Investigating the Impacts on Cyclist Casualty Severity at Give Way Roundabouts with Mixed Traffic

SCHOOL OF ENGINEERING

NEWCASTLE UNIVERSITY

By

Nurten Akgün

A thesis submitted for the degree of

Doctor of Philosophy in Transport at

Newcastle University

September 2019

Declaration

This thesis is submitted to Newcastle University for the degree of Doctor of Philosophy. I declare that this thesis and the original work presented in it is my own and has been generated by me. The material presented has never been submitted to any other educational establishment for purposes of obtaining a higher degree.

September 2019 Nurten Akgün

Abstract

Roundabouts are designed to improve the safety for all vehicles by decreasing the number of conflict points at intersections and reducing entry and circulating speed. P*revious research suggests that the design does not provide similar safety benefits for vulnerable users, particularly cyclists.* Local road authorities usually have very limited budgets for improving cycling facilities however, recently in the UK, there is more emphasis on policies to promote cycling particularly in urban areas. This means that cyclists increasingly are using give way roundabouts in mixed traffic, and therefore, there is a need for a fundamental understanding of which design parameters influence cyclist safety and what are the behaviour related contributory factors.

The global aim of this research is to investigate statistically significant variables, considering geometric design parameters, sociodemographic descriptors of cyclist, meteorological conditions, traffic characteristics and driver/rider behaviour related contributory factors that have impact on cyclist casualty severity at give way roundabouts with mixed traffic.

The first analysis explored the significant geometric design parameters, socio-demographic characteristics of cyclist, meteorological conditions and speed limit on casualty severity. Two components namely Approach Capacity (number of lanes on approach, half width on approach, number of flare lanes on approach and entry path radius) and Size of Roundabouts (number of arms, type of roundabout and number of circulating lanes) emerged from the Principal Component Analysis. The Multiple Logistic Regression suggested that a unit increase in number of lanes on approach, entry path radius and speed limit increase the probability of serious casualty occurrence with odd ratios 4.97, 1.04 and 1.02, respectively and a higher Approach Capacity increases the probability of serious casualty occurrence by 86% (odds ratio 1.86). Linear Regression suggested that if the entry path radius was more than 80 metres, the casualty severity was more likely to be serious.

The second analysis explored the impact of driver/rider behaviour related contributory factors on cyclist casualty severity. One-unit increase in cyclist age group, junction restart, failed to look properly and failed to judge other person's path or speed, increased the probability of Killed or Seriously Injured (KSI) casualty occurrence with odds ratio 1.15, 2.09, 2.82 and 1.64, respectively. Multilevel Logistic Regression showed that the regional variance between cities in England was not statistically significant at the 95% confidence level.

In the final analysis, a comparison was made between roundabouts in England and Belgium using three-way chi square test of independence, Multiple and Multilevel Logistic Regressions. The results showed that older cyclists were more likely to be involved in KSI than slight casualties in both countries. Cyclist's non respect of the priority to drivers increased the probability of KSI casualty. Speed limit emerged as a significant contributory factor in KSI casualties in England (tangential design); however, it did not show any significance in Belgium (radial design). In addition, country residual was statistically significant in the multilevel modelling.

This research has demonstrated that speed has a dominant impact on cyclist casualty severity but the novelty rests with identifying that it is the approach capacity, and more specifically entry path radius and number of lanes, that most influences vehicle speed. This has enabled generic predictors for the probability of severity, which are valid for specific countries and regions with similar design approach.

Acknowledgements

I would like to express my deepest gratitude to my supervisors Dr Dilum Dissanakaye, Dr Neil Thorpe and Professor Margaret Carol Bell CBE for their consistent guidance. I am grateful for their scientific support and emotional encouragements during all stages of this research. This thesis would not be the same without their suggestions and input.

I owe my sincere gratitude to Professor Stijn Daniels and Nina Nuyttens from VIAS Institute, Belgium for their contribution in this research. Special thanks are due to Gateshead Council, Department for Transport (the United Kingdom), EDINA and Data Library (University of Edinburgh) and VIAS Institute (Belgium) for providing the data used in this thesis.

I am grateful to the Republic of Turkey, Ministry of National Education for giving me this great opportunity by providing the financial assistance.

I was delighted to interact with both of the panel examiners, Dr Anil Namdeo and Mr. Roger Bird. I also would like to thank to the members of Transport Operation Research Group (TORG) and my PhD colleagues for their support and suggestions.

Many thanks to my dearest friends in Postgraduate Research Centre to become a part of my family and help me to spend amazing time in the UK. Our memories are unforgettable! And finally, thanks to my mother and father, for their unconditional love and support during this journey.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

DfT	Department for Transport
EFA	Exploratory Factor Analysis
EPV	Events per Variable
KMO	Kaiser Meyer Olkin
KSI	Killed or Seriously Injured
MLR	Multiple Logistic Regression
MMLR	Multiple Multilevel Logistic Regression
PCA	Principal Component Analysis
PPMC	Pearson Product Moment Correlation
SLR	Simple Logistic Regression
SMLR	Simple Multilevel Logistic Regression

Chapter 1 Introduction

1.1 Background

Transportation has economic, social and environmental influences, and amongst the various travel modes available in transport systems, non-motorised modes such as walking and cycling, and public transport are considered to be environmentally and economically friendly travel modes (Litman and Burwell, 2006). In particular, the main benefit of cycling is efficiency having journey-specific advantages such as being the fastest door to door travel mode in urban areas (Parkin, 2018). Given the benefits (including health, environment, economic etc.) transport authorities and policy makers continue to encourage the public to cycle through society's awareness programmes which promote the advantages of cycling such initiatives have increased in recent years. However, safety in traffic is one of the main barriers to significantly enhance the number of cyclists, and people still hesitate to choose cycling as a travel mode in their daily life. Road infrastructure plays a major role in creating a safer travel environment to cyclists; therefore, this research area came into prominence.

Roundabouts reduce or alter the conflict points and force a reduction in motor vehicle speed when entering the roundabouts by providing a deflection (Retting *et al.*, 2001; Gross, 2013; Silvano and Linder, 2017). In addition to safety, they also deliver capacity and environmental (such as air pollution) advantages (Silvano *et al.*, 2015), and delays are distributed more uniformly (Silvano and Linder, 2017). Therefore, roundabouts are known as being safer for motor vehicle drivers than signalised and priority junctions, and as a consequence many intersections have been converted to roundabouts in order to increase the capacity and reduce the number of crashes (Montella, 2011). This has led to a wide range of detailed designs of roundabouts, with numbers increasing every day.

However, the safety performance of roundabouts is questionable for vulnerable users, particularly cyclists (Daniels *et al.*, 2008; Jensen, 2017). Researchers suggest that roundabouts should be investigated in detail to identify the impacts on cyclist safety and eliminate these influences to keep encouraging people to cycle (De Brabander and Vereeck, 2007; Daniels *et al.*, 2010; Daniels *et al.*, 2011; Polders *et al.*, 2014). This suggestion was the starting point for the motivation of the study in this thesis and elaborate upon in the following section.

1.2 Motivation of Study

Converting priority and signalised junctions to roundabouts has been shown to increase the number of crashes and casualty severity for vulnerable users, particularly cyclists (Robinson *et al.*, 2000; Persaud *et al.*, 2001; Elvik, 2003; De Brabander and Vereeck, 2007). Therefore, roundabouts are considered to be high risk locations for cyclists. The common message in the literature on cyclist safety at roundabouts was that the contribution of geometric design and traffic related variables for vulnerable user safety needed to be investigated in a more detailed analysis.

Previous studies did not consider a wide range of variables, such as geometry, traffic, sociodemographic, environmental and behaviour related contributory factors, together in one model mainly due to data availability (De Brabander and Vereeck, 2007; Daniels *et al.*, 2010; Daniels *et al.*, 2011; Polders *et al.*, 2014). More specifically, the impact of these variables at tangential design style roundabouts is still not clear (Davies *et al.*, 1997; Lawton *et al.*, 2003; Jurisich *et al.*, 2011). More specifically the role and relative importance of these influences on cyclist casualty severity reduction at roundabouts has not been carried out.

With respect to the literature review, several research gaps were identified, namely cyclist casualty severity analysis, influence of geometric design parameters and driver/rider behaviour on casualty severity, investigating the consistency of casualty modelling whilst including different countries. It is important to address these gaps because cycling is increasing every year in response to local government policy. Gaining a much deeper understanding of the impact of a wide range of variables on cyclist casualty severity at roundabouts is what the research presented in this thesis aimed.

Considering the collective knowledge from the literature review, several research gaps embracing cyclist safety and analytical applications were identifies as follows:

- Cyclist casualty severity analysis, with logistic regression including comprehensive set of predictive variables, was not applied.
- Geometric design parameters were not fully considered. Some critical variables, such as speed, speed limit, sociodemographic characteristics of cyclist/driver and meteorological conditions were only partially included and analysed in a few of these studies, but the studies were not considering casualty severity reduction.

- The influence of behaviour was informed by yielding and perception related research. The impact of driver/rider behaviour on casualty severity was not considered.
- Additional statistical applications, such as descriptive statistics, test of normality, correlation analysis, dimension reduction, reliability analysis, were applied generally and not conducted to develop a reliable empirical model or determine internal relationships between variables.
- The interpretation of logistic regressions was very narrow and detailed analysis such as calculating the predictive margins was not considered.
- The results of previous studies pointed out the impacts on casualty severity. However, advice for policy makers and design engineers were very shallow. A reverse modelling approach (both logistic regression and linear regression) has not been attempted in previous studies.
- A multilevel modelling, which included different regions/counties, was not applied. Therefore, a comparative analysis and determining the consistency of the models were not identified.

1.3 Research Questions

Given the gap in the detailed state of art review in this thesis, five research questions can be posed:

1. What are the relative contributions of geometric parameters for give way roundabouts with mixed traffic, speed limit and traffic flow profile to cyclist casualty severity?
2. What are the relative contributions of sociodemographic characteristics of cyclist and environmental conditions to cyclist casualty severity?
3. What is the relative contribution of driver/rider behaviour related contributory factors to cyclist casualty severity?
4. What is the consistency of the results for cyclist casualty severity between tangential and radial roundabouts based on a comparative analysis?
5. What is the appropriate statistical method to analyse the safety impact of variables on cyclist casualty severity?

These five research questions indicate the direction of the study reported in this thesis and help to formulate the aim and objectives of the research as given in the following sections.

1.4 Research Aim

This study aims to investigate the relative contribution of variables such as geometric design, sociodemographic, environmental conditions and behaviour related contributory factors on cyclist casualty severity that occurred at give way roundabouts with mixed traffic, with the study also aiming to identify city/county based regional influence on the prediction models.

1.5 Research Objectives

The objectives of the research are given as follows:

1. To identify the relationship between considered variables using a combination of correlation analysis, dimension reduction and chi square tests;
2. To explore the statistically significant impact of geometric design parameters, traffic variables, sociodemographic, environmental condition and driver/rider behaviour related contributory factors using regression analysis;
3. To estimate a mathematical model which explores the city/regional influence on casualty severity analysis;
4. To identify the influence of considered variables in extended two country-based analysis.

1.6 Research Tasks

In order to achieve the research objectives, the following tasks were proposed:

1. Carry out a critical literature review of previous studies including roundabout safety performance for all road users and particularly cyclists in detail to identify the research gap in the knowledge and determine the predictive variable which may have influence on cyclist casualty severity.
2. Determine the role of existing cyclist casualty dataset and identify the method of data collection including access from available data resources, measuring and manipulating data into formats usable in statistical analysis.
3. Assemble and critically review a wide range of statistical methods used in previous studies to develop a reliable and comprehensive set of analytical and prediction methods that are appropriate to fill the gaps. Investigate the assumptions and limitations of each statistical method and make a decision to apply the most suitable method consistent with the structure of the dataset.

4. Investigate the impacts of the range of variables on cyclist casualty severity for crashes occurring at give way roundabouts with mixed traffic. Conduct several statistical methods to identify the interrelationship between variables and investigate individual/group influence on severity based on the derived prediction models.
5. Extend the study area across the country including a greater number of cyclist casualty records and carry out a comparative analysis internationally in order to determine consistency in results from the different steps in the analysis.
6. Draw conclusions, discuss limitations of the study and make recommendations for design engineers, policy makers and future research to improve the safety for cyclists at roundabouts by reducing the level of severity.

1.7 Thesis Outline

The thesis begins with a critical review of the literature presented in Chapter 2. This review starts with a general knowledge on road safety for all road users at roundabouts and continues with cyclist safety in detail. In addition, it focuses on determining the most appropriate statistical methods to conduct the analysis in this thesis. The conclusion of Chapter 2 provides the research gaps in the literature and informs the most suitable analytical methods to address the research gaps.

Chapter 3 outlines the steps in the methodology, data collection and details each stage of the analysis including the statistical methods employed. Chapter 3 provides a flow diagram which summarises the methodology and forms a basis for structuring the thesis. Chapter 4 includes details of data collection. Three types of data collection methods were involved in this study: i) obtaining cyclist casualty records from the STATS19 available from the local authorities; ii) direct measurement of geometric design parameters and iii) associated demographics and behavioural data direct from the DfT by special permission. Finally, details are given of the coding of the dataset to prepare for application in the three statistical analysis steps.

