteristics, (2) driver characteristics, (3) road characteristics, (4) socio-economic factors, (5) land-use factors, and (6) environmental factors. Table 2.1 below shows the categories and the relevant factors in each of these categories based on evidence to be presented in the remaining of this section.

Table 2-1 Factors influencing road safety in developed countries

Type	**Factors**
Traffic characteristics	• Traffic flow • Traffic density • Traffic speed • Traffic congestion
Driver characteristics	• Age • Gender • Experience • Marital status
Road characteristics	• Lane and shoulder width • Curvature • Gradient • Junctions • Roundabouts
Socio-economic factors	• Population • Employment • Income • Deprivation
Land-use factors	• Commercial areas • Residential areas • Open space • Industrial areas
Environmental factors	• Rainfall • Snowfall

The following section will examine the evidence in previous studies on all factors influencing road traffic crashes in developed countries.

2.3.1.1 Traffic characteristics

There is evidence of a positive association between crash occurrence and traffic flow (For instance, Gwynn and Baker, 1970; Jadaan and Nicholson, 1992; Peirson et al. 1998; Martin, 2002).

For instance, Peirson et al. (1998) examined the crash risk by additional road users and how users respond to it. They investigated the relationship between road crashes and traffic flow and found that when the traffic flow increases the number of crashes increases.

Other researchers have developed models to predict crashes and traffic flow using the volume–capacity ratio V/C of the road (Frantzeskakis and Iordanis, 1987). Others have tried to derive a safety performance function to predict the effect of the change in exposure on the number of crashes (Mensah and Hauer, 1998). It seems that traffic flow positively affects road traffic crashes.

For instance, Martin (2002) developed a relationship between crash rates and hourly traffic flow (Fig.2.3) using data from 2000 km of French inter-urban motorways over two years, and showed that when traffic flow is at a rate of 1000-1500 vehicle/h incident rates are the lowest but for incident rates increase steadily with the increase in traffic flow. He also found that for light traffic levels the number of crashes was higher on three lanes than two lanes and higher at weekends, while in the case of heavy traffic the number of crashes was higher on weekdays, and there was no significant difference in the number of crashes between daytime and night-time. In addition there was no difference in the severity of crashes by number of lanes or weekdays/weekends, but it was high at night with light traffic.

Vehicle speed is also found to be a contributing factor in the occurrence of crashes. By looking at speed as another of the traffic characteristics influencing road safety, Johansson (1996) looked at the effect of speed limit reductions on crashes based on data from 1982 to 1991 in a number of Swedish counties. He found that reducing the speed limit can reduce the number of crashes, including minor injuries and vehicle damage. His results were unexpected because he showed that the effects of speed on fatalities and serious injuries were insignificant; this might be due to the improper use of the Poisson model for time series analysis which causes serial correlation normally present in a time-series dataset (e.g. Quddus, 2008).

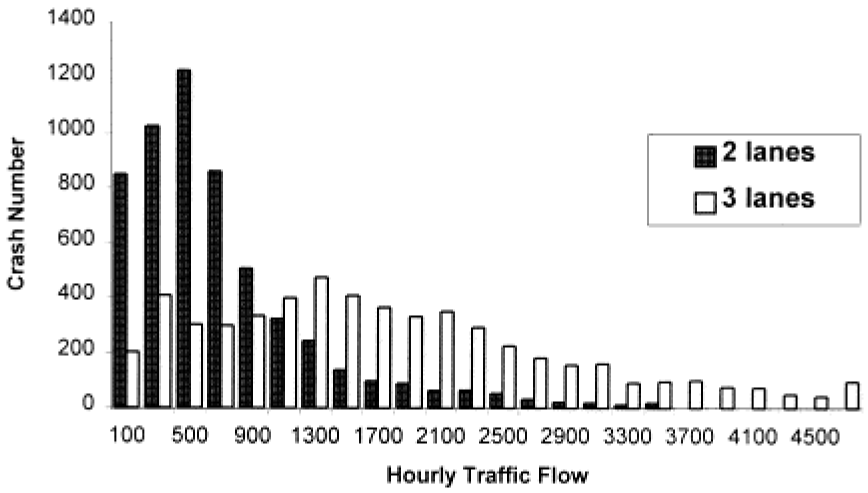

Figure 2-3 Number of crashes per number of lanes versus hourly traffic flow (Martin, 2002)

Later, Ossiander and Cummings (2002) examined the effect of the change of the freeway speed limit in Washington State using time series data (Figure 2.4). They found that increased speed limits are associated with a higher fatality rate even though the mean speed and speed limit were not equivalent. The spatial differences in road speeds between different spatial units may affect road crashes.

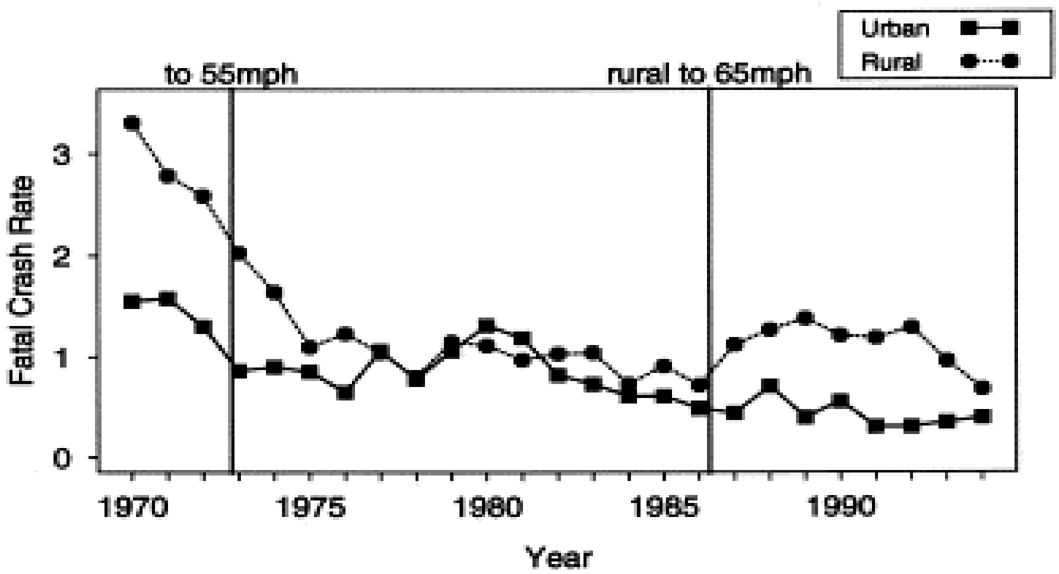

Figure 2-4 Fatal crash rates (fatal crashes per 100 million miles travelled) on Washington State interstate freeways Ossiander and Cummings (2002)

The vertical lines in Figure 2.4 show the date at which the 55 mph speed limit went into effect (November 1973), and the date at which rural interstates adopted the higher 65 mph speed limit (April 1987).

Elvik et al. (2004) have concluded that there is a strong statistical relationship between speed and road safety. When the mean speed of traffic is reduced, the number of crashes and the severity of injuries will almost always go down. When the mean speed of traffic increases, the number of crashes and the severity of injuries will usually increase. The relationship between changes in speed and changes in road safety holds for all speeds in the range between about 25 km/h and about 120 km/h. Speed is clearly a very important risk factor with respect to both crash occurrence and injury severity.

Most of the research has focused on establishing the relationship between crashes and different traffic flow characteristics, but little has focussed on the relationship with road density, volume and capacity ratio (V/C), level of service (LOS), or speed distribution because this type of data is not easy to obtain. Brodsky and Hakkert (1983) examined the effect of travel densities, defined as annual vehicle-miles per mile of road on the likelihood of crashes. They reported that crashes were not changed with travel densities for rural motorways. Lord et al. (2005) used data collected from rural and urban motorway segments inside and outside Montreal, Quebec, using crash prediction models of the relationship between crashes and the hourly traffic flow characteristics. They found that the predictive models which used traffic volume as the only variable may not capture the crash process on the motorway adequately, while the functional forms which included density and V/C ratio gave a better description of crashes.

Hauer (2009) fitted exponential functions to the data provided in Elvik et al. (2004) describing the effects of changes in speed on fatal crashes and injury crashes. The data provided in Elvik et al. (2004) have since been updated and expanded (Elvik, 2009) who updated and performed a new analysis of the Power Model of the relationship between changes in speed and changes in road safety. The updated analysis was based on 115 studies containing a total of 526 estimates. The Power Model provided a good description of the relationship between speed and road safety. One version of the model has been developed for roads in urban areas, another version for rural roads and freeways. The effects of changes in speed on road safety are smaller in urban areas than rural areas. The report analysed the normative foundations of speed limits, it has concluded that speed limits are needed.

From the above, it appears that speed is associated with the number of fatalities, and if speed is reduced crashes will be reduced. It is also clear from the above that traffic characteristics influence road safety, and the number of crashes will be increased when traffic flow and speed increase; a few research studies have focused on density because of the availability in data. It is clear that traffic flow is positively affecting road safety, in that the number of crashes increases with the increase of traffic flow.

Another important traffic characteristics factor, traffic congestion which is however studied less in the literature. Martin (2002) found that light traffic (i.e. less congested) is a safety problem both in terms of crash rate and severity. Noland and Quddus (2005) investigated congestion and safety in London using an area-wide spatial analysis approach. London was divided into 15,366 spatial units, real-world data in each area were collected and analysed using Negative Binomial models by controlling for other contributing factors. Their results suggested that there is little effect of traffic congestion on road safety. Kononov et al. (2008) investigated the relationship between traffic congestion and road crash rates on urban freeways using the data from California, Colorado and Texas. They found that fatal and injury crash rates increase with the increase in traffic congestion. Wang et al. (2009) investigated the effects of traffic congestion on road crashes of the M25 London orbital motorway in England while controlling for other contributing factors. A series of relevant statistical models have been employed to develop the relationship between the frequency of crashes and the congestion index. It has been found that traffic congestion has little or no impact on the frequency of road crashes (either for fatal or serious injury crashes).

2.3.1.2 Driver characteristics

The general findings in the literature show that high speed driving and more aggressive driving were related to gender and age, and especially to young and male drivers.
Since human who is travelling and making crashes; therefore, it is important to study their behaviour, because many researchers have investigated demographic factors as risk factors in crashes.

For example, Matthews and Moran (1986) found that young male drivers (26 years old and younger) are over-represented in traffic crashes because they are overconfident and overestimate their ability. Mercer (1989) showed that, without correcting for annual mileage, young and male drivers were strongly represented in terms of both crashes and traffic-related fines, while after correcting for mileage female drivers were higher in number of crashes. Garber and Srinivasan (1991) concluded that crashes involving old drivers were higher at intersections outside cities than inside cities. Evans (1991) found that elderly drivers (60 years old and above) were less likely to be involved in crashes. Simon and Corbett (1996) also found a relationship between committing violations and crash involvement in which male and young drivers were more involved. With the aim of investigating other driver characteristics, Kim et al. (1995b) estimated a log-linear model to investigate the role of driver characteristics and behaviour in the crashes leading to more severe injuries. They found that alcohol consumption and driving without wearing a seat-belt increased with the severity of crashes.

Other findings by Chen (1997) showed the relationship between age and crash involvement, which followed a U-shape, with a high crash rate for drivers below the age of 25. The injury–crash involvement rate followed the same trend, but fatal crash rates were very high for drivers aged 80 and above in the state of Florida. These results show that young drivers are strongly involved in crashes that are related to speed.

Later, Abdel-Aty et al. (1998) used the 1994 and 1995 Florida crash database and also employed log-linear regression models to estimate the relationships between the driver’s age and several crash-related factors such as location, severity and characteristics of the crash. Other variables that were related to driver behaviour, e.g. speeding and alcohol usage were also included in the study. They showed that there was a significant relationship between the driver’s age and involvement in traffic crashes indicating that in some situations middle-aged drivers were more likely to be involved in crashes.

This result was not consistent with the findings of a study conducted by Kim et al. (1995b) who showed that younger drivers were more likely to be classified as fatal than older drivers involved in crashes, but it was consistent with the findings of Stamatiadis et al. (1991). Vachal and Malchose (2009) showed that 14-year-old drivers are three times more likely to die or be disabled in a crash involving injury than 17-year-old drivers in North Dakota; the likelihood for death or disablement is 165% greater for unbelted teenage drivers than for those who are properly belted; and teenagers are six times more likely to be severely injured in crashes on rural roads than on urban roads.

Gray et al. (2008) analysed the characteristics of male drivers in relation to the severity of traffic crashes and concluded that younger male drivers (aged 17-19) from London are likely to be involved in more severe crashes than older male drivers (aged 23-25). Their study, however, found that the impact of the driver's age on the severity of traffic crashes is negative when all male drivers (aged 17 to 25) from Great Britain are considered. Lardelli-Claret et al. (2009) studied traffic crashes registered in Spain between 2000 and 2004, and found that the risk of early death associated with the severity of the crash in drivers of passenger cars decreases as age increases, but increases with age when the risk of early death was associated with resilience to the energy released by the crash.

However, this study has some limitations arising from the database, where some eligible crashes were not included, and there may be some residual as it was not possible to adjust for the effect of other factors which were not registered (e.g. size or weight of the vehicle or the presence of an airbag in the vehicle), which may cause a bias leading to underestimation of the effect of seat belts.

In addition, information on the quality of medical care received by the victims was not available. Therefore, it can be said that the impact of the driver's age on the severity of crashes is not consistent in the literature. It is clear that the age of the driver affects road safety; young drivers are more likely to be involved in crashes than older drivers.

Simon (2001) analysed road crashes in Slovenia involving pedestrians, cyclists or motorcyclists and cars from 1994 to 1998, using a logistic regression method and concentrating on factors that distinguish a fatality or severe injury from a slight injury or absence of injury caused by crash. He found that cyclists and motorcyclists that used helmets were at lower risk and pedestrians and motorcyclists were at higher risk than cyclists. It was also found that pedestrians, cyclists and motorcyclists hit by a car with a driver aged 25 or less, additionally, victims who getting injured by cars drivers were old. Crashes that happened at night were found to be much more serious than the daytime ones, and Saturday and Sunday were the most risky days.

Evidence by Noland (2003) suggested that seat belt usage is beneficial to safety. Similarly, Curtis et al. (2007) investigated the effect of seat belts on crashes by examining seat belt law in New Hampshire in the US. More extensive research on the effects of seat belt use on safety can be seen in the study by Kim (2003), who investigated belted and unbelted road users and crash characteristics; his intention was to obtain a better enforcement design and education programmes. These authors all found that wearing seat belts will reduce the severity of crashes and will be beneficial for safety. While Lourens et al. (1999) strongly supported previous findings as they presented some analyses on a Dutch database that contains disaggregated data on both the traffic system input variables of the driver population as a characteristics of drivers including their annual mileage and the output variables in terms of habitual driving behavior and crash involvement the number of fines which driver received and male and female crash involvement were positively correlated (Figure 2.5).

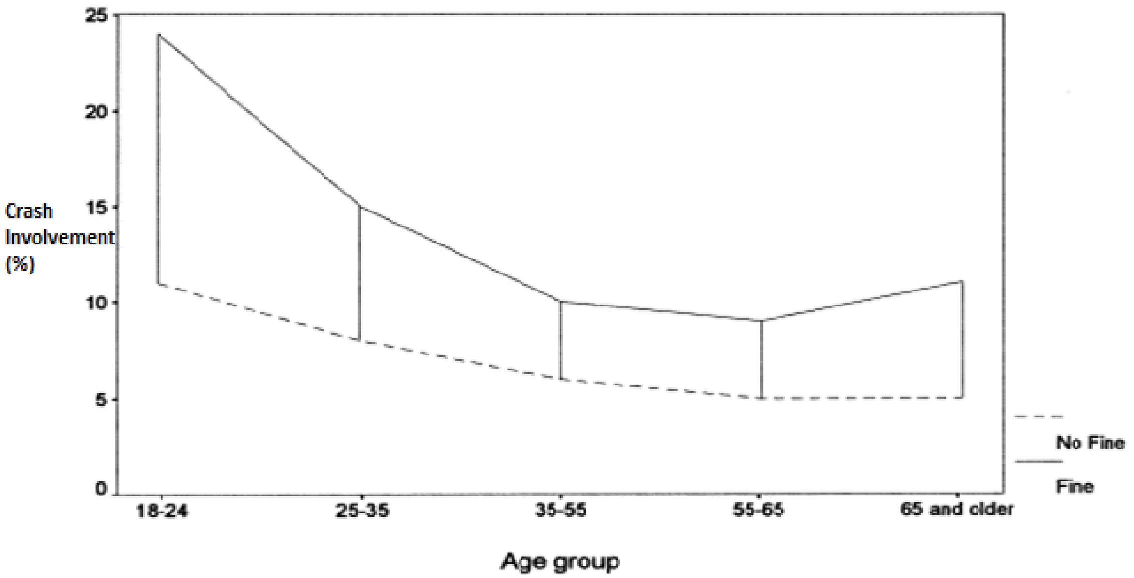

Figure 2-5 Crash involvement (%) for drivers with and without a fine over the five age groups (Lourens et al., 1999).

It is clear that, among the driver characteristics, alcohol consumption and driving without wearing a seat belt increase the severity of crashes. In other words, driver behaviour affects road crash involvement, and enforcement and education programmes are necessary.

Quddus et al. (2002) examined factors affecting the severity of motorcycle crashes and vehicle damage in Singapore. They used nine years of crash data categorised into three levels of injury severity and found that non-Singaporean riders are associated with high crash severity compared to Singaporean riders, which is consistent with Lardelli-Claret et al. (2002), who estimated the association between driver nationality and the risk of causing vehicle collisions in Spain during the period from 1990 and 1999 and found that foreign drivers in Spain are at a greater risk of causing collisions than Spanish drivers in urban areas.

Vourvahakis et al. (1997) investigated the epidemiology of road traffic crashes during pleasure travelling in the island of Crete; they found that victims in road traffic crashes among travellers have a different epidemiologic profile compared with crashes of a similar nature among locals. Their findings showed that tourists coming from countries that drive on the left to countries that drive on the right may be at increased risk of motor vehicle injuries. Alcohol abuse was reported as a primary cause of crashes, with a highly significant proportion among foreign nationals. Flegr et al. (2002) showed that in central Prague, Czech Republic, subjects with latent toxoplasmosis had a significantly greater risk of traffic crashes than non-infected subjects.

The effect of the gender variable disappeared when annual mileage was taken into account. The effect of driver’s age group on risk was strong in the analyses and present in all classes of annual mileage. After correcting for annual mileage the level of education of the driver showed no significant relation with crash involvement.

A comparison conducted by Sun et al. (2004), who studied the characteristics of deaths related to traffic crashes in Japan and Ireland between 1950 and 2000, analysed mortality and the negative effect on life expectancy and generated a multivariate model. The time trends showed an increase in mortality followed by a decrease. The mortality rates were about 13 and 5 per 100,000 for males and females, respectively, while in 2000 the correlation coefficients for gender were over 0.9. Mortality rates were high among the young groups, mainly among males. They found that the characteristics in the two countries were similar, which means traffic crash-related deaths are consistent in Japan and Ireland.

In another study about the characteristics of the driver, Clarke et al. (2006) used four types of crash covering a two year period to analyse a sample of over 3000 crash cases from midland British police forces involving drivers aged 17–25 years. They concluded that there are specific propensities for crashes of different types, depending on driver's age, gender, and experience. Young male drivers aged 17–19 years were found to be significantly more likely to be involved in crashes during the hours of darkness on rural curves. Cross-flow turn crashes seem to decline with increased driver experience while crashes in darkness (without street lighting) showed the slowest improvement. Rural curves seemed to show a difference because of the single vehicle factor but when this was corrected they showed a large decrease over the first three years of driving experience. Generally, analyses showed that, when corrected for annual mileage, male and female drivers did not differ in crash involvement, while young drivers had the highest rate of crashes, and level of education was not related to crashes.

The gender and experience of the driver influence road safety; young and male drivers were found to be significantly more involved in crashes, and enforcement and education programmes are very important for drivers.

In conclusion, age, nationality, gender, experience, wearing seat-belts and alcohol consumption are associated with road crashes and driver characteristics affecting road safety.

2.3.1.3 Road characteristics

As mentioned in Table 2.1, the road characteristics are: road length, curvature, gradient, junctions and roundabouts. Road infrastructure and geometry affect the severity of crashes and improved road infrastructure could reduce road crash severity and frequency. Findings from different researchers support this hypothesis.

Joshua and Garber (1990) used multiple linear and Poisson regressions to estimate truck crash rates based on factors related to traffic and road geometry. Miaou et al. (1992) also used a Poisson regression model on traffic data on 8779 miles of roads from the Highway Safety Information System (HSIS) in the USA to establish quantitative relationships between truck crash rates and geometric characteristics of roads. Their results indicated that for horizontal alignment the alternative measures for mean absolute curvature and the mean absolute grade for vertical alignment are the most important variables for crash rate estimation. Knuiman et al. (1993) examined the effect of median width of four-lane roads on crash rates using a negative binomial distribution. They indicated that crash rate decreases with increasing median width.

Shankar et al. (1995) found that an increased number of horizontal curves per kilometre increased the possibility of physical injury compared to property damage alone occurring in a crash. Shah et al. (1993) used a linear regression to predict the truck crash involvement rate per mile per year, based on average annual daily traffic (AADT) and truck AADT per lane, shoulder width, horizontal curvature, and vertical gradient. The results suggested that truck involvement rate increases with AADT and truck AADT, degree of curvature and gradient. Another study undertaken by Hadi et al. (1993) used data from the Florida Department of Transportation's Roadway Characteristics Inventory (RCI) system, and used a negative binomial (NB) regression for crash rates on different types of rural and urban roads with different traffic levels. Their results suggested that higher AADT levels and the existence of intersections are associated with higher crash rates, while wider lanes and shoulders are helpful in reducing crash rates.

Abdel-Aty and Radwan (2000) also used a negative binomial regression model to predict crash rates as a function of AADT, degree of horizontal curvature, section length, and lane, shoulder and median widths, as shown in Table 2.2. This study also included gender (male or female) and age (young, middle-aged and old). They showed that the crash rate increased with AADT, degree of horizontal curvature and section length. The crash rate decreased with lane, shoulder and median width.

Table 2-2 Elasticity estimates for the crash frequency model (Abdel-Aty et al., 2000)

Variable	Elasticity
Section length (km)	0.33
AADT per lane	0.62
Degree of horizontal curve	0.07
Shoulder width (m)	-0.13
Median width (m)	-0.12
Lane width (m)/number of lanes	-0.38

Ivan and O'Mara (1997) used an NB regression model employing 1991–1993 crash data from the Traffic Accident Surveillance Report of Connecticut, USA. They found that AADT was a significant crash predicted variable, while geometric design and speed differential measures were not found to be effective predictors of crash rates.

Navin et al. (2000) suggested that, when crashes take place, making improvements in road safety engineering and not only in vehicle design can reduce the severity of injuries in these crashes, as road infrastructure plays an important role in road safety. Noland and Oh (2004) analysed panel data from the state of Illinois in the USA and showed that increasing lane width and the number of lanes will increase fatalities, while increasing the width of outside shoulders will reduce crashes. This is supported by Perez (2006) who found that highway upgrading has a significant positive effect on road safety. Navin et al. (2000) and Perez (2006) pointed out that improved infrastructure design and engineering work play an important role in road safety.

On the other hand, Haynes et al. (2007) conducted two studies on the effect of road curvature on traffic casualties, one based on New Zealand data and the other for England and Wales. They found in both studies that fatal crashes have an inverse relationship with road curvature, which may be due to the difference in demographic characteristics and spatial units they used. Kim et al. (2007) found that crashes on curved roads result in more severe outcomes, especially in the case of bicycle-motor vehicle crashes.

In conclusion, improving the geometric design and characteristics of roads will reduce fatalities and the severity of injuries of crashes, which means road design, has an inverse relationship with road crashes.

2.3.1.4 Vehicle characteristics

The type, age, and design of vehicle have been found to be contributory factors in road traffic safety.

Chang and Mannering (1999) collected state crash data during 1994 in Washington State, USA, and applied nested logit modelling techniques to study injury severity according to different vehicle occupancies. They studied the relationship between occupancy and injury severity in truck- and non-truck-related crashes. Large truck effects showed a significant impact on the most severely injured vehicle occupant by separately estimating nested logit models for truck-related crashes and for non-truck-related crashes, and, when the crash characteristics of truck-involved crashes and non-truck-involved crashes were compared, comparison results indicated that several variables which significantly increase injury severity were found only for truck-involved crashes, including high speed limits, crashes occurring when a vehicle is making a right or left turn, and rear-end collisions. The results also indicated that the effects of trucks on crash-injury severity were more significant for multi-occupant vehicles than single-occupant vehicles. Using data for a large number of multi-occupant vehicles in truck-involved crashes will allow more comprehensive comparison of crash characteristics between truck-involved crashes and non-truck-involved crashes. It would also be interesting to study the factors that affect severity levels for crashes involving different types of trucks.

Navin et al. (2000) suggested that better vehicle design can reduce the severity of whiplash injuries when crashes occur. Kweon and Kockelman (2003) found that pickups and sport utility vehicles are less safe than passenger vehicles in single-vehicle crashes while in two-vehicle crashes these vehicle types are associated with less severe injuries for their drivers and more severe injuries for occupants. Yau (2004) examined risk factors affecting the severity of single-vehicle traffic crashes in Hong Kong using three types of vehicles, private, goods and motorcycles, during the 2-year period 1999–2000. For private vehicles and motorcycles, age of vehicle was a significant factor associated with injury severity. For goods vehicles, seat-belt usage and weekday occurrence were significant factors associated with injury severity.

Tay et al. (2008) used a large and highly disaggregate set of traffic collision data obtained from the traffic police. They identified the factors that were associated with the likelihood of hit-and-run crashes, including driver characteristics, vehicle types, crash characteristics, roadway features and environmental characteristics, using a logistic regression model to distinguish hit-and-run crashes from non-hit-and-run crashes. They found that crashes involving two-wheeled vehicles would result in a higher likelihood of running after the crash.

Arenas et al. (2009) illustrated the influence of heavy goods vehicle traffic on crashes on different types of Spanish urban roads by taking the annual average daily traffic as the most important variable, followed by the percentage of heavy goods vehicles. They found a reduction in the total number of crashes as a result of a drop in the number of heavy goods transport vehicles; However, the higher traffic intensity reduces the effects on the number of crashes on single carriageway road segments compared with high capacity roads. It is recommended that the effect of other types of variable should be studied, including speed, geometry and road layout.

Cooper et al. (2009) studied the safety of vehicles imported from right-hand-drive (RHD) vehicle configuration countries when operated in a left-hand-drive (LHD) vehicle environment. They took vehicles over 15 years of age imported into Canada from countries that drive on the left side of the road (such as Japan). They investigated whether the RHD configuration led to increased risk of crash involvement. They utilised three different approaches: a relative crash culpability analysis where RHD and (LHD) crash rates were compared for the same group of drivers, another that focused on the time before the first crash event following initial policy date for RHD versus LHD vehicles, and a third approach which compared RHD drivers and vehicles in crashes to a comparison group of drivers in the general population in a multiple regression model using a number of driver–vehicle control variables. They found that all three analyses were consistent. RHD vehicles had a significantly greater risk of at-fault crash involvement than similar LHD vehicles. However, crashes involving RHD vehicles were no more severe than those involving LHD vehicles only. These analyses suggest that the level of injury severity is dependent on the different vehicle occupancies, vehicle types, presence of heavy goods vehicles and vehicle design.

It is noticeable from the above that injury severity is conditional on different vehicle occupancies, vehicle types, presence of heavy goods vehicles and vehicle design, all of which have a significant influence in traffic crashes.

2.3.1.5 Land-use characteristics

Land-use characteristics could affect the severity of traffic crashes. These include percentage of commercial, residential, open-space and industrial areas. According to urban safety management theory, land-use policy is one of the strategies used to prevent and reduce crashes (The Institution of Highways and Transportation, 1997). A few previous studies have investigated the links between land use and road traffic crashes; this section will discuss the effect of land use on road safety in some of the developed countries.

Several studies have been carried out in the US and the UK; for example, Levine et al. (1995) investigated the relationship between zonal land use and road traffic crashes. They found that residential population density, manufacturing, retail trade and services industries were positively related to the number of road traffic crashes. Different types of land use generate different numbers of trips making different driver behaviour which finally cause road crashes. Peled et al. (1996) looked at many other aspects of road safety analysis which were also associated with different forms of land use and their associated activities. Later, Aultman-Hall and Kaltenecker (1999) found that land-use characteristics were less significant than traffic levels and attitudes for predicting bicycle safety.

Petch and Henson (2000) examined child casualties in Salford, England, covering an area of about 100 km^2 over a 40-month period. Factors related to land use like the number of trips or trip generators, the percentage of terraced housing and the amount of open space were found to be significant variables, and in terms of cyclist safety, little has been found related to land use, so increasing urbanisation is associated with an increase in non-motorised transport casualties, through demographic factors, road and traffic environment factors such as road length and junction density, and land use.

Similarly, Noland and Quddus (2004) also conducted a spatially disaggregated analysis, and found that urbanised areas have fewer casualties while areas with higher employment suffer more casualties, while controlling for other exposure factors such as population and traffic flow. Other research which was not consistent with Noland and Quddus (2004) was conducted by Kim et al. (2006), who performed a spatial analysis of pedestrian crashes in Hawaii and found that both residential and commercial areas were positively associated with pedestrian crashes.

Another study performed by Kim and Yamashita (2002) compared the vehicle-to-vehicle crashes with vehicle-to-pedestrian and vehicle-to-bicycle crashes per acre land-use category for Honolulu over a 10-year period. They found that vehicle-to-vehicle crashes were highest in commercial and industrial areas (6.62/10 acre year) (number of crashes over ten years per acre of land), visitor lodging (5.15/10 acre year) and manufacturing and industry (3.67/10 acre year); the results were different for vehicle-to-pedestrian crashes, with visitor lodging (0.43/10 acre year), commercial and industrial (0.30/10 acre year) and public services (0.20/10 acre year) being the three highest categories.

These findings were not consistent with Wier et al. (2008), who described the development of a multivariate area-level regression model of vehicle–pedestrian injury collisions based on environmental and population data in 176 San Francisco, California, census tracts. They applied this model to predict area-level change in vehicle–pedestrian injury crashes in relation to land-use development and transportation planning decisions. They explained approximately 72% of the variation in census-tract vehicle–pedestrian injury crashes including traffic volume, commercial and residential land uses, employee and resident populations, proportion of people living in poverty and proportion aged 65 and older. These results were consistent with previous findings, but this study provided additional evidence that traffic volume was a primary cause of vehicle–pedestrian injury crashes at the area level, while land areas zoned for neighbourhood commercial use and residential/neighbourhood commercial use were statistically significant in San Francisco.

Wedagama et al. (2006) looked at the relationship between non-motorised road traffic casualties and land use in two zones of approximately 8 km^2 in Newcastle upon Tyne, England; they developed generalised linear models using non-motorised casualties as the response variable, with primary functional land use, population density and junction density as descriptive variables, and investigated pedestrians and cyclists in relation to land use types such as retail and commercial areas. They found that pedestrian casualties in the city centre zone during working hours were mostly associated with an increase in retail and community land use, while an increase in pedestrian casualties out of working hours and an increase in cyclist casualties during working hours was associated with an increase in retail land use. Kim et al. (2007) examined factors associated with the severity of injuries to cyclists in bicycle–motor vehicle crashes, using police-reported crash data between 1997 and 2002 from North Carolina, USA. They found that there were largely no statistically significant differences between the different land development characteristics.

In conclusion, residential population density and manufacturing, commercial, and industrial areas were positively related to the number of road traffic crashes, and areas with more employment suffered more casualties. So land-use types and their characteristics influence road traffic safety in developed countries.

2.3.1.6 Environmental factors

Environmental factors such as lighting and weather are other important factors which could affect road crashes through vehicle speed and behaviour of the driver. These factors will be discussed in this section.

In looking at environmental factors, Shankar and Mannering (1996) found that road surface and weather conditions were significant factors in predicting five levels of severity of rider injuries in Washington D.C. Edwards (1998) investigated the relationship between weather and road crashes in England and Wales. The weather information recorded on Police Accident Report Forms was taken as the current weather at the time of the crash; crash severity for the poor weather categories of rain, fog, and high winds was compared with the non-hazardous condition of fine weather. He found that crash severity decreased significantly in rain compared with fine weather, while severity in fog showed geographical variation, with different findings in different local authorities. High winds and fog were not found to be significant, but speed during these weather conditions will increase the severity of crashes. Abdel-Aty (2003) predicted driver injury severity in Central Florida, finding that crashes occurring at signalised intersections with bad weather and poor street lighting had a significantly higher probability of severe injury.

Depending on the data used for weather effect in crashes both in time and space and also depending on the kind of models used. Brijs et al. (2008) used the Poisson Integer

Autoregressive Model (INAR) for counting data, and used a number of covariates related to exposure and weather aspects such as wind, temperature, sunshine, air pressure and precipitation in order to study the effect of weather conditions on the number of daily crashes in the Netherlands using a discrete time-series model. They showed that apart from exposure weather aspects such as rain, fog, high winds and temperature have an influence on the number of crashes.

Using meteorological data for weather instead of using crash weather records, especially when modelling the number of crashes in a large area, will cause a measurement problem, if the model used does not differentiate between different types of crashes and rain or snow.

2.3.1.7 Summary

As can be seen from the above, in developed countries there are a range of different factors affecting road traffic safety, which are categorised as traffic, driver, road, vehicle, land use, socio-economic and environmental characteristics. As indicated, traffic flow and speed are significant factors in road crashes. Driver characteristics (age, gender, and experience) influence road safety; on the other hand, driver behaviour (seat-belt usage, alcohol consumption) has a positive relationship with road crashes. It is also found that improving the geometric design will reduce fatalities and the severity of injuries of crashes, which indicates that improvements in road design have an inverse relationship with road crashes. According to the facts shown above, land-use types and their characteristics influence road traffic safety, which indicates that residential population density and manufacturing, commercial, and industrial areas, are positively related to the number of road traffic crashes, and areas with more employment suffer more casualties. Furthermore, environmental factors (rain, fog, wind, temperature, etc) are not the only causes of road crashes, but speed during these weather conditions will increase the severity of these crashes.

Developed countries have been able to reduce the number and severity of road traffic crashes by education, engineering and enforcement, as in the United States, Europe, and Australia; in particular, highway and traffic engineering could be easily implemented and very cost-effective (Berhanu, 2000).

2.3.2 Factors influencing road traffic crashes in developing countries

Traffic crashes have been considered as one of the major causes of human losses in both developed and developing countries, but more in developing countries because of their seriousness and the limited resources and lack of data to reduce these road crashes by developing realistic research. As can be seen from the case of developed countries, data associated with traffic, driver, vehicle and environmental characteristics are essential in order to conduct a robust crash analysis that will assist in developing road safety policies and educational training and campaigns. The major problems in developing countries are the lack of availability of such data and the quality of the available data. First of all, most of the crashes that happen in developing countries are not reported by the authority. The recorded crash data only have limited information from which it is difficult to conduct a thorough crash analysis. There is also a lack of resources in terms of researchers and facilities to carry out a proper crash analysis. Moreover, the findings from the studies on developed countries may not be applicable to a developing country because of the difference in all aspects of road environments. This section reports on a literature review of various factors affecting road traffic crashes in the case of developing countries.

Vasconcellos (1999) looked at urban development and traffic crashes in Brazil by analysing the patterns of traffic crashes in urban areas, using the data from the two largest cities in the country. Hospital data were analysed to understand the social impacts of such crashes. Several causal factors were identified: political, in the sense of rights and duties, in that drivers think as political human being that they have the priority in occupying the space while pedestrians think the opposite; cultural, in that people see crashes as a cost of development they still not feeling the problem, and technical, in that there is technical expertise to design, build, and maintain roads in an advanced way, but little expertise to analyse traffic crashes. Moreover, planners are not trained to consider safety as a priority issue. As human factors people involved in traffic crashes were relatively young as a result of the higher mobility of younger people. Vasconcellos found that reorganisation of the built environment was a very important role which means the built environment that emerged from transformations. He described the environment as a natural producer of traffic and crashes, and concluded that educational programmes and an increase in enforcement were necessary.

In another study, Hijar et al. (2000) examined the risk factors in traffic crashes in Mexico. They identified the risk factors related to the driver, the vehicle and the environment which were associated with motor vehicle crashes on motorways. The risk factors used were age under 25 years, alcohol consumption, and driving during the daytime on a weekday under poor weather conditions. Results obtained with respect to age agreed with other studies that reported that the risk was higher for young people and lower for adults but increased again in the older age group where crashes became more likely to be fatal. With respect to alcohol consumption, results showed that it was not a major cause of traffic crashes. For vehicle characteristics no significant differences were found. Poor environmental conditions (rain, fog, and wet roads) also showed a strong association with the risk of a crash.

Later, Bastos et al. (2005) looked at seat belt and helmet use among victims of traffic crashes in a city of Southern Brazil between 1997 and 2000 after the implementation of a new traffic code in Brazil in January 1998. They analysed the use of safety devices (helmets and seat belts) among victims of motorcycle and car crashes during the above period. They found that the rate of seat belt non-use remained the same after 1998 while the rate of helmet non-use decreased continuously each year.

In another study, Jain et al. (2008) tried to find the trends among two-wheeler crashes on India's roads for a five-year period (2000–2004) with respect to age and gender of the victim, type of injury, type of vehicle involved and time distribution of crashes. They found that there was considerable morbidity and mortality due to two-wheeler road traffic crashes. Among a total of 1231 two-wheeler crashes recorded during this period, 77% of the victims were in the age group of 18-44 years. A higher percentage of males (83%) than females (17%) were involved in crashes.

2.3.2.1 Findings and evidence from Middle East countries

Due to the similarities in life style and cultural nature and the resemblance of people behaviour and their environment, which includes the road and the traffic management systems, it was worthwhile investigating the findings and evidence regarding the causes of road traffic crashes other countries within the Middle East region.

Zadeh et al. (2002) performed an epidemiological study to determine the causes of traffic crash-related deaths in Tehran, Iran, during 2000–2001. In a one-year survey the number of traffic crash-related deaths recorded was 2128; the male-to-female ratio was 4.1 to 1, the majority of deaths were between the ages of 21 and 30, and they did not notice any seasonal variation. They again recommended public education about traffic rules, compulsory use of seat belts, and wearing of crash helmets.

Another study in Tehran by Roudsari et al. (2004) studied gender and age distribution in transport-related injuries in Tehran; they analysed information on 8426 trauma patients over 13 months. Forty-five per cent of the injuries were related to car crashes; the male-to-female ratio was 4.2 to 1, while for motorcyclists the ratio was 16 to 1. They also found that 67% of the injuries to females and 44% of the injuries to males were related to pedestrian crashes. They discovered that, in the data they used; only 6% of the male motorcyclists were wearing a helmet and only 3% of the male car occupants were using a seat belt at the time of the crash, while among females the condition was worse where they were not using any protective devices.

Berhan (2004) carried out a study to develop predictive models relating traffic crashes with the road environment and traffic flows, based on the data collected on main roads in Addis Ababa, Ethiopia. He showed that the existing inadequate road infrastructure and poor road traffic operations were possible contributors to traffic crashes in Addis Ababa. The results also indicated that improvements in roadway width, pedestrian facilities and traffic engineering in general were effective in reducing road traffic crashes.

Another study was conducted by Abbas (2004), who studied traffic safety assessment and the development of predictive models for crashes on rural roads in Egypt. He analysed the causes of crashes on the basis of all crash records collected in 1998 for five rural roads in Egypt; more than 26 causes were included and categorised under five main categories: driver, pedestrian, vehicle, road, and environmental causes. The causes which contributed most highly were driver-related. In general, driver-related causes contributed around 59–73%, followed by vehicle-related causes, which contributed around 23%. Pedestrian-related causes also contributed around 4%, while road-related causes contributed only 3.5%. Environmental and other causes contributed around 3.5%. Abbas (2004) concluded that there was a lack of past reliable and detailed crash data collection programmes as well as a lack of crash prediction models.

Yereli et al. (2006) reported that latent toxoplasmosis in Turkey could cause extended reaction times and increase traffic crashes. Toxoplasma gondii can be located in neural or muscular tissues and cause prolonged reaction times of the muscles and cysts in the brain tissues: the extended reaction times lead to the deceleration of reflexes, which could be a major cause of traffic crashes. This was investigated among the population who were involved in traffic crashes while driving. A total of 185 people (100 men and 85 women), aged between 21 and 40 years, were enrolled in the study group (SG) during a period of six months, and the other group used was the control group (CG) which also consisted of 185 people (95 men and 90 women). They found that latent toxoplasmosis could cause a rate of traffic crashes higher than in the general population, and that the possible risk of traffic crashes in Turkey increased for drivers due to the high seroprevalence of latent toxoplasmosis.

2.3.2.2 Findings and evidence from Gulf Co-operation Council (GCC) countries

Ameen et al. (2001) developed a model for the analysis and forecasting of road crash fatalities in Yemen, considering the data restrictions common in most of the developing countries. They used data for the period 1978–1995 to build models to understand the nature and size of the causes of fatalities. They examined the influence on road users of consuming a locally grown stimulant called Qat, which can be considered as a type of drug, and found that it increased the risk of crashes (Figure 2.6), because, although Qat has a stimulating effect during the first few hours, the findings showed that after being consumed it will increase the risk of road crash fatalities, because Qat causes anxiety and tensions in users after several hours of use, and makes drivers busy drinking water, smoking or preparing and cleaning the Qat.

The changes in road crash fatalities in Yemen can also be explained through changes in some socio-economic and cultural variables: for example, political changes caused a huge arrival of Yemeni immigrants returning from Gulf countries after the Gulf crisis, leading to an increase in the road crash fatalities making GNP high (Ameen et al. 2001). Drivers are usually careful during economic recessions to avoid the expenses of repairing or replacing vehicles; this will also lead to reduced travelling, which would finally reduce the number of traffic crashes and the rate of fatalities.