Chapter 5 reports the results of Analysis 1, which investigates the impact of geometric design, sociodemographic characteristics of cyclists, environmental conditions and traffic characteristics on cyclist casualty severity that occurred in crashes at give way roundabouts with mixed traffic. The case study area was selected as Northumbria (North East of England) and the casualty records used were for the period between 2011 and 2016. The analysis starts with descriptive statistics and is followed by correlation analysis, dimension reduction and a reliability test. Finally, Multiple Logistic Regressions provide the influence of predictors on casualty severity.

The influence of driver/rider behaviour related contributory factors is demonstrated in Analysis 3 given in Chapter 6. The first section of the analysis develops a Multiple Logistic Regression model based on relaxing p-value criteria and in the second section the study is extended by including cities across England to investigate variance of city impact on the model using Multilevel Logistic Regression. A further comparative study for investigating the variance between the countries of England and Belgium is shown in Analysis 3, given in Chapter 7. This comparative analysis started with a three-way chi square test of independence and was followed by Multilevel Logistic Regression Models.

Finally, the thesis is concluded in Chapter 8. First, the main and secondary findings are detailed and discussed. The limitations of the study and recommendations to highway design engineers and policy makers are articulated based on the key messages. Lastly, suggestions for further studies are provided.

Chapter 2 Cyclist Safety at Roundabouts and Analytical Approaches for Investigating Cyclist Safety – A Critical Review

2.1 Introduction

In the UK, roundabouts are the location where 10% of the total cyclist crashes occur. This rate is 14 times that of motor vehicle crashes occurring at roundabouts. Moreover, the number of cyclist crashes at roundabouts is three times that at signalised intersections (Davies *et al.*, 1997). Regarding the casualty severity, roundabouts particularly are not safe for cyclists (Daniels *et al.*, 2010). This is the prime motivation for this study which aims to investigate the impact of variables, such as geometric design, traffic, sociodemographic, environment and behaviour related factors, on cyclist casualty severity at roundabouts.

This chapter provides a comprehensive review of the literature to develop an understanding of the design of roundabouts and former research on general safety issues for all road users, but more specifically cyclists. The chapter also evaluates previous studies in the problem statement section, in order to identify the research gap and illustrate how this study sets out to fill identified gaps in literature. This chapter also reviews methods of analysis employed in previous studies to identify methods appropriate for this research. In particular, correlation analysis to explore the relationship between variables, dimension reduction to address the assumptions and limitations of the regression modelling and finally, regression models to find out the most appropriate approach for this study.

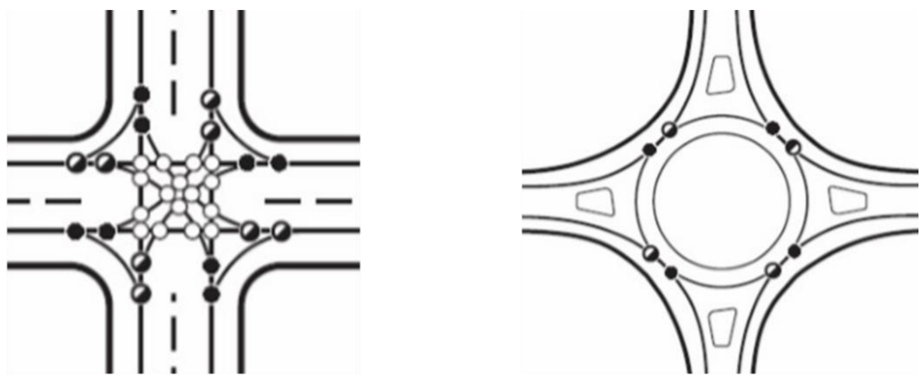
Regarding the structure of this chapter: Section 2.2 presents details of the geometric design of roundabouts; Section 2.3 provides a literature review of safety studies at roundabouts for all road users; Section 2.4 focuses on vulnerable road user, particularly cyclist, safety at roundabouts; the statistical methods and models are reviewed in Section 2.5 and the research gap is stated in Section 2.6. Finally, a conclusion of this chapter is presented in Section 2.7.

2.2 Understanding the Geometric Design of Roundabouts

In order to assist in a critical review of former studies on cyclist safety at roundabouts, this section provides in-depth knowledge of the design philosophies of different types of roundabouts. The first step is to explain ‘What is a roundabout?’ and ‘For what reasons do we use roundabouts rather than signalised intersections?’

2.2.1 *What is a Roundabout?*

Intersections are the main locations that cause traffic problems, such as vehicle crashes, emissions and queues. Mazari *et al.* (2008) suggested that roundabouts have a significant impact on the quality of life since road users experience traffic congestion. A roundabout is a type of intersection which has a one-way circulatory traffic flow around a central island (DfT, 2007). Roundabouts are safer than priority junctions for vehicle drivers because they reduce conflict points compared to signalised junctions (Montella, 2011). Conflict points are potential collision locations of traffic at junctions (i.e. eight conflict points exist at a four-arm roundabout, while this number is 32 at a four-arm priority junction) (Fromme, 2010) (See Figure 2.1). Therefore, roundabouts are designed to improve the traffic safety by decreasing the number of conflict points at junctions.



a) Four Arms Give-Way Junction b) Four Arms Give-Way Roundabout

Figure 2.1 - Conflict Points Illustration at Give Way Junctions (a) and Roundabouts (b) (FHWA, 2000)

There are two types of applications of traffic control systems at roundabouts: signalised and give way. A signalised roundabout has traffic lights to control the traffic movement. It should be used where a combination of factors, such as increased traffic flow, unbalanced traffic flow between arms, high circulatory speed and significant increased flows at peak hours, are present. In 1966, the United Kingdom applied the give way rule "priority to right" such that traffic entering a roundabout should yield, allowing priority to circulating traffic (Bruce *et al.*, 2000). At give way roundabouts, drivers must adhere to the road markings when entering and driving through a roundabout, always giving priority to the moving traffic from the right-hand side (DRIVINGED, 2018). This type of roundabout increases the capacity of the intersection by enabling continuous moving, correct positioning and managing movements at conflict points. However, if the traffic flow on arms is not balanced, these give way roundabouts should be signalised (Chard *et al.*, 2009).

Interestingly, (Tollazzi, 2015) claims that:

- There is no uniform guideline in terms of roundabout design because each country focuses on their own requirements.
- A safe design solution in one country might be very dangerous in another.
- Each country has their own design philosophy and dimensions, hence there is a difference in vehicle dimensions and human behaviour related factors
- Consequently, design guidelines are individual and specific for most of the countries.

For the UK, geometric design definitions and limitations are provided in the design manual, namely Geometric Design of Roundabouts TD 16/07 UK (DfT, 2007) (See Table 2.1). The standards are mainly developed for motor vehicles and heavy vehicles and the specifications do not apply to vulnerable users.

Roundabout design is a site-specific process for individual applications with their own characteristics, such as traffic flow, maximum speed requirement and construction space (Taylor, 2011). In the other words, design parameters are flexible to be quantified for each roundabout from different requirements in design; thus, several types of roundabouts have emerged. The following subsection provides different types of roundabouts.

Table 2-1 Geometric Design Parameters for UK Roundabouts (DfT, 2007)

Name	Definition
Central island	The circular island which is in the centre of roundabout
Splitter island	The kerbed island which separates entering and leaving traffic on each arm
Approach half width	The shortest distance between edge of the road and median line at the approach arm
Entry width	The shortest distance between the corner of the splitter island and edge of the road at the entry of a roundabout
Entry angle	The geometric proxy for the conflict angle between entering and circulating traffic streams
Average effective flare length	The average curve length which is parallel to the road edge curb
Entry kerb radius	The minimum tangential radius of the curve nearside the road
Entry path radius	The radius of the deflection to the left imposed at entering a roundabout
Exit width	The shortest distance between the corner of the splitter island and edge of the road at the exit of a roundabout
Exit kerb radius	The radius of the deflection to the left imposed at exiting a roundabout

2.2.2 *Types of Roundabouts*

In general, worldwide there are two types of design base, either radial or tangential (Patterson, 2010) (See Figure 2.2). The radial base is used mainly in continental European countries; thus, it also is called the 'continental design'. In radial base design, the legs of the roundabout are stated as radial to the centre. This brings a very big advantage of significant speed reduction since radial roundabouts have a tight geometry at entry locations (See Table 2.2). However, this also brings a disadvantage of less capacity. On the other hand, tangential roundabouts are applied mainly in the UK, New Zealand and Australia. The performance of a tangential base structure works in reverse to a radial base. In a tangential design, legs are tangential to the centre of the roundabout. Speed reduction is achieved with a deflection at the entry; however, both traffic speed and capacity remain high (Patterson, 2010).

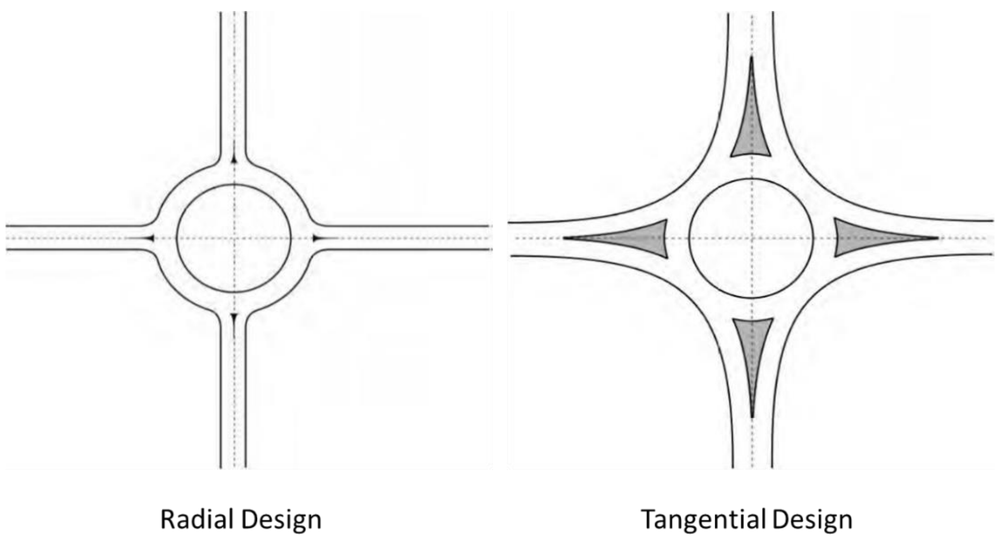


Figure 2.2- Radial and Tangential Design of Roundabouts (Patterson, 2010)

Table 2-2 - Upper and Lower Limits of Design Features for Radial and Tangential Geometry

Design Features	Radial				Tangential	
	Urban		Rural		Urban/Rural	
	Min.	Max.	Min.	Max.	Min.	Max.
Central island (m)	5	-	16	-	4	-
Splitter island (m)	-	-	-	-	-	-
Approach half width (m)	-	-	-	-	-	-
Entry width (m)	3	7	4	9	-	-
Entry angle (degree)	-	-	-	-	20	60
Average effective flare length (m)	-	-	-	-	25	100
Entry kerb radius (m)	8	15	10	15	6	100
Entry path radius (m)	-	100	48	100	-	100
Exit width (m)	4	-	4	7	7	11
Exit kerb radius (m)	15	20	15	20	20	100

Roundabouts in the UK primarily are designed for increasing their capacity for vehicles. Therefore, in the UK tangential structure with wide and deflected entry is a preferred geometric design for roundabouts (Lawton *et al.*, 2003). However, studies (Davies *et al.*, 1997; Lawton *et al.*, 2003) have shown that the tangential design application did not improve safety for cyclists after converting signalised junctions to roundabouts. These studies compared the radial and tangential designs in order to illustrate the differences between both design methods on capacity and safety. The results showed that radial design (with tighter entry geometry) increases the safety for all road users; however, the capacity was much lower compared to the UK design. The studies (Davies *et al.*, 1997; Lawton *et al.*, 2003) recommended further research was needed to identify the optimum design for higher capacity and safety for all road users. This recommendation was considered, to develop a comparative analysis between England and Belgium in this thesis (See Chapter 7).

Given these two main types of basic geometric design, several types of roundabouts have developed in application. Tollazzi (2015) stated that there are three main groups: i) roundabouts (normal, mini, grade separated, double) which have been already implemented in most of the countries, ii) modern roundabouts (turbo, dog bone, compact semi-two-lane roundabout) applied in some countries, and iii) under development solutions on roundabouts (turbo-square, flower, target, with segregated left-turn slip lanes) (See Appendix A).

The study in this thesis was conducted using data for the United Kingdom roundabouts. Therefore, the UK roundabout design standard definition and specifications will be used as the basis for data collection and interpretation of results. The UK roundabouts are designed mainly according to Volume 6 section 2 (Design of Mini-Roundabouts) and section 3 (Geometric Design of Roundabouts) of the Design Manual for Roads and Bridges (DfT, 2007). The manual defines the types of roundabouts, geometric design parameters, limitations, aspects and hierarchy of design, as well as safety at roundabouts. How these design parameters are measured and used in this study will be described in detail in Chapter 3.