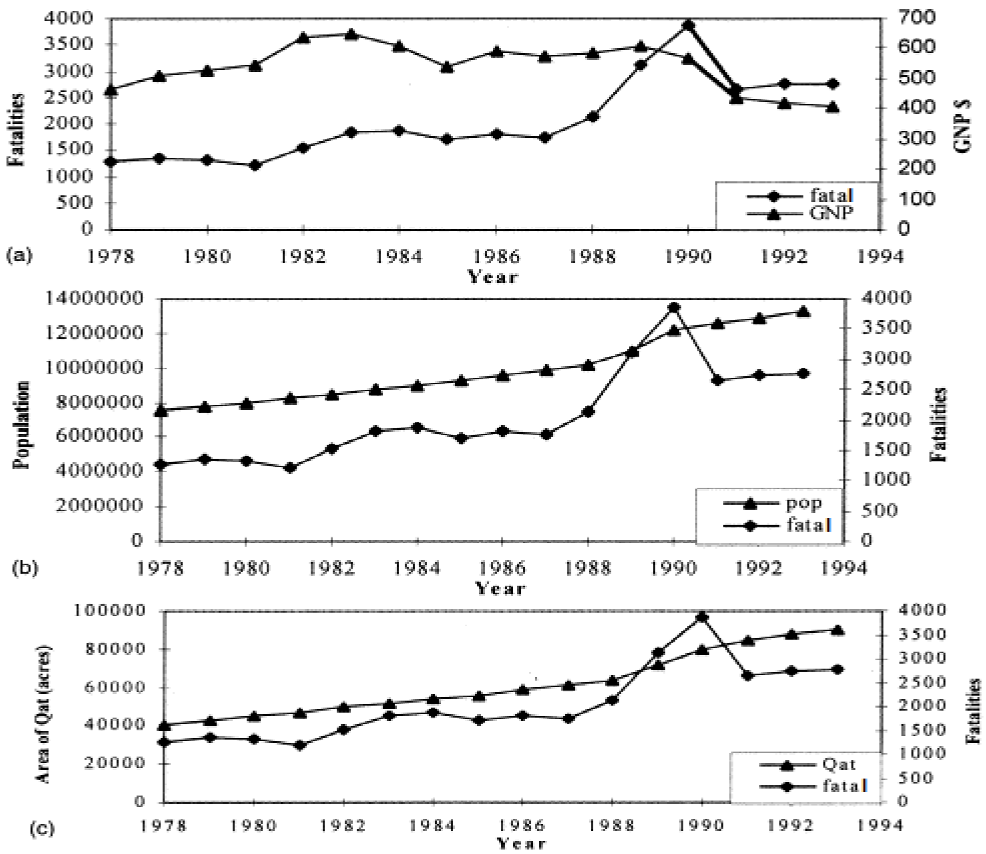

Figure 2-6 (a) Trends of fatalities and GNP; (b) trends of fatalities and population; (c) trends of Qat and fatalities.

In a further study in another GCC country, Abdalla (2002) investigated the fatality risk assessment and modelling of drivers' responsibility for causing traffic crashes in Dubai, and concluded that the possible characteristics and attributes of drivers involved in crashes, variations between different classes of road users and groups of the resident population, and the lack of safety and risk awareness among road users were causing the traffic crashes in Dubai.

In another GCC country, El-Sadig et al. (2002) studied trends of morbidity and mortality in road traffic crashes in the United Arab Emirates during 1977–1998. They investigated the effectiveness of traffic safety measures, including measures enforcing speed limits on motorways, and suggested that the use of fixed positions for speed radar on UAE motorways, and the possibility of using radar detectors which will warn drivers against speeding, might reduce the number of violating drivers. They recommended increasing police surveillance on the roads in order to enforce traffic safety regulations including seat belt laws and speed limits, and conducting effective education campaigns on safe driving, with legislation to impose high penalties on careless driving and driving at high speeds.

Al-Madani and Al-Janahi (2006) attempted to test personal exposure risk factors in pedestrian crashes in the Kingdom of Bahrain by investigating socio-economic characteristics such as gender, type, age, nationality, and educational background; the crashes were categorized according to these characteristics then compared to their exposure risk. They found that these characteristics have a significant influence on traffic crashes involving pedestrians. Bener et al. (2009) investigated the severity of head and neck injuries in victims of road traffic crashes in Qatar, using a total of 6,709 patients during the period 2001-2006. This study confirmed that road traffic crashes were a major cause of head and neck injuries, and found that the majority of the victims were non-Qataris. They showed that male pedestrians have more exposure risk to crashes than female, and that in terms of age groups the most vulnerable were the young and the old; in terms of nationality, non-locals have higher exposure risk than locals, and educated pedestrians were less likely to be involved in crashes. They indicated that the personal and social backgrounds of pedestrians have a great influence on their exposure risk in terms of being involved in crashes (Figure 2.7).

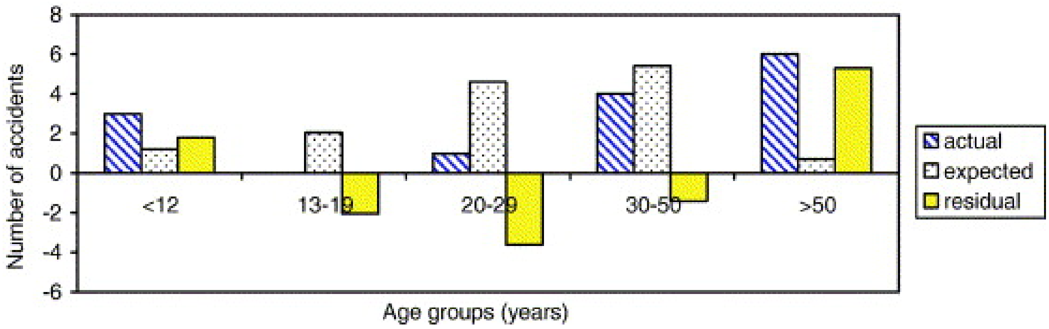

Figure 2-7 Pedestrians' fatal crashes by age group

Koushki and Bustan (2006) looked at smoking, seat belt use, and road crashes among youth in Kuwait by using a questionnaire survey of 1467 randomly selected young drivers in Kuwait. They examined the interrelationships between socio-economic factors, belt use, smoking behaviour while driving, and road traffic crashes, and determined the degrees of association between these variables. They found that young female drivers were safer drivers than young males, and that those who were smoking while driving and frequently used a seat belt had a higher involvement rate in road traffic crashes. It was also indicated that young drivers who smoked were more than twice as numerous as the non-smokers. The results showed no major differences in socio-economic factors between males and females. The young males experienced more than eight times as many injuries in road crashes as females. When the sample was classified by nationality the differences in the socio-economic and road crashes of the young adults were also large; the study demonstrated that less than one-fifth of the sample males but the majority of the sample females wore seat belts.

2.3.2.3 Findings from Saudi Arabia

Saudi Arabia has lower crash rates but higher casualty and fatality rates than Kuwait (Ofosu et al., 1988). Shanks et al. (1994) reviewed road traffic accidents (RTA) over a one-year period in Saudi Arabia and found that out of 361 victims 16% were less than 10 years old and 47% between 11 and 30 years old. None of those involved in crashes was wearing a seat belt. Half of the children injured were pedestrians, the male-to-female ratio was 4:1 because of the driving laws in Saudi Arabia, and tyres exploding because of the heat were identified as a common cause (39%) of crashes. RTAs are a major health hazard in Saudi Arabia, particularly during the holy month of Ramadan when all people fast from sunrise to sunset. Consequently trauma increased in direct proportion to the increase in the number of road vehicles.

Nofal and Saeed (1997) studied the seasonal variations and weather effects on road traffic crashes in Riyadh city which were recorded on rainy days for the period 1989–1993, looking at the time of day and lighting conditions. They found crashes increased with the increase of dry and wet bulb temperatures but decreased with the increase of humidity and the amount of precipitation of rain, snow and hail, which was not surprising because snow and hail are unusual in Riyadh; also for humidity Riyadh is not near the sea. It was found that the seasonal variation in RTAs was greatest in Riyadh during the summer, mostly between 12 noon and 3pm, when the traffic is heavy and the strong heat increases stress and decreases performance. They considered the extended exposure to heat to be a hazard to the safety and health of drivers and a factor leading to an increased incidence of RTAs.

Ansari et al. (2000) recommended the compulsory use of safety seat belts in vehicles, and setting up a database to collect, store and analyse information relating to road traffic crashes in Saudi Arabia, because of the high rate of fatalities and injuries due to traffic crashes. Al-Ghamdi (2002) used a sample of 560 subjects involved in serious crashes reported in traffic police records and occurring on urban roads in Riyadh to estimate the effect of statistically significant factors on crash severity. Unfortunately, police reports did not describe injuries in much detail at crash sites because of the lack of police qualifications and training. Police records categorised crashes into three classes: property damage only (PDO), injury (no injury classification was available), and fatal crashes. Such a categorisation makes it difficult to examine the factors affecting crash severity properly.

Al-Ghamdi and AlGadhi (2004) studied warning signs as countermeasures to camel–vehicle collisions in Saudi Arabia; they used seven camel-crossing warning signs to determine if they would reduce the number of camel–vehicle collisions on rural roads. Although most of the signs showed statistically significant reductions in mean speed, the speed reductions were quite small, ranging from 3 to 7 km/h. Furthermore, statistical analysis was used to rank the signs according to their effectiveness, and it was found that the influence of the warning sign on the driver starts nearly 500 m ahead of the sign and diminishes at 500 m downstream of the sign.

Similarly, Bendak (2005) studied seat belt use in Saudi Arabia and its impact on the number of road crash injuries during the first few months following the performing of the law in 1998 by using a questionnaire to investigate drivers' behaviour and characteristics and their relationship with the use of seat belts (Figure 2.8).

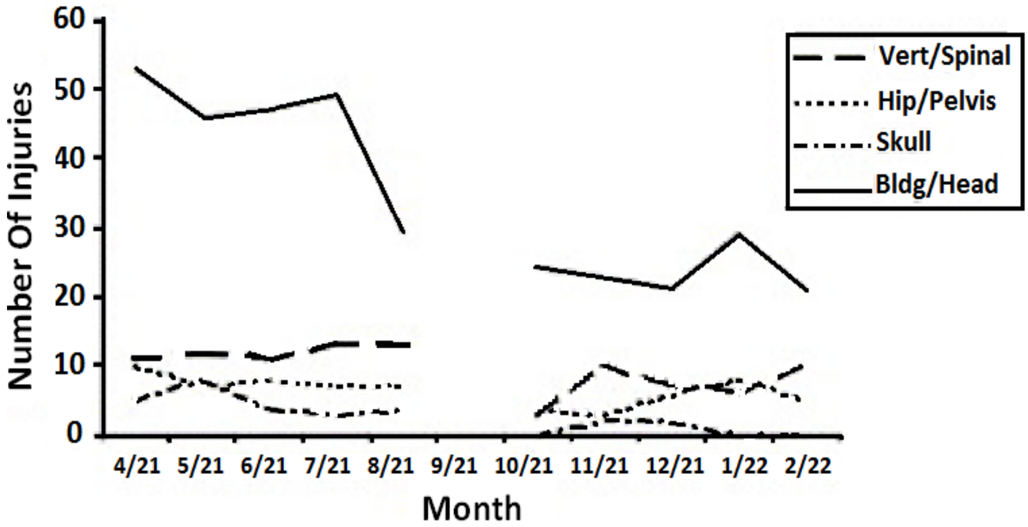

Figure 2-8 Number of injuries (four injury types) plotted over months (Bendak, 2005)

Results showed that there was a significant drop in injuries due to traffic crashes after implementation of the seat belt law. He also showed that personal characteristics were correlated with seat belt use. An average seat belt use rate of 60% for drivers and 22.7% for front seat passengers (FSPs) was recorded in Riyadh suburbs during the first few months after the seat belt laws were enacted. These rates are considered low when compared with rates in developed countries and compared with Kuwait, where the rate was almost 100% for drivers and dropped to 50% for Kuwaiti drivers and 65% for non-Kuwaiti drivers after the use of seat belts was enforced in 1994 (Koushki and Al-Anazi, 1998).

Therefore, in developing countries, a public campaign aiming to increase public awareness of the importance of using seat belts is very important, especially when a drop in injuries has been recorded in the above studies, which is encouraging.

Generally, the unavailability of suitable and detailed data for crash analysis is a common problem in most developing countries (Jacobs and Sayer, 1983; Mekky, 1984; Al-Saif, 1997); the previous researchers in traffic crash analysis who attempted to estimate the factors influencing crashes in Riyadh city used inaccurate, imprecise, and unreliable data, which was a critical problem in the country (Al-Ghamdi, 2002; Al-Ghamdi, 2004).

In conclusion, it is clear that developing countries need to give more attention to traffic safety, and further research is needed in crash analysis and identification of the significant influencing factors and causes in order to develop safety policies for reducing traffic crashes, giving attention to the privacy of some developing countries like the Kingdom of Saudi Arabia (KSA), which will be the study area of this research. For example, a high number of registered vehicles and almost one quarter of the population come from different nationalities and different cultures and backgrounds; this might deteriorate the road safety problem. On the other hand, drugs and alcohol have a significant effect on road safety, while in KSA it is prohibited by the Islamic law to sell or consume drugs or alcohol, and the penalty for doing so is very severe. In addition, females are not allowed to drive in KSA, and as a result male drivers are hired from abroad, which may make road safety worse in the country.

So, improving road safety is a crucial aim of transport policy. Therefore, crash analyses need to consider all factors related to traffic safety, such as the geometry of roads and their infrastructure, the number and width of lanes, shoulders, curvature, safety signs, and land use; aspects of vehicle design, including speed and the installation of air bags and anti-lock braking systems (ABS); and, in terms of road users, their behaviour, including seat belt use and alcohol consumption, age, and demographic factors. After the review of studies on road traffic crashes conducted in developing countries, it was noticeable that driver demographic factors and driver behaviour were discussed and examined in developing countries more than other factors such as traffic, land-use, and environmental factors. Driver behaviour is a complex phenomenon which largely depends on culture, habits, and attitudes of people in different countries. Therefore, the impact of driver behaviour on road safety would be different in developing countries. A literature review and discussion of statistical methods used to identify contributing factors in crashes will be provided in the next chapter.

3 REVIEW OF STATISTICAL METHODS USED IN CRASH MODELLING

3.1 Introduction

This chapter describes previous studies on the development of traffic crash prediction models. As discussed in Chapter 2, there are a wide range of methods used in crash analysis to identify factors affecting road crashes. These methods include econometric approaches and empirical evidence. Statistical methods used to analyse crashes were conducted at both an aggregate level, addressing the frequency of crashes, and a disaggregate level, addressing the severity of crashes. Econometric models employed to analyse crash severity are different from those used for analysing crash frequency, and models used for both can be used to establish a relationship between crashes and their contributing factors.

All the literature suggests that two key types of crash analysis are conducted to identify factors affecting road traffic crashes in developed countries. These are: (1) *counts of crashes for a specific entity,* in which the crash frequency (annual) for that entity, such as a road segment or a junction or an area, is considered unit of analysis, and factors affecting the frequency of crashes are then identified using crash prediction models **(**Donnell and Mason, 2006; Kim et al., 2007; Quddus, 2008); and (2) *disaggregate level analysis,* in which each individual crash is considered as an unit of analysis, and factors affecting individual crashes are then determined **(**Milton et al., 2008**).** In the second category of analysis, it is also possible to determine factors affecting the severity of crashes.

This Chapter reviews statistical methods that have been employed to develop crash prediction models for modelling both crash severity and crash frequency. This is followed by a review of existing studies which are mainly focused on modelling crash frequency, non-spatial models, road-segment level, intersection level, area-wide level, and spatial models. Finally, the conclusion of this Chapter is provided.

3.2 Modelling Crash Severity

A wide variety of statistical techniques have been applied to analyse crash-severity data. Over the years, researchers have used a wide range of methodological tools to assess the impact of traffic, driver, vehicle, road, environment, land use and socio-economic factors on disaggregate-level injury-severity data. Advanced methods have enabled the development of sophisticated models capable to determine precisely the influence of these factors. The following section will demonstrate previous research into crash severity, focusing on statistical methods employed in the analysis.

The literature suggests that the factors affecting the frequency of traffic crashes are different from those affecting the severity of traffic crashes (Kweon, 2003). The modelling techniques used to identify the factors affecting the severity of traffic crashes are also different from the modelling techniques used in the analyses of crash frequency. While modelling crash severity, one needs to consider the characteristics of individual traffic crashes, such as the category of crash in terms of severity (whether fatal, a serious or a slight injury crash), and the characteristics of the driver involved in the crash (e.g. age, gender, experience). Therefore, such studies can be regarded as disaggregated level analysis.

A disaggregate level analysis studies of crash severity, which has a special interest for researchers in traffic safety as they are not only concerned with prevention of crashes but also with reducing their severity. Crash severity can be classified as fatal, serious injury, slight injury, and property damage only. Logistic and probit models are suitable models for categorical data. There are some interesting findings in terms of modelling crash severity; for example, Mercier et al. (1997), Al-Ghamdi (2002), and Yau (2004) used logistic regression models to examine crash severity. Mercier et al. (1997) used a logistic regression model to determine whether either age or gender (or both) was a factor influencing severity of injuries suffered in head-on automobile collisions on rural highways.

A binary logistic regression model is suitable for modelling binary outcomes, such as fatal and non-fatal (e.g. Pitt et al., 1990; Al-Ghamdi, 2002; Hamitt, 2003). For example, Al-Ghamdi (2002) used a sample of 560 subjects involved in serious crashes reported in traffic police records and occurring on urban roads in Riyadh city. Logistic regression was used to estimate the effect of the statistically significant factors on crash severity. Crash severity (the dependent variable) with two categories of injury (non-fatal and fatal) crash records was considered. Out of nine independent variables, the two found to be most significant for crash severity were location and cause of crash. Al-Ghamdi found that a logistic regression was a promising tool in traffic safety improvements in Riyadh, providing a reasonable statistical fit. Unfortunately, the police reports did not describe injuries in much detail at crash sites because of the lack of police qualifications and training; police records showed that the crash was a property damage only (PDO) crash, injury crash (no injury classification was available), or fatal crash. Also, medical reports were hard to obtain because police crash data and medical data were not kept together. So, it was impossible to obtain details on the degree of severity of the crashes in Riyadh.

Nassar et al. (1997) developed an integrated accident risk model (ARM) in Ontario using factors affecting both crash occurrences on road sections and severity of injury to occupants involved in the crashes; using negative binomial regression and a sequential binary logit formulation, they developed models that are practical and easy to use.

Yau (2004) used stepwise logistic regression models for three types of vehicles, private, goods and motorcycles, during the 2-year period 1999–2000 to examine factors affecting the severity of single-vehicle traffic crashes in Hong Kong.

More advanced models have been used; two types of such models that have been proposed are: (1) ordered response models (ORM), such as ordered logit and probit; and (2) unordered nominal response models, such as multinomial, nested and mixed logit models.

Previous studies which have employed such models are discussed below.

Because crash severity is ordered in nature (ranging from non-injury to fatality), ordered response models are a suitable choice for analysing crash severity data. The ordered response models (ORM) include ordered logit and probit models and their various extensions, such as a generalised ordered logit model. For example, O'Donnell and Connor (1996) used an ordered logit and probit model to investigate how attributes of road users affect crash severity. According to O'Donnell and Connor (1996) and Abdel-Aty (2003), both logit and probit ordered models produce very similar results.

Many researchers have used the ordered probit model to examine crash severity (e.g. Kweon and Kockelman, 2003; Quddus et al., 2002; Zajac and Ivan, 2003; Abdel-Aty, 2003; Quddus et al, 2009).

Quddus et al (2002) used an ordered probit model to examine factors that affect the injury severity of motorcycle crashes and the severity of damage to the vehicle in Singapore. They found that factors which lead to increases in the probability of severe injuries include the motorcyclist having non-Singaporean nationality, increased engine capacity, headlights not turned on during daytime, collisions with pedestrians and stationary objects, driving during early morning hours, having a pillion passenger, and when the motorcyclist is determined to be at fault for the crash. In addition, they found that both injury severity and vehicle damage severity levels are decreasing over time. Pai and Saleh (2007) employed ordered probit models to investigate the factors that contribute to more severe motorcyclist injury severity at three-legged junctions in the UK. The modelling results uncover several combined factors that were deadly to motorcyclists: for example, injuries tended to be much more severe while motorcyclists involving in approach-turn collisions at signalised junctions than at un-signalised junctions.

The ordered logit and probit models have a primary assumption which is that the error variances are the same for all observations. To correct this assumption, Williams (2008) employed a heterogeneous choice model which relaxes the assumption by specifying the determinants of heteroskedasticity. A heteroskedastic model was used to investigate injury severity of age and pedestrians in motor-vehicle crashes in North Carolina, USA, using police-reported crash data between 1997 and 2000 (Kim et al., 2008). It was found that the age of pedestrians was significant to heteroskedasticity (random variables have different variances), and that the age-specific heteroskedasticity became important after the age of 65; darkness was also found to be associated with greater severity.

According to Long and Freese (2006), another important assumption associated with ordered logit and probit models is that the relationship between each pair of outcome groups is the same, which means, for example, the ordered model assumes that the slope coefficients for slight injury vs serious injury plus fatality are equal to the slope coefficients for serious injury vs fatality. This assumption is known as the parallel regression assumption in the literature, or, for the ordered logit model, as the proportional odds assumption. This assumption is frequently violated and can lead to inappropriate or misleading model estimation results (Long and Freese, 2006; Fu, 1998). To relax the parallel regression assumption on slope coefficients, a generalised ordered logit model can be used (Fu, 1998). This model allows the coefficients to vary across different outcome groups. Similarly, a partial proportional odds model has been proposed (Williams, 2006), to constrain only a subset of coefficients across different outcome groups. This model has been used by Wang and Abdel-Aty (2008) and by Quddus et al. (2009).

Srinivasan (2002) allows for random coefficients that are capable of capturing observation-specific differences in the effects of covariates on injury severity. Srinivasan showed the ordered mixed logit to provid superior fit to the traditional ordered logit model.

Research has been undertaken by Eluru and Bhat (2007), who employed a random coefficient (i.e. mixed) ordered logit model in order to allow randomness in the effects of independent variables which were due to unobserved factors. Later, Eluru et al. (2008) extended this work by developing a mixed generalised ordered response (MGORL) model to assess pedestrian and bicyclist injury severity level in traffic crashes. This model allows thresholds to vary according to both observed and unobserved factors. Their results suggested that the MGORL model became a generalised ordered logit model because no significant heterogeneity effects were found.

However, ordered response models (ORM) have been criticised. ORMs assume that the effects of a variable would either increase or decrease crash severity; for example, the likelihood of severity levels increases from “slight injury” to “serious injury” but does not increase to “fatality”. This problem may be solved by using a generalised ordered response model (GORM) which allows the coefficients to vary across different levels of severity. Another problem associated with ordered response models is related to the fact that crash data often suffer from the under-reporting problem, especially for lower categories of severity, such as “no injury” and “slight injury”. Hauer and Hakkert (1989) compared police-, hospital-, and insurance-reported injury data and concluded that approximately 20 percent of severe injuries, 50 percent of minor injuries, and up to 60 percent of no-injury crashes were not reported. Elvik and Myssen (1999) found underreporting rates of 30, 75, and 90 percent for serious, slight, and very slight injuries, respectively. A technical report by the National Highway Traffic Safety Administration (2009) estimated that 25 percent of minor injury crashes and half of no-injury crashes are unreported**.**

As a result of under-reporting, the data will be over-presented, which can lead to biased and inconsistent results when traditional ordered response models are used (Yamamoto et al., 2008). To tackle this issue an (unordered) nominal response model was proposed, such as a multinomial logit (MNL) model which does not distinguish ordinality in the model. This model is more flexible as the independent variables are not assumed to be identical across all degrees of severity in the model. Therefore, an MNL model allows different severity categories to be associated with different sets of independent variables (Shankar and Mannering, 1996; Carson and Mannering, 2001; Ulfarsson and Mannering, 2004; Khorashadi et al., 2005; Kim et al., 2007; Yamamoto et al., 2008).

The problem of an MNL model is that this model assumes that the unobserved effects associated with each crash severity category are independent; this assumption is referred to as the independence of irrelevant alternatives (IIA) property (Long and Freese, 2006). If the IIA assumption is violated, the model estimation results would be incorrect. To avoid this limitation, a nested logit (NL) model has been used (Shankar et al., 1996; Chang and Mannering, 1999; Lee and Mannering, 2002; Abdel-Aty and Abdelwahab, 2004). The NL model groups severity outcomes with shared unobserved effects in a nest; for example, “fatal” and “serious injury” could be grouped to form a high severity crash “nest”. For example, Shankar (2000) used a full-information maximum likelihood nested logit model (FIML-NL), which was helpful for severity modelling.

Nassar et al. (1997), Shankar (2000), Holdridge et al. (2005), and Milton et al. (2008) used nested logit models to study the severity of crashes. Holdridge et al. (2005)

analysed the injury severity in collisions with fixed roadside objects in Washington State, USA; by using multivariate nested logit models with full information maximum likelihood, they showed that leading ends of guardrails and bridge rails increased the probability of fatal injury, The face of guardrails is associated with a reduction in the probability of injury, and concrete barriers are shown to be associated with a higher probability of lower severities. Rock banks increase the propensity toward evident injury.

A more powerful approach with a great deal of flexibility has been proposed, which is the mixed logit model. The mixed logit model can be used to accommodate complex patterns of correlation between crash severity outcomes and unobserved heterogeneity (McFadden and Train, 2000). This model has been employed by Milton et al. (2008) who attempted an explanation of how multiple factors affect injury-severity distributions using a mixed (random parameters) logit model and data from Washington State; 22,568 crashes were combined with weather data which were used for geometric, pavement, roadside and traffic characteristics associated with roadway segments using highway-injury. They found that variables related to traffic volume and weather effects were best modelled as random variables, while road variables were best modelled as fixed parameters. They demonstrated a modelling approach that can be used to achieve a better understanding of the injury-severity distributions of crashes on highway segments and the effect of traffic volume and weather on these distributions. They also showed that the mixed logit model was a significant tool in traffic safety.

So, it is important to consider the possibility that the influence of variables affecting the number of crash injury-severity could be different along different segments because of variations in driver behaviour. However, the approach they used allows for the possibility that estimated model parameters can vary randomly across roadway segments to account for unobserved effects which potentially relate to roadway characteristics, environmental factors, and driver behaviour. Savolainen et al. (2011) concluded that models that do not account for the ordinal nature of injury data have also been popular in the analysis of crash injury-severity data. They also suggested that the appropriate methodological approach can often depend heavily on the available dataset, including the number of observations, quantity and quality of explanatory variables, and other data-specific characteristics.

3.3 Modelling Crash Frequency

In the eighties and early nineties, multiple linear regressions were used in modelling the frequency of crashes observed over a specific entity (e.g. road segments or junctions) for a specific period of time. Crash prediction was first modelled by using a linear regression model (for example, Jovanis and Chang, 1986; Joshua and Garber, 1990; Miaou and Lum, 1993), but this model was appropriate for continuous data, and because crashes are by their nature countable, discrete, random events, and an integer cannot be a negative value, this model is not suitable for crash data (Jovanis and Chang, 1986; Zeeger et al., 1990; Miaou and Lum, 1993). Then a Poisson regression model was used by assuming a Poisson distribution to establish a statistical relationship between traffic crashes and various factors which contribute to crash occurrence, but in order to use this model the mean and the variance must be equal (Jones et al., 1991; Miaou et al., 1992; Kulmala, 1994; Shankar et al., 1995). Shankar et al. (1995) and Noland and Quddus (2004) have argued that vehicle crashes were better represented by a NB distribution. This caused a problem of over-dispersion and under-dispersion in the crash data (Miaou, 1994; Shankar et al., 1995; Vogt and Bared, 1998). A negative binomial (NB) model was then established to solve the problem of over- and under-dispersion, and a test called the heterogeneity test was developed to check whether it is over- or under-dispersing (Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000; Lord, 2000; Ivan et al., 2000).

There are constraints and limitations with the NB model as it ignores spatial correlation and spatial heterogeneity. Therefore, researchers have developed have spatial econometrics models based on a Bayesian framework by employing conditional autoregressive (CAR) models (e.g. Sun et al., 2000; MacNab and Dean, 2001; Liu, 2008; Quddus et al., 2012).

This aggregate level of analysis models the frequency of crashes by developing a relationship between counts of crashes in regions, districts, or wards and the contributing factors to these crashes. These factors are usually land-use, socio-economic, traffic, driver, and road characteristics, and environmental factors. This method uses two types of models: (1) non-spatial models and (2) spatial models. Crashes are usually located at points measured with X and Y coordinates or represented by Eastings and Northings (E, N). The non-spatial models ignore spatial correlation and heterogeneity, while the spatial models which are relatively new take them into account.

3.3.1 Non-spatial models

This section will discuss different levels: (A) road-segment level; (B) intersection level; and (C) area-wide level. Each level uses a different set of data. Generally, using the appropriate statistical model depends on the types of data available, which means different types of data will be modelled by different statistical models. There are mainly three types of data: time-series, cross-sectional, and panel; panel data combine both time-series and cross-sectional data. Gujarati (1995) suggested that an autoregressive integrated moving average (ARIMA) model can analyse the time series data, while Poisson or NB models can be used to analyse the cross-sectional dataset; on the other hand, the panel dataset can be modelled using either a fixed effects Poisson or NB (FENB) model or a random effects Poisson or NB (RENB) model (Shankar et al., 1998; Noland and Oh, 2004). In traffic crash analyses, there are many unobserved explanatory variables that affect frequencies and severities.

To address the heterogeneity, many recent studies have used (univariate) panel count data models, such as random-effect negative binomial (RENB) and fixed-effect negative binomial (FENB) regression models (Kweon and Kockelman, 2003; Karlaftis and Rarko, 1998; Shankar et al., 1998; Chin and Quddus, 2003); to test the dataset whether it is a fixed or random effects a Hausman specification test was used. In some cases data contain many zeros; the literature shows that zero-inflated Poisson (ZIP) and zero-inflated NB (ZINB) models were used for cross-sectional data (Shankar et al., 1997; Chin and Quddus, 2003).

According to the road safety modelling literature, non-spatial models are found to apply at various spatial units, such as road segments (Shankar et al., 1995; Miaou and Lum, 1993; Miaou, 1994; Ivan et al., 1999; Donnell and Mason, 2006); junctions or intersections (Kulmala, 1994; Poch and Mannering, 1996; Vogt and Bared, 1998; Chin and Quddus, 2003; Kim et al., 2006; Kim et al., 2007); and areas (Levine et al., 1995b; Barker et al., 1999; Baum, 1999; Washington et al., 1999; Miaou et al., 2003; Amoros et al., 2003; Noland and Oh, 2004; Noland and Quddus, 2004; Noland and Quddus, 2005; Aguero-Valverde and Jovanis, 2006; Kim et al., 2006; Haynes et al., 2007; Haynes et al., 2008; Quddus, 2008).

In most cases, the dependent variable is the annual frequency of crashes, and the explanatory variables are traffic characteristics (i.e. traffic flow, volume, and traffic density); road characteristics (i.e. types of roadways, curvature, and gradient); and land-use and environmental factors. The dataset used in most studies is cross-sectional. However, some of the studies also employ panel data (Karlaftis and Rarko, 1998; Kweon and Kockelman, 2005), for which a Poisson or an NB model is not suitable. In order to take into account unobserved heterogeneity among spatial units, either a random-effect or a fixed-effect Poisson or NB model is used (Shankar et al., 1998; Noland and Oh, 2004; Chin and Quddus, 2003).

For instance, many recent studies have used random-effect negative binomial (RENB) and fixed-effect negative binomial (FENB) regression models to develop a relationship in the case of panel data (Karlaftis and Rarko, 1998; Shankar et al., 1998; Chin and Quddus, 2003; Kweon and Kockelman, 2005). Some of these studies use a Hausman specification test to determine whether a random-effect model is more suitable than a fixed-effect model or vice versa (crash studies use such a method). The Poisson and NB models discussed above are non-spatial models and such models do not account for spatial dependence among spatial units (Quddus, 2008). However, crashes are usually located at points measured with X and Y coordinates or represented as eastings and northings (E, N). The non-spatial models ignore the spatial effect known as spatial correlation (Miaou, 2003; Song, 2004; Aguero-Valverde and Jovanis, 2006; Quddus, 2008).

In some cases, cross-sectional crash data contain a lot of zero observations. In order to deal with excess zero observations, another class of models are used. These are known as zero-inflated Poisson (ZIP) or zero-inflated NB (ZINB) models (Shankar et al., 1997; Chin and Quddus, 2003).

However, Lord et al. (2005) argued that excess zero observations in crash data are due to small spatial entities and short temporal units rather than an underlying mechanism that generates both zero and non-zero counts. Their study warns researchers not to use ZIP and ZINB models in crash modelling. Despite this, there are a number of recent studies on the use of such models in crash analyses (Li et al., 2008).
The following sections will examine levels of analysis, including the statistical models in each level of analysis, with different types of datasets. Some of the existing studies using non-spatial models are presented below.

3.3.1.1 Road-segment level

Several other researchers have advocated the use of either Poisson or Negative Binomial regression models to predict crash frequency as a function of geometric design variables, in addition to the median-related accident models described previously. For instance, Miaou and Lum (1993) demonstrated how the Poisson regression model can be used to estimate the effects of motorway geometric design on truck crash involvement. Later, Miaou (1994) investigated the performance of the negative binomial and zero-inflated Poisson regression models when estimating truck crash involvement on interstate highways in Utah, USA.

Shankar et al. (1995) investigated the effect of roadway geometry and environmental factors on rural freeway crash frequencies. An NB model was used to develop models relating road geometry (i.e. length, width, number of lanes) and environmental factors with the frequency of crashes. They discovered a relationship between the weather and geometric elements which showed that the weather had only a little effect on motorway sections which are characterised by risky geometrics such as sharp roads.

In another study, Ivan et al. (1999) investigated differences in causality factors for single- and multi-vehicle accidents on two-lane roads, including the effects of site characteristics, using Poisson regression models for predicting single- and multi-vehicle crashes separately. Their analysis focused on rural two-lane freeways with hourly level of service (LOS), traffic composition, and highway geometric characteristics as independent variables. They concluded that single-vehicle crash rates decrease with the increase of traffic intensity, shoulder width and sight distance, while multi-vehicle crash rates increase with the increase in the number of signals, the daily single-unit truck percentage, and shoulder width in rural roads.

Zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression models have also been applied in analysing crash frequency at road segment level (see Shankar et al., 1997; Garber and Wu, 2001; Lee and Mannering, 2002; Kumara and Chin, 2003; Miaou and Lord, 2003; Rodriguez et al., 2003; Lord et al., 2005).

A study by Donnell and Mason (2006) predicted the frequency of median barrier crashes on Pennsylvania interstate highways using an NB regression model. Traffic volume, horizontal alignment, interchange ramp presence, and median barrier offset distance from the travel lanes were used as explanatory variables. They found that all the above factors influenced median barrier crash frequency.

There are limitations with this model. Because crash data are referenced with locations measured as X and Y coordinates, two problems occur (LeSage, 1998). The first problem is the existence of spatial correlation in observations among neighbourhood areas, and the second limitation is spatial heterogeneity. NB models used in crashes ignore the spatial correlation and assume observations are independent so the result might be biased.

3.3.1.2 Intersection level

There are a number of studies that have used Poisson or NB models to develop a relationship between annual crash frequency at junctions and junctions' characteristics (e.g. Poch and Mannering, 1996; Vogt and Bared, 1998; Chin and Quddus, 2003; Kim et al., 2006; Kim et al., 2007).

For example, Chin and Quddus (2003) examined traffic crash occurrence at signalised intersections in the South-Western part of Singapore, using a total of 52 four-legged intersections with panel crash data from 1992 to 1999. A total of 832 observations were used in the analysis. There were 3000 crashes in their data (3% fatal, 5% serious injuries and 82% slight injuries). Their study used random-effect negative binomial (RENB) models which could deal with the spatial and temporal effects in the data to investigate the relationship between crash occurrence and the geometric, traffic and control characteristics of signalised intersections. They found that eleven variables significantly affected safety at the intersections and confirmed that the RENB model was appropriate.

Kim et al. (2007) examined the relationships between roadway, environmental, and traffic factors and motor vehicle crashes using NB and Poisson regression models. Data on 548 motor vehicle crashes were collected from 91 two-lane rural intersections in the state of Georgia. Crash prediction models were estimated for angle, rear-end, and sideswipe crashes. They employed the hierarchical data structure and a suitable method of analysis approach for multilevel data into intersection-related crashes by crash type. They found that crashes at rural intersections in Georgia were hierarchical in structure and the modelling indicated that the effects of geometric characteristics and environmental factors can be modelled using multilevel modelling techniques. They examined the random effects and suggested that there was a significant variation in the probability of specific types of crashes occurring at intersections.

3.3.1.3 Area-wide level

There is a lack of literature on traffic crash analyses at an area-wide level, which might be due to the fact that area-wide data (e.g. an area-wide measure of speeds or road density) were more difficult to collect, as one needs to integrate data from different sources (i.e. different government agencies). With recent development in GIS software (e.g. ArcGIS, MapInfo) such area-wide data are becoming more available, and, as a result, area-wide analyses of traffic crashes are also becoming more frequent.

There are a number of studies based on an aggregated area-wide level (e.g. Barker et al., 1999; Baum, 1999; Washington et al., 1999; Amoros et al., 2003; Noland and Oh, 2004; Noland and Quddus, 2004; Noland and Quddus, 2005; Kim et al., 2006; Haynes et al., 2007; Haynes et al., 2008).

For example, Baum (1999) found that there was a strong correlation between rates of drink-driving offenders and particular social and demographic factors at the aggregate level in Australia.

Later, Noland and Quddus (2004) used NB models at ward (a census tract) level to analyse the association between area-wide factors of road casualties and traffic fatalities, serious injuries, and slight injuries. This was carried out by using 8414 wards in England which were put into a geographic information system using four different categories of predictor variables, which are land-use indicator variables, road characteristics, demographic characteristics (age cohorts), and traffic flow proxies. They found that urbanised areas have fewer traffic casualties while areas with high employment have more traffic casualties. A further study by Noland and Oh (2004) used country-level data for Illinois, USA, to estimate the expected number of crashes, using infrastructure characteristics and demographic indicators as independent variables, and using a fixed-effect NB (FENB) model. They found that increased lane widths and the number of lanes were associated with increased fatalities, and that crashes were decreased by increasing the shoulder width, but changes in the use of seat belts and alcohol consumption were not considered. It is clear from this study that time trend was important in the model for this type of analysis.

In another example, Hadayeghi et al. (2006) conducted an area-wide traffic zone crash analysis of the city of Toronto, Canada, to examine the temporal transferability of crash prediction models, in which various factors such as Volume/Capacity (V/C) ratio, population, and road length were considered. In their study, NB models were used to examine the effects of these factors on the crashes that occurred from 1996 to 2001. In this study, spatial dependence or spatial correlation was not considered, which one of its limitations is. A study by Delmelle and Thill (2008) used an ordinary least squares (OLS) regression analysis to investigate young and adult bicycle crashes in the city of Buffalo, NY. In terms of methods, there is also a limitation of spatial autocorrelation in this study, as well as no consideration of count nature of the crash data.

Wang et al. (2009) used 8019 census wards in England as spatial units of analysis to investigate the effect of speed and road curvature on ward-level traffic casualties by employing NB models. Different types of crash frequencies were considered, such as the frequency of fatal and serious injury crashes and the frequency of slight injury crashes. Crashes were also disaggregated according to categories of road users such as motorised transport (MT), non-motorised transport (NMT), and vulnerable road users (VRUs). They found that average road speed was positively associated with total fatalities and MT serious injuries within a ward while it was negatively associated with seriously and slightly injured casualties among NMT and VRUs.

On the other hand, road curvature was negatively associated with casualties. However, because of the lack of data, they could not examine other contributing factors associated with area characteristics such as weather conditions, signalised/un-signalised junctions, and roundabouts, and socioeconomic factors. One of the limitations in terms of methods in this study is the lack of consideration of spatial dependence. While there is a large body of literature on the analysis of traffic crash occurrences at road segment or junction levels, there is a lack of literature on traffic crash analysis at an area-wide level. This might be due to the fact that area-wide data (e.g. speeds or road density) are more difficult to collect as one needs to integrate data from different sources (i.e. different government agencies).

3.4 Summary

This chapter has demonstrated series of ordered and nominal response models which have been used to analyse and identify the factors affecting the severity of traffic crashes, such as an ordered response model and nominal response model. The ordered response model includes ordered logit/probit models; generalised ordered logit (GOLOGIT) model; partially constrained generalised ordered logit (PC-GOLOGIT) model; the nominal response model includes multinomial logit (MNL) model; nested logit model; mixed logit (ML) model, and mixed binary logistic regression model. As discussed in this chapter, nominal response models such as a mixed logit model may be preferred to an ordered response models. Both ordered and nominal response models will be tested in this thesis. A summary of various crash severity models, their advantages and the previous researchers who examined or used these models is presented in Table 3.1.

Table 3-1 Summary of the crash severity models reviewed in this chapter (Savolainen et al. 2011).

Methodological approach	Key features	Reported by the following researchers
Logistic regression model	Suitable for categorical data	Mercier., 1997; Al-Ghamdi, 2002; Yau, 2004
Binary logistic regression model	Suitable for binary outcomes	Pitt et al., 1990; Al-Ghamdi, 2002; Toy and Hammitt, 2003
Ordered logit ordered probit	Suitable for ordinal outcomes	O'Donnell and Connor, 1996; Kockelman and Kweon, 2002; Quddus et al., 2002; Zajac and Ivan, 2003; Abdel-Aty, 2003; Quddus et al., 2009
Heterogeneous choice model	Relaxes the assumption of the error variances are the same for all observations imposed by ordered logit/probit models	Williams, 2008; Quddus et al., 2009
Generalised ordered logit model	Relaxes the parallel regression assumption to allow the coefficients to vary across different outcome groups	Fu, 1998
Partial proportional odds model	A subset of coefficients across different outcome groups are constrained to be the same	Williams, 2006; Wang and Abdel-Aty, 2008; Quddus et al., 2009
Random coefficient (i.e. mixed) ordered logit model	Allow randomness in the effects of explanatory variables due to unobserved factors	Eluru and Bhat, 2007
Mixed generalised ordered response model	Thresholds are allowed to vary according to both observed and unobserved factors; the model can accommodate heterogeneity in both explanatory variables and threshold values	Eluru et al., 2008
Multinomial logit (MNL) model	Suitable for nominal outcomes. More flexible functional form and consistent coefficient estimates except constant terms when under-reporting occurred in the data (compared to ordered response models)	Shankar and Mannering, 1996; Carson and Mannering, 2001; Ulfarsson and Mannering, 2004; Khorashadi et al., 2005; Kim et al., 2007
Nested logit model	Relaxes the assumption of independence of irrelevant alternatives (IIA) by MNL models	Shankar et al., 1996; Nassar, 1997; Chang and Mannering, 1999; Shankar et al., 2000; Lee and Mannering, 2002; Abdel-Aty and Abdelwahab, 2004; ; Holdridge et al. 2005; 2007, Milton 2008
Mixed logit model	Can accommodate complex patterns of correlation among crash severity outcomes and unobserved heterogeneity	McFadden and Train, 2000; Milton et al., 2008

As for modelling crash frequency, models are primarily used for examining the factors affecting the frequency of traffic crashes. The crash data will be disaggregated into different categories by their severity levels (fatal and serious injury crashes), and each category of crash is modelled separately. It has been found that models used to develop a relationship between frequency of crashes and their contributing factors include, Poisson, Negative Binomial (NB), fixed or random effect NB, zero-inflated Poisson (ZIP) and zero-inflated NB (ZINB).

Table 3-2 Summary of the crash frequency models reviewed in this chapter (Lord and Mannering, 2010).

Methodological approach	Key features	Reported by the following researchers
Linear regression model	Suitable for continuous data	Jovanis and Chang, 1986; Joshua and Garber, 1990; Miaou and Lum, 1993
Poisson model	Most basic model; easy to estimate and cannot handle over- and under-dispersion. Suitable for non-negative count data; the mean is equal to the variance	Jovanis and Chang, 1986; Joshua and Garber, 1990; Jones et al., 1991; Miaou et al., 1992; Miaou and Lum, 1993; Kulmala, 1994, Shanker, 1995; Ivan et al., 1999
Negative Binomial (NB) model	Easy to estimate and can account for over-dispersion but cannot handle under-dispersion	Miaou, 1994; poch and and Mannering, 1996; Shankar et al, 1995; Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000; Lord, 2000, Ivan et al., 2000; Amoros et al. 2003; Graham and Glaister, 2003; Noland and Quddus, 2004; Noland and Quddus, 2005; Donnell and Mason, 2006; Hadayeghi et al, 2006; Wang et al., 2009
Fixed- or random-effects Poisson/NB model	Suitable for panel data	Karlaftis and Rarko, 1998; Shankar et al., 1998; Chin and Quddus, 2003a; Kweon and Kockelman, 2003; Noland and Oh, 2004
Zero-inflated Poisson (ZIP) or zero-inflated NB (ZINB) model	Handles datasets that have a large number of zero-crash observations	Shankar et al., 1997; Garber and Wu, 2001; Lee and Mannering, 2002; Kumara and Chin, 2003; Miaou and Lord, 2003; Rodriguez et al., 2003; Chin and Quddus, 2003b; Lord et al., 2005b

Both classical count outcome models Poisson and Negative Binomial regression models will be employed. A summary of various crash frequency models, their advantages and the previous researchers who examined or used these models is presented in Table 3.2.

4 METHODOLOGY

4.1 Introduction

Modelling primarily refers to the development of mathematical expressions that describe the behaviour of a random variable (RV); such a variable is known as the dependent variable. Other variables that are thought to provide information on the behaviour of the dependent variable are incorporated into the model as predictors or explanatory variables. A regression analysis is used to develop the model that examines the relationship between a quantitative dependent variable and one or more quantitative or qualitative independent variables. It traces the conditional distribution of the dependent variable or some characteristics of this distribution, such as its mean, as a function of the independent variables. This functional relationship is known as a crash prediction model. In crash research, it is a fairly routine practice to use such crash prediction models to establish links between (on the one hand) the road infrastructure and environmental, traffic, and socioeconomic conditions in spatial units (e.g. road segment, junctions or whole areas) with (on the other hand) the severity or counts of traffic crashes.

The aim of this thesis is to identify the factors affecting the severity and frequency of traffic crashes that have occurred as dependent variable. In order to accomplish this, both severity (disaggregated) and frequency (aggregated) analyses of road traffic crashes need to be considered. Therefore, a disaggregated analysis will reveal the factors affecting the severity of road traffic crashes, while an aggregated analysis will help to find the factors affecting the frequency of road traffic crashes, using two types of analysis: an individual crash level analysis which will explore the factors affecting the severity of traffic crashes; and an area- level analysis which will explore the factors affecting the frequency of traffic crashes. Statistical methods to be used in both analyses are explained below.

This chapter will present the methods employed for this thesis as shown in Figure 4.1, with a description of the statistical methods for modelling both severity and frequency of crashes, followed by a summary.

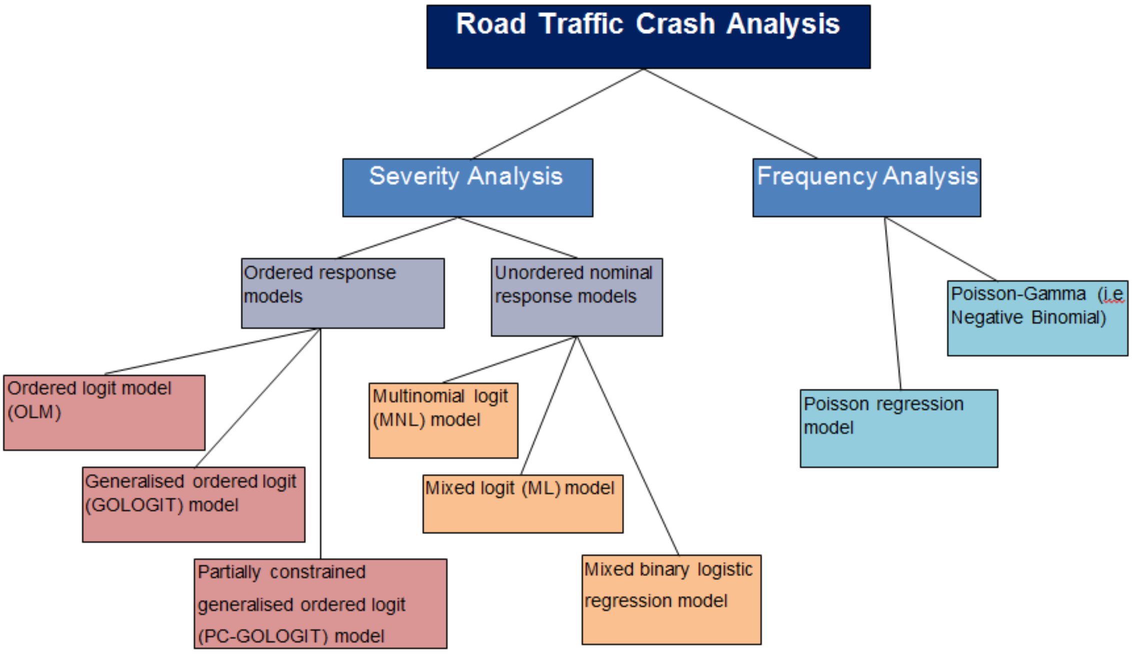

Figure 4-1 Methods employed for the road traffic crash analysis in Riyadh city

4.2 Crash severity models

The modelling techniques used to identify the factors affecting the severity of traffic crashes are different from the modelling techniques used in the analyses of crash frequency. Crash severity modelling needs to consider the characteristics of individual traffic crashes, such as the category of crash in terms of severity (whether fatal, a serious or a slight injury crash), given that the crash has occurred, as discussed in Chapter 3.

Therefore, a model that is suitable for categorical data has been explored and tested such as ordered response models and nominal response models. The two modelling techniques are presented in the following sections.

In this analysis, an individual crash will be used as the unit of observation for the model. The dependent variable will be the severity of a crash (fatal, serious and slight). The explanatory variables will be other factors associated with the crash. Since the dependent variable is ordinal in nature, the suitable model will be an ordered response model such as an ordered logit model. Different specifications and assumptions associated with these models will be tested in the section below.

4.2.1 Ordered response models

Crash severity (the dependent variable) has three categories: fatal, serious injury, and slight injury. Since the severity of the crash is ordinal in nature, an ordered response model such as an ordered logit model is suitable for analysing crash severities and examining the contributory factors affecting road traffic severity, given that a crash has occurred (O'Donnell and Connor, 1996). Crash severity models usually develop the relationship between different levels of crash severity (e.g. fatal, serious and slight injury) and the characteristics of each crash for the purpose of identifying factors affecting crash severity. The ultimate goal of crash severity models is to aid in deriving safety policies aimed at reducing the severity of traffic crashes.

The general representation of this model using a latent variable model (Long and Freese, 2006) is:

$$y_i^* = \beta X_i + \varepsilon_i \tag{4.1}$$

Where $y*_i$ is the latent variable; β is a vector of coefficient to be estimated; X_i is a vector of explanatory variables related to the crash; ε_i is the error.

The observed crash severity level y is determined by the value of the latent variable $y*$ as follows:

$$y_i = \begin{cases} 1\,(\text{slight}) & \text{if } -\infty \le y_i^* < \tau_1 \\ 2\,(\text{serious}) & \text{if } \tau_1 \le y_i^* < \tau_2 \\ 3\,(\text{fatal}) & \text{if } \tau_2 \le y_i^* < +\infty \end{cases} \tag{4.2}$$

where y_i is the observed crash severity, and τ_j is the cut-points (threshold) to be estimated.

The probabilities of observing each crash severity are:

$$\Pr(y_i > j) = \frac{\exp(\beta X_i - \tau_j)}{1 + \exp(\beta X_i - \tau_j)}, \qquad j = 1,2 \tag{4.3}$$

As discussed in Chapter 3, this model considers two assumptions (Long and Freese, 2006):

1. The error variances are the same for all observations; the model assumes the residuals are homoskedastic.
2. The relationship between each pair of outcome groups is the same; this is known as the proportional odds assumption. For example, for the three levels of severity (fatal, serious injury and slight injury), the ordered model assumes that the slope coefficients for slight injury vs serious injury and fatal crashes are equal to the slope coefficients for slight and serious injury vs fatal crashes. In other words, β values are the same for all values of j.

The violation of the assumptions would lead to misleading results (Long and Freese, 2006; Fu, 1998).

To relax the proportional odds assumption, two additional ordered response models have been estimated: a generalised ordered logit (GOLOGIT) model can be employed, which will have different coefficients across different severity outcomes, and a partially constrained model (PC-GOLOGIT). In the OLM model, coefficients are constrained to be the same, while the coefficients in GOLOGIT and PC-GOLOGIT models are allowed to vary across different equations. For the PC-GOLOGIT model, a series of Wald tests were performed on each variable individually to see whether its coefficients differ across equations. If the Wald test is statistically insignificant for one or more variables, the variable with the least significant value on the Wald test is constrained to have equal effects across equations. The model is then re-estimated with constraints, and the process is repeated until there is no more variables that meet the parallel lines assumption (Williams, 2006). The unconstrained GOLOGIT model can be written as:

$$\Pr(y_i > j) = \frac{EXP(\beta_j X_i - \tau_j)}{1 + EXP(\beta_j X_i - \tau_j)}, \qquad j = 1,2 \qquad (4.4)$$

It is noticeable that the expression for GOLOGIT is very similar to the expression of the OLOGIT model. The only difference is that in GOLOGIT the β's differ across different severity outcomes. This model estimates more parameters than needed (Williams, 2006), which sometimes makes it difficult to interpret the results. Williams (2006) then proposed a partially constrained PC-GOLOGIT model which constrained only a subset of coefficients to be the same across different severity outcomes and when the independent variables violate the partial proportional odds (PPO) model can be used which is less restrictive than the GOLOGIT. This model can be expressed as:

$$\Pr(y_i > j) = \frac{EXP(\beta_1 X_{1i} + \beta_{2j} X_{2i} - \tau_j)}{1 + EXP(\beta_1 X_{1i} + \beta_{2j} X_{2i} - \tau_j)}, \qquad j = 1,2 \tag{4.5}$$

Where the coefficients β_1 associated with X_{1i} are constrained to be the same and coefficients β_{2j} associated with X_{2i} differ across different severity outcomes.

Because the crash data have severe under-reporting of slight injury crashes, as discussed in Chapter 5, the results may be biased in terms of slight injury crashes. Therefore, more advanced statistical models are applied to overcome this problem, i.e., initially, the multinomial logit (MNL) and mixed logit (ML) models. Nominal response models have the advantage of flexibility, in terms of under-reported data; as a result, nominal response models may be appropriate for such data, as they are capable of providing more robust estimation results. Two nominal response models have been estimated: a standard multinomial logit model (MNL) and a mixed logit model (ML).

The Mixed Logit model can take into account the unobserved correlated effects and the unobserved heterogeneity between different crash severity categories (Milton et al., 2008). Accordingly, the mixed logit model is expected to provide more coherent results. The multinomial logit model and the mixed logit model have been developed with three levels of severity: slight, serious and fatal.

Crash data normally suffer from under-reporting, especially in the case of slight injury crashes. The occurrence of under-reporting means that high-severity crashes such as 'fatality' and 'serious injury' crashes are over-represented in the data, which can lead to biased and inconsistent results when using ordered response models (Yamamoto et al., 2008). Thus, nominal response models are proposed, such as a multinomial logit (MNL) model, because it deals with under-reporting better, as discussed.

4.2.2 Nominal response models

Two nominal response models are considered in this thesis: a standard multinomial logit (MNL) model and a mixed logit model (ML). The occurrence of under-reporting means that high level severity crashes such as 'fatality' and 'serious injury' crashes are over presented in the data which can lead to biased and inconsistent results when using ordered response models (Yamamoto et al., 2008). Thus, nominal response models are proposed such as a multinomial logit (MNL) model because it is able to handle the underreporting. Both a multinomial logit model and a mixed logit model have been developed with the three level of severity: slight, serious and fatal. Multinomial logit model is described in the following section:

4.2.2.1 The Multinomial Logit Model (MNL)

This model is more suitable for nominal outcomes and more flexible when under-reporting occurs in the data than ordered response models, and it is more flexible when the independent variables are not assumed to be identical across all severities in the model (Shankar and Mannering, 1996; Ulfarsson and Mannering, 2004).

This model allows different severity levels to be associated with different independent variables. In addition, an MNL model provides consistent coefficient estimates except constant terms when under-reporting occurs (Yamamoto et al., 2008). One limitation of the MNL model is that it assumes that the unobserved effects associated with each crash severity category are independent (Long and Freese, 2006).

A multinomial logit model can be expressed as follows (Washington et al., 2003):

$$\Pr(y_i = j) = \frac{EXP(\beta_j X_i)}{\sum_{\forall m} EXP(\beta_m X_i)}, \quad j = 1,2,3,............m \tag{4.6}$$

where $\Pr(y_i = j)$ is the probability of observation i having discrete outcome j; β_j is a vector of coefficients to be estimated for discrete outcome j; and X_i is a vector of the observed characteristics (covariates) that determine discrete outcomes for observation i, m denoting all possible outcomes for observation i. The coefficient of the severity outcome β_m is the coefficient of observations.

4.2.2.2 The Mixed Logit Model (ML)

The MNL model assumes that the unobserved terms associated with each crash severity category are independent. In some cases some severity categories may share unobserved effects. To take the unobserved correlated effects and heterogeneity between severity categories into account, the mixed logit model can be used to accommodate complex patterns of correlation among crash severity outcomes and unobserved heterogeneity (McFadden and Train, 2000). This can be expressed by integrating the standard MNL model as follows (Train, 2003):

$$\Pr(y_i = j) = \int \frac{EXP(\beta_j X_i)}{\sum_{\forall m} EXP(\beta_m \mathrm{X}_i)} f(\beta) d\beta \quad j = 1,2,3,............m \tag{4.7}$$

Where $f(\beta)$ is a density function.

The standard MNL model is a special case of the mixed logit model. In the MNL model, β values are fixed parameters, whereas in a mixed logit model these coefficients are allowed to vary across different crashes and are assumed to be randomly distributed.

Coefficients in the mixed logit model are considered to be random parameters if they produce statistically significant standard deviations. The mixed logit model will be tested using 150 Halton draws, since the sample size was quite large. Train (2003) suggested that the estimation of a parameter in the mixed logit model would be more consistent in the maximum simulated likelihood (MSL) if a high number of Halton draws could be used. An initial test has suggested that numbers above 100 produce reasonably stable estimates, and the results are generally consistent (between 100 and 150 draws), in terms of the set of significant estimators. Haan and Uhlendorff (2006) also showed that 100-150 Halton draws may be sufficient for stable results. To compare goodness-of-fit and complexity of different models estimated using the maximum likelihood method, the Akaike information criterion (AIC) will be used. It is defined as:

$$\text{AIC} = -2\log L + 2P \qquad (4.8)$$

Where L is the likelihood of the model and P is the number of parameters to be estimated in the model. The lower the AIC, the better is the model (Spiegelhalter et al., 2002).

4.2.2.3 Mixed binary logistic regression model

As mentioned previously, existing studies suggest that the MNL and mixed logit models are capable of handling the problem of under-reporting associated with crash data (Yamamoto et al., 2008; Milton et al., 2008). However, this is not the case for the data employed in this thesis, mainly because the crash data are affected by a severe problem of under-reporting, in terms of slight injury crashes (only 3.8% of the total crashes were reported as slight injury crashes). Mixed binary logistic regression models will be developed for fatal and serious injury crashes.

This model can be expressed as follows (Washington et al., 2003):

$$\Pr(y_i = j) = \frac{EXP(\beta_j X_i)}{\sum_{\forall m} EXP(\beta_m \mathrm{X}_i)} \;, \; j = 1 \text{ or } 2 \qquad (4.9)$$

This model is a special case of an MNL model. The above equation estimates the probability of outcome 1 occurring for observation i considering two discrete outcomes denoted 1 or 2 where j only two outcomes have.

4.3 Crash frequency models

Since crashes are countable, discrete and random, the use of a linear regression model is found to be inappropriate when the frequency of crashes is considered as the dependent variable. Models based on a Poisson distribution, including Poisson and Poisson-Gamma (i.e. negative binomial) regression models, will be employed (Miaou and Lum, 1993; Miaou, 1994; Shankar et al., 1995; Poch and Mannering, 1996; Shankar et al., 1998; Abdel-Aty and Radwan, 2000; Lord and Mannering, 2010).

The most suitable model for crash data appears to be a Poisson regression model that assumes a Poisson distribution (Jovanis and Chang, 1986; Shankar et al., 1995; Ivan et al., 1999; Miaou and Lum, 1993; Miaou, 1994; Kim et al., 2007; Thakuriah and Cottrill, 2008).

A Poisson regression model is expressed as:

$$Y_i \sim \text{Poisson}\left(\mu_i\right) \tag{4.10}$$

$$\log(\mu_i) = \alpha + \beta X_i \tag{4.11}$$

Where

Y_i is the observed number of accidents occurred on area-wide i;

μ_i is the expected Poisson accident rate at an area-wide i;

α is the intercept;

$\mathbf{X}_i$ is the vector of explanatory variables for area-wide i;

$\boldsymbol{\beta}$ is the vector of coefficients to be estimated.

However, Poisson regression models are not always appropriate for modelling traffic crash occurrences (count) and are not without limitations. The important limitation is that such a model assumes that the mean of crash counts must be equal to the variance of the crash count i.e. equidispersion (Jones et al., 1991; Miaou et al., 1992; Kulmala, 1994; Shankar et al., 1995; Lord and Mannering, 2010). If this assumption does not hold for the case of dispersed (either over- or under-dispersed) crash data, a Poisson Regression model produces undesirable results (Miaou, 1994; Shankar et al., 1995; Vogt and Bared, 1998).

To address the problem of over- and under-dispersion often found in crash data, researchers employ a negative binomial (NB) model which allows for over-dispersion (Miaou, 1994; Kulmala, 1994; Poch and Mannering, 1996; Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000; Lord, 2000; Ivan et al., 2000). A negative binomial model is given by:

$$\left.\begin{aligned} Y_i &\sim Poisson(\lambda_i) \\ \log(\lambda_i) &= \beta_0 + \boldsymbol{X}_i\boldsymbol{\beta} + u_i \\ \exp(u_i) &\sim Gamma\ (\theta, \theta) \end{aligned}\right\} \qquad (4.12)$$

Where:

$\boldsymbol{Y_i}$ is the total number of observed traffic crashes recorded in a spatial unit *i* at a given time period;

$\boldsymbol{\beta_0}$ is the intercept term;

$\boldsymbol{X}_i$ is the vector of explanatory variables;

$\boldsymbol{\beta}$ is the vector of parameters to be estimated;

u_i is a random term capturing heterogeneity effects for spatial unit *i;*

and $\exp(u_i)$ follows a gamma distribution with mean 1 and variance $\frac{1}{\theta}(= k)$ in which k is known as an over-dispersion parameter.

The model presented in equation 4.12 can be estimated using the maximum likelihood method (Cameron and Trivedi, 1998). Both log-likelihood ratio (i.e. pseudo R^2) and Akaike information criterion (AIC) can be employed to measure the model goodness-of-fit (GoF).

4.4 Summary

This chapter has shown the statistical methods for modelling crash severity and frequency which will be followed in this thesis. It has also provided a detailed discussion on the statistical models to be used in severity and frequency crash analysis.

For modelling the severity of crashes, an ordered response model and a nominal (unordered) response model have been detailed and considered to determine whether they are suitable for categorical data. As for modelling the frequency of crashes, classical count outcome models have been details and considered.

The econometric models described in this chapter will be used to analyse the data that are described in the next chapter (Chapter 5).

5 DATA DESCRIPTION

5.1 Introduction

In order to accomplish the aim and objectives of this research, the analysis of the crash dataset and road network, population and land-use data was required, in order to reveal the factors affecting the severity and frequency of road traffic crashes in Riyadh city.
Data which are a crucial part of this research were not easy to obtain, because they are scattered among different governmental organisations in Riyadh city. A number of different datasets are required to develop the models; these include:

- Data on crashes which were obtained from Riyadh General Department of Traffic (RGDT) and the Higher Commission for the Development of Riyadh (HCDR)
- Land use data which were also collected from HCDR
- Road network data which were obtained from Riyadh Department of Transport (RDT)
- Socio-economic data obtained from Saudi General Directorate of Statistics (SGDS).

A series of meetings were held with the above organisations during two visits to Saudi Arabia in April 2009 and December 2009 to request and gather the data.
Since 2004 the Riyadh Traffic Department (RTD) has collected road crash data for the Riyadh region. These data were then made available to the Higher Commission for the Development of Riyadh (HCDR) for further processing. The road crash data for this study were obtained from the HCDR. These data cover a period of five years, namely AH 1425, 1426, 1427, 1428, and 1429 (equivalent to 2004, 2005, 2006, 2007 and 2008).

Some of the data gathered include age of the people involved in the crash, the time of day and the day of the week when the crash occurred. Figure 5.1 shows the distribution of age of the people killed in traffic crashes in Riyadh city for 2007. It is noticeable that the number of fatalities due to road crashes is relatively higher for the age groups 16-20, 21-25 and 26-30, knowing that this is not the majority age group (SGDS, 2005).

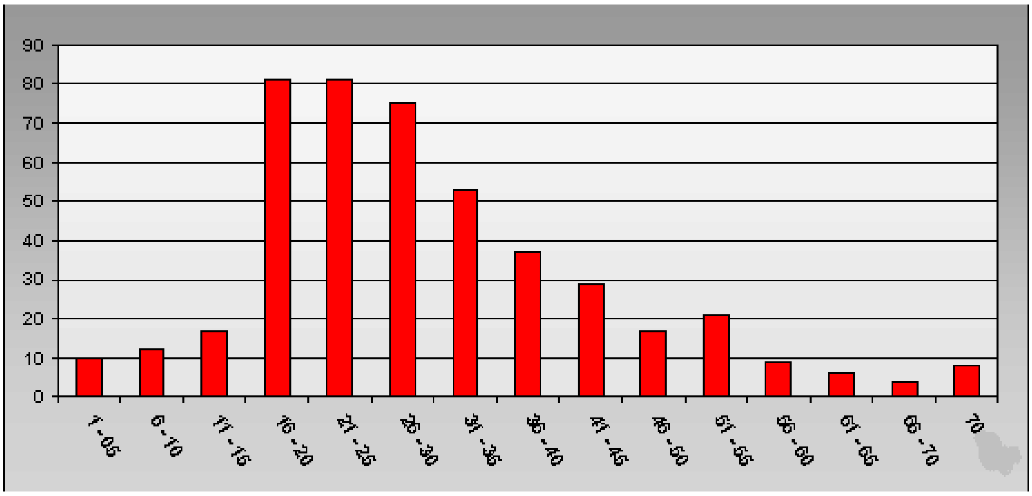

Figure 5-1Number of fatalities according to the age group in Riyadh for 2007 (RGDT, 2007).

As can be seen from Figure 5.2, it is clear that the period from 0600 to 0800 has the highest fatality rate during the day, which is not surprising as it is the time for trips to school and work in Riyadh.

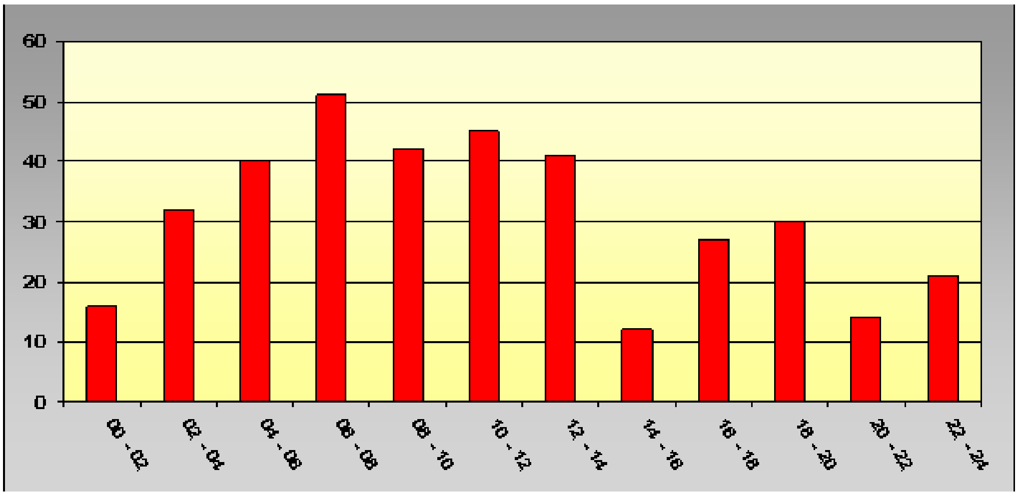

Figure 5-2 Fatal crashes according to time of day in Riyadh city for 2007 (HCDR, 2008)

Referring to the statistical data from 2004-2007 produced by the Higher Commission for the Development of Riyadh in 2007, it is found that most crashes took place during Wednesdays and Thursdays, which are the days of the weekend in Riyadh, when most people travel for recreation and social activities (see Figure5.3). Furthermore, most crashes took place in the East and West parts of the city because of the high density of traffic and population in these areas.

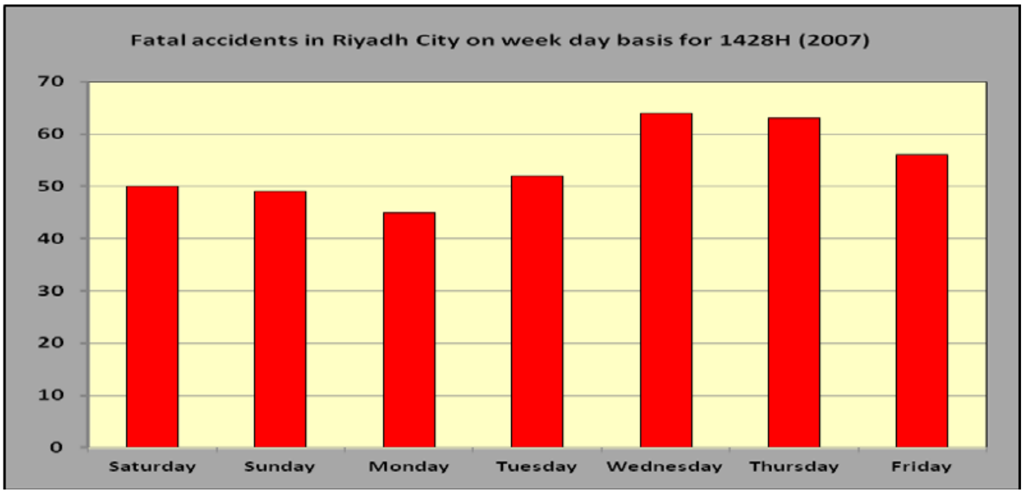

Figure 5-3 Fatal crashes by week day in Riyadh city for 2007 (Source: HCDR 2008).

The details of the final data sets used in this study, including descriptive statistics of the variables to be employed in both crash frequency and crash severity models and the validation of data, are presented in this chapter.

5.2 Data description

5.2.1 Crash data

Crash data are contained in three different datasets. The first file contains the details of the participants involved in the crash, the second file contains the details of the crash and the third file contains the details of the vehicles involved in the crash. The variables in each dataset are detailed below:

The participant data file has the information on age, nationality, percentage of blame among the drivers involved in the crash, and participant type: whether the participant was the driver, passenger, pedestrian, fugitive driver, fugitive passenger, or fugitive pedestrian.

The crash data file contains the information about the crash. This includes crash location (latitude, longitude), crash severity, date and time, number of casualties, number of vehicles involved in the crash, cause of the crash (which was divided into causes relevant to the driver, passenger, pedestrian, road, and vehicle), type of the crash, lighting condition, road surface condition, whether day or night, and type of damage, to either public or private property.

The vehicle data file shows the details of the vehicles involved in the crash, with respect to their type, colour, model, make, and country exporting the vehicle.

This chapter describes the data used in this study, followed by a validation of the data and descriptive statistics of the variables to be employed in both crash severity and crash frequency models, and, finally, the potential variables used and the hypotheses.

5.2.1.1 Participant data (file 1)

1. Crash key, which is used as a unique identifier for each crash, and can be used to link different crash data.
2. Participant number, which represents the number of participants involved in the crash, ranging from 2 to 100.
3. Vehicle sequence which is for each participant number involved in the crash.
4. Age of the participant.
5. Nationality of the participant
6. Participant type, coded as 1 to 6 (where 1=Driver, 2=Passenger, 3=Pedestrian, 4=Fugitive driver, 5=Fugitive passenger, 6=Fugitive pedestrian).
7. Health conditions (coded as 1=none, 2=Treated in hospital, 3=Transfer to hospital, 4=Death on site, 5=Death on the way to hospital, 6=Death in hospital, 7= unknown).
8. Percentage of blame (0, 25, 50, 100 per cent). Each crash has different participants who have different percentages of blame depending on the error made by each participant, which can be decided and reported by the policeman at the time of the crash.

5.2.1.2 Crash data (file 2)

This data file shows the details of the crash.

1. Crash key, which is used as a unique identifier for each crash, and can be used to link different crash data.
2. Number of vehicles involved in the crash.
3. Number of casualties in each crash.
4. Crash severity, coded as 1 to 4 (where 1=fatal crash, 2=serious injury crash, 3=slight injury crash, 4=damage in either public or private premises).

5. The sector, territory, and police centre (police station) to which the crash belongs (divided into four sectors).
6. Time of the crash.
7. Date of the crash.
8. Day of the week, coded as 1 to 7 (where 1=Saturday, 2=Sunday, 3=Monday, 4=Tuesday, 5=Wednesday, 6=Thursday, 7=Friday).
9. Day or night (0, 1, where 0=Day, 1=Night).
10. Date and time of informing the police about the crash.
11. Date and time of police arrival.
12. Date and time of closing the file of the crash.
13. The direction of the crash 1 to 4 (where 1=North, 2=South, 3=East, 4=West)
14. Crash location, coded as 1 to 11 (where 1=Straight, 2=Curve, 3=Intersection with traffic light, 4=Intersection without traffic light, 5=High or hill, 6=Railway junction, 7=Slope, 8=T-junction, 9=Roundabout, 10=Other, 11=Unknown).
15. Crash coordinates (Northing and Easting).
16. Road surface (1, 2) (where 1=Dry, 2=Wet).
17. Lighting condition (1, 2) (where 1=With lighting, 2=Without lighting).
18. Crash causes which are relevant to:

- The driver: 101 Sleeping, 102 Tiredness or Fatigue, 103 Distraction, 104 Excessive speed, 105 Red light violation, 106 Non-compliance with the stop sign, 107 Reverse traffic or Driving in the wrong direction, 108 Illegal overtaking, 109 Violating the priority, 110 Illegal parking, 111 Showing off, 112 Driving under the effect of alcohol, 113 Violating the traffic light or Violating the traffic signs, 114 Sudden overtaking or Sudden turn.
- The passenger: 201 Getting out before the vehicle stops, 202 Getting in before the vehicle stops, 203 Clinging to a vehicle, 204 Sitting on the vehicle, 205 Sitting on the vehicle boot, 206 Other mentioned in report.
- The pedestrian: 301 Illegal crossing, 302 violating the pedestrian sign, 303 walking in the middle of the street, 304 playing in the street, 305 other mentioned in report.
- The road: 401 Slippery road, 402 Uphill (gradient), 403 Downhill, 404 Turning, 405 Motorway without fence, 406 Lack of warning signs, 407 Lack of a sign or Lack of traffic light, 408 Insecure work site, 409 Other mentioned in report.

- The vehicle: 501 Lack of lights, 502 Tyre damage, 503 Brake damage, 504 Engine breakdown, 505 Electric breakdown, 506 Fault in gear system, 507 Fault in steering, 508 Fault in windscreen wipers, 509 Overweight, 510 Extra dimensions, 511 Other mentioned in report.

19. Point of collision (1, 2, 3, 4, 0) (Angle, Head-on, Rear-end, Side-swipe, Unknown).
20. Weather condition (1-7) (Clear, Fog, Dust, Snow, Rain, Cloud, Unknown).
21. Crash type (1-17) (Hitting a moving vehicle, Hitting a parked vehicle, Hitting a motorcycle, Hitting a bicycle, Hitting an animal, Hitting a road fence, Hitting a pylon, Hitting a pedestrian, Coup vehicle, combustion vehicles, Hitting a side barrier, Hitting a tree, Hitting a traffic light, Hitting a sign, Hitting a garbage container, Hitting a fixed object, Other).
22. Private damage (1-7) (None, Cars, Cycles, Walls, Gates, Shops, Other)
23. Public damage (1-13) (None, Lighting columns, Columns of optical signals, Electricity station, Phone booth or box, Traffic plates, Street name plates, Advertising plates, Metal fence, Metal barriers, Trees, Other).

5.2.1.3 Vehicle data (file 3)

This data set shows the details of the vehicles involved in the crash with respect to their type, colour, model, make, and country exporting the vehicle.

1. Crash key, which is used as a unique identifier for each crash, and can be used to link different crash data.
2. Vehicle sequence.
3. Flow direction, coded as 1-4 (in which 1=North, 2=South, 3=East, 4=West)
4. Type of vehicle registration, coded as 1-10 (in which 1=Motorcycle, 2=Sedan, 3=Transport, 4=Bus, 5=Trailer, 6=Car for private use, 7=Heavy equipped vehicle, 8=Bicycle, 9=Unknown vehicle, 10=Other).
5. Country of manufacture of the vehicle.
6. Colour of the vehicle.
7. Model of the vehicle, which indicates the year in which the vehicle was manufactured.
8. Make of the vehicle (i.e. Honda, Mazda, Ford, etc).

A key piece of information contained in the crash data is the location (in terms of X and Y coordinates) of the crash. This allows the geo-coding of crash data and enables the integration of crash data with other relevant datasets such as land use and population. Once a crash occurs, the traffic police for RTD attend the crash scene and fill in an information form related to the crash.

5.2.2 Population data

The demographic data for Riyadh city were based on the data collected through a land-use survey. This was conducted in 2004, considering the household as the sampling unit. The data samples 1.75% of the population, which represents a size close to 11,500 households. Data were obtained from SGDS (Saudi General Directorate of Statistics) and include population data for each HAI (equivalent to a ward) with a total of 171 HAIs.

The household survey included the following variables:

1. Age of each member of the household (age = 0 for less than one year old).
2. Nationality of each member of the household.
3. Gender of each household member (male/female).
4. Marital status of each member, coded as 1, 2, 3, 4 (in which 1=Single, 2=Married, 3=Divorced/separated, 4=Widowed).
5. Employment status of each member in each household in that HAI, coded as 1 to 7 (Employed, Unemployed but looking for work, Student, Housewife, Retired, Unable to work/handicapped, Not employed and not looking for work).
6. Education status of each member living in that household, coded as 1 to 7 (Illiterate, Able to read only, Able to read and write, Intermediate, Secondary, College/University, MA/MSc/PhD).
7. Average income per month (from all sources) of each member of the household who was indicated as employed. Average income was classified into twenty categories.
8. Mode of travel, coded as 1-5 (in which 1=Automobile, Pickup, Jeep, 2=Truck, 3=Private Van or Bus, 4=Taxi/Limousine, 5=Public Bus).

5.2.3 Land-use data

Land-use data for Riyadh city were obtained from HCDR (Higher Commission for the Development of Riyadh) who made an extensive survey in 2004 and categorised spatial level land-use data into nine main categories, which are:

Residential manufacturing, Transportation, Communication and utilities, Trade (commercial), Services, Cultural, Entertainment and recreational, Agricultural, and Resource production and extraction.

The land-use study covered all areas within the urban boundary of Riyadh city as a spatial-unit level with an area of more than 3000 square kilometres. This survey included land use, the number of parcels and the area of each parcel in each spatial unit (e.g. HAI), population and population density for each HAI, and the total area of each HAI.

5.2.4 Road network data

Data on the road network were provided by Riyadh Department of Transport (RDT) and include road length and area for Riyadh city. Variables in population data were merged using STATA; land-use and road network data were in a GIS format. The geo-coding of crash data allows its integration with other datasets such as land use and population for frequency analysis. Figure 5.4 shows the histogram of fatal and serious injury crashes:

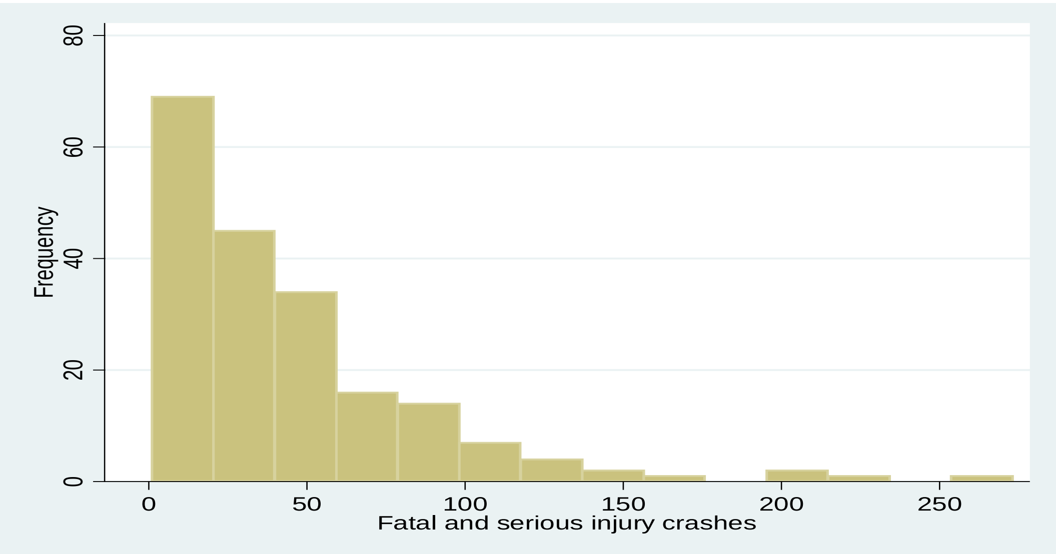

Figure 5-4 Histogram of fatal and serious injury crashes

5.3 Data validation

It is important to validate the data to ensure their accuracy and quality in order to avoid biased modelling results. Before carrying out any crash analysis it is necessary to validate the quality of the data by comparing different statistics generated from the data with the published statistics from other sources. This is to check the degree of similarity and compatibility between the generated and the published statistics before proceeding to the use of advanced statistics, analysis, and modelling. The following section will explore the quality of the data which will be used in this research.

As discussed, the idea is to generate some statistics from the data which are to be compared with the published statistics. HCDR publish statistics on monthly fatal and serious injury crashes for Riyadh city only.

However, several variables were looked at, in order to find the criteria that the publisher might have used. It was discovered that the published statistics used the crash data set to calculate the number of crashes by years and by months for Riyadh city only.

According to the information given by the consultant (Mr. M. Al-katib, August 2009) who designed Riyadh's traffic safety strategy with TRL (Transport Research Laboratory), this difference was due to the fact that the publisher considered only crashes inside Riyadh city and the consultant added that the statistics published were number of crashes. Some of the crashes in the dataset used in this thesis were, however, located in roads in the Riyadh region but outside the boundaries of the urban planning district of Riyadh city. Also, in some cases the crash data were not recorded correctly and accurately, which might be due to the fact that the police officer was busy calling or arranging for the ambulance or managing the crash. Figure 5.5 shows the spatial distribution of geo-coded crashes on the Riyadh region electronic map and Riyadh city marked on the map (obtained using ArcGIS 9.2 tool) to map the location of crashes. This will be discussed in detail in Chapter 7.

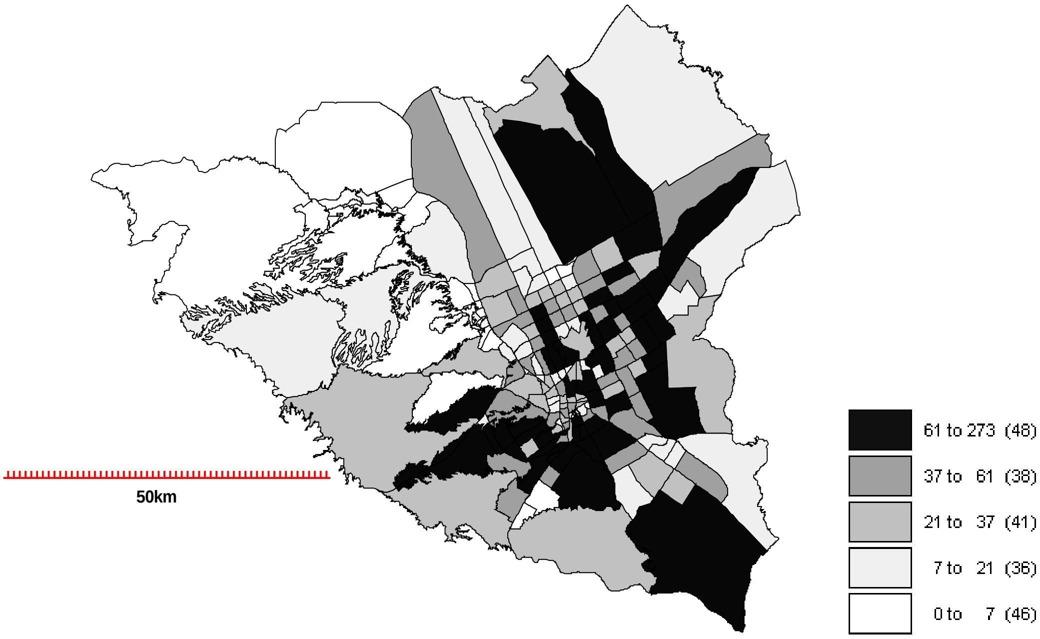

Figure 5-5 Spatial distributions of geo-coded crashes for Riyadh region and Riyadh city

Geographical information system (GIS) technology is becoming increasingly popular tool for visualisation and analyses of crash data in motorways. GIS has the ability to hold a vast amount of data that can be easily stored, shared, analysed and managed. It provides a platform for spatial data analyses and visualization to explore relationships between spatial and non-spatial data (Erdogan et al., 2007). When using a GPS device, one can obtain (x,y) coordinates at regular time steps. The trajectory is then automatically recorded, by joining the locations collected in time, and geo-referenced into a GIS. Different digital GIS maps can be plotted in different layers in the GIS software (e.g. roads, crashes), and GIS software (i.e. MapInfo in this case) can display them at the same time in one view, with each layer to be displayed in a desired way, such as making lines (roads) look wider. An example is shown in Figure 5.6, which represents a multi layered GIS map of Riyadh city, it integrates road networks as lines, road crashes as dots, land use and boundaries of HAI.