The types of roundabouts are primarily aimed to reduce delay and provide a service to motor vehicles. In order to accommodate cyclists into the traffic stream at roundabouts the capacity for vehicles is compromised and has led to different types of infrastructure solutions (See Figure 2.3). The first and very common one in the UK is a roundabout with mixed traffic. In this situation, both motor vehicles and cyclists are sharing the road. The second type of solution is by applying cycle lanes either adjacent to the main carriageway or a completely separated infrastructure.

Regarding the roundabout geometric design and several types of roundabouts, former studies have been reviewed to gain deeper understanding of the safety aspect, initially for all road users and then cyclists in particular. The following subsections focus on reviewing this previous research.

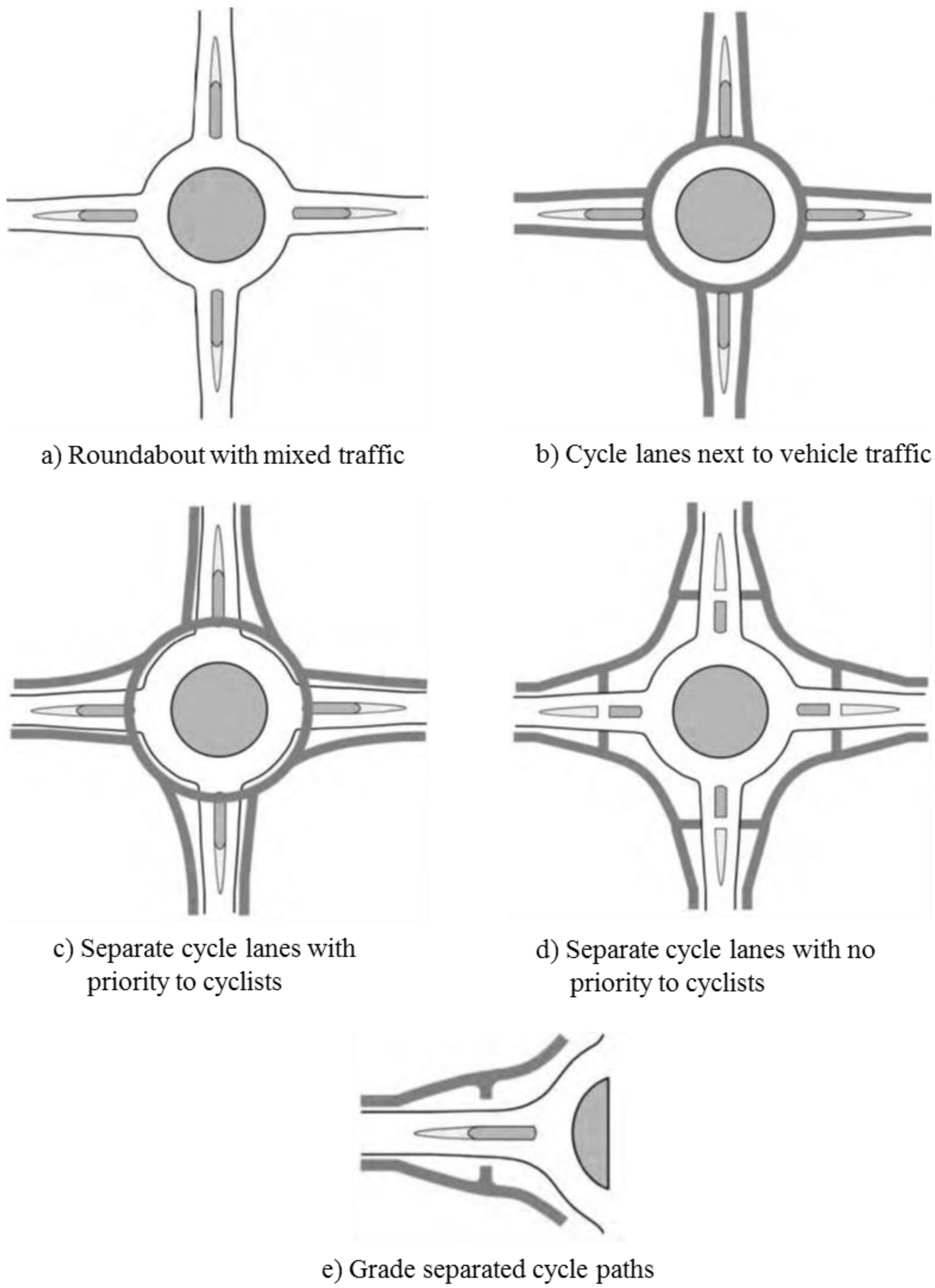


Figure 2.3 – Different Types of Cycling Infrastructure (Dark Grey Coloured) at Roundabouts (Daniels et al., 2009)

2.3 Safety at Roundabouts for All Road Users

A roundabout is known as a safer intersection for vehicle drivers than signalised and priority junctions, since roundabouts are eliminating or altering the conflict points and all vehicles are forced to reduce their speed while entering the roundabouts (Retting *et al.*, 2001; Gross, 2013; Silvano and Linder, 2017). In addition to safety, they also have capacity and environmental (such as air pollution) advantages (Silva *et al.*, 2014) and delays are distributed more uniformly (Silvano and Linder, 2017). Therefore, many intersections have been converted to roundabouts in order to increase the capacity and reduce the number of crashes (Montella, 2011).

In detail, former studies investigated the safety impact of converting signalised junctions to roundabouts (Robinson *et al.*, 2000; Persaud *et al.*, 2001; Elvik, 2003; De Brabander and Vereeck, 2007). The study carried out by Robinson *et al.* (2000) stated that roundabouts are associated with a reduction in the number of crashes after converting from signalised junctions. This observation was supported by Mazari *et al.* (2008) as roundabouts provide greater safety than signalised junctions.

The study carried out by Persaud *et al.* (2001) in the USA observed a 40% reduction in all types of number of crashes at 23 roundabouts converted from stop signed and signalised junctions. Moreover, an 80% reduction in injury crashes was claimed. This reduction was 72% for number of crash occurrences and 88% for injury crashes at single lane roundabouts. On the other hand, the number of crash occurrences and injury crashes reduced by only 5% at multilane roundabouts. The safety impact of converting stop sign and signalised junctions to roundabouts was more significant at roundabouts with single lane than with multilane. This result showed that multilane roundabouts with their increase in number of conflict points still carry the potential risk of crash and injury occurrence at a level similar to stop sign or signalised junctions. The study (Persaud *et al.*, 2001) recommended that roundabouts may not be the best option, if the volume of users is high.

De Brabander and Vereeck (2007) developed a comparison analysis between 95 roundabouts and 230 signalised intersections in Belgium. The total number of 325 samples was grouped and sub-grouped by the authors according to construction years and speed limits, which are 50 km/h, 70 km/h and 90 km/h. The aim of the classification was to obtain a precise comparison of the speed limit impact whether inside or outside of a built-up area had an effect on vulnerable user safety. The total crash number and severity details recorded at these intersections were collected and regression to the mean effect was calculated for each subgroup. The results showed that the total crash and number of serious injuries after roundabout conversion reduced by 39% and 17% respectively; however, roundabouts protect vulnerable users less effectively than signalised intersections. The reduction of number of injuries was 49% at give way roundabouts and 32% at signalised roundabouts, which concluded that give way roundabouts were performing better in total safety improvement than signalised roundabouts. The largest reduction occurred in high speed limit areas rather than low. This result led to the hypothesis of 'roundabouts perform less at inside of a built-up area than outside of a built-up area'. De Brabander and Vereeck (2007) stated that this study had limitations such as lack of traffic volume data which might have an important impact on the validity of crash analysis results. The authors (De Brabander and Vereeck, 2007) recommended that estimation of the reduction of crashes needs further study.

However, roundabouts are known to be safer for vehicle drivers, since there is evidence that they are reducing the number of crashes and that the severity of injury increases after converting signalised junctions to roundabouts is questionable (Mazari *et al.*, 2008). The study carried out by Mazari *et al.* (2008) claimed that the number of crashes cannot be used as a measure of safety when comparing junction types and it was recommended that more in-depth investigation needs to be carried out in order to understand the real impact of roundabouts on safety. Roundabouts are not the appropriate intersection design under all circumstances and sometimes they might be proactively avoided (Lenters, 2004), who stated the main reasons why as follows:

- *The space is not available for an acceptable outside diameter and the cost of the construction is high.*
- *Profile and the grade on entries are more than 4%.*
- *Traffic flows on each arm are severely unbalanced.*
- *Signal coordinated networks have impact on platooned traffic flow*
- *Horizontal or vertical impediments do not provide an available driver sight of the yield line at entry locations.*

The research previously suggested that roundabouts were safer for vehicle drivers after being converted from stop sign or signalised junctions. However, roundabouts still might be considered as risk locations for traffic. Safety for road users depends highly on many factors, such as geometry, pavement, markings, signing, driver education, public awareness and enforcement (Furtado, 2004). Only a few studies were carried out to show the impact on safety with the use of roundabouts (Nambisan and Parimi, 2007). The following studies tried to clarify the main reason for crash occurrence at roundabouts and looked for possible solutions (Daniels *et al.*, 2010; Daniels *et al.*, 2011; Montella, 2011; Polders *et al.*, 2014; Silva *et al.*, 2014).

Montella (2011) studied contributory factors on crash occurrence for all types of road users at 15 Italian roundabouts. The research identified 62 highly significant contributory factors with 2094 secondary contributory factors. The independence between all contributory factors was examined carefully in order to understand the main reason for high numbers of vehicle crashes at roundabouts. The results of the study showed that the most common crash type occurred at entry and circulating locations. In terms of contributory factors, geometric design parameters were involved in 60% of the crashes with the most significant geometric variable being radius of deflection (entry path radius) causing rear end crashes. More than half of crashes were associated with the lane marking factor; in addition, one third with the impact of pavement condition. Vehicle impact was negligible in crash occurrence at roundabouts. The study suggested that geometric design of a roundabout has significant impact on crash occurrence for all types of road users; however, improving geometric design may not be an economic solution. Therefore, marking and signs should be considered to reduce the number of crashes (Montella, 2011).

A well-designed roundabout brings a benefit of speed reduction which usually leads to homogenous behaviour (Turner and Roozenburg, 2009). However, roundabouts with multi lanes increase the capacity (Lindenmann, 2006), but they reduce the effectiveness of speed reduction which in turn influences driver behaviour (St-Aubin *et al.*, 2013). Bastos Silva *et al.* (2006) claimed that the higher the number of lanes the greater the freedom for drivers increasing the potential conflicts. Higher crash rates not only result from a higher number of conflict points, but also insufficient deflection which controls the speed while entering the roundabouts (Bastos Silva and Seco, 2005).

Silva *et al.* (2014) stated that driving behaviour at roundabouts was influenced by three main levels: i) speed profiles; ii) lateral acceleration profiles; iii) roundabout geometry. Therefore, the study examined driving behaviour at roundabouts with two lanes in an arterial road in order to describe the relationship between a roundabout's geometry, speed and lateral acceleration profiles. It was proven that roundabouts have a significant impact on speed reduction (between 26% and 37%) and the impact area was between 400m and 500m. The size of the impact area depended on approach speed and the deflection, which showed the importance of geometric design on speed reduction once more. Speed reduction consistency by using geometric design parameters helped in reducing the vehicle crash possibility and controlling a vehicle's speed at approach, entry and circulating locations on a roundabout. Vehicle crash probability at entry locations of roundabouts is higher than circulating and exit regions. The approach speed had significant impact on entry speed. The roundabouts' impact on speed reduction was found to be statistically significant; however, a homogenous behaviour was not observed between drivers because roundabouts with double lanes gave more freedom of movement compared to single lane roundabouts. Therefore, Silva *et al.* (2014) recommended that the geometric design speed and the impact of entry geometry needs to be investigated in more detail with a larger number of samples or an alternative detailed methodology in order to clarify the speed reduction effect of the approach lane of a roundabout and the associates with increase in the roundabout's safety.

Entry geometric design impact was mentioned in the Road Design Guide for Roundabouts by Austroads (2009). This report stated that the main reason for vehicle crashes is inconsistency in speed reduction behaviour of drivers at approach and entry locations of roundabouts. Road safety consistency is "the conformance between road geometric design and driver's expectancy" (Lamm *et al.*, 1999). Therefore, any speed reduction behaviour inconsistency increases crash probability. Austroads (2009) states that speed should be reduced to the correct expected limits when approaching the roundabouts. This means not only entry path deflection, which is the most important determinant for safety which controls speed by geometry of the roundabout (DfT, 2007), but also the entire entry geometry which should be examined in roundabout safety studies (Austroads, 2009).

Crash contributory factors at roundabouts were studied by Daniels *et al.* (2010) and Polders *et al.* (2014). Both studies have several similarities, such as using crash severity analysis for all road users, similar lighting conditions and being Belgian based case studies. On the other hand, Polders *et al.* (2014) considered cycling facilities and explored connections between crash severities to roundabout geometric parameters. They studied 28 roundabouts each divided into 11 segments to determine crash locations along with details of casualty types. Pearson's chi square test was used to investigate the statistical relationships between variables. The results showed that the number of injury crashes of vulnerable road users is higher than for vehicles and the highest serious injury risk group of road users at roundabouts is cyclists and moped riders. 80% of crashes occur at circulating and entry locations of roundabouts (Polders *et al.*, 2014).