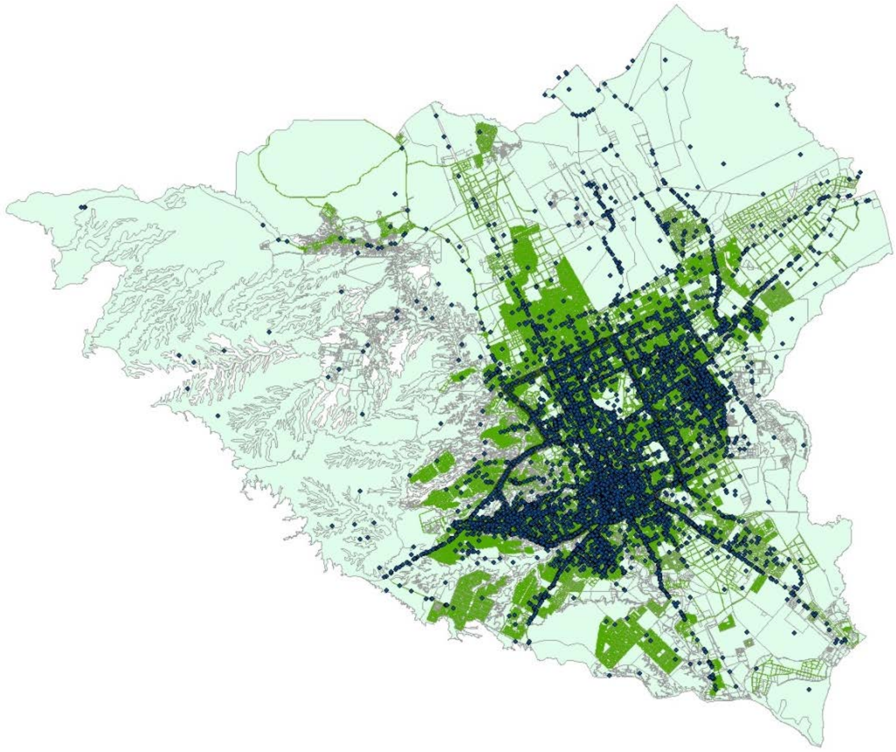

Figure 5-6 Example of GIS maps in multiple layers

Many researchers have used GIS to display crash locations on digital maps and perform various spatial analyses (including hot spot analysis) of crashes (Petch and Henson, 2000; Wang et al., 2009). In addition, GIS enables researchers to link crash data with travel information, land use, and socio-economic information to better capture the relationship between crash occurrence and contributing factors. There are several ways to locate crashes onto digital maps. Crashes can be directly added if the exact geographic references of crash locations are known, such as easting and northing coordinates which can be obtained by a GPS device. Address geocoding can also be conducted when the exact address (e.g., street name and number, city, state, post code) is available.

In terms of map construction MapInfo tool developed by ESRI was used for GIS modelling by cross mapping (overlaying) data. There are several ways to locate crashes onto digital maps; crashes can be directly added if the exact geographic references of crash locations, like coordinates, are available. First, base digital map was obtained from HCDR, which includes land use and road network. The next step was adding the geocoded data of the crash locations.

In this research, land-use and road network data were in a GIS format. The geo-coding of crash data allowed integration with other GIS datasets such as land use for frequency analysis. Figure 5.6 showed the spatial distribution of geo-coded crashes on the Riyadh region electronic map and Riyadh city marked on the map. Furthermore, road density data are calculated at spatial levels of HAI, the equivalent of a ward in England, and crash frequency per HAI are then determined and integrated with the HAI-level road density data using GIS.

5.3.1 Statistical tests

A t-test is used to compare two sets of data for validation purposes. The t-test calculates a p-value (probability value) that shows the possibility that these results could occur by chance or to decide whether we have enough evidence to reject the null hypothesis and say our research hypothesis is supported by the data. If there is a less than 5% chance of getting the observed differences by chance, we reject the null hypothesis and say we have found a statistically significant difference between the two groups when the confidence interval is 95%. The t-test is used for problems that deal with the population mean or for problems which involve comparative samples (i.e. in cases where we are trying to determine if means from different samples are significantly different) (Walpole et al., 2002).

When using the t-test with two paired samples, the following steps are followed (Bluman, 2004):

The *t* test can be calculated when the two samples are of equal size as follows:

1. Calculate the difference for each pair, $r_1 = x_1^i - x_2^i$

2. Calculate $\sigma_r^2 = \dfrac{\sum d_i^2}{n-1}$

   $$\sigma_n = \frac{\sigma_r}{\sqrt{n}}$$

3. Test statistic $t = \dfrac{x_1^- - x_2^-}{\sigma_n}$

Where n is the sample size σ is the population standard deviation, n-1 is the degree of freedom, and x is the sample mean.

The three files of data were received in Microsoft Access format as raw data. All of these three files have a unique identifier as a “crash key”. The data were transferred to STATA to proceed with the crash analysis. Before any analysis was carried out it was necessary to validate the quality of the data by, firstly, comparing them with the statistics generated by the HCDR. Published statistics were available for Riyadh city only so this forms the basis of the comparison. The results are shown in Table 5.1.

It can be observed from the above table that the data for fatal and serious crashes over the specified time period are close to the published results. Statistics obtained for fatal and serious crashes are over-estimated by 2.1% and 3.9% respectively compared to the published results.

Table 5-1Crash year with fatal and serious injury crashes from published and obtained statistics for Riyadh city

	Published statistics		**Statistics obtained for the crash data base**		**% difference fatal**	**% difference serious injury**
Crash Year	Fatal	Serious	Fatal	Serious		
1425	430	1555	424	1573	-1.4	1.2
1426	408	1481	419	1585	2.7	7.0
1427	353	1276	361	1225	2.3	-4.0
1428	357	1178	346	1092	-3.1	-7.3
1429	315	959	320	1026	1.6	7.0
Total	**1863**	**6449**	**1870**	**6501**	**2.1**	**3.9**

A t-test was conducted to compare the two sets of data for validation purposes. The t-test allowed us to determine a p-value (probability value) that shows the probability that these results could occur by chance or to decide whether we have enough evidence to reject the null hypothesis and conclude that our research hypothesis is supported by the data (Walpole et al., 2002).

A paired t-test is applied to look at the differences between the published mean and the mean of the results obtained in monthly results for fatal and serious-injury crashes in Riyadh. By following the steps of the t-test stated above:

H_0: the null hypothesis: the mean difference is equal to zero, i.e. $(M_o\text{-}M_p) = 0$

H_1: the alternative hypothesis: the mean difference is not equal to zero, i.e. $(M_o\text{-}M_p)$ #0

M_o= Obtained mean whereas M_p= Published mean.

By looking at the data and the statistical paired t-test results, we found that for the monthly fatal crashes the t value is 0.8571, which is less than the critical t obtained from a t-distribution table, which is: 2.0. This means we do not reject our null hypothesis (see Table 5.2).

Table 5-2 The statistical paired t-test for fatal crashes

Variable	Observations	Mean	Std. Error	Std. Dev.	(95%	Conf.)
Res-Fatal	60	31.53	1.05	8.12	29.44	33.63
Pub-Fatal	60	31.05	0.98	7.56	29.10	33.00
Difference	60	0.48	0.07	0.56	0.34	0.63

And for the serious-injury crashes, t statistics equals to 0.8769 (see Table 5.3), which is less than the critical t with a degree of freedom of 60-1=59 at 95% confidence level is level. This means we do not reject our null hypothesis.

Table 5-3 Statistical paired t-test for serious-injury crashes

Variable	Observations	Mean	Std. Error	Std. Dev.	(95%	Conf.)
Res-serious	60	108.05	3.58	27.76	102.88	117.22
Pub-serious	60	107.48	3.50	27.11	100.48	114.49
Difference	60	0.57	0.08	0.65	2.40	2.73

The (Rejection H_o) areas in the curve will be the area on the right of critical t (+2.0) and left of the critical t (-2.0) while the (Do not reject H_o) will be the area between the values of critical t (+2.0 and -2.0).

This suggests that we do not have enough evidence to reject the null hypothesis, which indicates that the means are the same for the fatal crashes and the serious injury crashes. The data set is usable.

5.4 Descriptive statistics

Descriptive statistics are reported for the four levels of severity (fatal, serious, slight, and property damage only), with different variables from the three dataset files (participant, crash, and vehicle files), such as age, nationality, cause, location, time of the crash, road surface, lighting condition, day of the week, and day or night.

Table 5.4 shows the frequency of different levels of crash severity, coded as 1, 2, 3, and 4, for fatal, serious injury, slight injury, and property damage only (PDO**)** respectively**:**

Table 5-4 Frequency of different levels of crash severity (AH 1425 to 1429)

Severity Category	**Frequency**	**Percentage**
Fatal	2498	0.36
Serious injury	7835	1.12
Slight injury	1127	0.16
PDO	686785	98.36
Total	**698245**	**100**

It is noticeable that the vast majority of the crashes were recorded as property damage only (98.36%). In the USA in 2003 the figures were 1% fatal, 11.6% serious injury, 10.5% minor injury, 12.3% possible injury and 64.6% PDO (Eluru and Bhat, 2007). In the UK, figures for injuries were 2% fatal, 15% serious injury and 83% slight injury (1991 – 2003) as UK crash data do not include ‘property damage only’ crashes (Gray et al., 2008). It is therefore clear that the percentage of slight injury crashes recorded in Riyadh city appears to be very low compared with other countries. The low percentage of slight injury crashes recorded in Riyadh city may be due to the following reasons:

- Crashes involving both slight injury to people and damage to property may be recorded as ‘property damage only’ crashes.
- Slight injury crashes may suffer from under-reporting (Al-Katib, 2009).
- The government target is to reduce fatal and serious injury crashes and hence there is less focus on collecting data on slight injury crashes.
- There might actually be a much lower percentage of slight injury crashes. However, this is unlikely as slight injuries are the most common injury outcome.

In KSA there is a need to register each crash for insurance and repair purposes. Saudi law does not allow vehicle workshops to receive or repair any damaged vehicle without a letter from the Traffic Police Department, according to the Saudi Council of Ministers distinguished decisions numbers 222 and 271 dated AH 25.12.1422, on compulsory insurance on the vehicle and the need for a repair paper. Therefore, all ‘property damage only’ crashes will be reported. The traffic safety strategy in Riyadh for road crashes concentrates on fatal and serious injury crashes, which take the highest priority when reported in the annual published statistical reports. Since the crash data are heavily biased towards ‘property damage only’ crashes, the inclusion of this category of crashes may influence the analysis. Therefore, ‘property damage only’ crashes are not considered in this research. The total number of injury crashes over the five years is 12,039.

A comparison between neighbouring countries and Riyadh on the numbers of crashes with different levels of severity was also made. Then, a t-test was used to compare the two sets of data for validation purposes. The t-test allowed us to determine a p-value (probability value) that shows the possibility that these results could occur by chance, or to decide whether we have enough evidence to reject the null hypothesis and say our research hypothesis is supported by the data.

A comparison between Bahrain, which is one of the Gulf Cooperation Countries (GCC), and Riyadh on the numbers of crashes with different levels of severity in the year AH 1429 is shown in Table 5.5 below:

Table 5-5 A comparison between Riyadh and Bahrain on number of crashes for the year 1429AH

Severity	**Riyadh**	**Percentage**	**Bahrain**	**Percentage**
Fatal	320	0.25	75	0.10
Serious injury	1026	0.82	522	0.69
Slight injury	51	0.04	1446	1.91
Damages	124852	98.89	73538	97.30

(Reference: Bahrain General Directorate of Traffic and Licensing - Ministry of the Interior) http://www.traffic.gov.bh/arabic/index.asp

The Kingdom of Bahrain was chosen for comparison because it is similar to Riyadh in area and shares a similar culture and style of life. From the above table there is a noticeable difference in the number of slight injury crashes between Riyadh and Bahrain which could be arguable knowing that the population density of Riyadh city is about twice that of Bahrain (according to: www.wikipedia.org). This difference can be associated to the under-reporting problem, or the HCDR has decided to not count the slight injury crashes and rather consider them as PDO crashes. While in Bahrain these types of crash injuries are taken more seriously.

In addition, damage and slight injury crashes were recorded for the purpose of estimating the costs resulting from these crashes to the national economy of the Kingdom of Saudi Arabia and were not included in the traffic safety strategy. The traffic safety strategy in Riyadh for road crashes concentrated on the fatal and serious injury crashes, which received the highest priority and the most concern when reported in the annual published statistical reports.

It should be noted that the observation in severity analysis is the crash rather than the participants involved in the crash. This makes it difficult to obtain data related to age and nationality of driver (or other participant) involved in the crash. To address this issue, data should be obtained on the age and nationality of participants who have 100% of the blame (are wholly at fault) for the crash. Data suggest that participants with 100% of the blame are mostly drivers. This would, however, reduce the number of observations.

The review of literature indicates that two common factors affecting the severity of traffic crashes are the age and nationality of the drivers (at fault) involved in the crashes. The datasets supplied by the HCDR do not implicitly identify whether a driver is at fault or not for the crashes, and therefore it was not straightforward to distinguish crashes with drivers at fault. However, the participant data file contains information on the distribution of blame attributed to drivers (0%, 25%, 50%, 75% and 100%) involved in a crash. The crash data analysed in this thesis were created by extracting cases where a driver was allocated 100% of the blame for the crash. The sample with 100% of the blame was used because there will be a single value for age and nationality for each driver. The number of such crashes is 3,648.

Based on the literature review (see Chapter 2) the functional relationship between the severity of a traffic crash and its contributing factors can be expressed as:
The severity of a crash = *f* (age, nationality, time of the crash, cause of the crash type of collision, location of the crash, road surface, lighting condition, weather condition, number of vehicles, crash year, and number of casualties). Among these contributory factors, gender will not be included as females are not allowed to drive according to the Saudi law.

The most widely used measure of exposure is the number of kilometres travelled for each travel mode. Additional useful insight can be is provided by taking into account the speed of travel, in which case exposure is expressed as the amount of time spent in the traffic system. One of the developments in recent years has been the installation of electronic and telecommunication equipment inside vehicles and along roads. Another development is the increasingly widespread use of mobile telephone data. As a result it is becoming easier to collect up-to-date and reliable information on a variety of parameters that could be of importance in the calculation of vehicle exposure and risk. Additional information on the distribution of speeds, types of vehicles and following distances also seems to be a future possibility to measure aspects of exposure. In many road safety analyses, different exposure measures are used, according to data availability and quality, as well as the particular objective of the analysis. These measures may vary significantly in terms of the potential level of desegregation and the possible underlying bias in their estimates. No general rule is available concerning the preferred measures of exposure. Vehicle- and person kilometres of travel, as well as the time spent in traffic, are conceptually closer to the theoretical definition of exposure and can be theoretically available to a satisfactory level of detail. In this particular research the dataset used didn’t contain the parameters described above to measure exposure rates (such as traffic volume, travelled distances, fuel consumption or driver’s mobile phone movement data) which lead to omission of this factor from the study.

In the frequency analysis the full sample of the crash dataset is used. However, the dataset to be used in the severity analysis contains only those drivers who were 100% to blame for the crash based on information in the participants file from 573,206 cases. The crash data contains personal data for all drivers. In order to allow for age and nationality in the analysis it is necessary to have data for one driver only. Therefore, this analysis is based on cases where one driver was 100% to blame for the crash. Three variables will be used from the participant file as follows: age, nationality, and percentage of blame. Each variable will be used in the following manner: The nationality will be either Saudi or non-Saudi and the age of the driver is classified into one of five age groups.

Table 5-6 Crashes by age group and level of severity

	Age group				
Severity	**13-17**	**18-24**	**25-39**	**40-64**	**65+**
Fatal	60	342	374	166	18
Serious injury	171	946	1,204	398	29
Slight injury	11	48	64	22	4
PDO	37	355	602	219	4
Total	279	1,691	2,244	802	55
Percentage	5.5	33.35	44.25	15.82	1.08

The legal driving age in Saudi is eighteen years but as shown in Table 5.6 the under-age drivers represent 5.5% of the total number of crashes.

The above table, with data on drivers with 100% of the blame, shows that lower age groups and especially group two (25-40) are involved in a high number of crashes at all levels of severity compared to the other age groups. According to HCDR, in the latest census, made in October 2004, the population of Riyadh was 4,260,192; 66% of the population are Saudis and 34% of the population are Non-Saudis. According to HCDR, 40% of the Saudis in Riyadh are below 15 years of age, while the figure is only 23% for the non-Saudis because they come to Saudi to work, which means they should be within the working age. The tabulated values for age group with nationality are given below in Table 5.7.

As noted in Table 5.7, the non-Saudis have more crashes than the Saudis in the age groups 25-39 and 40-64 while Saudis have many more crashes between the ages of 14 and 24. In addition, the average age among the Saudis in Riyadh is 18 years while it is 30 years for the non-Saudis, and the percentage of males among the Saudis is 53% while it is 63% for the non-Saudis, which are normal because the most non-Saudis who are employed as drivers are males.

Table 5-7 Crashes by age group and nationality only for those with 100% of the blame

	Nationality	
Age group	**Saudi**	**Non-Saudi**
14-24	1,480	215
25-39	1,076	1,170
40-64	304	502
65+	44	11
Total	2,904	1,898
Percentage	60.48	39.52

Generally, the percentage of Saudis in all age groups who are involved in crashes is more than the percentage of non-Saudis.

5.5 Variables

Table 5.8 shows potential independent variables which will be used in the severity models presented in Chapter 6. The weather condition variable was categorised as fine and 'other' which included fog, rain, dust, snow and cloud. After investigation it was decided to omit the weather variable because 99.67% of the crashes occurred during 'fine' weather conditions. In addition, after the development of an initial model the weather conditions variable was not significant in any models and therefore excluding it had little effect.

Table 5-8 Potential variables for severity models

	Description of the variable	**Related statistics**
Severity of the crash	Slight injury=1, Serious Injury=2 and Fatal=3	Slight=3.82%, Serious=71.29%, and Fatal=24.90%.
Age of the driver (Years)	A continuous variable	Mean=30.21, Standard deviation=11.24
Road density (m/km^2)	A continuous variable	Mean=14.45, Standard deviation=7.86
Nationality of the driver	Non-Saudi=0; Saudi=1	Saudi=61.46% and Non-Saudi=38.54%
Time of day		
00:00 to 03:59		Yes=15.56% and No=84.44%
04:00 to 07:59		Yes=17.08% and No=82.92%
08:00 to 11:59		Yes=20.16% and No=79.84%
12:00 to 15:59		Yes=15.17% and No=84.83%
16:00 to 19:59		Yes=17.16% and No=82.84%
20:00 to 23:59		Yes=10.22% and No=89.78%
Cause of the crash		
Overtaking	Overtaking=1, otherwise=0	Yes=33.02% and No=66.98%
Distraction	Distraction=1, otherwise=0	Yes=22.72% and No=77.28%
Excessive speed	Excessive speed=1, otherwise=0	Yes=16.39% and No=83.61%

Other (e.g. tiredness and sleeping)	If the cause is Other=1, otherwise=0	Yes=27.87% and No=72.13%
Type of collision		
Angle	angle=1, otherwise=0	Yes=15.81% and No=84.19%
Head-on	Head-on=1, otherwise=0	Yes=10.63% and No=89.37%
Rear-end	If the type is Rear-end=1, otherwise=0	Yes=24.57% and No=75.43%
Side-swipe	Side-swipe=1, otherwise=0	Yes=16.50% and No=83.50%
Other (unknown)	Other=1, otherwise=0	Yes=27.82% and No=72.18%
Location of the crash	Straight=1, Other=0	Straight=88.01% and Other=11.99%
Road surface	Dry=1, Wet=0	Dry=99.67% and Wet=0.33%
Lighting condition of the road	Lighting=1, without-lighting=0	Lighting=92% and Without lighting=8%
Weather	Fine=1, Other=0	Fine=99.67% and other=0.33%
Single vehicle	Single=1, Other=0	Single=43.27% and Other=56.73
Crash year		
Year 1425		Yes=20.99% and No=79.01%
Year 1426		Yes=28.48% and No=71.52%
Year 1427		Yes=22.50% and No=77.50%
Year 1428		Yes=18.74% and No=81.74%
Year 1429		Yes=9.78% and No=90.22%
Number of casualties	A continuous variable	Mean=2.5282 Standard deviation=1.553

It is anticipated that road density (length of all roads per unit area) may also influence the severity of traffic crashes as vehicle average speeds are more likely to be low in areas with high road density owing to high traffic volumes and more pedestrian activity. Traffic crash data, however, do not include any information on road density in the vicinity of the crash location. Road density data are therefore calculated at spatial levels such as HAI, the equivalent of a ward, and geo-coded crash data are then integrated with the ward-level road density data using GIS. This allows ward-level road density data to be obtained for each of the crashes, which means that if a traffic crash occurs in a ward, the calculated road density of that ward is taken as the road density attributed to the crash.

In addition to these variables, interactions between age and nationality, age and excessive speed, and nationality and excessive speed will be tested in the models to be developed as shown in the next chapter.

A unique crash key was used to integrate these data files. They were collected over a period of five years, namely 1425, 1426, 1427, 1428, and 1429AH, roughly equivalent to 2004, 2005, 2006, 2007, and 2008, respectively. The severity of traffic crashes was categorised as: fatal (2,498 crashes; 0.36%), serious (7,835 crashes; 1.12%), slight (1,127 crashes; 0.16%) and property damage only (686,785 crashes; 98.36%).

A dataset for use in the severity analysis was created containing only those cases where one driver has 100% responsibility, using the participant file; there were 573,206 cases. The sample with 100% of the blame was used because it has a single value for each variable (i.e. age and nationality for each participant), it shows the participant who is 100% responsible for the crash, and it shows all crash details. In the frequency analysis the full sample of the crash dataset is used.

Table 5.9 shows the dependent and independent variables which will be used in the frequency models presented in Chapter 7.

Table 5-9 Potential variables for frequency models

Variables	Description
Dependent variables	
Fatal crashes	Number of fatal crashes in the HAI.
Serious injury crashes	Number of serious injury crashes in the HAI.
Demographic characteristics	
Population	Number of people living in the HAI.
Vehicle registered	Number of vehicles in the crash, including all types (car, taxi, pickup/jeep, van, bus, truck).
% male	Percentage of males in the HAI.
% non-Saudi	Percentage of non-Saudi people in the HAI.
Income per capita	Total income divided by total population.
Income per adult >18	Total income divided by number of adult people >18 years old in the HAI
% low income	Percentage of people with income less than 2500 Saudi Riyals per month.
% aged 0-18	Percentage of population from 0-18 years of age living in the HAI.
% aged 65+	Percentage of people over 65 years in the population.
Employment	Number of people in employment who live in the HAI.
% illiterate	Percentage of illiterate people in the HAI.
Road density	Road density of each HAI estimated by dividing road lengths by the area of the HAI in km/km^2.
Area	Area or size of the HAI in km^2.
Land use	
% residential	Percentage of residential areas in the HAI.
% health utilities	Percentage of health facilities areas in the HAI, such as clinics, hospitals (public and private), and GPs.

% educational utilities	Percentage of educational utilities areas in the HAI, such as universities, colleges, institutes, and schools (primary, intermediate and high schools).
% religious services	percentage of religious areas in the HAI, such as mosques
% cultural utilities	Percentage of cultural areas in the HAI, such as libraries, museums, exhibition halls, and any social gathering halls.
% agricultural	Percentage of agricultural areas in the HAI, such as palm trees farms and all activities related to agriculture.
% industrial	Percentage of industrial areas in the HAI.
% transport utilities	Percentage of transport utilities areas in the HAI, such as car parks, transport stations in and outside the city, and taxis and limousines stations.
% communications and % public utilities	Percentage of communications and public utilities areas in the HAI, such as transmitting stations for land line telephones or wireless communications, radio and TV stations, and public utilities like electricity, gas, water, sewage disposal, and water treatment plants.
% recreation and parks	Percentage of the recreation and parks areas in the HAI.

5.6 Summary

This chapter has discussed the data to be employed in the following chapters. This includes crash data involving details of the crash, participant data, and vehicle data.
The data have been checked and validated for quality for fatal and serious injury crashes, but not for slight injury crashes, which were under-reported. As slight injury crashes were under-reported the analysis will be based on 1) fatal, serious injury, and slight injury crashes, and 2) binary (fatal and serious) which will be discussed in the next chapter.

Data were validated to ensure their accuracy and quality by performing statistical tests. It was concluded that the data used in this thesis are reliable and of good quality.
The descriptive statistics of the variables such as number of crashes, age, and nationality in both crash severity and crash frequency models were also presented. Potential variables which will be used in severity analysis were also presented, as well as population land use and road network data which will be used in frequency analysis in Chapter 7.

6 CRASH SEVERITY MODELS

6.1 Introduction

The aim of this chapter, and one of the objectives of this thesis, is to identify the factors affecting the severity of traffic crashes that have occurred within Riyadh city; appropriate statistical models have been developed in an attempt to explore the relationship between crash severity and contributing factors. Crash severity refers to the level of severity of a road crash outcome (e.g. fatal, serious or slight).

In order to conduct this analysis, different econometric models were employed for modelling crash severity at the individual crash level. Crash severity is usually measured in categories (i.e. fatal, serious injury and slight injury). Therefore, an econometric model suitable for categorical data is needed.

Here two techniques are applied:

- Ordered response models:
    - Ordered logit model (OLM)
    - Generalised ordered logit (GOLOGIT) model
    - Partially constrained generalised ordered logit (PC-GOLOGIT) model
- Unordered nominal response models:
    - Multinomial logit (MNL) model
    - The mixed logit (ML) model
    - The mixed binary logistic regression model.

In Table 5.8 in Chapter 5 are shown the potential independent variables which will be used in the severity models.

A number of hypotheses were formulated on the basis of the literature review in Chapter two and local knowledge:

- Younger drivers are associated with more severe crashes.
- Non-Saudi drivers are involved in more severe crashes as they are less familiar with local conditions.
- Excessive speed is positively associated with the severity of the crashes.

- Single-vehicle crashes are more severe than multi-vehicle crashes.
- Crash severity decreases over time.
- Time period 16:00 to 19:59 has more severe crashes as people go out for shopping and recreation.

When the sign of coefficient and t-statistics is positive, this means the crash is more likely to be severe, whereas when it is negative it is less likely.

6.2 Results from the Ordered Response Models

As severity outcomes are ordinal in nature, different ordered response models (ORM) have been developed, in order to examine the factors affecting crash severity in Riyadh city. Crash severity is coded as follows: 1 for a fatal crash, 2 for a serious injury crash and 3 for a slight injury crash.

Three ordered models have been estimated: an ordered logit model (OLM), a generalised ordered logit (GOLOGIT) model, and a partially constrained (PC-GOLOGIT) model; these are presented in Table 6.1.

As discussed in Chapter 3, to relax the proportional odds assumption, two additional ordered response models have been estimated: a generalised ordered logit (GOLOGIT) model can be employed, which will have different coefficients across different severity outcomes, and a partially constrained model. In the OLM model, coefficients for equations $y>1$ and $y>2$ are constrained to be the same, while the coefficients in GOLOGIT and PC-GOLOGIT models are allowed to vary across different equations. For the PC-GOLOGIT model, a series of Wald tests are carried out on each variable individually to see whether its coefficients differ across equations. If the Wald test is statistically insignificant for one or more variables, the variable with the least significant value on the Wald test is constrained to have equal effects across equations. The model is then re-estimated with constraints, and the process is repeated until there are no more variables that meet the parallel lines assumptions (Williams, 2006).

The results of the generalised ordered logit (GOLOGIT) model and partially constrained (PC-GOLOGIT) model, including the OLM, are shown in Table 6.1.

Table 6-1 Model estimation results for the OLM, PC-GOLOGIT and GOLOGIT

	(OLM)		(PC-GOLOGIT)		(GOLOGIT)	
Explanatory Variable	**Coefficient**	**t-statistic**	**Coefficient**	**t-statistic**	**Coefficient**	**t-statistic**
			Threshold between slight and serious injury crashes (y>1)			
Age of driver	0.0170	4.17	0 .0166	4.19	0.0081	0.85
Nationality of participant	0.3500	3.85	0.3503	3.84	0.1938	0.94
Time of day						
00:00 to 03:59 (Reference)						
04:00 to 07:59	0.1806	1.28	0.1846	1.30	0.2567	0.76
08:00 to 11:59	0.0094	0.07	0.0054	0.04	0.0455	0.15
12:00 to 15:59	0.0348	0.24	0.0372	0.26	0.3807	1.13
16:00 to 19:59	-0.2584	-1.82	-0.2556	-1.80	0.1711	0.55
20:00 to 23:59	-0.0444	-0.28	-0.0497	-0.31	0.4494	1.17
Cause of the crash						
Other cause (Reference)	-0.2015	-1.79	-0.2096	-1.85	-0.1029	-0.38
Overtaking	-0.13090	-1.07	-0.1310	-1.07	-0.4066	-1.55
Distraction	0.3575	2.67	0.3471	2.63	0.1525	0.49
Excessive speed						
Type of collision						
Other collision types (Reference)	-0.05143	-0.40	-0.0319	-0.24	-0.7014	-1.88
Angle	-0.6762	-4.27	-1.5960	-5.75	-1.9461	-5.68
Head-on	-0.4073	-3.23	-1.1625	-4.16	-1.5977	-4.63
Rear-end	-0.2128	-1.61	-0.8981	-3.17	-1.3101	-3.73
Side-swipe						
Location of the crash	0.1729	1.24	0.1724	1.26	-0.0411	-0.15
Road surface	-1.2731	-1.90	-1.2540	-1.91	0.0564	0.05
Lighting condition of the road	-1.1309	-3.37	-1.1074	-3.32	-0.4851	-0.47
Single vehicle crash	0.6496	6.91	-0.0749	-0.37	0.0720	0.32
Crash year						
Year 1425 (Reference)						
Year 1426	0.0919	0.74	0.1029	0.86	0.2883	1.28
Year 1427	0.2718	2.02	1.4901	4.72	1.5994	4.72
Year 1428	0.7284	5.14	1.5967	4.49	1.7468	4.59
Year 1429	0.7384	4.20	3.2813	3.23	3.4258	3.32
Number of casualties	0.1737	6.14	0.1760	6.19	0.2810	2.81
Constant			4.5104	5.70	2.9205	1.85
			Threshold between serious injury crashes and fatal crashes (y>2)			
Age of participant			0.0166	4.19	0.1821	4.39
Nationality of participant			0.3503	3.84	0.3824	3.91
04:00 to 07:59			0.1846	1.30	0.1630	1.11
08:00 to 11:59			0.0054	0.04	0.0008	0.01
12:00 to 15:59			0.0372	0.26	-0.0230	-0.15
16:00 to 19:59			-0.2556	-1.80	-0.3588	-2.33
20:00 to 23:59			-0.0497	-0.31	-0.1444	-0.85
Overtaking			-0.2096	-1.85	-0.2315	-1.90
Distraction			-0.1310	-1.07	-0.0613	-0.47
Excessive speed			0.3471	2.63	0.3827	2.76
Angle			-0.0319	-0.24	0.0307	0.23
Head-on			-0.4325	-2.57	-0.4022	-2.37
Rear-end			-0.2869	-2.14	-0.2450	-1.81
Side-swipe			-0.0975	-0.71	-0.0518	-0.37
Location of the crash			0.1724	1.26	0.2453	1.59

Road surface			-1.2540	-1.91	-1.4547	-2.26
Lighting condition of the road			-1.1074	-3.32	-1.1112	-3.31
Single vehicle crash			0.7746	7.81	0.7673	7.68
Year 1426			0.1029	_ 0.86	0.03589	0.27
Year 1427			0.0655	0.46	0.02654	0.18
Year 1428			0.5921	4.10	0.5490	3.71
Year 1429			0.5434	2.97	0.5081	2.72
Number of casualties			0.1760	6.19	0.1696	5.78
Constant			-0.4818	-0.63	-0.3612	-0.48
Other Statistics						
Log-likelihood value	-2002.79		-1971.445		-1961.27	
Pseudo R-squared	0.0486		0.0635		0.0683	
Observations	2,946		2,946		2,946	

It is clear from the results of the OLM that age, nationality, excessive speed, number of vehicles involved, year, number of casualties, type of collision (head-on or rear-end), and lighting conditions are all significant. The interpretation of the results is as follows:

Age of the driver: The hypothesis is that younger drivers are associated with more severe crashes. However, the model shows that, as the age of the participant increases, the severity of the crash also increases. This result does not confirm the hypothesis.

Nationality of the driver: The results showed that Saudis were positively associated with the severity of crashes, compared to non-Saudis; again, this does not confirm the hypothesis. This may be due to the fact that non-Saudi drivers are expatriates (employed as family or company drivers) and may drive carefully out of fear of losing their jobs.

Time of day: As data on traffic flow are unavailable, the time-of-day variable was used as a proxy for traffic flow. This categorical variable consists of six categories and the time period 00:00 to 3:59 was used as a reference case. In relation to this time period, none of the five time periods was significant in the severity of injury crashes, except the time period 16:00 to 19:59 (which is negatively significant, with a 90% confidence level, relative to the reference time period).

Cause of the crash: The results indicated that excessive speed has the biggest impact on the severity of traffic crashes, and this was found to be statistically significant compared to the category of ‘other’ causes, which primarily consists of violating the priority at a junction, illegal parking, and tiredness. This confirms the hypothesis stated above. However, overtaking was found to be negatively significant, with a confidence level of 90%, and distraction was found to be insignificant, compared to the 'other' category.

Type of collision: The type of collision is a categorical variable with five categories, in which the ‘other’ category, which consists of ‘run-off-the-road’ collisions, was used as a reference case. It was clear from the results that head-on and rear-end collisions were statistically significant, with a 95% confidence level, and were found to be less severe, compared to the ‘other’ category. Angle and side-swipe collisions were found to be statistically insignificant.

Crash location: Location of the crash was measured using a dummy variable, and consisted of two categories, ‘straight’ and ‘other’, which included curve, intersections and gradient. The results suggested that crash location does not affect severity.

Road surface: With regard to road surface conditions, the model showed that a wet surface has a negative significant impact with a 90% confidence level compared to a dry road surface. This is because it does not usually rain in Riyadh city and when it rains drivers tend to drive carefully.

Lighting conditions: When the ‘without lighting’ condition of the road was used as a reference category, it was clear from the model that roads with adequate lighting conditions were associated with less severe crashes, when compared to roads without lights.

Single vehicle: The model showed that the variable representing single-vehicle crashes had a positively significant effect on severity compared to multi-vehicle crashes, which confirms the hypothesis above.

Time trend: Dummy variables representing different years were used to control time trends in the model. Five years of data were used in this model and the year AH1425 was taken as a reference case. In the first year, it seems that there was no statistically significant effect on crash severity when compared to the reference case, whilst the last three years showed a positively significant effect on severity, in comparison to the reference case. This result is inconsistent with the hypothesis above.

Number of casualties per crash: This variable had a highly positive significant effect on severity.

The weather condition variable was categorised as ‘fine’ and 'other', which included fog, rain, dust, snow and cloud. After investigation it was decided to omit the weather variable in all models because 99.67% of the crashes occurred during 'fine' weather. In addition, after an initial model was tested, the weather conditions variable was not significant in any models, so excluding it seems to have little effect.

The developed ordered logit model is used to estimate how the probabilities of different crash severities vary with age (Figure 6.1):

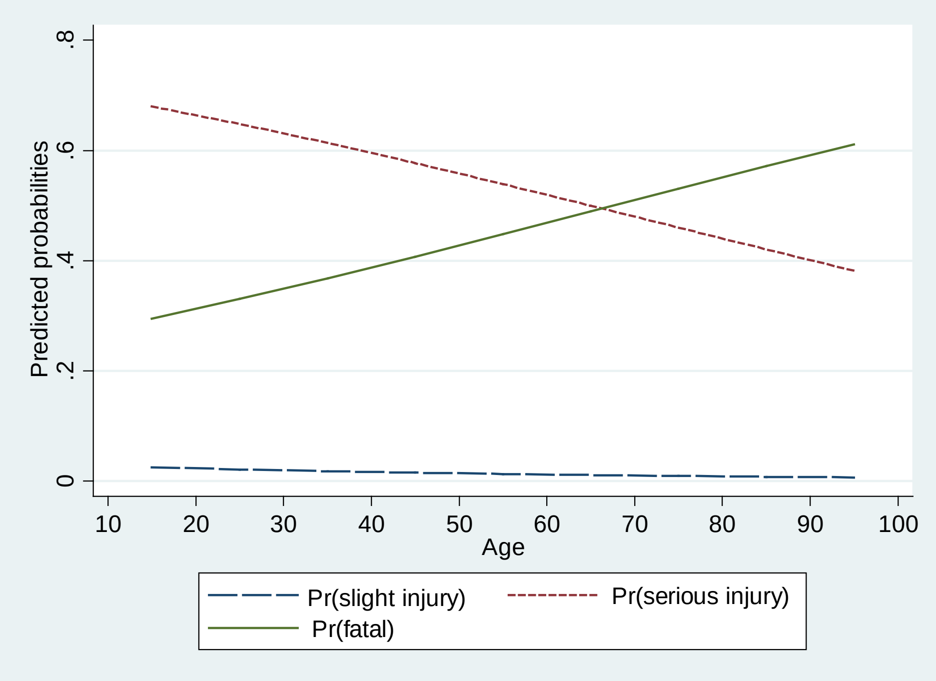

Figure 6-1 Probabilities of different degrees of crash severity vs age

The above figure shows that age increases the probability of fatal crashes, while the likelihood of serious injury decreases. Age has little impact on the probability of slight injury crashes occurring.

Table 6.1 shows that the coefficients of all independent variables are different across thresholds for the GOLOGIT model, whereas only the coefficients of head-on, rear-end, and side-swipe collisions, single-vehicle crashes, and the years 1427, 1428 and 1429 were different across thresholds for the PC-GOLOGIT model, which did not meet the proportional odds assumption.

Although the value of the likelihood ratio (LR) Chi square was higher in the GOLOGIT model than in the PC-GOLOGIT model (see Table 6.2), the difference was not statistically significant: the value in the GOLOGIT model was only 10.2 units more for 7 degrees of freedom, which is less complex (parsimonious). This suggests that the model goodness-of-fit is better in the PC-GOLOGIT model than in the GOLOGIT model. The PC-GOLOGIT model will be used to interpret the effects of the independent variables on crash severity.

Table 6-2 Summary of the three models

Model	Number of parameters	Log likelihood	Pseudo R^2
Ordered Logit Model (OLM)	23	-2002.79	0.0486
Partial Proportional (PC-GOLOGIT)	30	-1971.445	0.0635
Generalised Ordered Logit (GOLOGIT)	46	-1961.2651	0.0683

The interpretation of variables influences the cut-off points (thresholds), which mean that the coefficient value of the following variables differs across thresholds:

Head-on: This variable was found to be statistically significant, with a 95% confidence level in all models. In general, it was found that, compared to the 'other' type of collisions, ‘head-on’ collisions were less severe. It is noticeable that the percentage of coefficients is different between the thresholds. The negative in the threshold suggests that it is likely that higher values in head-on crashes increase the likelihood of such crashes yielding slight injuries.

Rear-end: The threshold between slight and serious injury crashes (y>1) was found to be statistically significant; compared to the 'other' category of collisions, rear-end collisions were less severe, with a 95% confidence level. At the threshold between serious injury crashes and fatal crashes (y>2), this was only significant in the PC-GOLOGIT model. It is noticeable that the value of the coefficient differed between the thresholds. The negative in the threshold suggests that it is more likely that higher values associated with head-on collisions increase the likelihood of slight injury.

Side-swipe: In terms of the threshold between slight and serious injury crashes (y>1), it was found to be statistically significant, when compared to the 'other' category of collisions: side-swipe collisions yield less severe crashes, with a 95% confidence level, whereas the threshold between serious injury crashes and fatal crashes (y>2) was not significant in all models. It is noticeable that the value of the coefficient differed between the thresholds. The negative in the threshold suggests that it is more likely that higher values associated with head-on collisions increase the likelihood of slight injury.

Single-vehicle crash: This was found to be positively associated with the severity of crashes, in the threshold between serious injury crashes and fatal crashes (y>2): this means that a single-vehicle crash is likely to be more severe than a multi-vehicle crash.

Year 1427: For the threshold between slight and serious injury crashes (y>1), this was found to be statistically significant, when compared to year 1425, and yielded more severe crashes, with a 95% confidence level. The threshold between serious injury crashes and fatal crashes (y>2) was not significant for all models. It is noticeable that the value of the coefficient differed between the thresholds, which mean that the effect on severity of crashes which occurred in the year 1427 was not uniform.

Years 1428 and 1429: These variables were found to be statistically significant, with a 95% confidence level in all models. It was found that, compared to the year 1425; these years yielded more severe crashes.

The other variables, which are constrained to be the same across different severity outcomes as they met the proportional odds assumption, are interpreted in the same way as in the OLM.

Because the crash data have severe under-reporting of slight injury crashes, as discussed in Chapter 4, in terms of slight injury crashes, the results may be biased. Therefore, more advanced statistical models are applied to overcome this problem, i.e., initially, the multinomial logit (MNL) and mixed logit (ML) models.

6.3 Results from the Unordered Nominal Response Models

The purpose of this research was to identify the factors affecting the severity of traffic crashes, through the development of appropriate statistical models. Nominal response models have the advantage of flexibility, in terms of under-reported data; as a result, nominal response models may be appropriate for such data, as they are capable of providing more robust estimation results. Two nominal response models have been estimated: a standard multinomial logit model (MNL) and a mixed logit model (ML).

The mixed logit model can take into account the unobserved correlated effects and the unobserved heterogeneity between different crash severity categories (Milton et al., 2008). Accordingly, the mixed logit model is expected to provide more coherent results. The multinomial logit model and the mixed logit model have been developed with three levels of severity: slight, serious and fatal. The results are presented in Table 6.3. At present, insignificant explanatory variables are retained in the models, as the primary purpose of this study is not to predict or forecast crash severity; rather, it is to identify the factors affecting this. Coefficients in the mixed logit model are considered to be random parameters if they produce statistically significant standard deviations. In this study, the results for the mixed logit model were obtained using 150 Halton draws, because the sample size was quite large. Train (2003) suggested that the estimation of a parameter in the mixed logit model would be more consistent in the maximum simulated likelihood (MSL), if a high number of Halton draws could be used. An initial test has suggested that numbers above 100 produce reasonably stable estimates and the results are generally consistent (between 100 and 150 draws), in terms of the set of significant estimators. Haan and Uhlendorff (2006) also showed that 100-150 Halton draws may be sufficient for stable results.

Serious injury crashes were used as the base outcome. Every variable was tested for random effect; coefficients in the mixed logit model are considered to be random parameters if they produce a statistically significant standard deviation, and only two variables (namely nationality and crash location) were found to have random effects in the mixed logit model, in which random parameters were assumed to be normally distributed. As expected, the Akaike Information Criterion (AIC) for the mixed logit model is lower than that of the MNL model, suggesting that the mixed model fits the data better.

In Table 6.3, the estimation results for the categories of fatal and slight crash injuries are presented, with the category of serious injury being the base outcome. Broadly speaking, the estimation results for the MNL and mixed logit models were similar, in terms of the statistically significant variables and the signs of their coefficients. However, a couple of exceptions were noted: the variable ‘crash location’ in the slight injury function was insignificant in the MNL model, yet significant in the mixed logit model, with a 95% confidence level. Also, nationality specifically in relation to the fatal crash category was highly and statistically significant in the MNL model. When the coefficient of nationality was treated as a random parameter, the standard deviation was found to be statistically significant (at the 95% confidence level).

Table 6-3 Model estimation results for multinomial and mixed logit models

	MNL				**Mixed Logit**			
	Specific to slight injury		**Specific to fatal injury**		**Specific to slight injury**		**Specific to fatal injury**	
Variable	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Age of the participant	0.0090	0.59	0.0394	5.24	0.0037	0.12	0.0407	5.31
Nationality (fixed)	0.4043	0.61	1.26678	3.95	0.0451	0.03		
Time of day								
00:00 to 03:59 (Reference)								
04:00 to 07:59	-0.2186	-0.64	0.1389	0.94	-0.2822	-0.37	0.1307	0.75
08:00 to 11:59	-0.0548	-0.18	-0.0151	-0.10	0.0850	0.13	-0.0310	-0.18
12:00 to 15:59	-0.4106	-1.21	-0.0461	-0.30	-0.3985	-0.56	-0.0742	-0.41
16:00 to 19:59	-0.2661	-0.86	-0.3719	-2.39	-0.1339	-0.20	-0.4828	-2.45
20:00 to 23:59	-0.5082	-1.32	-0.1631	-0.94	-0.6072	-0.77	-0.2337	-1.12
Cause of the crash								
Other cause (Reference)								
Overtaking	0.0415	0.15	-0.2113	-1.73	-0.1995	-0.33	-0.2406	-1.70
Distraction	0.3760	1.43	-0.0379	-0.29	0.9568	1.61	-0.0288	-0.19
Excessive speed	0.0098	0.01	0.2591	0.49	0.5145	0.18	0.1786	0.29
Type of collision								
Other collision types (Reference)								
Angle	0.7386	1.99	0.0594	0.43	1.5280	1.75	0.0297	0.18
Head-on	1.8546	5.38	-0.3010	-1.75	3.9655	3.65	-0.3742	-1.75
Rear-end	1.5637	4.51	-0.1957	-1.43	3.6867	3.09	-0.2527	-1.52
Side-swipe	1.2593	3.57	-0.0331	-0.23	2.8786	2.61	-0.0557	-0.33
Location of the crash (fixed)	0.0600	0.21	0.2260	1.45			0.2906	1.55
Road surface	-0.5149	-0.45	-1.5221	-2.29	-0.0329	-0.01	-1.7078	-2.31
Lighting condition	0.3394	0.33	-1.1520	-3.40	1.0523	0.52	-1.4607	-3.17
Single-vehicle crash	0.1284	0.57	0.7631	7.57	0.1777	0.38	0.8941	6.22
Crash year								
Year 1425 (Reference)								
Year 1426	-0.2587	-1.13	0.0531	0.40	-0.2250	-0.48	0.0738	0.47
Year 1427	-1.5913	-4.67	-0.0298	-0.20	-3.7014	-2.53	-0.0036	-0.02
Year 1428	-1.5967	-4.16	0.5055	3.36	-4.0812	-2.54	0.6152	3.17
Year 1429	-3.3276	-3.23	0.4209	2.23	-8.1653	-2.19	0.5242	2.26
Number of casualties	-0.2368	-2.32	0.1649	5.59	-0.3843	-1.91	0.1815	5.09
Age and nationality	-0.0191	-0.95	-0.0297	-3.27	-0.0166	-0.39	-0.0288	-2.84
Age and exspeed	-0.0130	-0.40	-0.0105	-0.79	-0.0576	-0.69	-0.0086	-0.53
Nationality and exspeed	0.5046	0.76	0.6120	2.04	1.0195	0.69	0.8458	2.17
Constant	-2.7835	-1.68.	-0.8708	-1.08	-4.9884	-1.34	-0.5694	-0.62
Random parameters								
Nationality								
Mean							0.9102	1.96
Standard deviation							1.3544	2.03
Location of the crash								
Mean					-5.8102	-1.72		
Standard deviation					5.6787	2.26		
Model statistics:								
Log-likelihood	-1951.7126				-1945.2195			
AIC	4011.425				4002.439			
Observations (N)	2,828				2,828			

The values of the coefficients, however, differed between these two models: for some variables associated with fatal crashes, such as age, angle, collision type and distraction as a cause, the difference was small. For other variables associated with slight injury crash injuries, such as distraction and excessive speed as causes, angle and head-on collision types and lighting conditions, the difference was more marked. This is expected because they are different models giving different results, but the AIC value is lower in ML than MNL, so ML is more robust. In addition, the value of log-likelihood in ML is higher than in MNL.

The interpretations of the variables are presented below:

Age of the driver: Both models suggest that, if the age of the driver increases, the probability of having a fatal crash increases compared to the probability of a serious injury crash (at the 95% confidence level), whilst there is no effect of age in slight injury crashes compared to serious injury crashes. This means that age is positively associated with the probability of fatal crashes, which suggests that both models are consistent. Different functional forms of this variable were also investigated, including both linear and quadratic functions of age and age as a categorical variable. In all models, increased age was found to be associated with more severe crashes. Although it seems that the result is surprising, it can be speculated that if, the age of the driver increases, the 'resilience' against the released energy decreases; this is in line with existing studies (Lardelli-Claret et al., 2009).

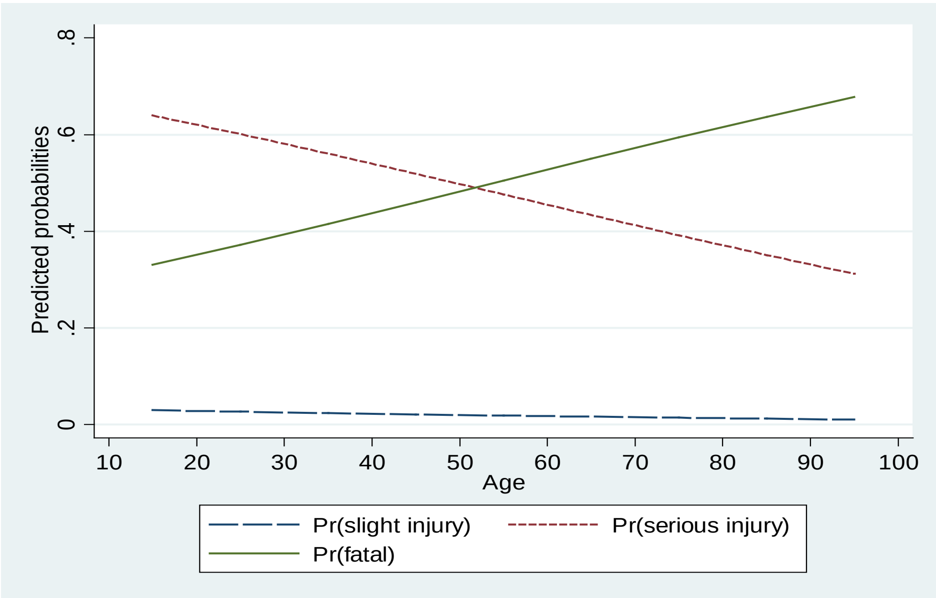

Figure 6-2 Probabilities of different degrees of crash severity vs age

Figure 6.2 shows how the probability of having a specific category of crash changes with an increase in the age of the driver, using the estimated mixed logit model.

Age has little impact on the probability of slight injury crashes occurring. In terms of fatal crashes, it was noted that the probability of a fatal crash occurring increases with age, while the likelihood of a car crash causing serious injury decreases in comparison to a fatal crash, which is consistent with the OLM.

Nationality of the driver: In the MNL model, Saudis were found to be positively associated with fatal crashes (compared to serious injury crashes), when compared to non-Saudis. Nationality had no significant effect on slight injury crashes, in comparison to serious injury crashes. In the mixed logit model, the coefficient for nationality in the fatal crash function was normally distributed, with a mean of 0.9102 and a standard deviation of 1.3544: both were significant (at the 95% confidence level), suggesting that the effect of nationality varies over the sample of drivers. With the estimated parameters, 25% of the distribution is less than 0 and 75% of the distribution is greater than 0, which suggests that a quarter of the Saudi drivers are negatively associated with fatal injury crashes while the rest of the Saudi drivers are positively associated with fatal crashes (compared to serious injury crashes) (see Figure 6.3). The model is likely to capture the complex driving behaviour with respect to driver demography and crash severity. This result suggests that the effect of nationality on injury crash severity cannot be assumed to be uniform (as suggested by the MNL model) across drivers. However, compared with non-Saudi drivers, the majority of Saudi drivers (75%) are found to be associated with fatal crashes (compared to serious injury crashes).

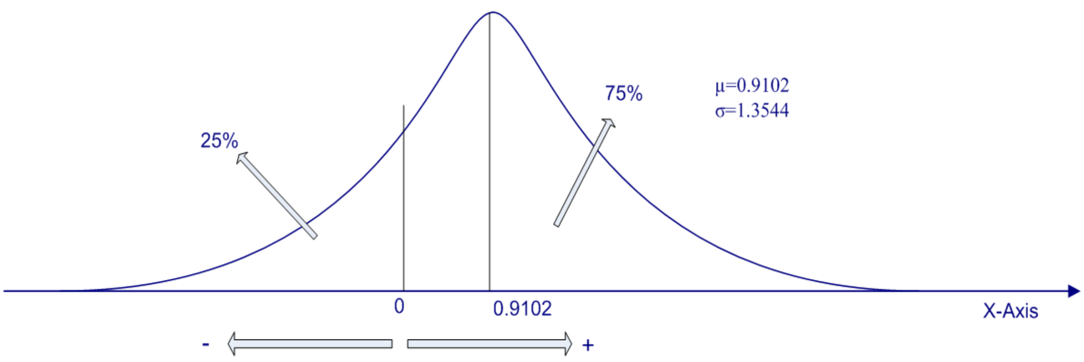

Figure 6-3 Distribution of nationality effect

Time of day: This variable is examined as a categorical variable with six categories, and the time period from midnight to 4 am is the reference category. Both models show that, compared to the reference time period, none of the other five time periods is statistically significant in either the fatal crash function or the slight injury function compared to the serious injury function, except 16:00 to 19:59, which is statistically significant in the fatal crash function and negatively associated with fatal crashes compared to serious injury crashes. This result may capture the fact that in Riyadh city traffic speeds during this period are relatively low as this is a peak period (just after the end of office hours at 14.30) for shopping and recreation.

Cause of the crash: This variable was thought to be an important factor affecting the severity of traffic crashes. A categorical variable is used to examine the impact of causes on crash severity. The results indicate that none of the causes is significant for both fatal crashes and slight injury crashes compared to serious injury crashes, owing to the effect of interactions with excessive speed in the models.

Type of collision: Collision type is also investigated using dummy variables and has five categories. The base case is ‘other’ collision type, which mainly includes ‘unknown’ category and other types such as run-off-road or rollover collisions. The results indicate that the probability of a slight injury crash increases if the type of collision is other than the base category.

Crash location: This variable is examined as a dummy variable with two categories: straight road sections and other sections including curved roads and junctions. This variable is found to be significant in the slight injury category (at the 95% confidence level) whereas it was marginally statistically significant (at the 90% confidence level) in the mixed logit mode. The normally distributed coefficient in the slight injury function has a mean of -5.81 and standard deviation of 5.68, both of which are statistically significant (at the 85% confidence level). This suggests that 85% of the distribution is less than 0 and 15% of the distribution is greater than 0. For 85% of the time, the likelihood of a slight injury crash decreases if the crash occurs on a straight road section and for 15% of the observations the likelihood of a slight injury crash increases. The model is picking up the effect of spatial difference that exists in different parts of Riyadh’s road network.

Road surface: With regard to road surface conditions, the models show that a wet surface has a negative significant impact compared to a dry road surface in the fatal crash function compared to serious injury crashes. This is because rain is unusual in Riyadh city, and when it rains drivers tend to drive carefully.

Lighting conditions: Roads with lighting show a negative effect compared to roads without lighting for fatal crashes compared to serious injury crashes, while there is no effect for slight injury crashes compared to serious injury crashes.

Single vehicle: The number of vehicles involved in a crash may have an impact on the severity outcome. This has been examined using a dummy variable: a single-vehicle crash (e.g. hitting an object or pedestrians) and multi-vehicle crashes. The results indicate that the probability of a fatal crash increases if a crash is a single-vehicle crash. This is consistent with the finding of existing studies from developed countries (Quddus et al., 2009).

Time trend: The effect of time on crash severity was examined using a categorical variable in which crashes which happened in AH1425 are used as a reference category. The results show that the coefficient of year 1429 is high and the z value is 0.00 because of high missing values in this variable for both fatal crashes and slight injury crashes.

Number of casualties: If the number of casualties is high then it is more likely that the crash is a fatal crash than a serious injury crash.

As mentioned previously, existing studies suggest that the MNL and mixed logit models are capable of handling the problem of under-reporting associated with crash data (Yamamoto et al., 2008; Milton et al., 2008). However, this is not the case for the data employed in this research, mainly because the crash data are affected by a severe problem of under-reporting, in terms of slight injury crashes (only 3.8% of the total crashes were reported as slight injury crashes). The reporting of fatal and serious injury crashes is more reliable, as the Saudi government’s road safety strategy focuses on serious and fatal crashes only; as a result, binary and mixed binary logistic regression models were developed for these two categories of severity. Table 6.4 presents the results from these models. Serious injury crashes are used as the base outcome.

Because only 111 out of 2,946 total observations were slight injury crashes, the results of the binary logit models were not significantly different from those associated with the MNL and mixed logit models. Serious injury crashes were used as a base outcome in both datasets, and the results were fairly consistent, in terms of the set of explanatory variables and the magnitudes of the coefficients. Therefore, the interpretations of these binary models would be the same as those of the multinomial response models presented above.

Table 6-4 Binary logistic and mixed binary logistic regression models

Variable	Binary logit		Mixed binary logit	
	Coefficient	t-stat	Coefficient	t-stat
Age of the participant	0.0397	5.29	0.0406	5.31
Nationality(fixed)	1.2673	3.95		
Time of day				
00:00 to 03:59 (Reference)				
04:00 to 07:59	0.1242	0.84	0.1230	0.72
08:00 to 11:59	-0.0232	-0.16	-0.0356	-0.21
12:00 to 15:59	-0.0640	-0.42	-0.0853	-0.49
16:00 to 19:59	-0.3930	-2.52	-0.4780	-2.44
20:00 to 23:59	-0.1809	-1.05	-0.2371	-1.16
Cause of the crash				
Other cause (Reference)				
Overtaking	-0.2181	-1.78	-0.2410	-1.74
Distraction	-0.0242	-0.18	-0.0206	-0.14
Excessive speed	0.2630	0.50	0.1987	0.33
Type of collision				
Other collision types (Reference)				
Angle	0.0690	0.50	0.0444	0.28
Head-on	-0.2909	-1.69	-0.3506	-1.70
Rear-end	-0.1908	-1.39	-0.2349	-1.45
Side-swipe	-0.0166	-0.12	-0.0369	-0.23
Location of the crash (fixed)	0.2360	1.51	0.2813	1.53
Road surface	-1.5957	-2.37	-1.7120	-2.34
Lighting condition	-1.1510	-3.38	-1.4001	-3.07
Single-vehicle crash	0.7735	7.66	0.8771	5.99
Crash year				
Year 1425 (Reference)				
Year 1426	0.0290	0.22	0.0572	0.38
Year 1427	-0.0363	-0.25	-0.0142	-0.09
Year 1428	0.4878		0.5813	3.08
Year 1429	0.4095	3.24	0.4953	2.20
Number of casualties	0.1653	5.6	0.1780	5.08
Age and nationality	-0.0297	-3.27	-0.0291	-2.94
Age and exspeed	-0.0110	-0.82	-0.0092	-0.59
Nationality and exspeed	0.6222	2.08	0.8035	2.09
Constant	-0.8056	-0.99	-0.5873	-0.65
Random parameters				
Nationality				
Mean			0.9947	2.19
Standard deviation			1.1788	1.64
Model statistics:				
Log-likelihood	-1509.3577		-1507.0692	
AIC	3072.715		3070.138	
Observations (N)	2,828		2,828	

6.4 Interaction between Variables

On the basis of the literature review, the key hypotheses were formulated, for example that younger drivers are associated with more severe crashes.

As shown in the results of the different models, it is suggested that age is positively associated with the probability of fatal crashes, which means that, if the age of the participant (the driver who has 100% of the blame for the crash) increases, then the severity of the crash also increases, and the age variable is statistically significant, with a confidence level of 95%. This unusual result does not confirm the hypothesis derived from the literature review on developed countries. Thus, it has been suggested that interactions between different variables should be explored.

From the results of the estimated models it was found that age and nationality were always significant and that they affect severity. The review of literature indicates that two common factors affecting the severity of traffic crashes are age and nationality of the drivers at fault in the crashes.

Interactions between variables were tested, in order to investigate the relationship between them (by looking at how they interact with each other and how they affect the severity of other variables, using different severity models). The interactions between age and nationality, age and excessive speed, and nationality and excessive speed were included in the multinomial logistic regression model, the mixed logit model, the binary logistic regression model, and the mixed binary logit model as random parameters, and the results are interpreted as follows:

Age and nationality: Models show that the interaction between age and nationality were found to be negatively associated with fatal crashes relative to serious injury crashes. It has no significant effect when slight injury crashes compared to serious injury crashes. The model suggests that, if the age of the driver increases, the probability of Saudis having a fatal crash decreases compared to non-Saudi drivers (at the 99% confidence level), whilst there is no effect of age in slight injury crashes compared to serious injury crashes for both nationalities. This means that the probability of non-Saudis having a fatal crash increases more rapidly with respect to age compared with Saudi drivers (see Figure 6.4). Interaction between age as a categorical variable (four categories) and nationality as a dummy variable was also tested; the age category between 14 and 24 was taken as a reference case. MNL and ML models were estimated and it was found that the category between 40 and 64 years of age was positively associated with crash severity in both models compared to the reference category (at the 95% confidence level), which means that this age category is associated with the probability of having a fatal crash relative to the reference case in the Saudi drivers compared to the non-Saudi drivers, whilst there is no effect in slight injury crashes for all categories in both models compared to serious injury crashes.

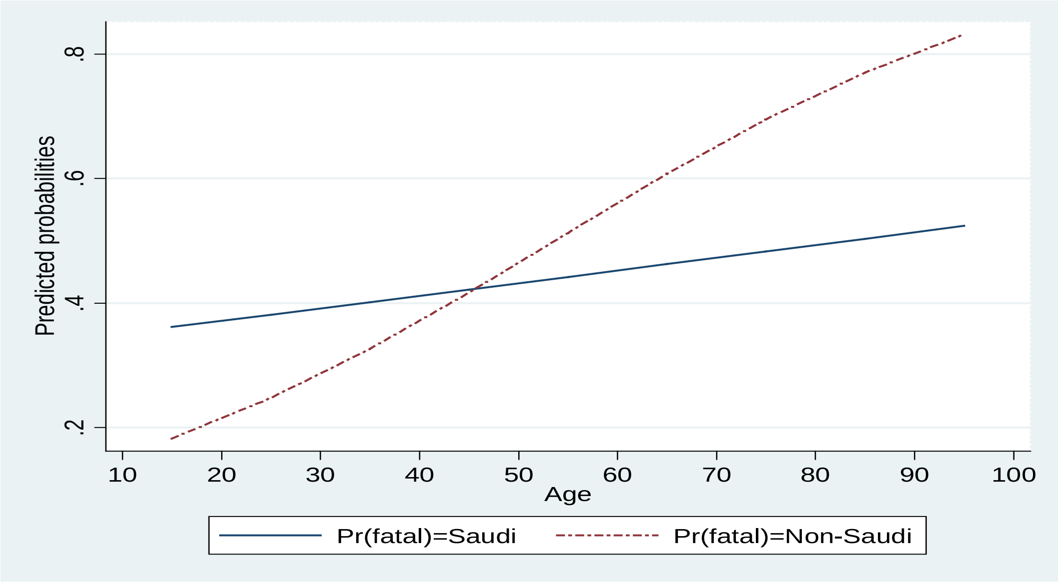

Figure 6-4 Probabilities of fatal crashes vs age by nationality

Age and excessive speed: This shows the interaction between age and excessive speed, which is found to be insignificant for fatal crashes and slight injury crashes, compared to serious injury crashes in all models.

Nationality and excessive speed: Interaction between nationality and speed was found to be positively associated (at the 95% confidence level) with the probability of fatal crashes whilst there is no effect in slight injury crashes due to excessive speed compared to serious injury crashes. Saudis are more likely to be involved in fatal crashes rather than serious injury crashes compared to non-Saudi drivers.

Road Density: It is expected that road density may also influence the severity of traffic crashes as vehicle average speeds are more likely to be low in areas with high road density owing to high traffic volumes and more pedestrian activity. This allows ward-level road density data to be obtained for each of the crashes.

The variable of road density has not been included in the models reported thus far as it did not exist in the dataset. However, a hypothetical road length was taken in each Hai, with its area and the crash coordinates, using the Map-Info GIS package. A road density variable was created by dividing the road length by the area, in order to obtain the road density, which can be merged into the STATA dataset and then included in the models with other variables.

This variable has 20% missing values because some crashes do not have geographical coordinates, which means that these crashes do not have a location when superimposed on the geocoded data.

In all models, road density has a negative effect for fatal crashes, while there is no effect for slight injury crashes compared to serious injury crashes (see Table 6.5).

Both models suggest that the probability of having a fatal crash decreases in comparison to serious injury crashes (at the 95% confidence level); whilst there is no effect of road density in slight injury crashes compared to serious injury crashes. This means that higher road density is associated with fewer fatalities and less severe crashes.

The results were fairly similar to those from models without road density (Table 6.3) in terms of the statistically significant variables and the signs of their coefficients, with the category of serious injury crashes being the base outcome. A few exceptions are noticed, however. The level of significance was higher in most of the variables in the models without road density (i.e. age, head-on, rear-end, side-swipe, lighting, single vehicle, and number of casualties). Random variables (nationality and location of the crash) were significant (at the 95% confidence level) in the models without road density whereas they were insignificant in models with road density.

Table 6-5 Model estimation results for multinomial and mixed logit models with road density

Variable	MNL				Mixed Logit			
	Specific to slight injury		Specific to fatal injury		Specific to slight injury		Specific to fatal injury	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Age of the participant	-0.003	-0.18	0.03604	4.22	-0.0193	-0.47	0.0368	4.24
Nationality(fixed)	-0.1836	-0.25	1.2714	3.47	-2.4928	-1.09		
Time of day								
00:00 to 03:59 (Reference)								
04:00 to 07:59	-0.2352	-0.61	-0.0162	-0.09	-0.5566	-0.5	-0.0208	-0.11
08:00 to 11:59	-0.0361	-0.11	-0.1767	-1.05	0.1522	0.17	-0.2217	-1.15
12:00 to 15:59	-0.1877	-0.52	-0.1748	-1.01	0.3304	0.33	-0.2172	-1.08
16:00 to 19:59	-0.1974	-0.57	-0.373	-2.14	-0.1345	-0.15	-0.458	-2.15
20:00 to 23:59	-0.9148	-1.87	-0.2933	-1.5	-2.9577	-1.39	-0.3749	-1.59
Cause of the crash								
Other cause (Reference)								
Overtaking	0.19332	0.65	-0.1675	-1.19	-0.0769	-0.08	-0.1931	-1.22
Distraction	0.40003	1.36	0.0031	0.02	1.3788	1.66	0.0168	0.10
Excessive speed	-0.2373	-0.17	0.2788	0.48	1.7964	0.32	0.207643	0.31
Type of collision								
Other collision types (Reference)								
Angle	0.3817	0.96	0.1246	0.82	0.6931	0.51	0.125	0.72
Head-on	1.4615	4.01	-0.2799	-1.45	4.5722	2.8	-0.3379	-1.48
Rear-end	1.2223	3.37	-0.1708	-1.06	4.0124	2.59	-0.1864	-1.02
Side-swipe	0.9799	2.66	0.0501	0.31	3.7710	2.02	0.0449	0.25
Location of the crash (fixed)	0.2315	0.72	0.2931	1.71			0.3447	1.73
Road surface	15.9273	0.01	-1.5464	-2.18	21.1207	0.04	-1.5891	-2.07
Lighting condition	16.2472	0.01	-1.1358	-2.52	24.8418	0.04	-1.3367	-2.47
Single-vehicle crash	0.0681	0.28	0.6357	5.54	0.2425	0.37	0.7241	4.84
Crash year								
Year 1425 (Reference)								
Year 1426	-0.0913	-0.36	0.0944	0.67	0.1211	0.17	0.1251	0.77
Year 1427	-1.5533	-4.02	0.0025	0.02	-5.0564	-1.95	0.0176	0.1
Year 1428	-1.2423	-3.01	0.493943	2.98	-5.3294	-1.59	0.5873	2.85
Year 1429	-31.2937	0.00	-30.4266	0.00	-32.5809	0.00	-28.771	0
Number of casualties	-0.2416	-2.17	0.1155	3.57	-0.3615	-1.31	0.1226	3.28
Road density	-0.000024	0	-0.0362	-5.44	-0.0031	-0.08	-0.0393	-4.52
Age and nationality	-0.0046	-0.21	-0.0317	-3.05	0.0443	0.66	-0.0322	-2.87
Age and exspeed	-0.0154	-0.41	-0.0132	-0.91	-0.1536	-0.82	-0.0113	-0.66
Nationality and exspeed	0.9475	1.22	0.6504	1.96	2.4303	0.86	0.8192	1.96
Constant	-34.7622	.	-0.1209	-0.13	-50.0128	-0.06	0.0425	0.04
Random parameters								
Nationality								
Mean							1.023	2.05
Standard deviation							1.1643	1.51
Location of the crash								
Mean					-10.9929	-1.25		
Standard deviation					9.522	1.52		
Model statistics:								
Log-likelihood	-1546.5669				-1536.8493			
AIC	3203.134				3189.699			
Observations (N)	2,384				2,384			

Binary logistic and mixed binary logistic regression models with road density were also estimated (Table 6.6).

Table 6-6 Binary logistic and mixed binary logistic regression models with road density

Variable	Binary logit		Mixed binary logit	
	Coefficient	t-stat	Coefficient	t-stat
Age of the participant	0.03622	4.25	0.03658	4.24
Nationality(fixed)	1.2839	3.51		
Time of day				
00:00 to 03:59 (Reference)				
04:00 to 07:59	-0.0211	-0.12	-0.02399	-0.12
08:00 to 11:59	-0.2053	-1.21	-0.22111	-1.17
12:00 to 15:59	-0.1867	-1.07	-0.21751	-1.10
16:00 to 19:59	-0.3877	-2.23	-0.45292	-2.14
20:00 to 23:59	-0.3058	-1.56	-0.36883	-1.58
Cause of the crash				
Other cause (Reference)				
Overtaking	-0.1817	-1.29	-0.19157	-1.23
Distraction	0.01	0.07	0.01714	0.11
Excessive speed	0.2798	0.48	0.21583	0.33
Type of collision				
Other collision types (Reference)				
Angle	0.1325	0.87	0.12935	0.76
Head-on	-0.281	-1.46	-0.32588	-1.46
Rear-end	-0.169	-1.05	-0.18021	-1.00
Side-swipe	0.0402	0.25	0.04657	0.26
Location of the crash (fixed)	0.293	1.71	0.33284	1.69
Road surface	-1.5548	-2.20	-1.58437	-2.10
Lighting condition	-1.1449	-2.54	-1.30723	-2.44
Single vehicle crash	0.6509	5.66	0.71713	4.78
Crash year				
Year 1425 (Reference)				
Year 1426	0.094	0.66	0.11333	0.71
Year 1427	0.011	0.07	0.01035	0.06
Year 1428	0.5054	3.04	0.56826	2.80
Year 1429	.	.	-18.121	0.00
Number of casualties	0.1145	3.54	0.12112	3.30
Road density	-0.0347	-5.22	-0.03847	-4.42
Age and nationality	-0.0317	-3.06	-0.03199	-2.89
Age and exspeed	-0.013	-0.89	-0.01144	-0.68
Nationality and exspeed	0.6428	1.94	0.79808	1.91
Constant	-0.1357	-0.15	0.0242	0.02
Random parameters				
Nationality				
Mean			1.0609	2.15
Standard deviation			1.0563	1.30
Model statistics:				
Log-likelihood	-1172.8727		-1172.5572	
AIC	2399.745		2403.114	
Observations (N)	2,285		2,285	

The results were fairly similar to those of models without density in terms of the statistically significant variables and the signs of their coefficients, with the category of serious injury crashes being the base outcome. Again, a few exceptions are noted; the level of significance was higher in most of the variables in the models without road density (i.e. age, head-on, rear-end, side-swipe, lighting, single vehicle, and number of casualties). A random variable (nationality) was insignificant in models with road density whereas it became significant (at the 90% confidence level) in the models without road density.

The values of the coefficients in these two models are different, however: for some variables the difference is small, such as age, overtaking (cause type) and head-on (collision type) associated with fatal crashes.

The mixed binary logit model with road density has lower AIC and a larger value of log-likelihood than that without road density, and it has no value for year 1429, which has many missing values because some crashes do not have geographical coordinates. The mixed binary logit model without road density has lowered AIC and a larger value of log-likelihood than the MNL and mixed logit models (both with and without road density); the results are fairly consistent in terms of the set of explanatory variables and the magnitudes of the coefficients.

It is suggested that the mixed binary logit model without road density fits the data better and is found to be the better model for crash severity in Riyadh city.

6.5 Preferred model

For the purpose of simplifying the interpretation of the interaction between age and nationality, age was divided into four categories (14-24, 25-39, 40-64, and above 65), with the age group 14-24 taken as a reference case.

In the previous sections, MNL and ML models were estimated by taking the significant and insignificant variables in the models, whereas in this section only the significant variables will be retained in the models. This is because the simpler the model is, and the fewer variables there are, the more it is preferred. The method followed was to take out the insignificant variables one by one and observe the effect on other variables until the final model was achieved. The rule followed for level of confidence was 90%, with the exception of age and the interaction of age and nationality, as according to the literature they are the two commonest factors affecting the severity of traffic crashes.

Generally, MNL and ML were consistent in terms of results, as shown in Table 6.7, for both slight injury crashes and fatal crashes, except that the MNL model has a higher level of confidence for most of the variables. The results and findings from both models are interpreted below:

Age of the driver: Both models suggest that drivers aged between 40 and 65 have a higher probability of having a fatal crash than a serious injury crash (at the 95% confidence level). In all models, the age category 40-65 has been found to be associated with more severe crashes. The age group 25-39 has a higher probability of having a fatal or slight injury crash than a serious injury crash (at the 85-90% confidence level). The age group above 65 has a greater probability of having a fatal crash than a serious injury crash (at the 90% confidence level).

Age and nationality: Models show that Saudi drivers aged between 40 and 64 are less likely to be involved in fatal crashes than in serious injury crashes in comparison to the reference age group (at the 99% confidence level). Saudi drivers aged between 25 and 39 are less likely to be involved in slight injury crashes than in serious injury crashes in comparison to the reference age group, whereas Saudi drivers aged above 65 are no different from the reference age group in their involvement in fatal or slight injury crashes compared to serious injury crashes.

Nationality of the participant: In the MNL model, Saudis were found to be positively associated with fatal crashes (compared to serious injury crashes), when compared to non-Saudis. Nationality had no significant effect on slight injury crashes, in comparison to serious injury crashes. In the mixed logit model, the coefficient for nationality in the fatal crash function was normally distributed, with a mean of 0.3715 and a standard deviation of 1.2761: both were significant (at the 90% confidence level), suggesting that the effect of nationality varies over the sample of drivers. With the estimated parameters, 39% of the distribution is less than 0 and 61% of the distribution is greater than 0, suggesting that 39% of Saudi drivers are negatively associated with fatal injury crashes while the rest of the Saudi drivers are positively associated with fatal crashes (relative to serious injury crashes). However, compared with non-Saudi drivers, the majority of Saudi drivers (61%) are found to be associated with fatal crashes relative to serious injury crashes.

Time of day: Both models show that, in comparison to the reference time period, 16:00 to 19:59 in the fatal crash function which is statistically significant and negatively associated with fatal crashes compared to serious injury crashes. This result may reflect the fact that, in Riyadh city, traffic speeds during this period are relatively low as this is a peak period (just after the end of office hours at 14:30) for shopping and recreation.

Cause of the crash: This variable was thought to be an important factor affecting the severity of traffic crashes. A categorical variable is used to examine the impact of causes on crash severity. The results indicate that, in comparison to the category 'other causes', overtaking was found to be negatively significant in specific to fatal crashes (at the 90% confidence level) compared to serious injury crashes, whereas distraction was found to be positively significant in specific to slight injury crashes compared to serious injury crashes (at the 90% confidence level).

Type of collision: For both models, the results indicate that, compared to serious injury crashes, the probability of a slight injury crash increases for all types of collision if the type of collision is other than the base category. For head-on and rear-end collisions the probability of fatal crashes decreases compared to serious injury crashes (at the 90% confidence level).

Crash location: This variable was found to be significant in slight injury crashes compared to serious injury crashes (at the 95% confidence level) in the mixed logit model whereas it was insignificant in the MNL. The normally distributed coefficient in the slight injury function has a mean of -4.8188 and standard deviation of 4.8593, both of which are statistically significant (at the 95% confidence level). This suggests that 84% of the distribution is less than 0 and 16% of the distribution is greater than 0. For 84% of the time, the likelihood of a slight injury crash decreases if the crash occurs on a straight road section, and for 16% of the observations it increases. The model is picking up the effect of spatial difference that exists in different parts of Riyadh's road network.

Road surface: With regard to road surface conditions, the models show that a wet surface has a negative significant impact compared to a dry road surface in the fatal crash function in comparison to serious injury crashes. This is because rain is unusual in Riyadh city, and when it rains drivers tend to drive carefully.

Table 6-7 Model estimation results for multinomial and mixed logit models with categorical age

	MNL				Mixed Logit			
	Specific to slight injury		Specific to fatal injury		Specific to slight injury		Specific to fatal injury	
Variable	Coefficient	t-stat	Coefficient	t-stat	Coefficien	t-stat	Coefficient	t-stat
Age of the participant								
Age 14-24 (Reference)								
Age 25-39	1.0541	1.71	0.4033	1.53	1.5443	1.37	0.4057	1.53
Age 40-64	0.9387	1.42	1.0849	3.91	1.2035	0.98	1.1204	3.99
Age 65>	1.7764	1.42	1.2524	1.65	1.7379	0.73	1.2828	1.67
Age category and Nationality								
Age14-24 and Nationality (Reference)								
Age 25-39 and nationality	-1.3658	-2.00	-0.4157	-1.43	-2.2786	-1.78	-0.4215	-1.37
Age 40-64 and nationality	-1.0576	-1.34	-0.9994	-2.93	-1.2712	-0.87	-1.01426	-2.70
Age 65> and nationality	-1.710	-1.03	-0.3187	-0.37	-1.3397	-0.37	-0.0962	-0.10
Nationality (fixed)	0.9733	1.58	0.6796	2.62	1.3146	1.16		
Time of day								
00:00 to 03:59 (Reference)								
16:00 to 19:59			-0.3680	-2.97			-0.4524	-2.91
Cause of the crash								
Other cause (Reference)								
Overtaking			-0.1742	-1.65			-0.1892	-1.57
Distraction	0.3660	1.67			0.9824	1.93		
Type of collision								
Other collision types (Reference)								
Angle	0.7512	2.04			1.4886	1.92		
Head-on	1.8873	5.50	-0.3010	-1.95	3.8257	3.98	-0.3509	-1.89
Rear-end	1.5254	4.44	-0.2138	-1.87	3.2823	3.43	-0.2505	-1.88
Side-swipe	1.2990	3.71			2.6207	2.93		
Location of the crash (fixed)	0.0111	0.04						
Road surface			-1.3878	-2.19			-1.6176	-2.25
Lighting condition			-1.1605	-3.47			-1.4412	-3.18
Single-vehicle crash			0.7598	7.70			0.8771	6.32
Crash year								
Year 1425 (Reference)								
Year 1426	-0.3109	-1.38	0.0274	0.21	-0.2687	-0.62	0.0336	0.22
Year 1427	-1.6031	-4.79	-0.0262	-0.18	-3.4243	-3.06	-0.0181	-0.11
Year 1428	-1.5474	-4.13	0.5448	3.69	-3.3842	-2.97	0.6369	3.39
Year 1429	-3.2524	-3.18	0.4592	2.46	-7.5187	-2.48	0.5554	2.42
Number of casualties	-0.2744	-2.91	0.1621	5.56	-0.3865	-2.23	0.1784	5.08
Nationality and exspeed			0.6348	4.36			0.8285	3.27
Constant	-3.6576	-5.00	0.0108	0.01	-5.1693	-3.45	0.4040	0.46
Random parameters								
Nationality								
Mean							0.3715	0.97
Standard deviation							1.2761	1.92
Location of the crash								
Mean					-4.8188	-2.11		
Standard deviation					4.8593	2.86		
Model statistics:								
Log-likelihood	-1956.3792				-1951.8436			
AIC	3992.758				3987.687			
Observations (N)	2,946				2,946			

Lighting conditions: Roads with lighting have a negative effect compared to roads without lighting for fatal crashes in comparison to serious injury crashes.

Single vehicle: The number of vehicles involved in a crash may have an impact on the severity outcome. This has been examined using a dummy variable: a single-vehicle crash (e.g. hitting an object or pedestrians) and multi-vehicle crashes. The results indicate that the probability of a fatal crash increases if a crash is a single-vehicle crash. This is consistent with the finding of existing studies from developed countries (Quddus et al., 2009).

Time trend: The results show that, compared to the reference case (year AH1425), year 1428 and year 1429 were more likely to have fatal crashes compared to serious injury crashes but less likely to have slight injury crashes, whereas year 1427 was less likely to have slight injury crashes than serious injury crashes while there was no effect on fatal crashes. Year 1426 had no effect for both fatal and slight injury crashes compared to serious injury crashes.

Number of casualties: If the number of casualties is high then it is more likely that the crash is a fatal crash compared to a serious injury crash, while the number of casualties is negatively associated with slight injury crashes compared to serious injury crashes.

Nationality and excessive speed: Interaction between nationality and speed was found to be positively associated (at the 95% confidence level) with the probability of fatal crashes. Saudis are more involved in speed in fatal crashes compared to serious injury crashes than non-Saudi drivers.

Binary logistic and mixed binary logistic regression models (Table 6.8) were estimated; the results are fairly similar to those of the multinomial response models presented above in terms of the statistically significant variables and the signs of their coefficients, with the category of serious injury crashes being the base outcome.

Two exceptions are noticed, however: overtaking and lighting conditions became insignificant variables.

Table 6-8 Binary logistic and mixed binary logistic regression models with categorical age

Variable	Binary logit		Mixed binary logit	
	Coefficient	t-stat	Coefficient	t-stat
Age of the participant				
Age 14-24 (Reference)				
Age 25-39	0.4489	1.80	0.4718	1.88
Age 40-64	1.0248	3.91	1.0607	4.01
Age 65>	1.1222	1.54	1.3154	1.73
Age category and Nationality				
Age 14-24 and Nationality (Reference)				
Age 25-39 and nationality	-0.4220	-1.55	-0.4650	-1.68
Age 40-64 and nationality	-0.9272	-2.96	-0.9600	-2.97
Age 65> and nationality	-0.3573	-0.44	-0.4778	-0.56
Nationality(fixed)	0.6992	2.87		
Time of the day				
00:00 to 03:59 (Reference)				
16:00 to 19:59	-0.3245	-2.81	-0.3551	-2.55
Cause of the crash				
Other cause (Reference)				
Overtaking	-0.1097	-1.15	-0.1090	-1.08
Type of collision				
Other collision types (Reference)				
Head-on	-0.2553	-1.77	-0.1826	-1.16
Rear-end	-0.1666	-1.63	-0.1511	-1.35
Road surface	-1.3355	-2.11	-1.5484	-2.24
Lighting condition	-0.0585	-0.40	-0.0576	-0.37
Single-vehicle crash	0.7502	8.40	0.7969	5.54
Crash year				
Year 1425 (Reference)				
Year 1426	-0.0153	-0.13	-0.0516	-0.41
Year 1427	0.0695	0.53	0.0018	0.01
Year 1428	0.4325	3.19	0.3787	2.36
Year 1429	0.3870	2.34	0.3158	1.71
Number of casualties	0.1553	5.79	0.1544	4.76
Nationality and exspeed	0.5739	4.57	0.6301	3.01
Constant	-1.2163	-1.75	-0.9607	-1.3
Random parameters				
Nationality				
Mean			0.6324	1.70
Standard deviation			0.6232	0.56
Model statistics:				
Log-likelihood	-1872.079		-1837.7782	
AIC	3786.158		3719.556	
Observations (N)	2,946		2,946	

To compare different models, the Akaike Information Criterion (AIC) is used to compare model goodness-of-fit and complexity. The lower the AIC value, the better the model. Although the mixed binary logit model in Table 6.4 has produced a lower AIC value and a larger value of log-likelihood than that of Table 6.8, when flexibility, performance, and lower complexity with fewer parameters are considered, it is suggested that the mixed binary logit model without the insignificant variables (Table

6.8) fits the data better, and is found to be the best model for crash severity in Riyadh city.

6.6 Conclusions

This chapter has presented the estimation results and findings of the different crash severity models, which examined and identified a range of factors affecting road crash severity in Riyadh city. Initial models used, such as the ordered logit model and its extensions, including the generalised ordered logit model, proposed a partially constrained model, and more advanced models were used, such as a multinomial logit model and a mixed logit model.

To relax the proportional odds assumption in the ordered logit models, a generalised ordered logit model was employed that would have different coefficients across different severity outcomes. The results showed that the coefficients of all independent variables were different across thresholds in the GOLOGIT model, whereas only the coefficients of head-on, rear-end, and side-swipe collisions, single-vehicle crashes, and years 1427, 1428 and 1429 were different across thresholds in the same model: this suggests that the goodness-of-fit is better in the PC-GOLOGIT model than in the GOLOGIT model. The PC-GOLOGIT model was the best to use, rather than the OLM and GOLOGIT models.

Multinomial logit and mixed logit models were estimated and found to be consistent with variables specific to slight injury crashes, except for location of the crash, which was significant in ML only. As for variables specific to fatal crashes, both models were consistent.