Alternatively, Daniels *et al.* (2010) in their model considered vehicle traffic flow, age, gender, lighting conditions and alcohol consumption as the main parameters related to safety issues. They developed severity analyses of 1491 crashes at 148 roundabouts using logistic regression and hierarchical logistic regression models. The latter study stated that the crash severity is highly related to road user types and showed that vulnerable users are in the high-risk group of serious injuries. However, the study showed that fatality and serious injury are rare for vehicle drivers. In addition, the majority of fatal or serious injuries in multiple vehicle collisions are cyclists and the severe and fatal injury probability increases for the older road users (Daniels *et al.*, 2010).

Whilst the results of Daniels *et al.* (2010) and Polders *et al.* (2014) were credible, the studies suffered from several important limitations such as lack of knowledge of speed limit or vehicle speed and geometric design parameters. Daniels *et al.* (2010) recommended that vehicle speed impact and geometric design parameters should be considered in roundabout safety analysis studies. Both studies show that the main casualty risk group at roundabouts is vulnerable road users, especially cyclists.

The validity of this result was supported in a follow-on study by Daniels *et al.* (2011). Vulnerable users were found to be more likely to be involved in injury crashes. In terms of number of crashes, separate cycle paths emerged as being safer than other types of cycle facilities, such as road share and roadside cycle paths. Roundabouts with four or more arms had a higher number of crash occurrence for all road users than three arms, and single vehicle crashes were more likely to occur at roundabouts with larger central islands. The study (Daniels *et al.*, 2011) recommended that risk factors, such as geometric design and traffic volume, on vulnerable user injuries at roundabouts should be considered.

The section given above has investigated the safety impact of roundabouts for all types of road users. The previously mentioned studies stated that geometric design, speed reduction and driver behaviour emerged as significant factors which need to be studied in more detail. In addition, the safety performance of roundabouts also should focus on vulnerable road users' casualty analysis because roundabouts perform badly for this specific group. Therefore, the following sections will focus on the relationship between stated risk factors and vulnerable users.

2.4 Vulnerable Road User Safety at Roundabouts

As mentioned earlier, roundabouts are designed for vehicle safety (Gross, 2013) and the safety impact of roundabouts from the perspective of vulnerable users is unclear. Therefore, this needs more attention in the future (Silvano and Linder, 2017). Only few studies have been conducted on vulnerable user behaviour and safety at roundabouts (See Table 2.3). An early study carried out by Brown (1995) comprehensively summarised the safety for vulnerable users at roundabouts. Brown (1995) found that a queue of two to three between yield line and pedestrian crossing increased the safety for pedestrians at roundabouts. However, the study argued that roundabouts could not be considered as a safe intersection design for cyclists, a result supported by further studies (De Brabander and Vereeck, 2007; Daniels *et al.*, 2010; Daniels *et al.*, 2011; Polders *et al.*, 2014). The common message on vulnerable road users' safety at roundabouts was that roundabouts are not safe for vulnerable users and the contribution of geometric design and traffic related variables for vulnerable user safety need to be investigated in more detailed analysis. In addition to this outcome, a study was conducted by Safe Transportation Research and Education Centre (Arnold *et al.*, 2010). This study identified factors affecting pedestrians and cyclists involved in collisions at multilane roundabouts. This comprehensive research investigated the vulnerable user travel behaviour, travel demand and re-signage at multilane roundabouts. The research was carried out on five selected roundabouts in the State of California, in the United States. Vulnerable user facilities were identified within 300 metres of each arm of the roundabouts and collision data were collected in this region. Path choice and route change of cyclists and pedestrians were observed from video records. Additionally, a survey along corridors around the selected roundabouts was carried out at nine locations. The hypothesis behind the corridor survey was that the large volume of cyclist and pedestrian movements might influence travel demand and behaviour of users. The results showed that 25% of cyclists and 14% of pedestrians were changing their routes to avoid a multilane roundabout. This result was supported some years earlier by Davies *et al.* (1997) who found that cyclists avoided using roundabouts since they do not feel comfortable in terms of safety.

The study (Arnold *et al.*, 2010) also concluded that traffic considerations were less important than land use, connectivity and directness which means that safety studies also should consider other variables in addition to traffic issues. Vulnerable user number displayed an inverse relationship with traffic volume and the study concluded that levels of both should be counted in behaviour analysis. The results of a questioning survey showed that 18% of pedestrians feel uncomfortable at roundabouts whilst this rate was significantly higher for cyclists at 32%. Age, geometric design and all road users flow influenced the level of comfort for cyclists. The limitation of this study was the lack of relevant data and the study evaluation was only based on European studies, although the study was carried out in the USA. The study (Arnold *et al.*, 2010) recommended that future studies should investigate the relationship between vulnerable user comfort and socio demographic variables.

Table 2-3 – Former Studies on Vulnerable Road User Safety at Roundabouts

Author(s), Year, Title	Study details	Objective & Method of data collection and analysis	Limitations/ Recommendations for further research
De Brabander & Vereeck (2007) Safety effects of roundabouts in Flanders: signal type, speed limits and vulnerable road users	- Belgium - Vulnerable user - 95 Roundabouts - 230 Intersections - Before and after study	***Objectives:*** - Comparison between signalised intersections and roundabouts; - Determined the speed limit impact on vulnerable user safety. ***Methodology:*** - Odds ratio, expected number of crashes, effectiveness ratio, Meta-analysis, regression to the mean; - Clustered intersections according to the speed limits (50 km/h, 70 km/h, 90 km/h), traffic signals.	***Limitations:*** - No measured safety performance; - No separation of vulnerable users such as pedestrians and cyclists. ***Further Study:*** - Estimated reduction of number of crashes should be analysed.
Daniels *et al.* (2010) Externality of risk and crash severity at roundabouts	- Belgium -148 roundabouts -1491 samples -Vulnerable user	***Objectives:*** -Investigated the factors which affect severity of crashes and injuries at roundabouts; -Related these injury factors from the literature. ***Methodology:*** -Logistic Regression; -Hierarchical Binomial Logistic Regression; -Information of the construction year of roundabout (Roads and Traffic Agency Database), traffic data collection at entry, classifying traffic modes Average Daily Traffic, GIS.	***Limitations:*** -Systematic differences in the reporting were calculated but not the correlation between variables. ***Further Study:*** -Impact speeds of vehicles should be observed in relation to the location of crashes such as entry and exit lanes and other roundabout characteristics; -Collision points and impact angles should be diagrammed; -Investigate speed in the model.
Arnold *et al.* (2010) Identifying Factors that Determine Bicyclist and Pedestrian: Involved Collision Rates and Bicyclist and Pedestrian Demand at Multi-Lane Roundabouts	-U.S.A. -Vulnerable user -Roundabout -Five sample	***Objectives:*** -Identified cyclist collision factors and demand at multilane roundabouts; -Recommended design treatments of multi lane roundabouts in order to improve cyclist safety. ***Methodology:*** -In-field counts and surveys on focus groups, pedestrian and cyclist volume counting for 2 hours at peak times; -Video analysis and collision data collection; -Corridor count for 9 locations and compared; -User facilities within 1000 feet of roundabout.	***Limitations:*** -Lack of relevant data; -Based on only European multilane roundabout experiences. ***Future study:*** -The relationship between comfort and socio demographic data should be explored.
Continued to the next page			

Table 2.3 (continued)			
Daniels *et al.* (2011) Extended prediction models for crashes at roundabouts	- Belgium -148 roundabouts -1491 samples	***Objective:*** -Investigated the factors which affect severity of crash occurrence at roundabouts; -Related these factors to literature of injury factors. ***Methodology:*** -Poisson and gamma modelling; -Crash Database (Statistics Belgium); -Information of the construction year of roundabout (Roads and Traffic Agency Database), traffic data collection at entry, classifying traffic modes Average Daily Traffic, GIS.	***Limitations:*** -Underreporting the crash in police records; -Small sample; -Limited time of data collection for ADT-values; -Roundabout design might change and this might lead an inconsistency in results. ***Further study:*** -Other risk factors need to be considered (such as geometric design parameters); -Larger sample of roundabouts; -Cross-county perspective should be considered.
Polders *et. al.* (2014) Identifying crash patterns on roundabouts: an exploratory study	-Belgium -28 roundabouts -399 samples	***Objectives:*** -Roundabout safety improved by determined crash patterns such as crash types, locations and factors ***Methodology:*** -Crash records from police reports; -Creating collision diagrams.	***Further study:*** -Further study should investigate the relationship between crash type and roundabout characteristics which are speed limit, type of cycle facility, locations and entry related geometric design features.
Harkey and Carter (2006) Observational analysis of pedestrian, bicyclist and motorist behaviours at roundabouts in the United States	-U.S. -Vulnerable road user -Seven roundabouts -769 pedestrian, 690 cyclists	***Objectives:*** -Examining the interaction between motor vehicles and vulnerable users. ***Methodology:*** -Descriptive statistics; -Study area selection based on vulnerable user flow, geometric and operational conditions; -Video recording (event time, location, geometric, yielding behaviour and number of conflicts).	***Limitation:*** -Limited number of roundabouts. ***Further study:*** -The result was not consistent with the previous step of the study; therefore, it needs further investigation; -Countermeasures required to change because change in design and operations may change the results.

The literature illustrated that roundabouts are safer than signalised junctions; however, the safety of cyclists is questionable (Furtado, 2004). For instance, Jensen (2017) suggested that converting signalised junctions to roundabouts reduced the safety for cyclists. Insufficient safety performance of roundabouts on vulnerable users, in particular cyclists, led to studies focusing specifically on cyclist safety. According to research conducted by the Transport Research Laboratory (TRL) (Davies *et al.*, 1997), the number of cyclist crash occurrence is 14 times higher than vehicles and, given the increase in cycling since the year of this study, a better understanding of the safety issues for cyclists is becoming more important. Increase in the number of roundabout constructions reduces cyclist safety (Daniels *et al.*, 2009). Also, the emphasis on local government policies which promote more shift to sustainable transport, in particular cycling, places some urgency on the need for a comprehensive study of cyclist safety at roundabouts.

Former studies (Lawton *et al.*, 2003; Hels and Orozova-Bekkevold, 2007; Møller and Hels, 2008; Daniels *et al.*, 2009; Sakshaug *et al.*, 2010; Silvano *et al.*, 2015; Jensen, 2017) of cyclist safety at roundabouts investigated the impacts of contributory factors on either number of casualty occurrence or severity. Infrastructure (cycle facility and roundabout geometry) and traffic related parameters (speed, speed limit and user volume), as well as cycling and driving behaviour, were the main parameters considered in the analysis (See Table 2.4).

A more detailed study of cyclist safety was conducted by Daniels *et al.* (2008) in a before and after study of roundabout conversions from signalised intersections inside built-up areas. The effectiveness index was proposed for crash probability of cyclists at 91 randomly selected intersection conversions. An effectiveness index of 1.48 means that the probability of a cyclist crash increased by 48% after conversion. Regarding casualty severity analysis, fatal and serious injury increased 41-46% and total injury rose by 27% in all locations after roundabout construction. The lack of information concerning Average Annual Daily Traffic (AADT), number of lanes and type of cyclist facilities were the limitations of the analysis. The study (Daniels *et al.*, 2008) strongly recommended that geometric design features should be considered in further studies on cyclist casualty severity.