Because of a serious problem with the under-reporting of slight injury crashes, in terms of the data used in this thesis (see 5.4), binary logit and binary mixed logit models were also used to identify the factors affecting slight injury and fatal crashes. No significant differences were found in the results between the multinomial and binary response models except that the level of confidence was higher in the binary response models.
In all models, the highly statistically significant (at the 95% confidence level) factors were: the age and nationality of the driver who was at fault in a crash, the time period

from 16:00 to 19:59, excessive speed, road surface and lighting conditions, single vehicle and number of casualties. A few variables were marginally significant: overtaking manoeuvres, head-on collision and location of the crash. In addition, the time trend was interesting, in that it suggested that the probability of fatal crashes increases over time, which is not consistent with the findings of developed nations.

Older drivers (age category between 40 and 64) are associated with a higher probability of having a fatal crash, which implies that younger drivers are associated with less severe road crashes. Although this is a counter-intuitive finding with respect to developed countries in which younger drivers are associated with more severe crashes, this may be true for Riyadh city; here, there is a lack of appropriate driving regulations for older drivers. Improved eyesight checks and fit-for-driving health screening should be introduced, in order to regulate drivers over the age of 40. Results showed that Saudi drivers aged 40-64 are less likely to be involved in fatal crashes than the reference age group, whereas Saudi drivers aged 25-39 are less likely to be involved in slight injury crashes compare to serious injury crashes than the reference age group.

The research findings in terms of the nationality of the driver were also interesting: the mixed logit model highlighted that the effect of Saudi nationality on crash severity was not uniform. The finding that 75% of Saudi drivers were associated with more fatal crashes (compared to non-Saudi drivers) does not support the hypothesis presented earlier. Non-Saudi drivers may be more careful, due to the fact that the purpose of their driving is work-related and there are tough regulations put in place by their company (i.e. they may fear losing their jobs or fear repatriation).

Interactions between different variables were considered, including age and nationality, age and excessive speed, and nationality and excessive speed: these were tested using the multinomial logistic regression and the binary logistic regression models; it was found that the interaction between age and nationality has a negative significant impact whereas the interaction between nationality and excessive speed has a positive significant effect. The interaction between age and excessive speed has no effect.
Road density was also tested, and was found to have a negative significant effect on the severity of fatal crashes when compared to serious injury crashes in the MNL and

binary logistic regression models. However, it had no effect on the severity of slight injury crashes, when compared to serious injury crashes. This means that the higher the road density is, the fewer fatalities there are.

For the purpose of simplifying the interpretation of the interaction between age and nationality, age was divided into four categories and only the significant variables were retained in the models. MNL, ML, binary and binary ML models were estimated.

The results suggest that the mixed binary logit model without the insignificant variables fits the data better and is found to be the best model for crash severity in Riyadh city.

7 RESULTS FROM CRASH FREQUENCY MODELS

7.1 Introduction

The aim of this chapter is to present the results of crash frequency models while examining the factors affecting the frequency of traffic crashes within Riyadh city. This is in line with one of the objectives of this research: to develop appropriate statistical models to explore the relationship between crash frequency and contributing factors.

As discussed in Chapter 4, the count of crashes in each HAI may be examined by employing suitable statistical models to examine the relationship between crash frequency and its contributing factors, such as area-wide total population, total number of vehicles, nationality, income, age, and land use.

The classical models for count outcome data are Poisson and negative binomial (NB) models. The Poisson model has the restriction that the mean is equal to the variance, which may be relaxed by using the NB model, which allows for over-dispersion in the data.

As discussed in Chapter 6 (section 6.3), slight injury crashes will not be tested because of a severe problem of under-reporting. Property damage only (PDO) crashes are not geocoded in the data; therefore, they are omitted from the analysis. The primary focus of this chapter is to report the results of fatal and serious injury models.

This chapter is structured as follows: first, a definition of HAIs is presented, showing the boundaries of HAIs with the spatial distribution of crashes for the Riyadh region and Riyadh city. This is followed by definitions of dependent and independent variables with their summary statistics. The next section presents the results from different statistical models associated with findings, followed by the preferred models. Finally, a summary and conclusions are presented.

7.2 Definition of HAIs and distribution of crashes within HAIs

The total number of HAIs in Riyadh region is 210 and Riyadh city alone has 169. HAIs between the outer parts of Riyadh region and Riyadh city have not yet been fully developed, which leaves many empty areas as they are outside the urban planning zone of the city. In addition, these HAIs have few roads and are largely agricultural areas with light road traffic and very few crashes. Therefore, the analysis is based on the 167 HAIs (from which all crash reports were available) within Riyadh city, which is surrounded with the solid line in Figure 7.1, whereas Riyadh region covers the whole map with 210 HAIs.

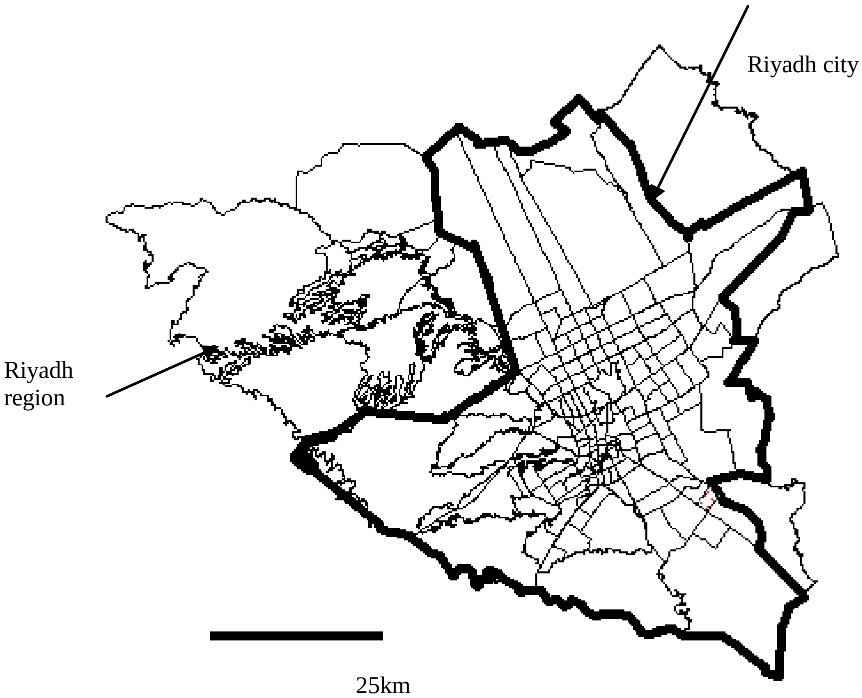

Figure 7-1Boundaries of HAIs for Riyadh region and Riyadh city

It can be noted from Figure 7.2, which is a thematic map of the spatial distribution of crashes for Riyadh region and Riyadh city, that HAIs with more roads have more crashes, and that these are concentrated in the centre of Riyadh city. Crashes are also found to be more frequent on the road to Riyadh international airport (Eastern ring road towards airport road) which serves the whole city. It also seems that crashes are located in places where young people gather for social activities (e.g. showing off and drifting with their cars). In addition, a number of crashes are found in places where people of

Riyadh city have picnics during the week and at weekends. More crashes are found to be in HAIs with low levels of income and education.

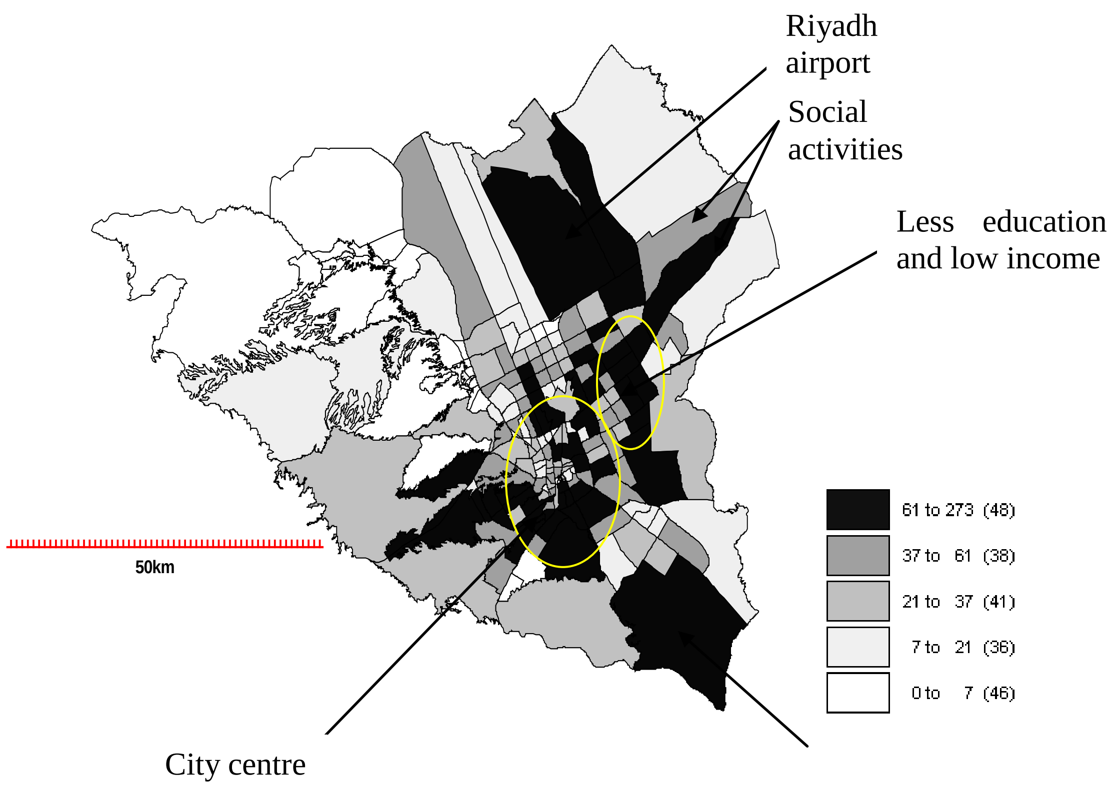

Figure 7-2Spatial distribution of crashes for Riyadh region and Riyadh city

7.3 Definition of variables

The aim of this section is to describe the dependent and independent variables which will be used in the models presented in section 7.4. In section 5.5 a description of the variables used is detailed as shown in Table 5.9.

7.3.1 Dependent variables

Table 7.1 contains all the variables which will be used for modelling the HAI-level frequency of crashes in Riyadh city, and provides a description for each variable for five years' data. The dependent variables are as follows:

- Fatal crashes variable: is the number of fatal crashes per HAI, which counts 167 observations (number of HAIs) with mean and standard deviation equal to 6.74 and 6.64 respectively. The minimum value is zero, which means there are no fatal crashes in that HAI, and the maximum value is 38 (Table 7.1 has summary statistics for all variables).

- Serious injury crashes variable: is the number of serious injury crashes per HAI which counts 167 observations (number of HAIs) with mean and standard deviation equal to 24.29 and 20.78 respectively. The minimum value is zero, which means there are no serious injury crashes in that HAI, and the maximum value is 113 (see Table 7.1).

The total number of HAIs studied in the city of Riyadh is 167. Therefore, the number of observations for fatal and serious injury crashes is 167 observations.

7.3.2 Independent variables

Table 5.9 in Chapter 5 contains the independent variables used for modelling HAI-level frequency of crashes in Riyadh city, and provides a description for each variable for five years' data.

Each member of the household in employment was asked about his/her monthly income during the household survey. The only available data for income are the number of people in each income category (described in Chapter 5) for each HAI and the number of households in each HAI. Area-wide income data at household level are not available in the dataset as described in Chapter 5 (section 5.2.2); data on the number of people within specific income groups are available, and this information is therefore used to generate a number of income variables.

For each category the mid-point is taken as the average.

Total income = number of people in an income category_i * middle-value of the income category:

$$\text{Total income per HAI} = \sum_{i=1}^{k} \text{Income}_i * \text{Middle-value}_i$$

Such that:

K= number of income categories (twenty categories).

Income_i= people within an income category $_i$

Middle-value_i= the middle value of income category_i

This information was used to estimate average income per HAI in three ways:

- Income per capita: total income in the HAI divided by total population in the HAI.
- Income per adult: total income in the HAI divided by number of adults >18 years old in the HAI.
- Percentage of low income people: percentage of people with income less than 2,500 Saudi Riyals per month from the total of people receiving an income.

According to the Saudi Central Department of Statistics and Information (2009) and the Saudi Ministry of Labour, low income is regarded as an income of less than 2,500 Saudi Riyals per month (roughly 400 GBP). The same source provided the following table Y which shows income categories in the Kingdom of Saudi Arabia. After a lengthy and intensive research it has been found that due to the job market style in KSA, the government has no interest in divulging the percentages related to proportion in each of the income categories. Therefore, the figures in Table 7.1 have been indirectly estimated by relating percentage of people in job categories which then were cross mapped to the salaries corresponding to these job categories.

Table 7-1 Model Categories of income with the Kingdom of Saudi Arabia

Income category	**Salary range**	Percentage		**Income category**	**Salary range**	**Percentage**
1	Under SR 1000	0.90%		11	SR 20000-29999	3.4%
2	SR 1000-1999	%1.75%		12	SR 30000-39999	2.2%
3	SR 2000-2999	%3.25%		13	SR 40000-49999	1.7%
4	SR 3000-3999	%7.8%		14	SR 50000-74999	0.90%
5	SR 4000-4999	11.2%		15	SR 75000-99999	0.63%
6	SR 5000-5999	18.7%		16	SR 100000-149999	0.45%
7	SR 6000-6999	12.3%		17	SR 150000-199999	0.25%
8	SR 7000-9999	17.55%		18	SR 200000-299999	0.15%
9	SR 10000-14999	10.32%%		19	SR 300000-399999	0.09%
10	SR 15000-19999	6.4%		20	SR 400000-499999	0.06%

Summary statistics for the data used in the analysis are presented in Table 7.2, which gives a list of variables used in the models. This table includes mean, standard deviation (S.D.), and minimum and maximum values of each variable.

In the original data, age refers to the number of people in each year of age from 1 to 99 for each HAI. Younger and older drivers experience more crashes than middle-aged drivers (Abdel-Aty and Radwan, 2000; Aguero-Valverde and Jovanis, 2006; Quddus, 2008). So it is expected that different age groups will affect the frequency of road crashes differently. Therefore, two age-related variables are defined: the proportion of the resident population under 18 and the proportion aged 65 or over

It is clear from the descriptive statistics in Table 7.2 for the percentages of young and old people that the average number of old people per HAI is very low while for young people it is very high. This is as expected since about 40% of the population are less than 15 years old.

There were seven categories of occupation in the data, defined as employed, unemployed but looking for work, student, housewife, retired, unable to work/handicapped, and not employed and not looking for work. Kim et al. (2006) and Noland and Quddus (2004) found that employment is positively associated with frequency of crashes. The category of employed people is used in the analysis.

Table 7-2 Summary descriptive statistics of all variables included in the models

Variables	Observations	Mean	S.D.	Minimum	Maximum
Dependent variables					
Fatal crashes	167	6.7365	6.6398	0	38
Serious injury crashes	167	24.2934	20.7757	0	113
Demographic characteristics					
Population	158	23438.29	25719.73	61	150396
Vehicle registered	158	6708.962	7203.995	0	33204
% of males	158	66.3579	16.4453	46.3940	100
% of non-Saudis	158	30.4807	32.3356	0	100
Income per capita	157	1366.267	1015.981	333.635	12095.16
Income per adult	157	2205.713	1182.104	667.9207	12619.13
% of low incomes	157	27.57813	23.86417	0	100
% <18	158	37.95223	14.1699	0	65.0655
% >65	158	1.4540	1.2218	0	7.3282
Employment	158	4682.19	5539.655	0	26385
% of illiterate	158	2.7341	2.8374	0	14.2304

education					
Road density km/km^2	167	16.3429	9.8089	0.0240	54.7872
Area in km^2	167	19.0180	41.6197	0.0836	260.2681
Land use					
% residential	167	14.3126	12.333	0.0002	53.9658
% health utilities	131	0.3530	0.8295	9.52E-05	7.5061
% educational Utilities	136	2.1478	3.7110	0.0017	32.6552
% cultural utilities	46	0.4868	1.2410	0.00064	6.383435
% recreation and Parks	154	1.4817	2.4367	0.002327	18.88962
% agricultural	103	3.4830	6.9772	0.00295	42.45158
% industrial	91	0.7159	3.8951	0.000602	36.55137
% transport utilities	161	2.6130	6.4787	0.0012	69.81165
% communications and public utilities	167	0.5909	0.9109	0.0009	6.522184
% commercial Utilities	163	1.7523	2.6710	0.0016	22.99477

Areas with low income and low education experience more crashes (Kim et al., 2006). Education has eight categories in the dataset (i.e. illiterate, able to read only, able to read and write, primary, intermediate, secondary, college/university, MA/MSc/PhD). The data show that more illiterate people are found to be in low income HAIs with fewer employed people. The category of illiterate people is used from the education categories.

Population has a minimum value of 61 in one HAI, which is not surprising because most of the area of this HAI is warehouses and stores; the only people living in this HAI would be guards or caretakers for the warehouses.

Vehicles registered in the HAI have a minimum value of zero because this HAI is a large university campus, with students who come to the campus for their classes and leave by the end of the day. Other HAIs have few registered vehicles because they are shopping centres and malls or industrial sites.

Some HAIs have zero non-Saudis because the HAIs are used as holiday homes for Saudis to spend their weekends; also, some of these HAIs have universities which are usually occupied by Saudi nationals.

For land use, Table 7.2 shows that the percentage of residential areas has the highest mean, with a low number of missing observations; the percentages of transport utilities, educational utilities, and commercial utilities have high mean values and high numbers of observations.

The percentage of each land use was calculated by dividing the area of each land use in a specific HAI by the total area of that HAI. Residential areas and the most economically active areas (such as retail, commercial and industrial) were taken as land use variables (Noland and Quddus, 2004; Graham and Glaister, 2005). The percentage of cultural utilities has the lowest number of observations as missing values and less mean, because libraries, museums, and exhibition halls are not numerous and not present in every HAI, and are spread across the city of Riyadh. The percentage of agricultural land use has a higher mean and standard deviation after residential, but with a low number of HAIs involved in agricultural land use; this type of land use is not of interest to this study because most of the agricultural areas are in the suburbs of the city, with fewer roads, less population and few road crashes (Kim et al., 2006).

7.4 Model development

The correlation coefficients of the variables considered are presented in Table 7.3. It is clear that numbers of vehicles and employment have been excluded from the models as these are highly correlated with population (correlation coefficients: 0.97 and 0.94). The correlation coefficients between other independent variables and population have also been examined, and the maximum value has been found to be 0.59, which suggests that multicollinearity is not a problem for the rest of the independent variables.

The percentage of males is highly correlated with the percentages of non-Saudis, low incomes, and young people (correlation coefficients: 0.98, 0.86 and -0.71). As Saudi law prohibits females from driving, and this study is only concerned with male drivers, therefore, the percentage of male variable was not considered.

Although the literature stated that employment is positively associated with frequency of crashes, initial models found this statistically insignificant. Moreover, when employment was included, other independent variables tended to become statistically insignificant (i.e. population, percentage of non-Saudis, percentage of older people >65, income per capita). Therefore, the employment variable will be discarded from models.

By using different mixes of demographic and land-use variables together in the models, and by looking at their performance through the value of the maximum likelihood, it was possible to estimate the value of Pseudo R^2, AIC value and the number of observations of each model, while observing the significance of each variable depending on the t-statistics value (1.65, 1.96, and 2.58, for 90%, 95%, and 99% confidence levels, respectively), and referring to what was stated in literature review such as:

Road density and area of HAI were discarded from models because an initial model was established, containing different independent variables such as population, percentage of non-Saudis, income, percentage of older people, and different land uses, which has been used to test other variables. They were insignificant in both Poisson and NB models, with low values of maximum likelihood and Pseudo R^2 and a higher value of AIC. This confirms the previous speculation.

Number of vehicles registered and percentage of young people were tested by including them with other independent variables such as population, percentage of non-Saudis, income, percentage of older people, road density, area of HAI, and different land uses; they were insignificant in the NB model, showing low values of maximum likelihood and Pseudo R^2 and a higher value of AIC. Therefore, they have been discarded from models.

Industrial, communications, cultural, commercial, and recreational and health land uses were statistically insignificant in all models estimated for fatal and serious injury crashes. Therefore, these variables will be discarded from the models. Residential land use was taken as an independent variable with transport and educational land uses because it can be noticed from Table 5.9 in the data chapter that the percentage of industrial areas has a low number of observations (low number of HAIs), with low mean and standard deviation, whereas residential, transport and educational land uses have a high number of observations and higher values of mean and standard deviation.

Although commercial and recreation and parks areas of land use have high numbers of observations and high values of mean and standard deviation, these variables became insignificant in the models when estimated, and had no effect; this may be because these land uses include large car parks and have intensive monitoring by traffic police controlling speed and traffic flow, which reduces the number of crashes in these areas of land use. Therefore, residential, transport utilities and educational land uses were retained in the models.

Table 7-3 Correlation coefficients between variables employed in the models

	population	**vehicles**	**% male**	**% non-Saudi**	**Income per capita**	**Income per adult**	**% low income**	**% youth age**	**% older people**	**Employ-ment**
Population	1									
Vehicles	0.97	1								
Percentage of males	-0.28	-0.24	1							
Percentage of non-Saudis	-0.28	-0.24	0.98	1						
Income per capita	-0.43	-0.33	0.45	0.56	1					
Income per adult	-0.47	-0.41	0.12	0.20	0.87	1				
Percentage of low incomes	-0.26	-0.20	0.86	0.92	0.46	0.05	1			
Percentage of young people	0.14	0.06	-0.71	-0.79	-0.60	-0.12	-0.86	1		
Percentage of older people	0.00	0.03	-0.29	-0.25	-0.09	-0.23	-0.10	-0.20	1	
Employment	0.94	0.94	-0.44	-0.42	-0.39	-0.41	-0.33	0.16	0.08	1
% illiterate	0.59	0.48	-0.34	-0.41	-0.56	-0.62	-0.35	0.13	0.41	0.50
Road density	0.48	0.53	0.28	0.33	0.13	-0.22	0.41	-0.56	-0.10	0.52
Area of HAI	0.31	0.28	-0.55	-0.62	-0.46	-0.07	-0.70	0.81	-0.27	0.29
% residential	0.48	0.54	0.06	0.15	0.03	-0.19	0.29	-0.33	-0.03	0.58
% commercial	0.20	0.24	0.62	0.65	0.19	-0.19	0.71	-0.68	-0.28	0.13
% health	-0.06	0.08	0.23	0.24	0.55	0.56	0.09	-0.13	-0.27	-0.06
% educational	-0.24	-0.27	-0.19	-0.26	-0.19	0.17	-0.41	0.63	-0.37	-0.28

% cultural	-0.34	-0.38	0.40	0.37	-0.11	-0.17	0.39	-0.06	-0.15	-0.41
% recreation	-0.54	-0.53	0.66	0.72	0.79	0.65	0.58	-0.55	-0.15	-0.62
% agricultural	-0.32	-0.36	0.35	0.33	-0.12	-0.16	0.35	-0.03	-0.12	-0.39
% industrial	0.00	0.03	0.04	-0.04	-0.06	-0.13	-0.08	-0.07	-0.17	0.02
% transport	-0.28	-0.29	0.53	0.59	0.79	0.70	0.42	-0.46	-0.36	-0.35
% communication	0.40	0.47	-0.00	-0.05	-0.24	-0.23	-0.06	0.14	-0.06	0.29

	% illiterate	**Road density**	**Area of HAI**	**% Residential**	**% commercial**	**% health**	**% Educational**	**% cultural**	**% recreation**	**% agricultural**	**% Indus-trial**	**% Transport**	**% communication**
% illiterate	1												
Road density	0.09	1											
Area of HAI	0.10	-0.47	1										
% residential	-0.07	0.84	-0.32	1									
% commercial	-0.06	0.76	-0.46	0.54	1								
% health	-0.40	0.11	0.08	0.20	0.09	1							
% educational	-0.36	-0.57	0.75	-0.41	-0.33	0.15	1						
% cultural	-0.18	-0.23	-0.31	-0.07	-0.10	-0.14	-0.14	1					
% recreation	-0.52	-0.09	-0.38	-0.27	0.23	0.17	0.06	0.04	1				
% agricultural	-0.16	-0.26	-0.28	-0.08	-0.16	-0.15	-0.14	0.99	0.02	1			
% industrial	0.27	0.22	-0.01	-0.18	0.06	0.02	-0.15	-0.15	-0.12	-0.14	1		
% transport	-0.39	0.09	-0.32	-0.16	0.27	0.16	-0.07	-0.07	0.87	-0.10	-0.03	1	
%communication	0.15	-0.12	0.25	-0.15	-0.01	-0.14	-0.12	0.07	-0.21	0.08	-0.10	-0.08	1

7.5 Modelling results

Spatially disaggregated HAI-level data for Riyadh city are used to analyse the impact of area-wide factors on road crashes in order to analyse the association of these factors with traffic fatalities and serious injury crashes. Two count outcome models, namely Poisson and negative binomial (NB) models were estimated, with fatal and serious injury crashes as dependent variables.

Based on the discussion in section 7.4, the independent variables found to be useful in modelling the traffic frequency of crashes in Riyadh city. The results and interpretations of both Poisson and NB models for fatal and serious injury crashes using data from one year (AH1425) and five years and the preferred models from among these estimated models are presented below.

As described in section 7.3.2, three income variables will be tested:

- Model 1 presents the income per capita.
- Model 2 presents the income per adult.
- Model 3 presents the percentage of low income.

All the above three models were estimated using both Poisson and NB models with fatal and serious injury crashes (see Tables 7.4, 7.5, 7.6, 7.7 below for five years' data).

As discussed in Chapter 5, land-use data for Riyadh city was obtained from HCDR (Higher Commission for the Development of Riyadh), who made an extensive survey in AH1425; the sampling design for demographic data for Riyadh city was based on the data collected through a land-use survey considering the household as the sampling unit. Therefore, it was decided to estimate the above models using only year 1425, which is the same year in which the land-use and demographic surveys were conducted.

Poisson and NB models for year 1425 dataset were undertaken. Poisson models for fatal crashes showed that population and percentage of illiterate people are positively significant for the frequency of fatal injury crashes in all the three models; income per capita is positively significant; the percentage of land use for transport is negatively significant in models 1 and 2. Poisson models for serious injury crashes population were positively significant for the frequency of serious crashes. The percentage of non-Saudis in models 1 and 2 is positively significant; the percentage of older people over 65+, the percentage of illiterate people and income per capita in model 1 are positively significant; and the percentage of older people over 65+ is significant. Residential land use is negatively significant in all models. NB models for fatal crashes showed the estimation results of the NB model for fatal crashes. The over-dispersion parameters in all three models were found to be statistically and significantly different from zero. As before, population is positively significant for the frequency of fatal crashes. The percentage of illiterate people is positively significant in models 1 and 2, whereas all types of land use were insignificant. NB models for serious injury crashes showed serious injury crashes. The over-dispersion parameters in all three models were found to be statistically and significantly different from zero. Population is positively significant for the frequency of serious crashes; the percentage of illiterate people is positively significant in models 1 and 2. The remaining variables are insignificant for all the three models. This means that when population and the percentage of illiterate people increase in the HAI, serious injury crashes will also increase in that HAI. The NB model was found to fit the data better and found to be the best model for crash frequency in Riyadh city. So the use of a simple Poisson model is insufficient. Since crashes are random events, observations of crashes at sites over short time periods may not reveal the true safety problems of sites.

Previous studies show that there are substantial regression-to-the-mean effects in 3 years' crash data (Persaud and Lyon, 2007), and the effects are reasonably large even for 5 years' crash data (Hauer and Persaud, 1983). Therefore, in this research the five years' data results will be used rather than the one year's data alone, which does not provide a big enough sample.

Table 7.4 shows Poisson models of fatal crashes by using five years' data. Population and the percentage of illiterate people are positively significant for the frequency of fatal injury crashes (at the 99% confidence level) in all the three models. Income per capita and the percentage of low incomes are positively significant (at the 95% confidence level). The percentages of residential, transport utilities and educational areas are negatively significant (at the 99% confidence level) in all models, except for educational land use in model 3 (at the 95% confidence level). This means that the greater the population and the greater the proportion of illiterate people in the HAI, the more fatal crashes will occur.

The higher the income per capita and the higher the percentage of low incomes in the HAI, the greater the number of fatal crashes. As the percentage of all types of land use (residential, transport, and educational land use) increases, the number of fatal crashes will decrease. The percentage of non-Saudis is negatively significant (at the 90% confidence level) in model 3, whereas it is insignificant in models 1 and 2. If Poisson models for fatal crashes in Riyadh city using five years' data are compared with the one-year data models, it is clear that the larger dataset is producing better models, as the majority of the socio-economic, income, and land-use variables in the five-year dataset models are more significant than those in the one-year dataset.

Table 7-4 Poisson models for fatal crashes in Riyadh city using five years' data

Variable	Model 1		Model 2		Model 3	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Log (population)	0.3225	8.41***	0.3113	8.25***	0.3202	8.41***
Percentage of non-Saudis	-0.00004	-0.02	0.0003	0.11	-0.0076	-1.90*
Percentage of older people age 65+	0.0418	1.21	0.04697	1.36	0.0467	1.35
Percentage of illiterate people	0.0566	3.60***	0.0490	2.97**	0.0447	3.44***
Income per capita	0.0002	1.96**	-	-	-	-
Income per adult	_-_	-	7.08E-05	1.01	-	-
Percentage of low incomes	-	-	-	-	0.0447	2.23**
Percentage of residential use	-0.0251	-5.7***	-0.0244	-5.54***	0.0137	-6.02***
Percentage of transport utilities	-0.0667	-3.77***	-0.0639	-3.6***	-0.0272	-3.1***
Percentage of educational use	-0.0568	-2.85***	-0.0571	-2.86***	-0.0527	-2.33**
Constant	-1.2010	-2.83***	-1.0138	-2.3**	-0.0429	-2.76***
Model statistics:						
Log-likelihood	-407.4910		-408.7946		-406.8424	
Pseudo R^2	0.2502		0.2478		0.2514	
AIC value	832.982		835.5891		831.6848	
Observations (N)	132		132		132	

* Statistically significant (at 90% confidence level) critical t=1.65
** Statistically significant (at 95% confidence level) critical t=1.96
*** Statistically significant (at 99% confidence level) critical t=2.58

The maximum likelihood value is highest in model 3 among the three models shown in Table 7.4, while the values of R^2 are similar in all models. Model 3 has the lowest value of AIC, which means that model 3 (Poisson model for fatal crashes with the percentage of low incomes) is the preferred model.

Figure 7.3 shows the relationship between the observed and the residual (= observed – predicted) values of crash frequency in Riyadh city for the Poisson model with the fatal crashes. There is a clear pattern to the residuals, in that they increase with respect to the increase in observed values. This increasing trend may be due to the regression-to-the-mean effect for fatal crashes due to the randomness of crashes; the observed value is not the true value as crashes are random phenomena (Elvik, 2007; Persaud and Lyon, 2007). However, this was the case for all models.

Because of the regression-to-the-mean effect, HAIs with a low crash count in one time period may actually have a high level of risk and could result in a high crash count in the following year, as crashes are random phenomena. HAIs with high crash numbers could have a reduction in the next year even if no treatment is applied. Even if five-year crash totals are considered at the worst crash HAIs, it is likely that the crash frequencies are at the high end of the naturally occurring random fluctuations, and in subsequent years these HAIs will experience lower numbers. The maximum value of residuals is nearly 25 for fatal crashes in all three Poisson models. The majority of residuals are between (10, -10). The observed count of fatal crashes was found to be over-predicted at the value of 17.

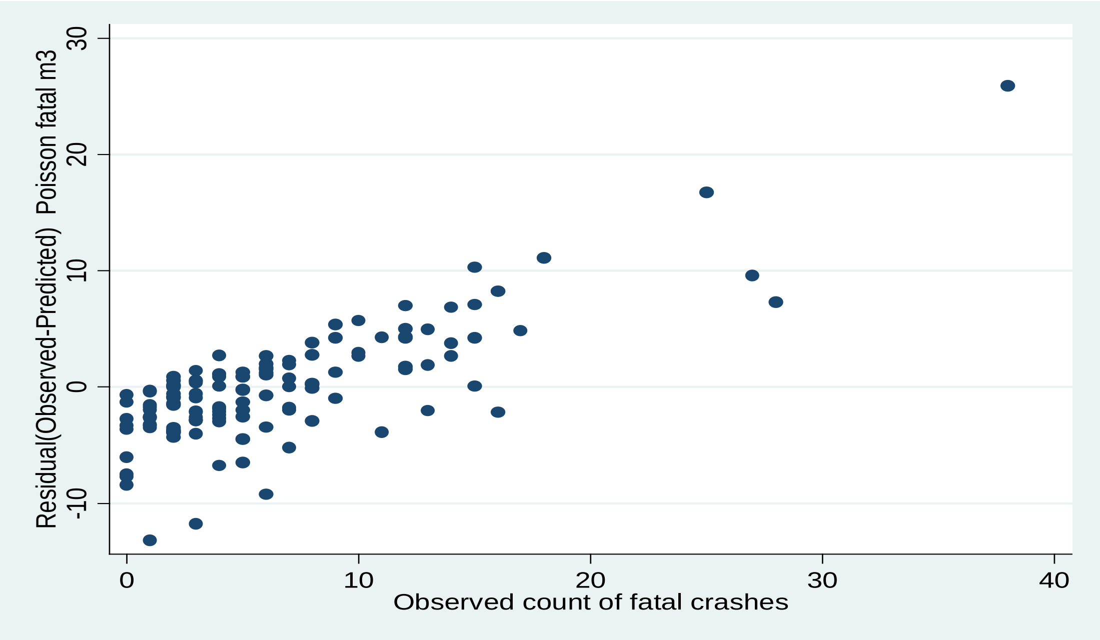

Figure 7-3 The relationship between the observed and the residual values of crashes for Poisson fatal count outcome model.

Table 7.5 shows Poisson models of serious injury crashes. All the demographic variables are positively significant for the frequency of serious injury crashes (at the 99% and 95% confidence levels), except for the percentage of non-Saudis in model 3, which was insignificant, whereas all types of land use (residential, transport and educational) are negatively significant (at the 99% confidence level), except for the percentage of educational land use in model 3, which was (at the 90% confidence level). This means that, when the population, the number of non-Saudi nationals, the percentage of older people (65+), the percentage of illiterate people, and all types of income increase, the level of occurrence of serious injury crashes will increase. While, low income picks up some of the illiteracy effect such that crashes are likely to occur more in poorer areas. In the opposite direction, when the percentages of all types of land use (residential, transport, and educational) increase, serious injury crashes will decrease.

Table 7.5 shows that R^2 values are similar among all models, whereas model 1 has the greatest value of maximum likelihood and lowest value of AIC. Therefore, the model 1 Poisson model for serious injury crashes with income per capita is the preferred model.

Table 7-5 Poisson models for serious injury crashes in Riyadh city using five years' data

Variable	Model 1		Model 2		Model 3	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Log (population)	0.5323	25.64***	0.5222	25.5***	0.5225	25.4***
Percentage of non-Saudis	0.0045	4.10***	0.004848	4.32***	0.0008	0.44
Percentage of older people age 65+	0.0708	3.77***	0.077675	4.15***	0.0790	4.22***
Percentage of illiterate people	0.0368	4.28***	0.029118	3.28***	0.0198	2.81***
Income per capita	0.0002	3.85***	-	-	-	-
Income per adult	-	-	7.66E-05	2.17**	-	-
Percentage of low incomes	-	-	-	-	0.0069	2.33**
Percentage of residential use	-0.0207	-9.77***	-0.02019	-9.5***	-0.0221	-9.96***
Percentage of transport utilities	-0.0313	-4.43***	-0.0289	-4.09***	-0.0231	-3.4***
Percentage of educational use	-0.0208	-2.65***	-0.02113	-2.68***	-0.0145	-1.92*
Constant	-2.1715	-9.35***	-2.00501	-8.4***	-1.8651	-8.96***
Model statistics:						
Log-likelihood	-763.58165		-768.3057		-767.8962	
Pseudo R^2	0.4149		0.4113		0.4116	
AIC value	1545.163		1554.611		1553.792	
Observations (N)	132		132		132	

* Statistically significant (at 90% confidence level) critical t=1.65
** Statistically significant (at 95% confidence level) critical t=1.96
*** Statistically significant (at 99% confidence level) critical t=2.58

Figure 7.4 shows the relationship between the observed and the residual (= observed – predicted) values of crash frequency in Riyadh city for the Poisson model with the serious injury crashes (the preferred model). There is a clear pattern to the residuals, in that they increase with respect to the increase in observed values. The maximum value of residuals is nearly 60 for serious injury crashes in all three Poisson models. The majority of residuals are between (20, -20). The observed count of fatal crashes was found to be over-predicted at the value of 50.

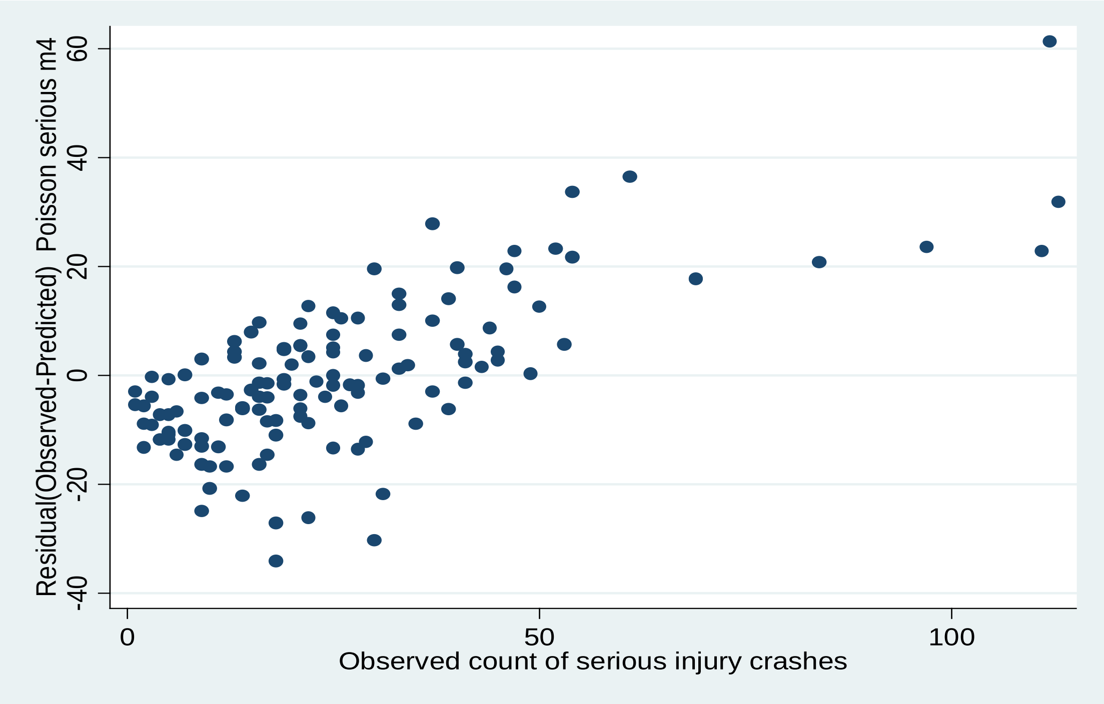

Figure 7-4 The relationship between the observed and the residual values of crashes for Poisson serious injury crashes count outcome model.

Table 7.6 shows the estimation results of the three NB models for fatal crashes. It was found that the over-dispersion parameters in all three models were statistically significantly different from zero (at the 99% confidence level), which means that the crash data were found to be significantly over-dispersed (i.e. the variance is much greater than the mean), so the use of the NB model is desirable. Population is positively significant for the frequency of fatal injury crashes (at the 99% confidence level) in all the three models. However, as can be seen from Table 7.6, the other socio-economic variables are all insignificant. The percentages of residential and transport land use are negatively significant (at the 99% confidence level) in all three models. The percentage of educational areas is negatively significant (at the 95% confidence level) in models 1 and 2, and (at the 90% confidence level) in model 3. This means that the greater the population and the greater the proportion of illiterate people in the HAI, the more fatal crashes will occur. As the percentages of all types of land use (residential, transport, and educational) increase, the number of fatal crashes will decrease.

Table 7-6 NB models for fatal crashes in Riyadh city using five years' data

Variable	Model 1		Model 2		Model 3	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Log (population)	0.3244	4.81***	0.0524	4.72***	0.322312	4.80***
Percentage of non-Saudis	0.0006	0.17	0.0418	0.23	-0.00464	-0.82
Percentage of older people age 65+	0.0452	0.66	6.45E-05	0.77	0.055751	0.82
Percentage of illiterate people	0.0502	1.64	-0.0255	1.37	0.0356	1.40
Income per capita	0.0002	1.09	-	-	-	-
Income per adult	-	-	6.45E-05	0.61	-	-
Percentage of low incomes	-	-	-	-	0.009784	1.10
Percentage of residential use	-0.0260	-3.57***	-0.0255	-3.5***	-0.02763	-3.71***
Percentage of transport utilities	-0.0759	-2.89***	-0.0730	-2.78***	-0.06562	-2.57**
Percentage of educational use	-0.0560	-2.24**	-0.0566	-2.23**	-0.04583	-1.84*
Constant	-1.1662	-1.67*	-1.0097	-1.42	-1.00938	-1.57
Over-dispersion parameter	0.3159	5.05***	0.3190	5.07***	0.315754	5.04***
Model statistics:						
Log-likelihood	-346.2337		-346.6533		-346.2433	
Pseudo R^2	0.0918		0.0907		0.0918	
AIC value	712.4674		713.3067		712.4865	
Observations (N)	132		132		132	

* Statistically significant (at 90% confidence level) critical t=1.65
** Statistically significant (at 95% confidence level) critical t=1.96
*** Statistically significant (at 99% confidence level) critical t=2.58

Table 7.6 shows that model 1 and model 3 are very similar in the maximum likelihood, the values of R^2, and the AIC value, which means that either model could be the preferred NB model for fatal crashes. Model 2 could be also a preferred model as its values are close to those of model 1 and model 2.

Figure 7.5 shows the relationship between the observed and the residual (= observed – predicted) values of crash frequency in Riyadh city for the NB model with the fatal crashes (the preferred model). There is a clear pattern to the residuals, in that they increase with respect to the increase in observed values. The maximum value of residuals is nearly 26 for fatal crashes for all three NB models. The majority of residuals are between (10, -10).

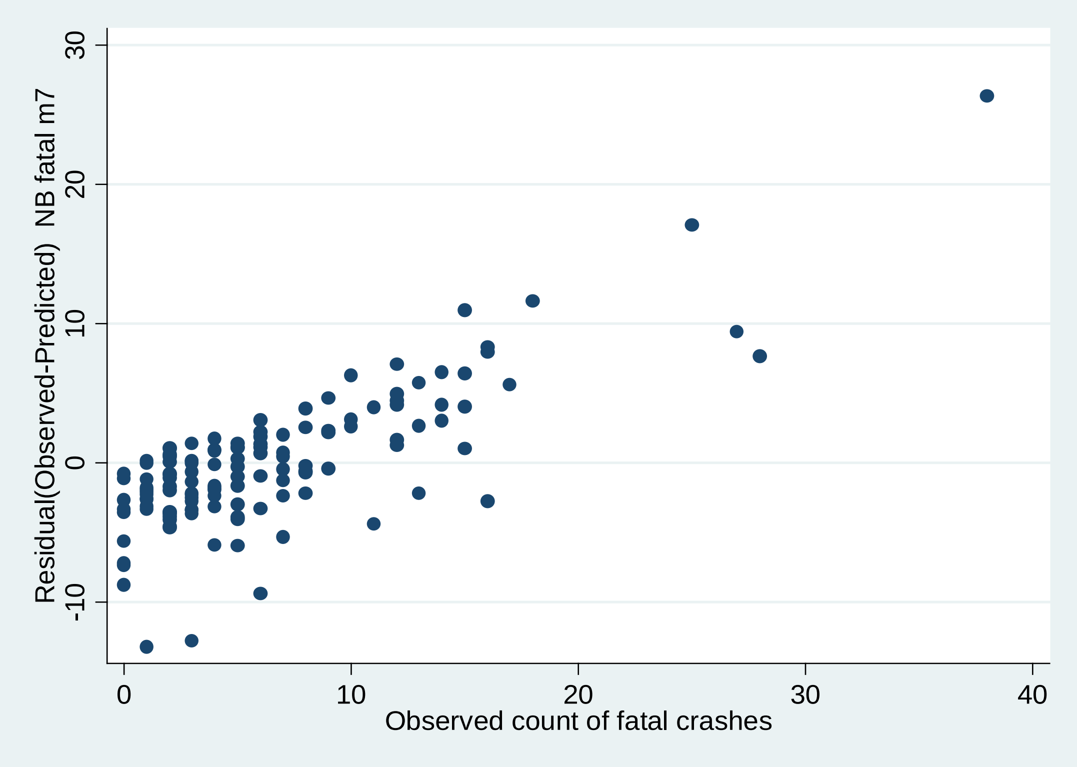

Figure 7-5 The relationship between the observed and the residual values of crashes for NB fatal count outcome model.

Table 7.7 shows the estimation results of the three NB models for serious injury crashes. It was found that the over-dispersion parameters in all three models were statistically significantly different from zero (at the 99% confidence level), which means that the crash data were found to be significantly over-dispersed (i.e. the variance is much greater than the mean), so the use of the NB model is sufficient. Population is positively significant for the frequency of serious crashes (at the 99% confidence level). The percentage of illiterate people is positively significant (at the 90% confidence level) in model 1, whereas land uses (residential and transport) are negatively significant (at the 99% and 95% confidence level respectively) in model 1. In model 2 land uses (residential and transport) are negatively significant (at 99% and 95% confidence level respectively).

Table 7-7 NB models for serious injury crashes in Riyadh city using five years' data

Variable	Model 1		Model 2		Model 3	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Log (population)	0.4755	9.26***	0.4661	9.12***	0.4623	9.12***
Percentage of non-Saudis	0.0040	1.52	0.0042	1.54	0.0012	0.31
Percentage of older people age 65+	0.0420	0.72	0.0530	0.91	0.0595	1.01
Percentage of illiterate people	0.0452	1.69*	0.0343	1.31	0.0215	1.00
Income per capita	0.0002	1.56	-	-	-	-
Income per adult	-	-	8.05E-05	0.94	-	-
Percentage of low incomes	-	-	-	-	0.0049	0.81
Percentage of residential use	-0.0190	-3.6***	-0.0186	-3.48***	-0.0198	-3.6***
Percentage of transport utilities	-0.0377	-2.17**	-0.0346	-1.99**	-0.0302	-1.76*
Percentage of educational use	-0.0229	-1.62	-0.0230	-1.61	-0.0181	-1.22
Constant	-1.5942	-2.91***	-1.4239	-2.54**	-1.2255	-2.56**
Over-dispersion parameter	0.2724	6.76***	0.2761	6.78***	0.2768	6.79***
Model statistics:						
Log-likelihood	-506.8352		-507.6501		-507.7723	
Pseudo R^2	0.0814		0.0799		0.0797	
AIC value	1033.67		1035.30		1035.545	
Observations (N)	132		132		132	

* Statistically significant (at 90% confidence level) critical t=1.65
** Statistically significant (at 95% confidence level) critical t=1.96
*** Statistically significant (at 99% confidence level) critical t=2.58

Residential and transport utilities land uses are found to be negatively significant (at the 99% and 90% confidence level, respectively) in model 3, whereas it is found that the percentage of older people, income per adult, and the percentage of low incomes are insignificant. This means that when population, the percentage of non-Saudi people, the percentage of illiterate people and income per capita increase in the HAI, serious injury crashes will also increase in that HAI; and when the percentages of all types of land use (residential, transport, and educational) increase, serious injury crashes will decrease.

It can be noted from Table 7.7 that model 1 has the highest values of maximum likelihood, the highest values of R^2, the lowest value of AIC, and more significant variables, which means that model 1 could be the preferred model (NB model for serious injury crashes with income per capita).

Figure 7.6 shows the relationship between the observed and the residual (= observed – predicted) values of crash frequency in Riyadh city for the NB model with the serious injury crashes (the preferred model).

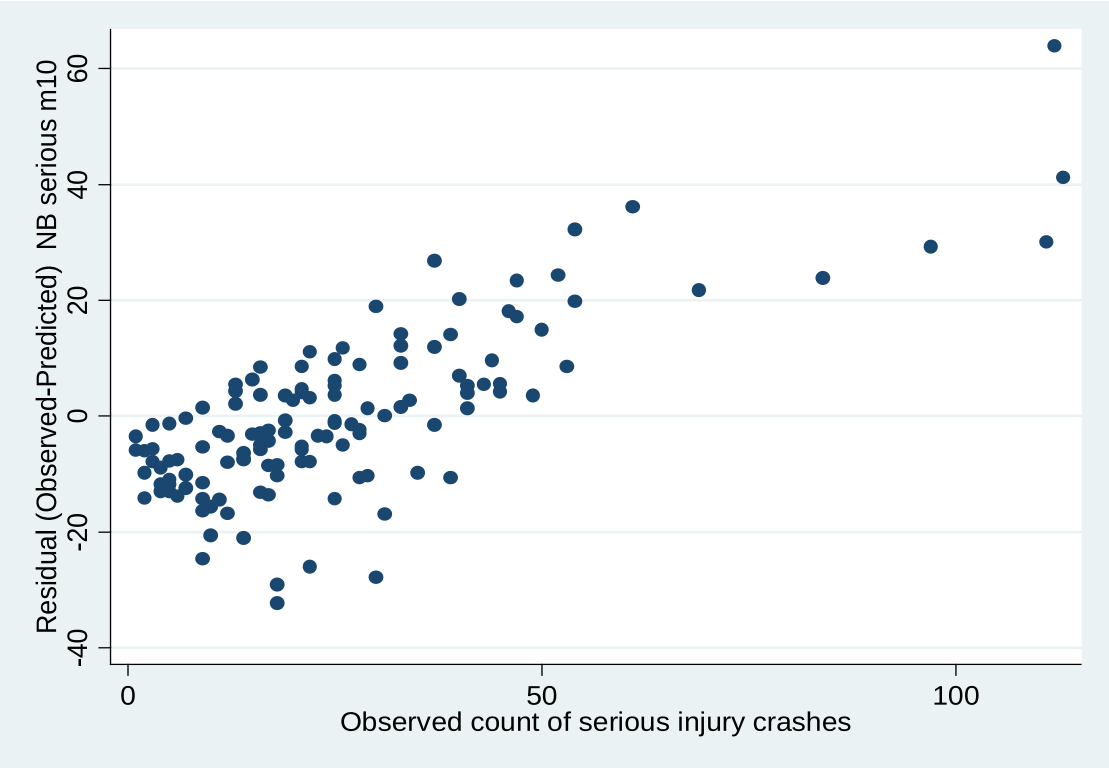

Figure 7-6 The relationship between the observed and the residual values of crashes for NB serious injury crashes count outcome model.

There is a clear pattern to the residuals, in that they increase with respect to the increase in observed values. The maximum value of residuals is nearly 60 for serious injury crashes for all three NB models. The majority of residuals are between (20, -20).

7.6 Conclusions

This chapter has presented the estimation results and findings from crash frequency models, which examined and identified a range of factors affecting road crash frequency in Riyadh city. Poisson and negative binomial models were employed in order to investigate the effect of factors affecting the frequency of fatal and serious injury crashes, using crash data from one year (AH1425) and five years.

In the case of fatal and serious injury crashes, in all models the population variable revealed the positive sign, suggesting that increased population is associated with the increased level of fatal and serious injury crashes. This result is in line with the findings of Kim et al. (2006).

For both categories of crashes, Poisson and NB models revealed the negative sign for all types of land use in all models when using five years' crash data, which suggests that the increase in residential, transport, and educational areas is associated with the decreased level of fatal and serious injury crashes. This may be due to the lower design speeds of roads in some HAIs, and higher levels of congestion resulting in reduced driving speeds on roads, whereas only the residential land use is significant (at the 90% confidence level). This result is consistent with the findings of Noland and Quddus (2005), who found that urbanised areas have fewer casualties, while Kim et al. (2006), have found that most crashes occur in an urban environment.

The percentage of non-Saudis was found to be statistically insignificant for fatal crashes in both models, but statistically significant for serious injury crashes in Poisson and NB models (at the 99% and 85% confidence level, respectively), whereas it is significant in Poisson models 1 and 2 for serious injury crashes when using only one year's crash data.

Income per capita and income per adult were found to be positively significant for the frequency of both fatal and serious injury crashes models (at the 90% and 95% confidence level, respectively) for Poisson models, whereas only income per capita is significant (at the 90% confidence level) for one year's data. This is because people who have a high monthly income are not aware about their vehicles and are encouraged to drive at a speed higher than the road is designed for.

The five years' data presented better results in all models, showing higher R^2 values and more significant variables. Socio-economic variables are insignificant in all NB models except for the percentage of illiterate people in model 1.

The negative binomial model fitted the data better; the better statistical fit, however, does not necessarily mean that the model could better reflect the actual effects of relevant factors on crashes (Lord et al., 2005b). Other data might help to improve the models, such as speed, traffic volume, road geometry, vehicle design, seat belt use, and the use of mobile phones.

Generally, NB models have lower significance levels than Poisson models in most variables. The over-dispersion parameters in all three models were found to be statistically and significantly different from zero, which suggests that the use of NB models is more appropriate than the use of Poisson models for the data analysed in relation to fatal and serious injury crashes.

There is a clear pattern to the residuals, in that they increase with respect to the increase in observed values. This increasing trend may be due to the regression-to-the-mean effect for both severity categories (fatal and serious injury crashes) and for both Poisson and NB count models. It was also found that the pattern of residuals in the three Poisson models for fatal crashes is similar to that in the NB model, and the pattern of residuals in the three Poisson models for serious injury crashes is similar to that in the NB model.

Model 1 and model 3 are very similar in the maximum likelihood, the values of R^2, and the AIC value, which means that either model could be the preferred NB model for fatal crashes; however, model 1 shows a higher t-stat than model 3. Model 2 could be also a preferred model as its values are close to those of model 1 and model 3. As a result model 1 is the preferred model. While model 1, could also be the preferred model (NB model) for serious injury crashes with income per capita.

Therefore for the case of Riyadh City the previous research didn’t report on any investigation using frequency modelling, hence the current study was undertaken and the obtained key results are now of prime importance to help reduce the number of fatal crashes and serious injury crashes. This result, which is about model 1 (income per capita) in the NB model with five years’ data for fatal and serious injury crashes that is chosen as the preferred model for crash frequency in Riyadh city. It has more significant variables, the highest maximum likelihood and R^2 values, and the lowest AIC value.

8 DISCUSSION AND POLICY IMPLICATIONS

8.1 Introduction

This thesis explores the factors affecting road crashes in Riyadh city. This has been achieved by examining the factors affecting the severity and frequency of traffic crashes through appropriate statistical models with the aid of GIS (Chapter 4). For crash severity analysis, model estimation results and findings were presented in Chapter 6 for crash frequency analysis, model estimation results were presented in Chapter 7. This chapter discusses the results and findings from Chapters 6 and 7; then, on the basis of those findings, several policy implications to reduce road crashes will be introduced and discussed.

8.2 Factors affecting road crash severity

The results of the model estimation presented in Chapter 6 show the contributing factors to crash severity in Riyadh city. Factors which were highly statistically significant (at the 95% confidence level) were the age and nationality of the driver who was at fault in the crash. The results show that as the age of the participant increases then the severity of the crash also increases: older drivers (in the age category between 40 and 64) are associated with a higher probability of having a fatal crash among Saudi drivers than non-Saudi drivers, whilst there is no effect in slight injury crashes, implying that younger drivers are associated with less severe road crashes. Although this is a counterintuitive finding, with respect to developed countries in which younger drivers are associated with more severe crashes, it may be true for Riyadh city as there is a lack of appropriate driving regulations to limit older drivers on the road. Tougher eyesight checks and fit-for-driving health screening could be introduced to regulate driving from the age of 40 to 65. Improved eyesight checks and fit-for-driving health screening should be introduced, in order to regulate drivers over the age of 65 too.

The results were surprising, but it can be speculated that if age of the driver increases, the ‘resilience’ against the released energy decreases; this is in line with existing studies (Lardelli-Claret et al., 2009). The results capture the complexities of driving behaviour with respect to driver demography and crash severity. Again, nationality was also interesting: it was clear that the effect of nationality on injury crash severity cannot be assumed to be uniform for all groups of drivers. However, compared with non-Saudi drivers, two-thirds of Saudi drivers were found to be associated with fatal crashes (relative to serious injury crashes).

In terms of the nationality of the driver, Saudis were positively associated with the severity of crashes, and were found to be more often involved in fatal crashes (compared to serious injury crashes) than non-Saudis. Nationality had no significant effect on slight injury crashes, in comparison to serious injury crashes; this may be due to the fact that non-Saudi drivers are expatriates, employed as family or company drivers, and may drive carefully out of fear of losing their jobs or being repatriated.

Results suggest that safety strategies aimed at reducing the severity of traffic crashes should take into account the structure of the resident population. Greater emphasis should be put on native residents; more attention should also be paid to older age groups in Riyadh city.

The time period from 16:00 to 19:59 is statistically significant, and is negatively associated with fatal crashes in comparison to serious injury crashes. This result may capture the fact that, in Riyadh city, traffic speeds during this period are relatively low as this is a peak period (just after the end of office hours at 14:30) for shopping and recreation.

As expected, excessive speed was consistently associated with fatal crashes. The results indicate that excessive speed has the biggest impact on the severity of traffic crashes in Riyadh city, and this was found to be statistically significant. Tougher enforcements (i.e. more police patrols in addition to introduction of average speed enforcement cameras) should be adopted to tackle the issue of excessive speed together with education scheme to make excessive speed socially unacceptable. However, overtaking was found to be negatively significant, with a confidence level of 90%, and distraction was found to be insignificant.

It was clear from the results that head-on and rear-end collisions were statistically significantly associated with a higher probability of having a slight injury crash, with a 95% confidence level, and were found to be less severe with fatal crashes, relative to the ‘other’ category. Driving on an undivided road with no barrier to separate vehicles from traffic moving in the opposite direction increases the risk of a head-on collision.

As discussed in Chapter 6, interaction between age and nationality was found to be negatively associated with fatal crashes compared to serious injury crashes. It has no significant effect when slight injury crashes are compared to serious injury crashes. The results suggest that, as the age of the driver increases, the probability of Saudis having a fatal crash decreases in comparison to non-Saudi drivers (at the 95% confidence level), whilst there is no effect of age in slight injury crashes compared to serious injury crashes for either nationality. This means that the probability of non-Saudis having a fatal crash increases more rapidly than that of Saudi drivers. So, older Saudis are more associated with fatalities.

The interaction between age and excessive speed was found to be insignificant for fatal and slight injury crashes compared to serious injury crashes. The interaction between nationality and speed was found to be positively associated (at the 95% confidence level) with the probability of fatal crashes, whilst there is no effect due to excessive speed in slight injury crashes compared to serious injury crashes. Saudis are more involved in fatal crashes compared to serious injury crashes than non-Saudi drivers.

The missing variables that may be related to the age (e.g. driving experience, licensing system, physical and mental health condition such as understanding the pre-crash circumstances of the older drivers) do have an effect that might be captured by the age variable and therefore there is a possibility that the sign of the age variable may change if all variables are included in the models. Previous studies in USA (AAA Foundation in Washington, 2011) found that approximately one in five fatal crashes involved an unlicensed or invalidly licensed driver. The proportion of fatal-crash involved drivers of any given age while unlicensed decreased as age increased across the entire age spectrum, whereas young adults ages 21–34 were the most likely to have had a suspended or revoked license. A similar situation is probably happening in KSA where unlicensed drivers are involved in fatal and serious injury crashes; however data to support this argument are missing. It was not possible to include all these factors in the models due mainly to a lack of data and the fact that the sample size is quite small. Both crash severity and frequency models may suffer from the problem of missing variables, since factors that may affect crash severity (such as vehicle design and seat belt usage) were not included in the crash severity models. Resolving the problem of missing variables requires much more data collection. As discussed in Chapter 1, it might not be possible to collect data for all crash severity and frequency factors and there may also be factors that were not previously known (i.e. imperfect information), which is the major driver for developing and using a sophisticated model. As discussed in Chapter 3, the mixed logit model is capable to take into account unobserved heterogeneity. Therefore the modelling techniques used in this thesis may reduce the impact of the problem of missing variables to a great extent. The models however cannot resolve the issue of missing variables completely.

Road density was also tested, and was found to have a negative significant effect on the severity of fatal crashes. However, it had no effect on the severity of slight injury crashes. This means that the higher the road density, the fewer fatalities and the fewer severe crashes there are in Riyadh city.

The results showed that a wet surface has a negative significant impact (with a weak confidence level of 90%) compared to dry road surface. This is because rain is not common in Riyadh city, and when it does rain drivers tend to drive carefully. Roads with adequate lighting conditions were associated with less severe crashes: roads with lighting have a negative effect compared to roads without lighting for fatal crashes in comparison to serious injury crashes, while lighting has no effect for slight injury crashes compared to serious injury crashes.

The variable ‘crash location’ was found to be significant in the slight injury crashes (at the 95% confidence level), whereas it was statistically significant (at the 90% confidence level); these findings pick up the effect of spatial difference that exists in different parts of Riyadh’s road network. These findings might lead to a greater focus on road crash locations through a careful examination of the road infrastructure, which will help agencies to focus their safety improvements by having roundabout systems instead of straight roads or intersections.

The time trend is also interesting in suggesting that the probability of slight injury crashes decreases over time whereas the probability of fatal crashes increases over time. This is not consistent with the finding elsewhere that the severity of traffic crashes decreases over time in the UK (Noland and Quddus, 2004).

The results indicate that the probability of a fatal crash increases if a crash is a single vehicle crash. This is consistent with the finding of existing studies from developed countries (Quddus et al., 2009). In addition, if the number of casualties is high, it has been found that it is more likely that the crash is a fatal crash than a serious injury crash. Furthermore, awareness about traffic regulations, safety education, and the consequences of violations to traffic safety rules in Riyadh should not be the duty of the government and official departments only, but also that of civil and public agents, as is the case in many of the developed countries that have succeeded in reducing road crash severity.

8.3 Factors affecting road crash frequency

The results presented in Chapter 7 identified a range of factors affecting road crash frequency across Riyadh city for fatal and serious injury crashes. These results suggest that the statistically significant variables in fatal crashes are the same as those in serious injury crashes.

The results show that the over-dispersion parameters in all models are statistically significantly different from zero (at the 99% confidence level), which means that the crash data are found to be significantly over-dispersed (i.e. the variance is much greater than the mean), justifying the application of an NB model rather than a Poisson model, so the use of NB model is sufficient and desirable; also, the five-year data present better results than the one-year data in all models, showing higher R^2 values and more significant variables.

Population is positively significantly associated (at the 99% confidence level) with the frequency of crashes at both severity levels, which suggests that the increased population is associated with the increased level of fatal and serious injury crashes. This result is in line with the findings of Kim et al. (2006). It can be speculated that the more people there are in the HAI, the more car ownership there is, and the more trips are generated; increased levels of street activity may cause more crashes in the HAI, especially under the effect of poor public transport, as is the case in Riyadh city.

The percentage of non-Saudis was found to be statistically significantly associated with serious injury crashes only at a weak level of confidence (85%). The percentage of illiterate people was found to be positively significantly associated with the frequency of both fatal and serious injury crashes (at the 90% confidence level); this may be because they live in low-income HAIs which may have low traffic safety standards.

Income per capita was found to be positively significantly associated with the frequency of serious injury crashes models (at the 85% confidence level). This may be because people who have high monthly incomes are not wary about their vehicles and are encouraged to drive at higher speed than the road design speed is and contrary to the safety measures. This is in line with some existing studies that suggest a positive association between poverty and traffic crash occurrences (Aguero-Valverde and Jovanis, 2006; Graham and Glaister, 2003).

In addition, crashes are affected by the nature of land use; the results suggest that the increase in residential, transport, and educational areas is associated with a decreased level of fatal and serious injury crashes. This may be due to the lower design speeds of roads in some HAIs, and higher levels of congestion which result in reduced driving speeds. This result is consistent with the findings of Noland and Quddus (2005), who found that urbanised areas have fewer casualties, while Kim et al. (2006), have found that most crashes occur in an urban environment.

Findings from this study could be useful in formulating safety policies aimed at reducing the severity and frequency of traffic crashes in Riyadh city. These policies will be presented in the next section.

8.4 Policy implications

The concept of road safety is not limited to the reduction of road crashes. It aims to encourage the adoption of road safety behaviour by establishing all necessary educational, engineering and medical programmes and plans, traffic regulations, and preventive measures. These would reduce or prevent crashes, ensure the safety of people and property, and preserve the human and economic wealth of Saudi Arabia.

Furthermore, as demonstarted in this research work a number of measures need to be considered by the policy organisations in Riyadh. As can be seen from Chapter 6 the significant factors contributing to road crash severity in Riyadh city includes age and nationality of the driver who is at fault in a crash, the time period from 16:00 to 19:59, excessive speed, road surface and lighting conditions, number of vehicles involved, location of the crash, time trend, and number of casualties. It is also clear from Chapter 7 that population, the percentage of illiterate people, income and land use are significant factors in road crash frequency. Therefore, it is desirable for transport policy makers in Riyadh city to take appropriate measures to reduce the impact of all these factors.

Per consequent, the findings from this study can be used to improve traffic safety in Riyadh city through implementation of measures to reduce the severity and frequency of road crashes. As shown in section 7.5, it has been found that the percentage of residential and educational land use are negatively related to the number of crashes. Time of the day does also influence, where between 16:00 and 19:59 the number of serious injury crashes increases. Furthermore, as the age of driver increases the severity of the crashes increases too. The time trend is also suggesting that the probability of slight injury crashes decreases over time whereas the probability of fatal crashes increases over time. In addition, if the number of casualties is high, it has been found that it is more likely that the crash is more of a fatal crash than a serious injury crash. Therefore, there is an urgent need to implement advanced prevention mechanisms such Traffic Management System (TMS) and Intelligent Transport Systems (ITS) in non-residential and none educational land areas of the city of Riyadh. As an example for the excessive speed and road density, an automated control and TMS linked with command

and control centres could be implemented by utilising the latest and most advanced technology in the field of ITS in order to create a safe traffic environment.

Strict, accurate and constant implementation of traffic regulations could be ensured by establishing an automated enforcement system through the installation of average speed enforcement cameras, which should be introduced to monitor crashes using digital cameras network. Fixed and mobile speed cameras (Figure 8.2) could monitor and control traffic violations such as excessive speed and could also provide information on traffic conditions which could be integrated into a system to help drivers avoid potential risks on roads by displaying real-time information about crashes through electronic warning signs. The aim of this system would be to reduce the severity and frequency of crashes; especially on roads with low density, as discussed in Chapter 6 higher road densities in Riyadh city are associated with fewer fatalities and less severe crashes.

In Riyadh city the presence of the police and medical services to the crash scenes has been found to be slow and not timely as shown by the previous literature (Al-Ghamdi, 1999 and Al-Ahmadi, 2011). Therefore, another promising system is the Auto Vehicle Location (AVL) system, which could be introduced to track the location of traffic police vehicles in order to increase the efficiency of traffic patrols and build good co-ordination with the medical and ambulance services.

http://www.transportxtra.com/files/8817-l.jpg http://citytransport.info/Digi/1364a.jpg

Figure 8-1 Automated enforcement system, fixed and mobile speed cameras.

It can be observed from Chapter 5 that the crash, driver, and vehicle data are scattered over different governmental organisations, resulting in a poor flow of data between the relevant institutions and researchers. Therefore, an initiative is required to create an integrated information centre for traffic safety data, linked to a supreme council for traffic safety or to higher council for traffic safety. Data collection and management could be improved by integrating sources into an easy-to-use and accessible database of all road crash data. In addition, national co-ordination of road safety strategies should involve all stakeholders (e.g. infrastructure providers, vehicle owners, other road-user groups, traffic police, and emergency response services), together with regional and local governments; they should participate in the development of the national road safety action plan.

The results show that older drivers are involved in more severe crashes; therefore appropriate driving regulations for older drivers should be introduced. It is important that a driving licence should not be issued unless the driver passes all the eyesight and medical checks, especially for those who are older than 60 years. Greater emphasis should be put on native residents because Saudi drivers are found to be associated with more severe crashes, as discussed in section 8.2.

The findings from this work suggest that head-on collisions and crash location are factors affecting road crash severity. Therefore, careful examination of the road infrastructure and dangerous intersections will help agencies to focus their safety improvements on road engineering, such as dividing roads with barriers to separate vehicles from traffic moving in the opposite direction, and having roundabout systems instead of straight roads or intersections.

According to the results of this study, the probability of fatal crashes increases over time, which is inconsistent with the findings in developed countries; this means there is an urgent need for a national traffic safety strategy covering all aspects of traffic safety and consisting of a set of programmes, plans, educational, engineering, medical, traffic regulations and laws, and preventive measures designed within the traffic system through the activation of the strict application of traffic law. Education may well be the key to address all the elements of this traffic safety strategy.

It is clear from Chapters 1 and 5 that in 2007 there were about six million trips generated per day in Riyadh city, 85% of which involved private vehicles; this is predicted to rise to about 15 million trips per day by 2020 (HCDR, 2008). Therefore, it is important to establish a proper public transport system to reduce car ownership in the city, which will result in a reduction of injuries and fatalities caused by car crashes. Furthermore, Riyadh has experienced a very high rate of population growth as its population reached more than 4.5 million in 2007 (which is about 18.5% of the total KSA population). The population of Riyadh city includes 66% Saudis and 34% non-Saudis in 2005, and is expected to reach 10.5 million by 2020 (HCDR, 2008). The findings of this study have also demonstrated that increased population is associated with an increased level of fatal and serious injury crashes, and, as discussed in Chapter 6, Saudi drivers were found to be more strongly associated with fatal crashes than non-Saudi drivers.

The findings shown in Chapter 7 suggest that an increase in residential, transport, and educational areas is associated with a decreased level of fatal and serious injury crashes. Therefore, it is essential that the Riyadh municipality reviews the urban planning, infrastructure, and land use of Riyadh city.

The measures described above (e.g. systems and schemes) could also be implemented not only in Riyadh city but in all major cities of Saudi Arabia, to achieve overall traffic safety and improve road safety in the city.

9 CONCLUSIONS, RECOMMENDATIONS AND FUTURE RESEARCH

9.1 Summary and conclusions

As discussed in the previous chapters, because of the negative impact of traffic crashes, which cause losses in the form of deaths, injuries and property damage, in addition to the pain and social tragedy affecting families of the victims, it is important to reduce the severity and frequency of road crashes in Riyadh city. Previous studies undertaken in Riyadh city lacked of quantitative evidence, and other factors affecting road traffic crashes have not been explored because there is a lack of suitable and reliable data for Riyadh city.

Therefore, this thesis aimed to explore factors affecting the severity and frequency of traffic crashes in Riyadh city using appropriate econometric models with the aid of GIS. This has been achieved by employing appropriate statistical models such as econometrics approach to analyse area-wide crash data and develop a relationship between crashes at area level and their contributing factors.

This thesis firstly conducted an in-depth review of current literature related to the various factors affecting road crashes. The review started by looking into theories of road safety: risk compensation theory, engineering theory, economic theory, public health theory and physiological theory, then looked at factors affecting road crashes which were related to (1) traffic characteristics, (2) driver characteristics, (3) road characteristics, (4) socio-economic factors, (5) land-use factors and (6) environmental factors. This included factors influencing road traffic crashes both in developed countries and in developing countries.

Some of these factors were considered in the econometric models in this thesis: they included age, nationality, time of day, cause of the crash, type of collision, location of the crash, road surface, lighting condition of the road, single vehicle crash, crash year, number of casualties, road density, population, where vehicle was registered, percentage of male, percentage of non-Saudi drivers, income, employment, and percentage of illiterate people.

Other factors were not included in the econometric models because of the unavailability of data.

In terms of econometric models, this research considered and reviewed the most popular models employed by road safety modellers. For crash severity analysis, a series of ordered response models (e.g. ordered logit and generalised ordered logit) and nominal response models (multinomial logit and mixed logit) were tested. The statistically significant factors (at the 95% confidence level) were the age and nationality of the driver who is at fault in a crash, the time period from 16:00 to 19:59, excessive speed, road surface and lighting conditions, single vehicle and number of casualties. The mixed binary logit model without the insignificant variables fitted the data better and was found to be the best model for estimating crash severity in Riyadh city because of its flexibility and model performance. According to the results from the mixed binary logit model, the effect of the nationality of the driver was interesting because the effect on crash severity was not uniform. 75% of the Saudi drivers were associated with more fatal crashes (compared to non-Saudi drivers). This was not a surprising result as non-Saudi drivers may be more careful, since the purpose of their driving is work-related and there are tough regulations put in place by their company (i.e. they may fear losing their jobs or fear repatriation).

In the City of Riyadh GIS is used to map crash locations in an integrated manner with land use, population and road networks, in particular at HAI level. For crash frequency analysis, two classical count outcome models, the Poisson and negative binomial (NB) models, were considered and used in this thesis. The findings from both models show that the statistically significant factors were population and percentage of illiterate people for both fatal and serious injury crashes, whereas the percentage of non-Saudis was found to be statistically significant only for serious injury crashes. Land use was found to be negatively significant for fatal and serious injury crashes, which is not surprising in view of the lower design speeds of roads in some HAIs and the higher levels of congestion resulting in reduced driving speeds.

The over-dispersion parameters in all models were found to be statistically and significantly different from zero, suggesting that the use of NB models is more

appropriate than the use of Poisson models for the analysis of data in fatal and serious injury crashes.
The data collected to conduct this thesis were obtained from several sources. HCDR (Higher Commission for the Development of Riyadh) provided the land-use data. The Riyadh traffic department, liaising with HCDR, provided the crash data, which included details of the crash and participant and vehicle data.

The demographic data for Riyadh city were provided by SGDS (Saudi General Directorate of Statistics), also liaising with HCDR, and data on the road network were provided by Riyadh Department of Transport (RDT). The geo-coding of crash data allows the data to be integrated with other datasets such as land use and population for frequency analysis. Data were investigated, checked and validated to improve the quality of the analysis. It has been found that the crash data suffer from a serious problem with the under-reporting of slight injury crashes.

On the basis of the results from the crash severity and frequency models, transport policy makers can improve road safety by introducing several policies aimed at reducing the effects of factors affecting the severity and frequency of traffic crashes in Riyadh city which were identified in this research. This includes the implementation of an automated control and traffic management system (TMS), utilising the latest and most advanced technology in the field of intelligent transport systems (ITS), establishing an automated enforcement system by installation of average speed enforcement cameras, and introducing an Auto Vehicle Location (AVL) system. These measures can assist transport policy makers in planning safety programmes (see Chapter 8, section 3).

9.2 Research contribution

The primary aim of the proposed research was to explore factors affecting the severity and frequency of traffic crashes in Riyadh city, which appear to have been rarely investigated in previous research, by using appropriate econometric models with the aid of GIS.

In order to achieve this aim a set of objectives were formulated (see section 1.3). Table 9.1 gives a brief description of each of the objectives and the corresponding Chapters in which they were achieved / addressed.

Table 9-1 Research objectives

Objective	Contribution	Chapter
Understand theories of road safety	Identified the most adequate theories of road safety which best apply to the case of Riyadh city	2
Conduct a review on the factors affecting traffic crashes	Conducted an in-depth review of literature relating to the various factors influencing road traffic crashes in developed and developing countries. It identified the most relevant factors influencing the road traffic crashes in Riyadh city	2
Integrate data from different sources using GIS	Used suitable, reliable, and recent data on road traffic crashes for Riyadh city It integrated data from various sources using GIS.	5
Develop models for the identification of contributory factors affecting traffic crashes	Identified adequate models to help study road traffic crashes in Riyadh city. Estimated the models and identified factors affecting the severity of traffic crashes that have occurred in Riyadh city. Estimated the models and identified factors affecting the frequency of traffic crashes that have occurred in Riyadh city.	3 and 4 6 7
Identify, explore and interpret factors affecting the severity and frequency of road injury crashes in Riyadh city	Provided empirical evidence on the factors affecting road traffic crashes based on appropriate statistical models.	6 and 7
Develop safety policies for reducing traffic crashes in the city of Riyadh.	Proposed safety policies for reducing traffic crashes in the city of Riyadh.	8

The main research contribution of this thesis is to identify the factors affecting the severity and frequency of traffic crashes that have occurred in Riyadh city. In addition, this thesis looked at the severity and counts of crashes with their contributing factors in an area-wide analysis which was not conducted for Riyadh city. There were no reported studies undertaken for Riyadh city in an area-wide analysis using advanced statistical models to analyse crash data and estimate the effect of the statistically significant factors at an aggregate level.

Furthermore, datasets were scattered among different organisations, which made the data difficult to obtain and demanded more time and effort from users and researchers. This study was able to integrate these datasets collected from various sources by employing GIS. In addition, the five years' data used in this study are recent and reliable, whereas data used in previous studies for Riyadh city were basic data, and other factors affecting road traffic crashes have not been explored because there is a lack of suitable data.

This thesis used advanced statistical models to provide more robust empirical evidence and develop safety policies based on rigorous analyses, whereas very few studies have used appropriate crash prediction models. For example, this thesis demonstrated how a mixed logit model can be used in examining crash severity at an individual crash level.

Finally, another important contribution of this research is that several policy implications were proposed, such as introducing a traffic management system (TMS) and an intelligent transport system (ITS). These measures would be useful for transport policy makers in improving road safety in Riyadh city.

9.3 Limitations and future research

There are, however, some limitations in this research. In the case of the data used in this research, crash data suffered from a severe under-reporting problem in the case of slight injury crashes (i.e. only 3.8% of the total crashes are reported as slight injury crashes). In addition, crash location data for property-damage only (PDO) crashes were not available in the crash dataset; it was not possible to develop models for this category of crashes. Furthermore, since the vast majority of the crashes were recorded as property damage only (98.36%), the crash data are heavily biased towards property damage only, so the inclusion of this category of crashes may influence the analysis. Therefore, property damage only crashes are not considered in this research, which has focused on the fatal and serious injury crashes only.

In this study only 100% blame subset data had been used in severity analysis instead of shared blame because the observation in severity analysis is the crash rather than the participants involved in the crash. This makes it difficult to obtain data related to age and nationality of driver (or other participant) involved in the crash.

When the road crash is defined in developed countries, personal injury involved becomes known to the police within 30 days of its occurrence (DfT, 2008), whereas in Riyadh city, crash injuries are known to the police within 24 hours (RDT, 2005).

In terms of future research, separate crash prediction models need to be developed for various road users such as pedestrians to reinforce some of the findings of this study. Data on other factors might help to improve the models, such as speed, traffic volume, road geometry, vehicle design, seat belts, and the use of mobile phones, which could produce a very interesting study.

Many spatial modelling techniques may be developed within a Bayesian approach or framework because of its flexibility in structuring complicated models and analyses. Previous studies have developed relationships between area-wide traffic crashes and various contributing factors using spatial econometrics to tackle the problem of unmeasured spatial correlation among neighbouring spatial units (Miaou et al. 2003; Ying and MacNab 2004; Song et al. 2006; Aguero-Valverde and Jovanis 2006 and Quddus 2008). From this, it is clear that one needs to consider the use of spatial models while developing a relationship between crashes and their contributing factors. In addition, Bayesian models were found more appropriate than any other models for crash analysis because they take spatial correlation and spatial heterogeneity into consideration, and are therefore called spatial models.

This research is based on Riyadh city. Therefore, further research is required to explore factors affecting the severity and frequency of road crashes in other cities and regions in the Kingdom of Saudi Arabia.

9.4 Recommendations

Based on the findings of this thesis, a number of recommendations have been formulated which aim at reducing the severity and frequency of traffic crashes in Riyadh city. These recommendations are as follows:

- Implement education as a basis to strengthen the culture of traffic safety, and support departments of education and schools across the Kingdom of Saudi Arabia to develop educational programmes with regards to traffic safety. Mechanisms to raise awareness about road safety issues such as safety awareness campaigns could be put in place too.
- Due to persistent denial of responsibility by different institution of the transport, it is essential to activate higher council for traffic safety. That is capable to use suitable, reliable, and recent data on road traffic crashes for Riyadh city and integrate data from various sources using GIS.
- To reconsider the current crashes reporting strategy adopted by HCDR, in particular the underreporting of slight injury crashes, which according to this study could contain crucial information to help reduce the number of crashes in Riyadh city.
- As time period from 16:00 to 19:59 was among the highly significant severity crash factor, therefore it is recommended to adjust working hours in schools and other places of work.
- Reduce the movement of Saudis from rural areas and other cities to Riyadh city and increase the distribution of population among the provinces and villages. As this work revealed that increased population is associated with the increased level of fatal and serious injury crashes.
- Increase the capabilities of emergency and medical care services and consider new ways such as air ambulance to support traffic safety.
- Apply the points system in traffic violations and improve the procedure in issuing driving licenses and develop appropriate driving regulations for older drivers.
- Target setting is necessary for road safety in Riyadh city, and realistic and achievable targets should be set for reducing traffic fatalities and serious injuries, e.g. by 50%.

- A strategy needs to be put in place for implementing penalties system such as driving license endorsements through penalty points or disqualification and introduction of heavy fines to effectively put off dangerous driving such as carelessness or usage of road unworthy vehicles. A strategy for driver's training could also be adopted to raise awareness and knowledge of road safety.

10 REFERENCES

AAA Foundation in Washington, D.C., Unlicensed to kill, 2011

Abbas, K.A. (2004) 'Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt', Accident Analysis and Prevention, 36(2), 149-163.

Abdalla, I. (2002) 'Fatality risk assessment and modeling of drivers responsibility for causing traffic accidents in Dubai', Journal of Safety Research, 33(4), 483-496.

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

Abdel-Aty, M., and Abdelwahab, H. (2004) 'Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure', Accident Analysis and Prevention, 36(3), 447-456.

Abdel-Aty, M., Chen, C., and Schott, J. (1998) 'An Assessment of the effect of driver age on traffic accident involvement using log-linear models', Accident Analysis and Prevention, 30(6), 851-861.

Abdel-Aty, M.A., and Radwan, E.A. (2000) 'Modeling traffic accident occurrence and involvement', Accident Analysis and Prevention, 32(5), 633-642.

Adams, G.F. (1981) 'The efficiency of seatbelt legislation: A comparative study of road accident fatality statistics', Transactions of the Society of Automotive Engineers, 28, 24-38.

Adams, G.F. and Hillman, M., (2001) 'The risk compensation theory and bicycle helmets', Injury Prevention, 7(2001), 89-91.

Aguero-Valverde, J. (2005) 'Spatial models of county-level roadway crashes for Pennsylvania', MSc thesis, The Pennsylvania State University.

Aguero-Valverde, J., and Jovanis, P.P. (2006) ‘Spatial analysis of fatal and injury crashes in Pennsylvania’, Accident Analysis and Prevention, 38(3), 618–625.

Al-Ahmadi, H. (2011) ‘Traffic accidents in Saudi Arabia: study and analyses’. Proceedings of the First Forum of Traffic Safety, Riyadh, Saudi Arabia.

Al-Gabbani, M. (2009) ‘Socio-economic impacts of foreign population on Saudi cities’, Fifth Biennial Population Geographies Conference, Dartmouth College, Hanover, New Hampshire.

Al-Ghamdi, A. (1993) ‘Forecasting traffic accidents in Saudi Arabia by using a time series model’. Proceedings of the 72nd Annual Meeting of the Transportation Research Board, Washington, DC.

Al-Ghamdi, A.S. (1996a) ‘Road accidents in Saudi Arabia: a comparative and analytical study’. Proceedings of the 75th Annual Meeting of the Transportation Research Board, Washington, DC.

Al-Ghamdi, A.S. (1998) ‘Spot speed analysis on urban roads in Riyadh’, Transportation Research Record, 1635, 162–170.

Al-Ghamdi, A.S. (1999) ‘Injury severity and duration of hospital stay for urban road accidents in Riyadh’, Transportation Research Record, 1665, 125–132.

Al-Ghamdi, A.S. (2000) ‘Analysis of Traffic Accidents at Urban Intersections in Riyadh’. Presented at the 79th Annual Meeting (TRB), Transportation Research Board, National Research Council, Washington, D.C. 2000

Al-Ghamdi, A.S. (2002) ‘Using logistic regression to estimate the influence of accident factors on accident severity’, Accident Analysis and Prevention, 34(6), 729-741.

Al-Ghamdi, A., and Al-Gadhi, S. (2004) ‘Warning signs as countermeasures to camel–vehicle collisions in Saudi Arabia’, Accident Analysis and Prevention, 36(5), 749-760.

Al-Katib, M. (2009) Personal communication with Mr. M. Al-katib, Managing Director, TSTN Limited, Reading, UK.

Al-Madani, H., and Al-Janahi, A. (2006) ‘Personal exposure risk factors in pedestrian accidents in Bahrain’, Safety Science, 44(4), 335-347.

Al-Saif, A. (1997) ‘The role of accident investigation forms in traffic safety’. Proceedings of First National Traffic Safety Conference. Riyadh: Ministry of Interior and King Abdulaziz City of Technology and Sciences, 41–55.

Al-Zahrani, A., and Jadaan, K. (1995) ‘Current traffic safety issues in Saudi Arabia’, Journal of Engineering and Science, 7(1), 151-162.

Ameen, J., and Naji, J. (2001) ‘Causal models for road accident fatalities in Yemen’, Accident Analysis and Prevention, 33(4), 547-561.