Table 2-4- Former Studies on Cyclist Safety at Roundabouts

Author(s), Year, Title	Study Details	Objective & Method of data collection and analysis	Limitations/ Recommendations for further research
Jansen (2017) Safe roundabouts for cyclists	-Denmark -Cyclist -255 single lane roundabouts	***Objectives:*** -Investigated how roundabout design parameters influence cyclist safety. ***Methodology:*** -Before and after study, comparison of signalised junctions with converted roundabouts; -Calculated correction for general crashes and injury trends; -Regression to the mean, safety effects; -Calculated expected crash rate after converting to roundabouts.	***Limitation:*** -Traffic volume could not be measured for most of the roundabouts.
Jensen (2013) Safety effects of converting intersections to roundabouts	-Denmark -332 roundabouts -2497 number of crashes -1328 KSI	***Objectives:*** -Investigated the safety impact of converting the signalised junctions to roundabouts on number of crash and severity. ***Methodology:*** -Before and after study, calculated the correction factors, general trends and regression to the mean, meta-analysis; -Urban/rural, county, speed limit, type of crash, number of arms, type of roundabout, central island height, cycling facility.	This study did not provide any limitation or recommendation.
Silvano *et al.* (2016) Analysis of vehicle-bicycle interactions at unsignalised crossings: A probabilistic approach and application	-Sweden -Cyclist -One roundabout	***Objectives:*** -Modelled cyclist-motor vehicle interactions at conflict points. ***Methodology:*** -Calculated probability of vehicle driver's perception of conflict location and yielding decision; -Discrete choice model; -Video recording and analysis software SAVA.	***Limitations:*** -Parameters were not directly measured; -Cyclist decision was not considered; -Interactions were considered only for one traffic direction; -Assumption of fixed intersection zones. ***Further study:*** -Interaction of both driver and cyclist; -Complete trajectory data should be included in analysis.
Continued on the following page			

Table 2.4 (continued)			
Daniels *et al.* (2009) Injury crashes with bicyclists at roundabouts: influence of some location characteristics and the design of cycle facilities	-Belgium -Cyclist -83 single lane and 7 double lane roundabouts -411 crashes at roundabouts, 649 crashes at control junctions	***Methodology:*** -Random roundabout selection; -Inside and outside built up area; -10 years data from 1991 to 2001; -Empirical Bayes, before and after study; -Regression analysis on effectiveness indicators.	***Further study:*** -Traffic conditions should be considered; -Larger samples should be used; -Different countries should be investigated for validity; -Extending knowledge about contributing factors; -Revealing possible casual mechanisms for crashes with cyclists at roundabouts should be investigated.
Moller & Hels (2008) Cyclists' perception of risk on roundabouts	-Denmark -Cyclists -1019 cyclist -Five roundabouts	***Objectives:*** -Determined the cyclist's perception of risk; -Identified factors on this perception risk; -Clarifying whether or not cyclists know the traffic rules. ***Methodology:*** -Questionnaire, age and gender data collection; -Descriptive analysis; Chi-square tests; Multiple linear regression; -Cronbach's alpha for internal consistency of questionnaire.	***Limitations:*** -No information about comparison of perceived and actual risk which has an influence on road using behaviour.; -The sampling of cyclists may also influence the results. ***Further study:*** -Larger number of roundabouts; -Barriers to cycling should be considered; -Relation between perceived risk and actual behaviour should be defined.
Sakshaug *et al.* (2010) Cyclists on roundabouts – Different design solutions	-Sweden -Cyclists -Two roundabouts	***Objectives:*** -Compared the two roundabouts which have similar traffic flow and vehicle speed with both different cycling facilities separated and mixed; -Determined the most appropriate roundabout design for cyclists and the yielding behaviour. ***Methodology:*** -Field study, video recording and automated video detection, crash analysis; -Swedish traffic conflict techniques, crash statistics, yielding recorded manually, measured actual speed for only one arms of two roundabouts, traffic flow counting manually.	***Limitations:*** -Only 2 sample of roundabouts; -The quality of video detection was not high; -Vehicle and cyclist flow collected partially. ***Further studies:*** -Comprehensive studies on behaviour should be carried out; -Reliability test is needed to prove that there is not a systematic detection error from video records.
Continued on the following page			

Table 2.4 (continued)			
Jurisich *et al.* (2011) Reducing Speed: The C-Roundabouts	-New Zealand	***Objectives:*** -Improving multilane roundabouts for cyclists in terms of capacity and safety. ***Methodology:*** -Survey on cyclists and video tape; -SIDRA modelling.	***Further studies:*** -Investigating the impact of C-Roundabouts on capacity when vehicle volume is high; -Safety for converting single lane roundabouts to C-roundabouts; -Refinement of C-roundabouts.
Silvano *et al*. (2015) When do drivers yield to cyclists at unsignalised roundabouts? Empirical evidence and behavioural analysis	-Sweden -Cyclist -One roundabout	***Objectives:*** -Determined the yielding factors of vehicle driver to cyclists; -Model 1 the yielding probability (vehicle speed); -Model II yielding probability (vehicle and cyclist speed); -Model III cyclist's proximity. ***Methodology:*** -T statistic test, logistic regression model, Discrete choice model; -Vehicle cycle interactions divided into four groups; -Zone division (conflict zone and interaction zone); -Video records and calculated vehicle and cycle trajectories; -Single lane roundabout; -SAVA video analysis program.	***Limitation:*** -Lack of number of samples. ***Further study:*** -Applicability of the results of this study should be investigated particularly outside of Northern European countries.
Hels & Orozova-Bekkevold (2007) The effect of roundabout design features on cyclist accident rate	-Denmark -Cyclist -171 crashes -88 roundabouts	***Objectives:*** -Investigated the statistical relationships between variables of roundabout geometry, roundabout age, traffic volume, cyclist volume and yearly rate of crashes; -Identified the prevalence and types of cyclist casualties; -Determined the degree of cyclist casualty missed reporting. ***Methodology:*** -Poisson regression and Logistic regression; -Cyclist crashes (4 years data).	***Limitations:*** -Small number of observations; -Higher percentage of non-reporting of cyclist casualties; -Limited data does not allow analysis of crashes risk per cyclist.
Continued on the following page			

Table 2.4 (continued)			
Daniels *et al.* (2008) Effects of Roundabouts on Traffic Safety for Bicyclists: An Observational Study	-Belgium -Cyclist -91 roundabouts	***Objectives:*** -Investigated the difference between inside and outside built up areas and the effects of converting roundabouts from signalised junctions compared to non-signalised junctions. ***Methodology:*** -Roundabout construction between 1994-2000; -Random roundabout selection; -Before and after study, regression to the mean, Meta-analysis, total crash & severity & location of crash; -Location determination of roundabouts (inside (50km/h) and outside (90-70 km/h) built up area), Speed limit.	***Limitations:*** -No information of AADT, number of lanes, type of bicyclist facility. ***Further study:*** -Geometric features should be considered.
Lawton *et al.* (2003) Cyclists at continental style roundabouts: report on four trial sites	-United Kingdom -Cyclist -Four Roundabouts -TRL Report	***Objectives:*** -Investigated the impacts of continental style roundabout on cyclist safety. ***Methodology:*** -Before and after video and interview survey; -Before and after crash statistics.	***Limitations:*** -Lack of data to analyse the impact on reducing the number of crashes.
Rasanen & Summala (2000) Car drivers' adjustments to cyclists at roundabouts	-Denmark -Cyclist -Six roundabouts located in Finland, Sweden and Denmark	***Objectives:*** -Analysing of driver's behaviour and adjustment to cyclists. ***Methodology:*** -3 hidden video records; -Vehicle approach speed, driver head movement, yielding to cyclists; -The size of central island, entry width, entry radius, central diameter, the distance of bicycle crossing, circulating width, sight distance; -Speed of the vehicle was measured by Mouse-Driven computer software; -ANOVA (vehicle speed, head movement and yielding analysis).	***Further Study:*** -Roundabouts with small central of islands (< 20m) and built up areas; -The best location and distance from circulating road for siting of cyclist crossing should be identified.
Continued on the following page			

Table 2.4 (continued)			
Davies *et al.* (1997) Cyclists at roundabouts — the effects of 'Continental' design on predicted safety and capacity	-UK -Six Roundabouts -Cyclist -TRL report	***Objectives:*** -Compared the continental European design of roundabout with UK design. ***Methodology:*** -Calculating the predicted crash index by ARCADY 3; -Vehicle flow, cyclist crash number -Classified the size of the roundabout (30-90) and modified based on continental design (before and after comparison)	***Limitation:*** -ARCADY/3 is a coarse tool which does not allow specific user group to be simulated.

Flared lanes, multilane roundabouts and higher speed decrease cyclist safety at roundabouts (AASHTO, 1999). Appropriate measures such as cycle facilities should be applied at roundabouts with these specifications and in the UK several cycling facility applications are described in the roundabout design guidelines (DfT, 2008). Ideally, a separate cycling path is recommended to improve safety because segregation of cycles is a safer alternative to cycling in mixed traffic. However, this type of cycling facility has some disadvantages such as cost and lack of land at existing roundabouts. Bypasses, underpasses or bridges are considered as other rather expensive cycling path solutions. Shared pavements are never recommended, although it is the most common solution in the UK. When a separate cycle path is not a possible option, continental geometry may be an alternative solution (CEGB, 2016). Highway authorities tried to look for a solution to reduce vehicle speed and improve cyclist safety at roundabouts and continental design geometry was developed to do this by having a radial design to decrease the vehicle approach and entry speed and increase the cyclist safety (Davies *et al.*, 1997). The Dutch style roundabout has an orbital cyclist circulating infrastructure around the roundabout that keeps the cyclist from the main circulating lanes as shown in Figure 2.4 (Yor *et al.*, 2015).

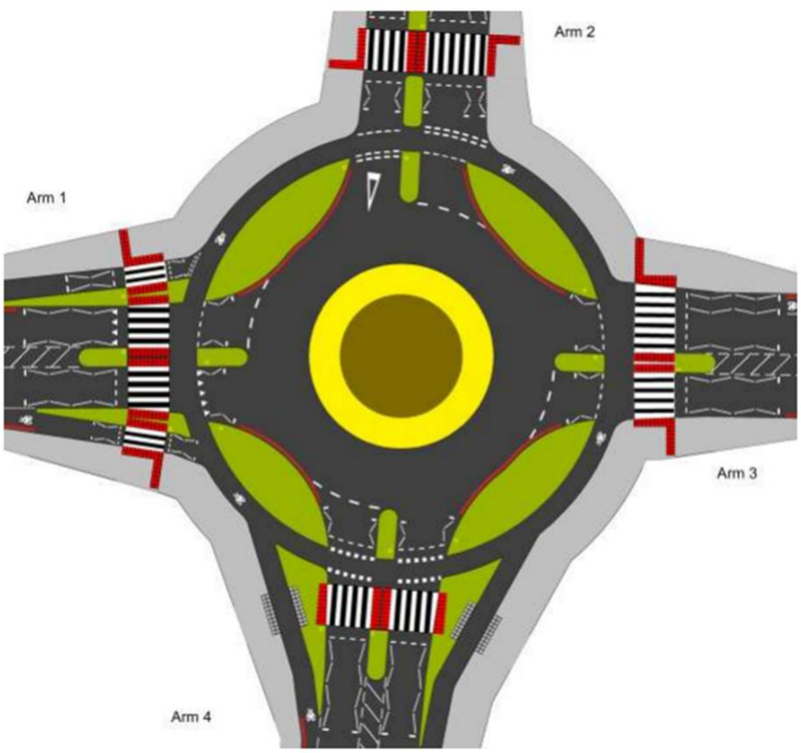


Figure 2.4 – Layout of the Dutch Style Roundabout with UK Road Markings (Yor et al., 2015)

The influence of cycling infrastructure on cyclist safety at roundabouts was investigated by several studies (Hels and Orozova-Bekkevold, 2007; Daniels *et al.*, 2009; Sakshaug *et al.*, 2010; Polders *et al.*, 2014) in the literature. The former four studies categorised results of cyclist facility impact on cyclist safety at roundabouts into categories of *no impact*, *impact* and *not clear results*. The earliest study was conducted by Hels and Orozova-Bekkevold (2007) in Denmark. Road geometry, cycling facility details and total crash data were collected from 88 roundabouts and all variables were analysed in Logistic regression and Poisson regression models. This study showed that the variables namely existence of a cyclist facility, number of legs and apron width have no significant effect on cyclist crash rates, which was not an expected result. On the other hand, number of cyclist crashes were found to increase with the increase in age. This study also mentioned an unexpected result that 75% of cyclist injuries recorded in a hospital database were not found in the police database (Hels and Orozova-Bekkevold, 2007). This shows that possibly there is a higher risk of cyclist crash at roundabouts and certainly researchers should be aware of potential bias caused by some data collection methods.

Polders *et al.* (2014) and Daniels *et al.* (2009) determined the performance of cyclist facilities at roundabouts. Both studies were in Belgium and were divided into four main groups, namely a) mixed traffic, b) cycle lanes within the roundabouts, c) separate cycling path and d) grade separate cycle paths. The studies presented similar results of the negative impact of cycling lanes within the roundabouts. Unlike Hels and Orozova-Bekkevold (2007), Daniels *et al.* (2009) used barriers, road markings and road signals as model variables and found that the Linear Regression model is the best fit model for the analysis. The results of the studies showed that cycle lanes next to the carriageway perform badly compared to mixed traffic, separated and grade separated cycling paths (Daniels *et al.*, 2009). According to the UK DfT cycling facility report (1997), mixed traffic is expected to be more dangerous for cyclists which is indeed opposite of the result from the study conducted by Daniels *et al.* (2009) and Polders *et al.* (2014). This conflict might be a result of cyclist and vehicle driver misunderstanding of yielding requirements at the roundabouts that have cyclist lanes marked on the carriageway. In other words, whether traffic is mixed or there is a separated cyclist facility next to the vehicle road, the yielding expectation between both users and likely vehicle speed varies and both users may or may not be more careful when entering and circulating at roundabouts.