Amoros, E., Martin, J.L., and Laumon, B. (2003) ‘Comparison of road crashes incidence and severity between some French counties’, Accident Analysis and Prevention, 35(4), 537-547.

Andreassen, D.C. (1985) ‘Linking deaths with vehicles and population’, Traffic Engineering and Control, 26(11), 547–549.

Ansari, S., Akhdar, F., Mandoorah, M., and Moutaery, K. (2000) ‘Causes and effects of road traffic accidents in Saudi Arabia’, Public Health, 114(1), 37-39.

Arenas, B., Ramírez, F. , Aparicio Izquierdo, C., González Fernández, and A. Gómez Méndez (2009) ‘The influence of heavy goods vehicle traffic on accidents on different types of Spanish interurban roads’, Accident Analysis and Prevention, 41(1), 15-24.

Aschenbrenner, M., and Biehl, B. (1994) ‘Improved safety through improved technical measures? Empirical studies regarding risk compensation process in relation to anti-lock breaking systems’, In Trim pop and Wild, Challenges to accident prevention: The issue of risk compensation behaviour. Groningen, the Netherlands: Styx publications.

Assum , T., Bjørnskau, T., Fosser, S., and Sagberg, F. (1999) ‘Risk compensation—the case of road lighting’, Accident Analysis and Prevention, 31(5), 545-553.

Aultman, H., Kaltenecker, A., Aultman-Hall, L., and Kaltenecker, M.G. (1999) ‘Toronto bicycle commuter safety rates’, Accident Analysis and Prevention**,** 31(6), 675–686.

Bahrain General Directorate of Traffic and Licensing (2009) ‘Traffic crash statistics in the Kingdom Of Bahrain’ <http://www.traffic.gov.bh/arabic/index.asp>, accessed November 2011.

Barker, J., Farmer, S., and Taylor, M. (1999) ‘The development of accident-remedial intervention levels for rural roads’. Crowthorne, Berkshire: Transport Research Laboratory.

Bastos, Y., Andrad, S., Soares, D., and Matsuo, T. (2005) ‘Seat belt and helmet use among victims of traffic accidents in a city of Southern Brazil’, 1997–2000’, Public Health, 119(10), 930-932.

Baum, S. (1999) ‘An aggregate level analysis of the socioeconomic correlates of drink driving offenders’, Accident Analysis and Prevention, 31(3), 213-220.

Bendak, S. (2005) ‘Seat belt utilization in Saudi Arabia and its impact on road accident injuries’, Accident Analysis and Prevention, 37(2), 367-371.

Bener, A., Abdul Rahman, Y., and Mitra, B. (2009) ‘Incidence and severity of head and neck injuries in victims of road traffic crashes: In an economically developed country’, International Emergency Nursing, 17(1), 52-59.

Berhan, G. (2004) ‘Models relating traffic safety with road environment and traffic flows on arterial roads in Addis Ababa’, Accident Analysis and Prevention, 36(5), 697-704.

Berhanu, G. (2000) ‘Effects of Road and Traffic Factors on Road Safety in Ethiopia’, Dr.Eng. Dissertation, Norwegian University of Science and Technology, Trondheim.

Bluman, A. (2004) Elementary statistics, 5th edn. New York: McGraw Hill.

Brijs, T., Karlis, D., and Wets, G. (2008) ‘Studying the effect of weather conditions on daily crash counts using a discrete time-series model’, Accident Analysis and Prevention, 40(3), 1180-1190.

Brodsky, H., and Hakkert, A.S. (1983) ‘Highway accident rates and rural travel densities’ Accident Analysis and Prevention, 15(1), 73–84.

Cabinet Office Financial Year (COFY) in Japan (2007) White paper on traffic safety in Japan, abridged edition. Cabinet Office in Japan.

Cameron, A.C., and Trivedi, P.K. (1998) Regression analysis of count data. New York: Cambridge University Press.

CARE Community Road Accident Database: <http://ec.europa.eu/transport/roadsafety/road_safety_observatory/care_en.htm> [7 February 2010].

Carolina, D., Elena, E., José Juan, J., Juan, D., Aurora, and Pablo, L. (2010) ‘The association of age, sex and helmet use with the risk of death for occupants of two-wheeled motor vehicles involved in traffic crashes in Spain’, Accident Analysis and Prevention, 42 (1), 297-306.

Carson, J., and Mannering, F. (2001) ‘The effect of ice warning signs on ice-accident frequencies and severities’, Accident Analysis and Prevention, 33(1), 99-109.

Chang, L., and Mannering, F. (1999) ‘Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents’, Accident Analysis and Prevention, 31(5), 579-592.

Chang, W., Guo, H., Lin, H., and Chang, Y. (2006) ‘Association between major injuries and seat locations in a motor coach rollover accident’, Accident Analysis and Prevention, 38(5), 949-953.

Chen, C. (1997) ‘Statistical analysis of the effect of demographic and roadway factors on traffic crash involvement’, MSc thesis, Department of Civil Engineering, University of Central Florida.

Chen, L., Chen, C., Ewing, R., McKnight, C.E., Srinivasan, R., Roe, M. (2012) 'Safety countermeasures and crash reduction in New York City—Experience and lessons learned', Accident Analysis and Prevention, in press.

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

Clarke, D., Ward, P., Bartle, C., and Truman, W. (2006) 'Young driver accidents in the UK: The influence of age, experience, and time of day', Accident Analysis and Prevention, 38(5), 871-878.

Cooper, P., Meckle, W., and Nasvadi, G. (2009) 'The safety of vehicles imported from right-hand-drive vehicle configuration countries when operated in a left-hand-drive vehicle environment', Accident Analysis and Prevention, 41(1), 108-114.

Curtis, K.M., Rodi, S.W., and Sepulveda, M.G. (2007) 'The lack of an adult seat belt law in New Hampshire: Live free and die?' Accident Analysis and Prevention, 39(2), 380-383.

Delmelle, E., and Thill, J. (2008) 'Urban bicyclists - A spatial analysis of adult and youth traffic hazard intensity', TRB Annual Meeting.

Department for Transport (2011) Road statistics 2011: traffic, speeds and congestion. Transport Statistics.

Dickerson, A., Peirson, J., and Vickerman, R. (2000) 'Road accidents and traffic flows: an econometric investigation', Economica, 67(265), 101-121.

Dimitriou, H.T. 1990. Transport and Third World City Development, in H. Dimitriou (ed.), 'Transport Planning for Third World Cities', London: Routledge. ISBN: 950-99593-0-8.

Donnell, E., and Mason, J. (2006) 'Predicting the frequency of median barrier crashes on Pennsylvania interstate highways', Accident Analysis and Prevention, 38(3), 590-599.

Duncan, C., Khattak, A., and Council, F. (1998) ‘Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions’, Transportation Research Record: Journal of the Transportation Research Board, 1635(1), 63-71.

Edwards, J. (1998) ‘The relationship between road accident severity and recorded weather’, Journal of Safety Research, 29(4), 249-262.

El-Sadig, M., Norman, J., Lloyd, O., Romilly, P., and Bener, A. (2002) ‘Road traffic accidents in the United Arab Emirates: trends of morbidity and mortality during 1977–1998’, Accident Analysis and Prevention, 34(4), 465-476.

Eluru, N., and Bhat, C.R. (2007) ‘A joint econometric analysis of seat belt use and crash-related injury severity’, Accident Analysis and Prevention, 39(5), 1037-1049.

Eluru, N., Bhat, C.R., and Hensher, D.A. (2008) ‘A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes’, Accident Analysis and Prevention, 40(3), 1033-1054.

Elvik, R., Myssen, A., (1999) ‘Incomplete accident reporting: meta-analysis of studies made in 13 countries’. Transportation Research Record 1665, 133–140.

Elvik, R., Christensen, P., Amundsen, A.H., (2004) ‘Speed and road accidents. An evaluation of the Power Model’. TØI Report 740. Institute of Transport Economics, Oslo.

Elvik, R., and Greibe, P. (2005) ‘Road safety effects of porous asphalt: a systematic review of evaluation studies’, Accident Analysis and Prevention, 37(3), 515-522.

Elvik, R. (2007) ‘State-of-the-art approaches to road accident black spot management and safety analysis of road networks’. TOI report 883. Oslo: The Institute of Transport Economics.

Elvik, R., 2009. ‘The Power Model of the relationship between speed and road safety’. Update and new estimates. Report 1034. Institute of Transport Economics, Oslo.

Flegr, J., Havlicek, J., Kodym, P., Maly, M., and Smahel, Z. (2002) ‘Increased risk of traffic accidents in subjects with latent toxoplasmosis: a retrospective case control study’, BMC Infectious Diseases, 2(1), 11–16.

Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2007). Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. Accident Analysis and Prevention. Volume 40, Issue 1, January 2008, Pages 174-181.

Frantzeskakis, J.M., and Iordanis, D.I. (1987) ‘Volume-to-capacity ratio and traffic accidents on interurban four-lane highways in Greece’, Transportation Research Record, 1112, 29–38.

Fu, V.K. (1998) ‘Estimating generalized ordered logit models’. Stata Technical Bulletin, 44, 27-30.

Fyhri, A., Bjørnskau, T., Backer-Grøndahl, A., (2012). ‘ Bicycle helmets – A case of risk compensation?’. Transportation Research, 15 (2012), 612–624.

Garber, N.J., and Wu, L. (2001) ‘Stochastic models relating crash probabilities with geometric and corresponding traffic characteristics data’. Charlottesville, VA: Center for Transportation Studies, University of Virginia, publication UVACTS-5-15-74.

Garder, P. (1995) ‘Rumble strips or not along wide shoulders designated for bicycle traffic’, Transport Research Record, 1502, 1-7.

Graham, D.J., Glaister, S., and Anderson, R. (2005) ‘The effects of area deprivation on the incidence of child and adult pedestrian casualties in England’, Accident Analysis and Prevention, 37(1), 125-135.

Grant and Smiley (1993) ‘Driver response to antilock brakes: a demonstration on behavioral adaptation’. Proceedings of the 8th Canadian Multidisciplinary Road Safety Conference, 14-16.

Gray, R.C., Quddus, M.A., and Evans, A. (2008) ‘Injury severity analysis of accidents involving young male drivers in Great Britain’, Journal of Safety Research, 39(5), 483-495.

Gujarati, D. (1995) Basic econometrics, 3rd edn New York: McGraw-Hill.

Gwynn, D.W., and Baker, W.T. (1970) 'Relationship of accident rates and accident involvements with hourly volumes', Traffic Engineering, 40, 42–47.

Haan, P., and Uhlendorff, A. (2006) 'Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood', StataJournal, 6(2), 229-245(17).

Hadayeghi, A., Shalaby, A., Persaud, B., and Cheung, C. (2006) 'Temporal transferability and updating of zonal level accident prediction models', 38(3), 579–589.

Hadi, M.A., Aruldhas, J., Chow, L.F., and Wattleworth, J.A. (1993) 'Estimating safety effects of cross-section for various highway types using negative binomial regression', Transportation Research Record, 1500, 169-177.

Haight, F.A. (1980) 'Traffic safety in developing countries', Journal Of Safety Research, 12 20, 50–58.

Hauer, E., and Persaud, B. (1983) 'Common bias in before-and-after accident comparisons and its elimination', Transportation Research Record, 905, 164-174.

Hauer, E., Hakkert, S., 1989. 'The extent and some implications of incomplete crash reporting'. Transportation Research Record 1185, 1–10.

Hauer, E., (2009) 'Speed and safety', Transportation Research Record, 2103 (2009), pp. 10–17

Haynes, R., Jones, A., Kennedy, V., Harvey, I., and Jewell, T. (2007) 'District variations in road curvature in England and Wales and their association with road-traffic crashes', Environment and Planning A, 39(5), 1222-1237.

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), 843-850 .

HCDR (2008) 'Phase III of ADA Strategic Traffic Safety Study Report', pp.40-42.

Higher Commission for the Development of Riyadh. www.arriyadh.com/ADA-Emp/StudiesDep/Traffic-Safety-for-Arriyadhhase-I2.doc [July 2011].

Hedlund J. (2000) 'Risky business: safety regulation, risk compensation, and individual behavior', Injury Prevention, 6(2000), 82–90.

Híjar, M., Carrillo, C., Flores, M., Anaya, R., and Lopez, V. (2000) 'Risk factors in highway traffic accidents: a case control study', Accident Analysis and Prevention, 32(5), 703-709.

Holdridge, J., Shankar, V., and Ulfarsson, G. (2005) 'The crash severity impacts of fixed roadside objects', Journal of Safety Research, 36(2), 139-147.

Horne, J.A., and Reyner, L.A. (1995) 'Sleep related vehicle accidents', British Medical Journal, 310, 565-567.

IRTAD (2005) Selected road accident data for 2005: <http://cemt.org/IRTAD/IRTADPUBLIC/we2.html> [June 2011]

Ivan, J., Pasupathy, R., and Ossenbruggen, P. (1999) 'Differences in causality factors for single and multi-vehicle crashes on two-lane roads', Accident Analysis and Prevention, 31(6), 695-704.

Ivan, J.N., and O'Mara, P.J. (1997) 'Prediction of traffic accident rates using Poisson regression'. Proceedings of the 1997 Meeting of the Transportation Research Board, Washington, DC.

Ivan, J.N., Wang, C., and Bernardo, N.R. (2000) 'Explaining two-lane highway crash rates using land use and hourly exposure', Accident Analysis and Prevention, 32(6), 787-795.

Jacobs, G.D., Aeron-Thomas, A., and Astrop, A. (2000) 'Estimating global road fatalities', Transportation Research Laboratory Report 445. Crowthorne: Transportation Research Laboratory.

Jacobs, G.D., and Cutting, C.A. (1986) ‘Further research on accident rates in developing countries’, Accident Analysis and Prevention, 18(2), 119-127.

Jacobs, G.D., and Sayer, I. (1983) ‘Road accidents in developing countries’, Accident Analysis and Prevention, 15(5), 337–353.

Jadaan, K.S., and Nicholson, A.J. (1992) ‘Relationships between road accidents and traffic flows in an urban network’, Traffic Engineering Control, 33(9), 507–511.

Jain, A., MBBS, Menezes, R., Kanchan, T., Gagan, S., and Rashmi, (2008) ‘Two wheeler accidents on Indian roads – a study from Mangalore, India’, Journal of Forensic and Legal Medicine, Original Research Article 16(3), 130-133.

Jiang, N., Sato, T., Hara, T., Takedomi, Y., Ozak, I., and Yamada, S. (2003) ‘Correlations between trait anxiety, personality and fatigue study based on the temperament and character inventory’, Journal of Psychosomatic Research, 55(6), 493-500.

Johansson, P. (1996) ‘Speed limitation and motorway casualties: A time series count data regression approach’, Accident Analysis and Prevention, 28(1), 73-87.

Jones, B., Janssen, L., and Mannering, F. (1991) ‘Analysis of the frequency and duration of freeway accidents in Seattle’, Accident Analysis and Prevention, 23(4), 239-255.

Jørgensen, F., and Wentzel-Larsen, T. (1999) ‘Optimal use of warning signs in traffic’, Accident Analysis and Prevention, 31(6), 729-738.

Joshua, S.C., and Garber, N.J. (1990) ‘Estimating truck accident rate and involvement using linear and Poisson regression models’, Transportation Planning and Technology, 15(1), 41-58.

Jovanis, P.P., and Chang, H. (1986) ‘Modeling the relationship of accidents to miles traveled’, Transportation Research Record, 1068.

Karlaftis, M.G., and Rarko, A.P. (1998) 'Heterogeneity considerations in accident modeling', Accident Analysis and Prevention, 30(4), 425–433.

Khadaroo, J., and Seetanah, B. (2008) 'Transport and economic performance: The case of Mauritius', Journal of Transport Economics and Policy, 42(130), 255-267.

Khorashadi, A., Niemeier, D., Shankar, V., and Mannering, F. (2005) 'Differences in rural and urban driver-injury severities in accidents involving large trucks: an exploratory analysis'. Accident Analysis and Prevention, 37(5), 910-921.

Kim, D., Lee, Y., Washington, S., and Choi, K. (2007) 'Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models', Accident Analysis and Prevention, 39(1), 125-134.

Kim, J., Ulfarsson, G., Shankar, V., and Kim, S. (2008) 'Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis', Accident Analysis and Prevention, 40(5), 1695-1702.

Kim, K., Brunner, I., and Yamashita, E. (2006) 'Influence of land use, population, employment, and economic activity on accidents', Safety Data, Analysis, and Evaluation, 1953, 56-64.

Kim, K., Nitz, L., Richardson, J., and Li, L. (1995) 'Personal and behavioral predictors of automobile crash and injury severity'. Accident Analysis and Prevention, 27(4), 469–481.

Kim, K., and Yamashita, E. (2002) 'Motor vehicle crashes and land use—empirical analysis from Hawaii', Transportation Research Record, 1784, 73–79.

Kim, S., and Kim, K. (2003) 'Personal, temporal and spatial characteristics of seriously injured crash-involved seat belt non-users in Hawaii', Accident Analysis and Prevention, 35(1), 121-130.

Kimberly, V., and Donald, M. (2009) 'What can we learn about North Dakota's youngest drivers from their crashes?', Accident Analysis and Prevention, 41(3), 617-623.

Knuiman, M.W., Council, F.M., and Reinfurt, D.W. (1993) ‘Association of median width and highway accident rates’, Transportation Research Record, p.1401.

Kopits, E., and Cropper, M. (2005) ‘Traffic fatalities and economic growth’, Accident Analysis and Prevention, 37(1), 169-178.

Koushki, P. A. (1988), ‘The effect of socio-economic development on household travel behaviour in Saudi Arabia’, Socio-Economic Planning Sciences, 22(3), 131-136

Koushki, P., and Al-Anazi, S. (1998) ‘Seat Belt Use Trend in Kuwait’, Research Report # CE-499. Kuwait: Department of Civil Engineering, Kuwait University.

Koushki, P., and Bustan, M. (2006) ‘Smoking, belt use, and road accidents of youth in Kuwait’, Safety Science, 44(8), 733-746.

Kononov, J., Bailey, B. and Allery, B., (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.

Kulmala, R. (1994) ‘Measuring the safety effect of road measures at junctions’, Accident Analysis and Prevention, 26(6), 781-794.

Kumara, S.P., and Chin, H.C. (2003) ‘Modeling accident occurrence at signalized T intersections with special emphasis on excess zeros’, Traffic Injury Prevention, 4(1), 53–57.

Kweon, Y., and Kockelman, K. (2003) ‘Overall injury risk to different drivers: combining exposure, frequency, and severity models’, Accident Analysis and Prevention, 35(4), 441-450.

Kweon, Y.J., and Kockelman, K. (2003) ‘The safety effects of speed limit changes: use of panel models, including speed, use, and design variables’, Transportation Research Record, 1908, 148–158.

Lardelli-Claret, P., and Luna-del-Castillo, J. (2003) ‘Risk compensation theory and voluntary use by cyclists in Spain’, Injury Prevention, 9, 128-132.

Lardelli-Claret, P., Espigares-Rudriguez, E., Amezcua-Prieto, C., Jiménez-Moleón, J., Luna-del-Castillo, J., and Bueno-Cavanillas, A. (2009) 'Association of age, sex, and seat belt use with the risk of early death in drivers of passenger cars involved in traffic crashes'. International Journal of Epidemiology, 38, 1128-1134.

Lardelli-Claret, P., Luna-Del-Castillo, J., Jiménez-Moleón, J., Trinidad Domínguez, R., García-Martín, M., Femia-Marzo, P., and Bueno-Cavanillas, A. (2002) 'Association of main driver-dependent risk factors with the risk of causing a vehicle collision in Spain, 1990–1999', Annals of Epidemiology, 13(7), 509-517.

Law, T., Andrew, W.E., and Noland, R. (2007) 'The Role of medical care and technology improvements in reducing motorcyclist fatalities'. Proceedings of the 39th Universities Transport Study Group Conference, Harrogate, UK, 3-5 January 2007.

Law, T.H, Noland, R.B, and Evans, A.W. (2009) 'Factors associated with the relationship between motorcycle deaths and economic growth', Accident Analysis and Prevention, 41(2009), 234–240.

Law, T.H, Noland, R.B, and Evans, A.W. (2011) 'The sources of the Kuznets relationship between road fatalities and economic growth', Journal of Transport Geography, 19(2011), 355–365.

Lee, J., and Mannering, F.L. (2002) 'Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis', Accident Analysis and Prevention, 34(2), 149–161.

Lee, K.W. (1986) 'An analysis of automobile accidents in Riyadh'. ITE J. Institute of Transportation Engineers, Washington, DC, pp 35–39.

LeSage, J.P. (1999) Spatial econometrics. Toledo: Department of Economics, University of Toledo.

Levine, N., Kim, K.E., and Nitz, L.H. (1995) 'Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators', Accident Analysis and Prevention, 27(5), 675-685.

Li, Z., Knight, S., Cook, J., Hyde, L., Holubkov, H., and Olson, L. (2008) ‘Modeling motor vehicle crashes for street racers using zero-inflated models’, Accident Analysis and Prevention, 40(2), 835-839.

Liu, J., Ravishanker, N., Ivan, J.N., and Qin, X. (2005) ‘Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov chain Monte Carlo modeling’, Journal of Transportation Engineering, 131(5) , 345–351.

Liu, Y. (2008) ‘Spatial statistical modelling of traffic accidents on A road network’, 40th UTSG Conference.

Long, J.S., and Freese, J. (2006) Regression models for categorical dependent variables using STATA, 2nd edn. College Station, Texas: STATA Press.

Lord, D. (2000) ‘The prediction of accidents on digital networks: characteristics and issues related to the application of accident prediction models’. Toronto: Department of Civil Engineering, University of Toronto.

Lord, D., Manar, A., and Vizioli, A. (2005) ‘Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments’, Accident Analysis and Prevention, 37(1), 185-199.

Lord, D., and Mannering, F. (2010) ‘The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives’, Transportation Research Part A: Policy and Practice, 44(5), 291-305.

Lord, D., Washington, S.P., and Ivan, J.N. (2005) ‘Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory’, Accident Analysis and Prevention, 37(1), 35-46.

Lourens, P., Vissers, J., and Jessurun, M. (1999) ‘Annual mileage, driving violations, and accident involvement in relation to drivers’ sex, age, and level of education’, Accident Analysis and Prevention, 31(5), 593-597.

McFadden, D., and Train, K. (2000) ‘Mixed MNL models for discrete response’, Journal of Applied Econometrics, 15(5), 447-470.

MacNab, Y.C., and Dean, C.B. (2001) ‘Autoregressive spatial smoothing and temporal spline smoothing for mapping rates’. Biometrics, 57(3), 949-956.

Martin, J.L. (2002) ‘Relationship between crash rate and hourly traffic flow on interurban motorways’, Accident Analysis and Prevention, 34(5), 619–629.

Matthews, M. L., and Moran, A. R. (1986) ‘Age differences in male drivers' perception of accident risk: the role of perceived driving ability’, Accident Analysis and Prevention 18(4), 299–311.

Mekky, A. (1984) ‘Road traffic accidents in rich developing countries: the case of Libya’. Accident Analysis and Prevention, 16(4), 263–277.

Mekky, A. (1985) ‘Effect of road increase in road motorisation levels on road fatality rates in some rich developing countries’, Accident Analysis and Prevention, 17(2), 101-109

Mensah, A., and Hauer, E. (1998) ‘Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow’, Transportation Research Record, 1635, 37–43.

Mercer, G.W. (1989) ‘Traffic accidents and convictions: group totals versus rate per kilometer driven’, Risk Analysis, 9, 71–77.

Mercier, C.R., Shelley, M.C., Rimkus, J., and Mercier, J.M. (1997) ‘Age and gender as predictors of injury severity in head-on highway vehicular collisions’, Transportation Research Record, 1581.

Miaou, S.P., Hu, P.S., Wright, T., Rathi, A.K., and Davis, S.C. (1992) ‘Relationship between truck accidents and highway geometric design: a poisson regression approach’. Transportation Research Record, 1376, 10-18.

Miaou, S., and Lum, H. (1993) ‘Modeling vehicle accidents and highway geometric design relationships’, Accident Analysis and Prevention, 25(6), 689-709.

Miaou, S. (1994) ‘The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions’, Accident Analysis and Prevention, 26(4), 471-482.

Miaou, S.P., and Lord, D. (2003) ‘Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and bayes versus empirical bayes’, Transportation Research Record, 1840, 31–40.

Miaou, S., Song, J.J., and Mallick, B. (2003) ‘Roadway traffic crash mapping: a space-time modeling approach’, Journal of Transportation and Statistics, 6(1), 33-57.

Miaou, S.P., and Song, J.J. (2005) ‘Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence’, Accident Analysis and Prevention, 37(4), 699–710.

Milton, J., and Mannering, F. (1998) ‘The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies’, Transportation, 25(4), 395-413.

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

Morrongiell, B.A., Lasenby, J., Walpole, B., (2007). ‘Risk compensation in children: Why do children show it in reaction to wearing safety gear?’ Journal of Applied Developmental Psychology, 28(2007), 56–63.

Mohamedshah, Y.M., Paniati, J.F., and Hobeika, A.G. (1993) ‘Truck accident models for interstate and two-lane rural roads’, Transportation Research Record, 1407, 35-41.

Morrongiello, B., Lasenby, J., and Walpole, B. (2007) ‘Risk compensation in children: Why do children show it in reaction to wearing safety gear?’, Journal of Applied Developmental Psychology

Ministry of Transport (2005) National Transport Strategy: First Report <http://www.mot.gov.sa/L_Strategy.asp> [date accessed] Nantulya, V.M., and Reich, M.R. (2002) ‘The neglected epidemic: road traffic injuries in developing countries’, British Medical Journal, 324, 1139-1141.

Nassar, S.A., Saccomanno, F.F., and Shortreed, J.H. (1997) ‘Integrated Risk Model (ARM) of Ontario’. Presented at the 76th Annual Meeting of the Transportation Research Board, Washington, DC.

National Highway Traffic Safety Administration, (2009) Traffic Safety Facts: Motorcycles, DOT HS 811 159, Washington, DC.

Navin, F., Bergan, A., and Qi, O. (1994) ‘Fundamental relationship for roadway safety: model for global comparisons’, Transport Research Record, 1441**,** 53–58.

Navin, F., Zein, S., and Felipe, E. (2000) ‘Road safety engineering: an effective tool in the fight against whiplash injuries’, Accident Analysis and Prevention, 32(2), 271-275.

Newstead, S., Linda, R., and Robert, R. (2008) ‘Trends in crashworthiness of the New Zealand vehicle fleet by year of manufacture to 2006’, Report No. 280 supplement. [place of pubn:] Monash University, Accident Research Centre. 1064 Max. Cameron

Nofal, F., and Saeed, A. (1997) ‘Seasonal variation and weather effects on road traffic accidents in Riyadh City’, Public Health, 111(1), 51-55.

Noland, R.B. (2003a) ‘Traffic fatalities and injuries: the effect of changes in infrastructure and other trends’, Accident Analysis and Prevention, 35(4), 599-611.

Noland, R.B., and Oh (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), 525-532.

Noland, R.B., and Quddus, M.A. (2004) ‘A spatially disaggregated analysis of road casualties in England’, Accident Analysis and Prevention, 36(6), 973-984.

Noland, R.B., and Quddus, M.A. (2005) ‘Congestion and safety: A spatial analysis of London’, Transportation Research Part A: Policy and Practice, 39(7-9), 737-754.

O'Donnell, C.J., and Connor, D.H. (1996) ‘Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice’, Accident Analysis and Prevention, 28(6), 739-753.

Ofosu, J., Abouammoh, A., and Bener, A. (1988) ‘A study of road traffic accidents in Saudi Arabia’, Accident Analysis and Prevention, 20(2), 1988, 95-101.

Ossiander, E.M., and Cummings, P. (2002) ‘Freeway speed limits and traffic fatalities in Washington State’, Accident Analysis and Prevention, 34(1), 13-18.

Pai, Chih-Wei., and Saleh, W., (2007) ‘An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK’, Safety Science, 45(8), 832-847.

Paulozzi, L.J, Ryan, G.W, Espitia-Hardeman, E.V, and Yongli Xi. (2007) ‘Economic development’s effect on road transport-related mortality among different types of road users: A cross-sectional international study’, Accident Analysis and Prevention 39(2007), 606–617.

Pawlovich, M.D., Carriquiry W. Li and Welch, T.M. (2006) ‘Iowa's experience with road diet measures: use of Bayesian approach to assess impacts on crash frequencies and crash rates’, Transport Research Record, 1953, 163–171.

Peirson, J., Skinner, I., and Vickerman, R. (1998) ‘The microeconomic analysis of the external costs of road accidents’, Economica, 65(259), 429-440.

Peled, A., Haj-Yehia, B., and Hakkert, A.S. (1996) ‘Arc Info-based Geographical Information System for road safety analyses and improvement’. ESRI User Conference.

Peltzman, S. (1975) ‘The effect of automobile safety regulation’, Journal of Political Economy, 83, 677-726.

Pérez, I. (2006) ‘Safety impact of engineering treatments on undivided rural roads’, Accident Analysis and Prevention, 38(1), 192-200.

Persaud, B., and Lyon, C. (2007) ‘Empirical Bayes before–after safety studies: lessons learned from two decades of experience and future directions’. Accident Analysis and Prevention, 39(3), 546-555.

Petch, R.O., and Henson, R.R. (2000) ‘Child road safety in the urban environment’, Journal of Transport Geography, 8, 197–211.

Pitt, R., Guyer, B., Hsieh, C., and Malek, M. (1990) ‘The severity of pedestrian injuries in children: an analysis of the Pedestrian Injury Causation Study’, Accident Analysis and Prevention, 22(6), 549-559.

Poch, M., and Mannering, F.L. (1996) ‘Negative binomial analysis of intersection accident frequencies’, Journal of Transportation Engineering, 122(2), 105-113.

Preston, B. (1980) ‘Road safety: international comparisons’. Transport Rev., 11, 75–98.

Quddus, M., Noland, R., and Chin, H. (2002) ‘An analysis of motorcycle injury and vehicle damage severity using ordered probit models’, Journal of Safety Research, 33(4), 445-462.

Quddus, M.A. (2008) ‘Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London accident data’, 40th UTSG Conference.

Quddus, M.A., Wang, C., and Ison, S.G. (2009) ‘Road traffic congestion and crash severity: An econometric analysis using ordered response models’, ASCE Journal of Transportation Engineering, 136(5), 424-435.

Ratrout, N. (2005) ‘Tire condition and drivers’ practice in maintaining tires in Saudi Arabia’, Accident Analysis and Prevention, 37(1), 201-206.

Richard C., Anne, A., Amy S., (1998) ‘Analysis of circumstances and injuries in 217 pedestrian traffic fatalities’, Accident Analysis and Prevention, 30(1), 11-20.

RGDT, Riyadh General Department of Traffic, Saudi Ministry of Interior, (2007), Traffic statistics: Annual Statistics Report from Riyadh General Traffic Department. http://www.rt.gov.sa/default.php

Rocío, S., Antonio, F., Aranzazu, J., and Antonio Martín, A. (2010) ‘The probability of death in road traffic accidents: How important is a quick medical response?’ Accident Analysis and Prevention, 42(4), 1048-1056.

Rodriguez, D., Rocha, M., Khattak, A.J., and Belzer, M.H. (2003) ‘Effects of truck driver wages and working conditions on highway safety: case study’, Transport Research Record, 1833, 95–102.

Roudsari, B., Sharzei, K., and Zargar, M. (2004) ‘Sex and age distribution in transport-related injuries in Tehran’, Accident Analysis and Prevention, 36(3), 391-398.

Sagberg, Fosser, and Saetermo (1997) ‘An investigation of behavioral adaptation to airbags and antilock brakes among taxi drivers’, Accident Analysis and Prevention, 29(3), 293-302.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., (2011) ‘The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives’ Accident Analysis and Prevention 43(2011) 1666– 1676.

Shankar, V., Albin, R., Milton, J., and Mannering, F. (1998) ‘Evaluating median crossover likelihoods with clustered accident counts: An empirical inquiry using the random effects negative binomial model’, Transportation Research Record, 1635(1), 44-48.

Shankar, V., and Mannering, F. (1996) ‘An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity’, Journal of Safety Research, 27(3), 183-194.

Shankar, V., Mannering, F., and Barfield, W. (1995) ‘Effect of roadway geometrics and environmental factors on rural freeway accident frequencies’, Accident Analysis and Prevention, 27(3), 371-389.

Shankar, V., Milton, J., and Mannering, F. (1997) ‘Modeling accident frequencies as zero-altered probability processes: An empirical inquiry’, Accident Analysis and Prevention, 29(6), 829-837.

Shankar, V.N., Albin, D., Milton, J., and Nebergall, M. (2000) ‘In service performance-based roadside design policy: Preliminary insights from Washington State's bridge rail study’, Transportation Research Record, 1720, 72–79.

Shanks, J., Ansari, M., and Kalai, A. (1994) ‘Road traffic accidents in Saudi Arabia’, Public Health, 108(1), 27-34.

Shawan, S., Al-Arfaj, A., Wosornu, L., Al-Habdan,I., and Hegazi, M. (1993) ‘Children as front seat passengers: Frequency of an unsafe practice, and its projected impact on car safety legislation in Saudi Arabia’, Saudi Medical Journal, 14(4), 312-314.

Simon, M. (2001) ‘Road accidents in Slovenia involving a pedestrian, cyclist or motorcyclist and a car’, Accident Analysis and Prevention, 33(2), 147-156.

Smeed, R.J. (1949) ‘Some statistical aspects of road safety research’, J. R. Statistical. Social. Ser. A, 112, 1–23.

SMOT (Saudi Ministry of Transport), (2007), National Transport Strategy Report. <http://www.mot.gov.sa/>

Song, J.J. (2004) Bayesian multivariate spatial models and their applications, Texas A&M University.

Song, J.J., Ghosh, M. S. Miaou, and Mallick, B. (2006) ‘Bayesian multivariate spatial models for roadway traffic crash mapping’, Journal of Multivariate Analysis, 97(1), 246-273.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Linde, A.v.d. (2002) ‘Bayesian measures of model complexity and fit’, Journal of the Royal Statistical Society, Series B (Statistical Methodology), 64(4), 583-639.

Srinivasan, K., (2002) ‘Injury severity analysis with variable and correlated thresh-olds: ordered mixed logit formulation’. Transportation Research Record 1784, 132–142.

Stamatiadis, N., Taylor, W.C., and Mckelvey, F.X. (1991) 'Older drivers and intersection traffic control devices', Journal of Transportation Engineering, 117(3), 311–319.

Stenks, J. (2007) 'Human factors and behavioural safety', Safety and Health Journal, 37(3), 323-332.

Sun, D., Tsutakawa, R.K., Kim, H., and He, Z. (2000) 'Spatio-temporal interaction with disease mapping', Statistics in Medicine, 19(15), 2015-2035.

Sun, J., Sun, P., and Misumi, J. (2004) 'Population characteristics of traffic accident-related deaths in Japan and Ireland, 1950–2000', Journal of Emergency Medicine, 72(4), 333-338.

Tay, R., Rifaat, S., and Chi, H. (2008) 'A logistic model of the effects of roadway, environmental, vehicle, crash and driver characteristics on hit-and-run crashes', Accident Analysis and Prevention, 40(4), 1330-1336.

Thakuriah, P., and Cottrill, C. (2008) 'Evaluating pedestrian risk in environmental justice areas', in Proceedings of the 87th Annual Meeting of the Transportation Research Board, Washington, DC, 13–17.

Toy, E.L., and Hammitt, J.K. (2003) 'Safety impacts of SUVs, vans, and pickup trucks in two-vehicle crashes'. Risk Analysis, 23(4), 641-650.

Train, K. (2003) 'Discrete choice methods with simulation'. Cambridge: Cambridge University Press.

Traynor, T.L. (2008) 'Regional economic conditions and crash fatality rates – a cross-county analysis'. Journal of Safety Research, 39(2008), 33–39.

Ulfarsson, G.F., and Mannering, F.L. (2004) 'Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents', Accident Analysis and Prevention, 36(2), 135-147.

Underwood, G., Jiang, C., and Howarth, C., (1993). 'Modelling of safety measure effects and risk compensation', Accident Analysis and Prevention, 25(3), 277-288.

Vachal, K., and Malchose, D. (2009) 'What can we learn about North Dakota's youngest drivers from their crashes?', Accident Analysis and Prevention, 41(3), 617-623.

Vasconcellos, E. (1999) 'Urban development and traffic accidents in Brazil', Accident Analysis and Prevention, 31(4), 319-328.

Verwey, W.B., and Zaidel, D.M. (2000) 'Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behaviour', Personality and Individual Differences, 28(1), 123-142.

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

Vourvahakis, D., Petridou, M., Psaromichalaki, E., Petrou, A., Agouridakis, P., Anastassaki, M., and Askitopoulou, H. (1996) 'Pleasure travelling road traffic accidents in Crete', Resuscitation, 31(3), S37.

Walpole, R., Myers, R., Myers, S., and Ye, K. (2002) 'Probability and statistics for engineers and scientists', 7th edn, New Jersey: Prentice Hall.

Wang, C., Quddus, M.A., and Ison, S.G. (2009) 'The effects of area-wide road speed and curvature on traffic casualties in England', Journal of Transport Geography, 17(5), 385-395.

Wang, X., and Abdel-Aty, M. (2008) 'Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models', Accident Analysis and Prevention, 40(5), 1674-1682.

Wang, C., Quddus, M.A., Ison, S. G., (2009) 'Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England'. Accident Analysis and Prevention, 41(4), 798-808.

Washington, S., Metarko, J., Fomunung, I., Ross, R., Julian, F., and Moran, E. (1999) ‘An inter-regional comparison: Fatal crashes in the southeastern and non-southeastern United States: preliminary finding’, Accident Analysis and Prevention, 31(1-2), 135-146.

Washington, S., and Oh, J. (2006) ‘Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: Example applied to at grade railroad crossings in Korea’, Accident Analysis and Prevention, 38 (2), 234–247.

Washington, S.P., Karlaftis, M.J., and Mannering, F.L. (2003) ‘Statistical and econometric methods for transportation data analysis’. Boca Raton, Fla.; London: Chapman and Hall/CRC press.

Wedagama, D.M.P., Bird, R., and Metcalfe, A.V. (2006) ‘The influence of urban land-use on non-motorised transport casualties’, Accident Analysis and Prevention, 38(6)**,** 1049–1057.

Wier, M., Weintraub, J., Humphreys, E., Seto, E., and Bhatia, R. (2008) ‘An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning’, Accident Analysis and Prevention, 41(1), 137-145.

Wilde, G. (1998) ‘Risk homeostasis theory: an overview’, Injury Prevention, 4, 89-91.

Wilde, G. (2002) ‘Does risk homoeostasis theory have implications for road safety?’, Education and Debate, 324, 1149-52.

Williams, Richard (2006) ‘Generalized ordered logit/partial proportional odds models for ordinal dependent variables,’ The Stata Journal, 6(1), 58-82.

Williams, R., 11 August, (2008)-last update. Estimating heterogeneous choice models with Stata. Available: http://www.nd.edu/~rwilliam/oglm/oglm_Stata.pdf [22/7/2011].

World Health Organisation (2008) World report on road traffic injury prevention, The Global Burden of Disease. Geneva. Retrieved from: http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf

WHO (World Health Organisation). (2009). Global Status Report on Road Safety, Geneva. Obesity and overweight. Retrieved from:

http://www.who.int/dietphysicalactivity/publications/facts/obesity/en/

Wouters, P., and Bos, J. (2000) ‘Traffic accident reduction by monitoring driver behaviour with in-car data recorders’, Accident Analysis and Prevention, 32(5), 643-650.

Yahia M. Al-Khaldi, ATTITUDE AND PRACTICE TOWARDS ROAD TRAFFIC REGULATIONS AMONG STUDENTS OF HEALTH SCIENCES COLLEGE IN ASEER REGION, J Family Community Med. 2006 Sep-Dec; 13(3): 109–113.

Yamamoto, T., Hashiji, J., and Shankar, V.N. (2008) ‘Underreporting in traffic accident data, bias in parameters and the structure of injury severity models’, Accident Analysis and Prevention, 40(4), 1320-1329.

Yau, K. (2004) ‘Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong’, Accident Analysis and Prevention, 36(3), 333-340.

Yereli, K., Balcioğlu, I., and Forensic, A. (2006) ‘Is Toxoplasma gondii a potential risk for traffic accidents in Turkey?’ Forensic Science International, 163(1-2), 10, 34-37.

Ying, C., and MacNab (2004) ‘Bayesian spatial and ecological models for small-area accident and injury analysis’, Accident Analysis and Prevention, 36(6), 1019–1028.

Zadeh, H., Vahabi, R., Nazparvar, and Amoei, M. (2002) ‘An epidemiological study and determination of causes of traffic accident-related deaths in Tehran, Iran (during 2000–2001)’, Journal of Clinical Forensic Medicine B, 9(2), 74-77.

Zajac, S.S., and Ivan, J.N. (2003) ‘Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut’, Accident Analysis and Prevention, 35(3), 369-379.

Zeeger, C.V., Stewart, R., Reinfurt, D., Council, F., Neuman, T., Hamilton, E., Miller, T., and Hunter, W. (1990) ‘Cost-effective Geometric Improvements for Safety Upgrading of Horizontal Curves’. Chapel Hill: University of North Carolina.

APPENDIX -A: PUBLICATIONS FROM THIS WORK

Following original research papers have been produced and presented at conferences by the author along with his supervisors as a result of this research:

Altwaijri, S. A. (2010) 'An exploration of factors affecting the severity of road injury crashes in Riyadh city, Kingdom of Saudi Arabia'. Paper presented at the Fourth Saudi International Conference, Manchester, UK, June 2010.

Altwaijri, S. A., Quddus, M.A., and Bristow, A. L. (2011a) 'Analysing the severity of traffic crashes in Riyadh city using statistical models'. Paper presented at the 43rd UTSG Annual Conference, Milton Keynes, UK, January 2011.

Altwaijri, S.A., Quddus, M.A., and Bristow, A. L. (2011b) 'Factors affecting the severity of traffic crashes in Riyadh city'. Paper presented at the 90th Annual Meeting of the Transportation Research Board (TRB), Washington, DC, January 2011.

Altwaijri, S.A., Quddus, M.A., and Bristow, A. L. (2011c) 'Analysing the severity and frequency of traffic crashes in Riyadh city using statistical models'. Paper presented at the First Saudi Traffic Safety Forum, Dammam, Kingdom of Saudi Arabia, December 2011.

Quddus, M.A., Bristow, A. L., and Altwaijri, S. (2012) 'A spatial analysis of traffic crashes in Riyadh city using GIS and statistical models'. Paper presented at the 91st Annual Meeting of the Transportation Research Board (TRB), Washington, DC., USA, January2012.

Altwaijri, S.A., Quddus, M.A., and Bristow, A. L., (2012). Analysing the severity and frequency of traffic crashes in Riyadh city using statistical models. International Journal of Transportation Science and Technology (IJTST), in process.