The study of Sakshaug *et al.* (2010) aimed to determine the safest roundabout design for cyclists and accordingly they selected two roundabouts in Sweden for safety comparison. The results did not provide a clear answer to cyclist facility impact. The two roundabouts serviced traffic with similar vehicle speed and levels of flow; however, the roundabouts utilised different cycling facilities with separate cycling paths rather than road sharing. Roundabout geometric design variables and total crash records for all road users were collected. Variables observed in video records included driver behaviour, cyclist behaviour, distance to the crash location, who yielded and the passing behaviour of both road users. These data were recorded during five days for each roundabout. According to the comparison analysis, it was not possible to determine which roundabout was safer because studying only two roundabouts is not sufficient to give reliable solutions. Nevertheless, it seems that the roundabout with separate cycling facilities is safer than road line sharing. The conflict points are higher at an integrated roundabout in mixed traffic; therefore, it is more complex compromising roundabout safety. On the other hand, indicators (road markings) which are effective in yielding of vehicle to cyclist at integrated roundabouts makes roundabouts safer than road sharing. However, this leads to another problem in that both vehicle driver and cyclist do not expect the yielding situation at integrated roundabouts and this infrastructure is responsible for safety. The study recommended that further studies should consider the requirement of reliability test in video analysis of the cyclist-driver interaction behaviour (Sakshaug *et al.*, 2010).

Cycling facilities may not be an appropriate solution for improving cyclist safety; therefore alternatively, examining the geometric design of the roundabout should be considered (Davies *et al.*, 1997). The design of a roundabout is a complex procedure which needs to address several design variables to ensure safety and higher capacity. Safety and capacity compete, therefore a balance between these two essential targets needs to be reached. A balanced design cannot be a one size fits all approach and a prescriptive design should be applied (Furtado, 2004). This approach has created several design solutions applied in different in regions such as in Europe (with continental design) and in the UK (with tangential design). There have been two main studies (Davies *et al.*, 1997; Lawton *et al.*, 2003) to understand the impact of these two different design approaches on the balance between capacity and safety. Both studies (Davies *et al.*, 1997; Lawton *et al.*, 2003) concluded that tighter geometry (Europe continental design) at approach increases the safety; however, it also reduces vehicle capacity. On the other hand, UK design, called tangential, allows a higher capacity but with less safety, particularly for vulnerable users, due to the higher entry speed. As far as the author of this thesis is aware that there has been no study yet which illustrates the balance of safety and capacity in detail for vulnerable users.

The most common collision for cyclists at roundabouts occurs when the cyclist is circulating and a vehicle driver entering (Davies *et al.*, 1997; AASHTO, 1999). According to the report by Davies *et al.* (1997), the reason for this type of crash is that of a driver's awareness and failure to yield to cyclists. Drivers tends to focus on positioning and negotiating with other vehicles, taking less notice of the smaller dangers such as cyclists present on the roundabout. This theory is supported with lower cyclist crash rates at roundabouts when cyclist volume is low (Davies *et al.*, 1997).

A more detailed study by Møller and Hels (2008) aimed to identify the factors related to cyclist's perceived risk at roundabouts. Five roundabouts were selected in Denmark and geometric design elements, vehicle flows, cyclist volume, age and gender were used as the analysis variables. A questionnaire survey resulted in 1019 responses from cyclists whose ages were between 18 and 85. Cyclist characteristics, variables and roundabout design features were analysed with chi-square test and descriptive analysis to determine the perceived risk levels in each condition. Finally, a simple linear model was constructed by using Linear Regression. The results showed that entry and exit of the roundabout were found to be the highest crash risk locations. Age, gender, traffic volume and design features highly influenced the perceived risk for cyclists and the perceived risk was found to increase when perceived control and predictability decrease. Some of the cyclists have a very good perception of risk while they are cycling at roundabouts; however, others do not. It is predicted that the lack of traffic knowledge in specific age groups and underestimating of risk might be taken as crash contributory factors in vehicle-cycle collisions. The limitation of the study stated by Møller and Hels (2008) is that there is no comparison between perceived risk and actual risk in the analysis. Additionally, the effect of these perceptions on cycling behaviour is not considered. The study recommends further studies on elder people's knowledge of traffic rules. The cyclist sample may influence results; therefore, the study should be extended to a larger number of cyclists and for more roundabouts and the relationship between actual behaviour and perceived risk should be determined. Also, it is highly recommended that barriers and physical limitations of street furniture should be considered in cyclist risk perception (Møller and Hels, 2008). According to Møller and Hels (2008), risk perception is not at the same level for every cyclist and crash probability depends on variable factors, such as socio demographic, traffic and geometric features.

Some of the studies claim that the possible main factor related to cyclist-vehicle crashes might be driver behaviour and yielding problems (Rasanen and Summala, 2000; Silvano *et al.*, 2015). Both of these studies are based on a hidden video recording data collection method from which speed, driver behaviour and cyclist behaviour were quantified. Silvano *et al.* (2015) calculated the probability of yielding by Logistic Regression. The number of yielding events, vehicle and cyclist speed and trajectory data were collected from one roundabout located in Sweden. The authors used the common classification of roundabout segments for analysis. The results showed that cyclist speed has a slight effect on vehicle yielding behaviour while any increase in vehicle speed causes a sharp decrease in the yielding probability. If the vehicle speed is under 20 km/h, yielding rate is expected to level off. Cyclists are very confident that vehicles will give priority to them; however, this presumption reduces the safety for cyclists. Vehicle driver behaviour has a strong impact on cyclist position at the roundabout Silvano *et al.* (2015). This study provides reliable results although the number of samples is limited to only one roundabout. The traffic volume and geometric features were not considered as variables in the study. The authors recommended that the analysis should be expanded to also include different variables and more samples at different case study areas.

Rasanen and Summala (2000) aimed to establish drivers' behaviour and adjustment to cyclists at six roundabouts which were in Finland, Sweden and Denmark. Three video cameras were installed at each roundabout and drivers' approach speed, drivers' yielding to cyclist, drivers' head movement and conflict locations where a cyclist enters the vehicle path were observed from video records. The results showed that the frequency of vehicle-cyclist crashes is high when drivers are entering roundabouts and cyclists are circulating. Also, 7-15% of drivers were found not to be aware of cyclists when the cyclists were approaching from the right. The main contributory factor to crashes is that drivers are not looking properly to the right side where cyclists appear unexpectedly. Another yielding problem was found to be high approach speed. If a driver's approach speed is higher, their yielding behaviour towards cyclists decreases. Large central islands, of around 40 m diameter, have less entry path deflection and this helps drivers to reduce their speed consistently; however, smaller central islands of around 13-16 m allow drivers freedom of a direct driving path encouraging higher speed. Therefore, it is highly recommended by Rasanen and Summala (2000) that the roundabout central island dimension should be considered in cyclist safety studies. Further studies also should consider smaller roundabouts which have less than 20 m diameter and are located not only in rural areas but also built up regions. According to Rasanen and Summala (2000), the main research question outstanding is, ‘Where should the cyclist path and crossing locations be located?’

2.5 Statistical Methods Used for Analysing Casualty Severity

The previous section has concentrated on the results of former research in order to identify the research gap. In this section, analytical methods adopted in previous studies on cyclist safety at roundabouts have been reviewed. As seen on the Table 2.5, several types of statistical analysis (i.e. Pearson’s chi-square, descriptive statistics, ANOVA, meta-analysis and comparison analysis) as well as models (linear, logistic, poisson, gamma, hierarchical binomial logistic, regression to the mean and Empirical Bayes model) and methods of data collection are given. The details of each methods are given in Appendix B and more general observations are expanded upon here.

As seen in Table 2.4, the studies investigating casualty severity used crash database as a data collection method. When the aim of the study was observing the driver/rider behaviour, a yielding analysis was based on video records. Safety index or danger perception related studies carried out a questionnaire in order to obtain the data. Regarding the analytical methods, the analysis of the relationship between two dependent variables is often carried out by testing a null hypothesis such as “A higher speed increases the crash rates”. These kinds of studies need basic statistical methods such as correlation analysis and Pearson’s chi square test rather than a regression model in order to analyse the dataset and it is normally applied when the dataset is limited for fitting into a selected regression model (Harrell, 2001). Polders *et al.* (2014) applied Pearson’s chi-square in order to observe the impact of independent variables, such as roundabout segments, weather, light condition, cycling facilities, and number of lanes, on distribution of cyclist and moped crashes. This test is applicable for investigating the impact of each categorical variables individually on cyclist casualties. However, whether the analysis has one or more predictors, if fitting a model is a requirement or the aim of the study is investigating the impacts on an outcome, regression models need to be considered.

Table 2-5 – Model Prediction Table Including Methods of Data Collection and Analysis

Authors, Years & Area	**Type of Model / Test**	**Predictive Variables of Model**	**Methods of Data Collection**	**Response Variable of Model/Outcome**
Lawton *et al.* (2003) UK	-Descriptive Statistics	Vehicle flow, Cyclist flow, Number of lanes, Central island diameter, More radial arms, Toucan crossing	Cyclist interview	- The change in safety
Jurisich et al. (2011) New Zealand	-SIDRA software	Video type, Multilane roundabouts	Questionnaire on cyclists, Video records	-Safety index
Jensen (2017) Denmark	-Regression to the mean	Speed limit, County, State of municipal, Central island height, Cycle facility, Urban/rural, Central island diameter	Crash records	-Compared real and estimated number of crashes after converting intersections to roundabouts
Jensen (2013) Denmark	-Regression to the mean	Number of crashes, Casualty severity, Urban/rural, county, Speed limit, Type of crash, Number of arms, Type of roundabout, Central island height, Cycling facility	Crash records	-Safety impact comparison
Polders *et al.* (2014) Belgium	-Pearson's chi-square tests	Roundabout segments, Weather, Crash severity, Lighting condition, Crash type, Number of lanes, Cycling facility, Road user type	Crash records	-Distribution of cyclist and moped crashes
Daniels *et al.* (2011) Belgium	-Poisson regression -Gamma regression -Descriptive statistics	Road user type, Cycling facility, Traffic flow, Outside diameter	Crash records	-Probability of crash occurrence -Probability of severity -The variance
Daniels *et al.* (2010) Belgium	-Logistic regression -Hierarchical binomial logistic regression -Descriptive statistics	Road user type, Cycling facility, Traffic flow, Outside diameter, Alcohol test, Gender and age, Urban/rural, Lighting Conditions	Crash records	-Probability of severity -Hierarchical structure between variables -The variance
De Brabander & Vereeck (2007) Belgium	-Odds Ratio -Regression to the mean -Meta analysis	Number of roundabouts, Speed limit, Casualty severity, Number of intersections, Years, With signalisation and without signalisation before roundabout implementation (before and after), With and without vulnerable road users (before and after)	Crash records	-Expected number of crashes -Effectiveness ratio
Continued on the following page				

Table 2.5 (continued)				
Daniels *et al.* (2008) Belgium	-Regression to the mean, -Meta analysis	Urban/rural, Construction year of roundabout, Casualty severity, Equipped with traffic signals or not in the before situation	Crash records	-Average yearly number of crashes -Effectiveness index
Moller & Hels (2008) Denmark	-Multiple linear regression -Descriptive analysis, Chi square	Gender, Near crash, Vehicle flow, Cyclist flow, Cycling facility	Questionnaire on cyclists	-Perception of Danger
Hels & Bekkevold (2007) Denmark	-Poisson regression -Logistic Regression	Number of legs, Central Diameter, Apron width, Urban/rural, Cycling facility, Entry path radius, Year of construction of roundabout, Vehicle flow, Cyclist flow, Circular roadway width	Crash records	-The variation of cyclist crash at roundabouts by predictors -The probability of the crash
Rasanen & Summala (2000) Denmark	-ANOVA statistical analysis	Vehicle speeds, Driver head movements	Video records	-Drivers and yielding and perception of cyclists
Daniels *et al.* (2009) Belgium	-Empirical Bayes -Linear regression -Regression to the mean -Meta analysis	Number of lanes, Cycling facility, Barrier, Casualty severity, Urban/rural, Construction year of roundabout, Traffic signals and marking	Crash records	-Estimated effectiveness -Estimated relationship between the estimated value for the effectiveness per location and some known characteristics of the roundabout locations.
Sakshaug *et al.* (2010) Sweden	-Comparison analysis	Number of motorists yielding, Number of cyclists yielding, Number of conflicts points, Number of crashes, Number of who should yield and who yields, Number of moving parallel and staying behind, Number of speed change, Adjust speed, Get off the bike, Stop and stand still	Video records	-Percentage of yielding number
Silvano *et al.* (2015) Sweden	-Logistic regression model	Vehicle speed, Cyclist speed, Segment 1 (if the bike is in (0-10 m) when the car arrives at decision point), Segment 2 (11-20 m), Segment 3 (21-30 m)	Video records	-Model of yielding probability -Model of conflict probability
Continued on the following page				

Table 2.5 (continued)				
Silvano et al. (2016) Sweden	-Discrete choice model	Vehicle speed, Cyclist speed, Travel distance, Yielding behaviour	Video records	-Calculating probability of vehicle driver's perceiving at conflict location and yielding decision
Arnold *et al.* (2013) U.S.	-Comparison analysis	Cyclist flow, Vehicle flow	Questionnaire of cyclists and pedestrians, Video records	-Percentage of behaviour and demand
Harkey and Carter (2006) U.S.	-Descriptive statistic	Number of cyclists, Number of motor vehicles	Video records	-Percentage of yielding to each other
Davies at al. (1997) UK	-ARCADY software	-Size of roundabout	Simulation	-Predicted safety index

With respect to the cyclist casualty severity analysis, former research mainly focused on Logistic Regression and Empirical Bayes modelling. Logistic Regression creates a probability prediction model regarding response and observed variables (Field, 2009), while Empirical Bayes develops the model by predicting the outcomes by comparing the observed data to prior knowledge in the literature (Efron, 2013). Scientists prefer to conduct Logistic Regression to investigate the influence of external impacts (in previous studies geometric design and sociodemographic characteristics) on casualty severity which is either a binary (slight/serious) or a categorical response (slight/serious/fatal). On the other hand, Empirical Bayes may be preferred in order to create a prediction model which considers variation achieved by Monte Carlo Simulation.

This thesis aimed to develop a model to investigate the impact of variables (including geometric design parameters, sociodemographic characteristic of cyclist, meteorological conditions, speed limit, traffic flow profile and driver/rider behaviour related contributory factors) on cyclist casualty severity at roundabouts. Cyclist casualty severity was used as the indicator because a study, which investigates the impact of roundabout geometry, environmental and human characteristics on casualty numbers, was not feasible. This type of study should include a measure of road user count (i.e. cyclist and vehicle flow) at each roundabout where casualty occurs. However, this is unavailable as a limitation. Therefore, comparative studies after converting signalised junctions to roundabouts can use number of casualties as a measure but investigating the impacts on casualties often use severity ratio. In addition, as mentioned earlier, former studies have already showed that converting signalised junctions to roundabouts increased the number of cyclist casualties and it was suggested that roundabouts were not safe for cyclists. This thesis aimed to investigate the influences on cyclist casualty severity at roundabouts in order to gain a further understanding.

Regarding this aim, Empirical Bayes was not an option since the aim was not to develop a prediction model based on prior data. Therefore, it was decided to apply Logistic Regression. However, Logistic Regression has assumptions and some limitations; therefore, a fundamental understanding of the data structure and how to address these assumption and limitations were a priority in developing the methods of analysis. The next sub-section focuses on understanding the principles, assumptions and limitations of a Logistic Regression.

Understanding the Principles, Assumptions and Limitations of Logistic Regression

The main challenge in data analysis is choosing and using the correct regression method to fit data and meet the aims of study. The researcher should ask the question, "Should this model be used in the study?" Harrell (2001) stated the method of the model choice in bullet points which are given below:

- *analyses the data efficiently*
- *fits the whole structure of study aim*
- *arises the problems in dataset*
- *is appropriate for further developing*
- *can be extended*

As mentioned earlier, it was aimed to apply Logistic Regression analysis since this analytical approach met with the required analysis regarding the structure of the data. Initially, it was focused on understanding the question 'What is a Logistic Regression model?' Logistic Regression has categorical variable of outcome with predictor variables which are either continuous or categorical or both. In Logistic Regression, the predicted outcome is the probability of Y occurring given the predictors of X_1, $X_2 \ldots X_{i.}$ Since the probability of an event should be between 0 and 1, the predicted outcome Y should be in this interval. If the outcome value Y is close to 0 (Probability~0%) it means that Y is *unlikely* to occur, meanwhile an outcome close to 1 (Probability~100%) means that Y is *likely* to occur (Field, 2009).

Similar to Linear Regression, in Logistic Regression the response is predicted by a linear combination of predictors. In Linear Regression, the coefficients are sufficient to explain the model; however, the coefficient of Logistic Regression cannot be explained by itself. Therefore, the odds ratio is usually used when interpreting the results. In other words, the impact of the predictor variables is usually explained in terms of the odds ratios. While coefficient estimates generate the linear equation in the regression, the log of odds of the outcome provide the equation of predictors in the Logistic Regression. Odds is the ratio of the probability of occurrence to the probability of non-occurrence "odds = $p/(1-p)$" where $p_n = p_1$ ($p_n = p_2$) which is the probability of success failure (Agresti, 2007). The odds ratio is the ratio of odds of success to odds of failure, (Agresti, 2007).

Odds Ratio $= Odds_1 / Odds_2 = (p_1/(1-p_1))/(p_2/(1-p_2))$

There are several types of Logistic Regression regarding response and predictors: i) Binary Logistic Regression (two response 0 and 1), ii) Ordinal Logistic Regression (with minimum three responses) and iii) Nominal Logistic Regression (with multilevel response without ordering) (Field, 2009). In this thesis, the response variable was slight and serious (coded as binary 0 and 1). Therefore, Binary Logistic Regression was applied into the modelling section of the analysis. Binary Logistic Regression investigates the change of dichotomous response (binary coded values 0 or 1) based on either continuous or categorical predictor variables. This type of regression is commonly applied when the dichotomous response variable is 'yes or no', 'yielded or not yielded' or 'slight or serious'.

The logit function of the binary outcome variable is given below.

$$logit(p) = \log(Odds) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

Where:

$X_1, X_2 \ldots X_n$:	Predictive variables
β_0:	Coefficients of the unknowns
$\beta_1, \beta_2 \ldots \beta_n$:	Coefficients of the predictive variables

With respect to the predictors, there are two main types of Logistic Regression in modelling: i) Simple Logistic Regression (SLR) and ii) Multiple Logistic Regression (MLR). SLR is a form of Logistic Regression with one response and one predictor. If the number of predictor variables is more than one, it is called MLR. Regression studies in road safety are carried out by applying a set of data into a model. When the number of variables should be reduced based on a selection method, recommended relaxing p-value criteria by Sperandei (2013) should be applied. This criterion is applying both SLR and MLR and selecting the statistically significant variables at 90% confidence level. The selected variables should be included in a final MLR.

In some cases, the data is nested in groups and the response variables (casualty severity) nested in the same groups are more likely to function in the same way than response variables nested a different group (Sommet and Morselli, 2017) (See Figure 2.5). For instance, the impact of weather on casualty severity may have a statistically significantly different level on casualty severity in different cities. In this situation, the nested cluster impact in the model occurs and Multilevel Logistic Regression should be applied (Sommet and Morselli, 2017).

The aim in Multilevel Logistic Regression is to estimate the effect of covariates at a regional level (Li *et al.*, 2011). Previous studies did not consider this type of regression model in their analysis probably because they did not conduct a study which has a large amount of nested data in groups.

The equation of the Multilevel Logistic Regression is given as follow (Steele, 2009):

$$log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 X_{ij} + u_j$$

$$u_j \sim N(0, \sigma_u^2)$$

Where:

β_0 = the log-odds that y = 1 when x = 0 and u = 0

β_1 = the effect on log-odds of one unit increase in x for individuals in same group

u_j = is the effect of being in group j on the log-odds that y = 1

σ_u^2 = is the level 2 (region) variance

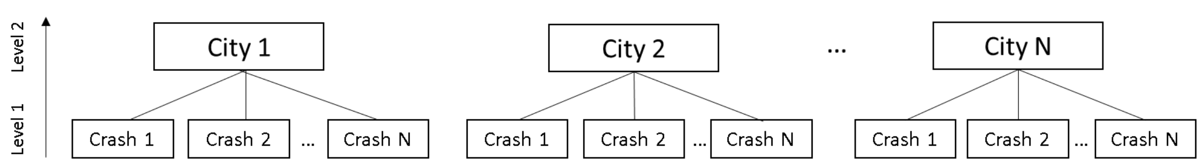


Figure 2.5 – Nested Grouped Data for Multilevel Logistic Regression

Adding variables in regression models increases the accuracy of the results. However, adding more variables than the maximum limit appropriate for the model causes inefficiency and over-fitting in the results. On the other hand, adding fewer variables than the minimum limit of the model would result in biased outcomes. According to Occam's razor approach, the model should be simple but not too simple; therefore, under and over fitting in regression models should be avoided. Therefore, in this respect, each model has its own optimum level for a given number of variables. According to Peduzzi *et al.* (1996), there are three types of error that occur in Logistic Regression: overfitting, underfitting and paradoxical fitting. Therefore, the number of predictors in the regression needs to be determined carefully. Peduzzi *et al.* (1996) recommended that the minimum of 10 events per variable (EPV) approach is an acceptable number in binary Logistic Regression studies. Agresti (2007) identified that adding too many variables leads to *poor standard errors*. The research reported in the thesis adopted the advice of (Peduzzi *et al.*, 1996) and used the minimum of 10 EPV. When the number of observations is not higher than the recommended limit (Peduzzi *et al.*, 1996), the variables can be applied individually in the regression model (Hels and Orozova-Bekkevold, 2007). However, the minimum limit of the EPV is a rule of thumb and Ogundimu *et al.* (2016) stated that "EPV≥20 generally eliminated the bias when low-prevalence of predictors in a regression model". Therefore, it is highly advised to consider the EPV in regression analysis.

Field (2009) stated that there are three main assumptions in Logistic Regression: i) Linearity, ii) Independence of errors and iii) Multicollinearity. The first assumption, linearity, is a general approach of all types of regression analysis. In Linear Regression, it is assumed that there is a linear relationship between outcome and response variables; however, as mentioned earlier, this assumption is violated for Logistic Regression. This is the main reason why logit function should be applied in Logistic Regression. Ultimately, in LR it is assumed that there is a linear relationship between continuous predictors and logit of the outcome variable. This assumption is not valid if the predictor variable is categorical. The significance of the interaction term between the predictor variable and its log transformation can test this assumption (Hosmer and Lemeshow, 2000). The second assumption, independence of errors, also is valid for all types of regression analysis. This assumption says that the samples of data should be not related; in other words, the same item cannot be measured at different points in time. If the data does not conform to this assumption, over dispersion is likely to occur. The third assumption, multicollinearity, whilst its absence is essential for Linear Regression, is not a must for Logistic Regression. In Linear Regression, it is assumed that predictors should not be correlated with each other; in other words, they should be independent. In Logistic Regression, there is no exact limitation for multicollinearity between predictors; however, if present the results should be interpreted carefully. Therefore, an in-depth understanding of the data by conducting a descriptive statistic, test of normality and correlation analysis should be the initial step of the regression modelling.

Descriptive Statistics and Test of Normality

Data can be classified into two main groups either continuous or categorical data (Field, 2009). Continuous data is either numeric (measured on a scale, such as size of the central islands of roundabouts) or discrete (a set number of values based on counts, such as number of crashes occurred at roundabouts). There are two types of categorical data: nominal (non-numeric, such as gender) and ordinal (data based on size, such as speed limit).

Statistical analysis is inferring information from data and the initial analysis starts with descriptive statistics. In other words, it quantitatively summarises the data based on statistical tests and graphs. It measures central tendency by mean (arithmetic means of the values), median (central value) and mode (most frequently occurring value) and measures of variability by standard deviation (amount of difference from the mean value) and minimum/maximum values. Descriptive statistics is followed by determining the test of normality. A Frequency distribution (histogram), shows how many times each value occurred in data on a horizontal axis (Field, 2009). The symmetrical bell-shaped curve is called a normal or Gaussian distribution. On the other hand, lack of symmetry (skew) and pointiness (kurtosis) causes a deviation from normal, which is referred to as non-normal distribution (Field, 2009). When the data is normally distributed, parametric statistical tests are applied; however, for non-normally distributed data, non-parametric statistical methods are employed. A first step in probability testing is to check whether or not the data is normally distributed.

Following the descriptive analysis, a correlation should be applied in order to determine any relationship between variables. In other words, correlation between sets of data is a measure of how well they are related. The correlation coefficient is a measure of the strength of the relationship between two random variables and ranges from -1 to +1. For instance, if the correlation coefficient is 0, there is no correlation between these variables and 1 with perfect correlation. The negative sign of the coefficient shows that higher value of one variable is associated with the smaller value of another variable, while when this relationship is the reverse the coefficient has a positive sign (Dalgaard, 2008).

Field (2009) states that there are two groups of correlations which are bivariate and partial. Bivariate correlation investigates the correlation between two variables, while partial correlation correlates two variables when one or more additional variables are included as controlling. Bivariate correlation has different types of analysis approach based on the data structure. These types of analysis are Pearson's, Spearman's, Kendall's Tau, Biserial and Point Biserial correlations, Phi and Cramer's V (Field, 2009).

The most common correlation analysis is the Pearson's correlation, with the full name Pearson Product Moment Correlation (PPMC), sometimes referred to as linear correlation (Dalgaard, 2008). Pearson's correlation investigates the linear relationship between two continuous variables (Field, 2009). The basic assumption in Pearson's correlation is that sample data should be normally distributed; however, this assumption can be neglected in only one case when one of the variables is categorical with two categories.

Spearman's correlation coefficient is calculated in non-parametric statistics when the data is not normally distributed (Field, 2009). Both biserial and point-biserial correlations are used when one of the variables is dichotomous (categorical with only two categories) and the other one is continuous. The difference between these two types depends on the type of dichotomous variable, whether discrete or continuous. If the variable is a continuous dichotomy, biserial correlation is used; while point biserial correlation is used when the dichotomy is discrete (Field, 2009). In some cases, both variables are dichotomous and in this situation Phi correlation needs to be applied. Moreover, when both variables are categorical with more than 2X2, Cramer's V is the most suitable correlation analysis approach. Clearly deciding the most appropriate correlation technique mainly depends on the structure of the data to which particular attention needs to be paid.

Suhr (2005) stated that a strong correlation between variables leads to errors in a regression model and recommended that three criteria should be followed in order to obtain more reliable results. These criteria are given below:

- *Some selected variables should not be included in the model*
- *Composite scores should be created based on measured variables in order to explain less variance*
- *Dimension reduction should be carried out to reduce the number of variables that explain more variance.*

Suhr (2005) argued that the best way to obtain better regression is through dimension reduction with fewer variables to explain more variance. The dimension reduction analysis delivers the requirement to reduce the number of variables. The following subsection addresses the details of the dimension reduction process.

Dimension Reduction

Dimension reduction is a statistical technique for simplifying complex sets of data (Kline, 1994). There are several methods that can be applied to achieve dimension reduction including principle components, unweighted least squares, generalised least squares, maximum likelihood, and factoring techniques including principal, alpha and image. The most commonly used, the oldest and best-known method, is PCA, Principal Component Analysis (Jolliffe, 2002). However, PCA cannot be applied to all types of data; therefore, exploratory factor analysis, the second commonly used method for dimension reduction, needs to be used in some cases (Suhr, 2005). Suhr (2005) stated that if the variables cannot be measured directly, or there is unreliability because of measurement error, or there is an influence response on measured variables, Exploratory Factor Analysis (EFA) needs to be used to reduce the number of variables.

In PCA, variables are grouped into factors with a loading level that gives the statistical significance of the variables in that factor. The loadings of variables in factors are given in a rotation matrix. The aim of the rotation is to obtain the least number of factors whilst increasing the weights of the variables. Rotations of the analysis are divided into two groups: orthogonal rotation (90^0 rotated) and oblique rotation (not rotated through 90^0) (Rummel, 1988). The oblique rotation is based on coordinates, which are the primary axes and reference axes. When the factors are highly correlated with each other, oblique rotation is used; whereas if there is no statistically significant correlation between factors, orthogonal rotation should be preferred. In selecting a low level of correlation between factors, coefficients should lie between -0.32 and +0.32 (Tabachnick *et al.*, 2014).

Reliability is an essential element in the interpretation of a measured variable in dimension reduction (Tavakol and Dennick, 2011). Reliability tests can be carried out by using Cronbach's Alpha (α) value which gives the internal consistency between variables (Yurdugül, 2008). There are two main requirements in order to observe a statistically acceptable value of Cronbach's Alpha. The alpha value should be equal to or more than 0.70 (Tavakol and Dennick, 2011). Furthermore, Yurdugül (2008) reported that the minimum number of samples for measuring the alpha could be determined based on the eigenvalue obtained from PCA. Yurdugül (2008) suggested that the required minimum number of samples should be 100, when the first eigenvalue was between three and six. If the first eigenvalue was higher than six, n=30 was sufficient to obtain a robust estimated coefficient alpha.

2.6 Identifying the Research Gaps

As mentioned earlier, the concern regarding cyclist safety at roundabouts was very high and the outcomes of the previous research on cyclist safety was unclear. Given that the number of cyclists on the roads have been increasing in recent years, safety for cyclists at roundabouts has become an important issue. Therefore, it is increasingly important to investigate the number and severity of cyclist casualties with attention to relevant contributory factors. Table 2.6 and 2.7 present a critique of previous studies to identify the research gap for cyclist safety at roundabouts, traffic behaviour and geometric design parameters. Unsigned boxes indicate that there was no evidence in the study that those aspects were covered in the data collection activities.

The initial and significant point in Table 2.6 was that studies did not cover all potential factors in the same study since either they had a limitation on data collection, or they only focused on either the impact of infrastructure or social or behaviour individually. None of the studies carried out to date used all relevant variables from infrastructure, socio-demographic and behaviour to generate fundamental understanding about the cyclist crashes. The 'Traffic Related' column in Table 2.6 represented vehicle speed, speed limit, vehicle flow and vulnerable user count. The majority of the studies considered that the vehicle flow as a variable for the analysis did not consider speed. On the other hand, only two studies measured vehicle speed and different five studies used speed limit, but none of these considered a measure or proxy for flow. However, as explained earlier in the literature review, there is a high correlation between cyclist casualty risk and vehicle speed.

The studies (Rasanen and Summala, 2000; Harkey and Carter, 2006; Møller and Hels, 2008; Sakshaug *et al.*, 2010; Silvano *et al.*, 2015; Silvano and Linder, 2017) addressed the impact of behaviour and focused on the probability of vehicles' yielding to cyclists. A study, which investigated the influence of driver/rider behaviour on cyclist casualty severity, appears absent from the literature, possibly because this type of data is not commonly collected and not straightforward to collect. The state of at review suggested that converting signalised junctions to roundabouts increases the casualty severity for cyclists (Daniels *et al.*, 2011). Some studies (De Brabander and Vereeck, 2007; Daniels *et al.*, 2008; Daniels *et al.*, 2009; Daniels *et al.*, 2010; Daniels *et al.*, 2011; Polders *et al.*, 2014; Jensen, 2013; Jensen, 2017) focused on cyclist casualty severity. However, the majority of the research, (De Brabander and Vereeck, 2007; Daniels *et al.*, 2008; Daniels *et al.*, 2009; Jensen, 2013; Jensen, 2017), examined the safety impact of converting signalised junctions to roundabouts regarding the cyclist casualty severity.

Table 2-6 – Research Gap Table for Cyclist Safety at Roundabout

Studies	Traffic				Data Collection				Infrastructure			Other				
	Speed	**Speed Limit**	**Vehicle Flow**	**Vulnerable User Flow**	**Questionnaire**	**Video Record**	**Total Crashes**	**Casualty Severity**	**Road Geometry**	**Sight Distance**	**Cycle facility**	**Behaviour or Yielding**	**Sociodemographic**	**Meteorological**	**Contributory Factors**	**Travel Distance**
Davies *et al.* (1997)			+				+		+							
Lawton *et al.* (2003)			+	+	+	+	+		+		+			+		
Jurisich *et al.* 2011)					+	+			+							
Brabander and Vereeck (2007)		+					+	+								
Daniels *et al.* (2008)		+					+	+								
Daniels *et al.* (2009)		+					+	+	+		+					
Daniels *et al.* (2010)			+	+			+	+	+		+		+	+		
Daniels *et al.* (2011)			+	+			+	+	+		+		+	+		
Polders *et al.* (2014)							+	+	+		+			+		
Hels and Bekkevold (2007)			+	+			+		+				+			
Moller and Hels (2008)			+	+	+						+	+	+			
Jensen (2017)		+					+	+	+		+					
Jensen (2013)		+					+	+	+		+					
Rasanen and Summala (2000)	+					+			+	+		+				+
Arnold *et al.* (2013)			+	+	+	+	+		+				+			
Sakshaug *et al.* (2010)			+			+	+					+				+
Silvano (2015)	+					+						+				
Silvano (2016)	+					+						+				
Harkey and Carter (2006)						+						+				
The study in this thesis		+						+	+				+	+	+	

Polders *et al.* (2014) investigated the impact of few variables on distribution of casualty severity between road users. In addition, Daniels *et al.,* (2010) and Daniels *et al.,* (2011) aimed to determine the high-risk users regarding the casualty severity. However, former research did not consider a comprehensive study on the impacts of variables on cyclist casualty severity at roundabouts.

As regards infrastructure represented by roundabout geometric design parameters, sight distance, barriers and cyclist facilities the literature is patchy. Safety barriers were not considered in any of the studies reviewed; in addition, the results on the impact of cycle facilities remain questionable. However, it can be clearly seen that road geometry has been considered in cyclist safety analysis because of the importance of geometric design parameters in safety analysis.

The particular attention paid to geometric design parameters led to further examination of previous studies; detail regarding design has been included in Table 2.6. A striking observation emerging is that so few studies are concerned with the effects of geometric design parameters on cyclist casualties along with all the other variables deemed to be important in previous studies. Former research, mainly carried out in Continental European countries such as Belgium and Denmark, conducted so far has placed emphasis on cyclist safety at roundabouts with due attention to radial geometric design type. Only Davies *et al.* (1997), Lawton *et al.* (2003) and Jurisich *et al.* (2011) considered the British approach with the fundamental tangential design structure. However, as illustrated by Table 2.6, none of this research was carried out to investigate the impacts on cyclist casualty severity. Therefore, a detailed study on the tangential design type of roundabout related to cyclist casualty severity is required.

Furthermore, regarding the geometric design, the tangential design related studies have only focused on two geometric design parameters, namely external diameter (Davies *et al.,* 1997) and number of lanes on approach arm (Lawton *et al.,* 2003; Jurisich *et al.,* 2011). However, former studies in Europe have considered some other geometric design parameters, but most previous studies recommended furthermore detailed studies of the impact of design, having provided evidence that roundabout geometry does influence risk to cyclist safety.

Table 2.7 illustrated that there has been significant attention to number of lanes, central island radius and number of arms. However, some critical design parameters, such as entry path radius (which helps to reduce the speed), type of roundabout and number of circulating lanes, were considered only in a few studies. Moreover, highly important variables related to capacity (i.e.

approach half width and number of flare lanes) have not been considered at all. The only study, which considered a wide range of parameters, was carried out by Hels and Bekkevold (2007), but this study was not a casualty severity investigation.

Finally, a critical review of the statistical applications used in previous research investigating cyclist safety at roundabouts was carried out to standardise the methodological approach designed for this study and a summary of the finding is presented in Table 2.8. Many of the studies considered descriptive statistics and regression to the mean because they focus on investigating the safety impact for *before* and *after* converting studies. Only few studies set out to develop a regression model.

Table 2-7 – Research Gap Considering Geometric Design Parameters

Research	Central Island Radius	Central Island Height	Roundabout Segments	Number of Lanes	Number of Arms	Apron Present	Apron Width	Entry Path Radius	Approach Half Width	Number of Circulating Lanes	Number of Flare Lanes on Approach	Type of Roundabout
Davies *et al.* (1997) (UK)	+											
Lawton *et al.* (2003) (UK)				+								
Jurisich *et al.* (2011) (New Zealand)				+								
Daniels *et al.* (2009) (Belgium)				+								
Daniels *et al.* (2010) (Belgium)	+			+	+	+						
Daniels *et al.* (2011) (Belgium)	+			+	+	+						
Polders *et al.* (2014) (Belgium)			+	+								
Hels and Bekkevold (2007) (Denmark)	+				+		+	+		+		
Jensen (2017) (Denmark)	+	+										
Jensen (2013) (Denmark)		+			+							+
Rasanen and Summala (2000) (Denmark)	+											
Arnold *et al.* (2013) (U.S.)				+								
The study in this thesis				+	+			+	+	+	+	+

Model selection is carried out based on the aim of the study and data structure and Logistic Regression is most suitable for modelling the impacts of variables for determining cyclist casualty severity reduction. However, only three studies (Hels and Bekkevold, 2007; Daniels et al., 2010; Silvano et al., 2015) applied Logistic Regression, but none of them related with investigation on casualty severity reduction. In addition, there has been no study which conducted margin calculation or Linear Regression in order to interpret the results of Logistic Regression analysis. The studies, which considered Linear Regression, only aimed to find out the safety impact on casualty occurrence.

Considering the created tables of the summary of research gap including cyclist safety and analytical applications, several research gaps were determined:

- *Cyclist casualty severity reaction analysis, with logistic regression including comprehensive set of predictive variables, was not applied.*
- *Geometric design parameters were not fully considered. Some critical variables, such as speed, speed limit, sociodemographic characteristics of cyclist/driver and meteorological conditions, were included but only partially and in only a few analyses and these studies were not considering casualty severity reduction.*
- *The influence of behaviour was informed by yielding and perception related research. The impact of driver/rider behaviour on casualty severity was not considered.*
- *Additional statistical applications, such as descriptive statistics, test of normality, correlation analysis, dimension reduction and reliability analysis, were applied generally and not conducted to develop a reliable empirical model or determine internal relationships between variables.*
- *The interpretation of logistic regressions was very narrow and detailed analysis, such as calculating the margins, was not considered.*
- *The results of previous studies pointed out the impacts on casualty severity. However, advice for policy makers and design engineers was very shallow. A reverse modelling approach (both logistic regression and linear regression) has not been attempted in previous studies.*

Table 2-8- Considered Statistical Applications in Former Studies

Study	**Descriptive statistic**	**Regression to the mean**	**Pearson's chi square test**	**ANOVA statistic**	**Poisson regression**	**Gamma regression**	**Logistic regression**	**Linear regression**	**Comparison analysis**	**Meta-analysis**	**Multilevel Models**	**Dimension Reduction**	**Other (Simulation/ discrete choice/ Empirical Bayes)**
Davies *et al.* (1997) (UK)	+												+
Lawton *et al.* (2003) (UK)	+												
Jurisich *et al.* (2011) (New Zealand)													+
De Brabander and Vereeck (2007) (Belgium)		+								+			
Daniels *et al.* (2008) (Belgium)		+								+			
Daniels *et al.* (2009) (Belgium)		+						+		+			+
Daniels *et al.* (2010) (Belgium)	+						+						
Daniels *et al.* (2011) (Belgium)	+				+	+							
Polders *et al.* (2014) (Belgium)			+										
Hels and Bekkevold (2007) (Denmark)					+		+						
Moller and Hels (2008) (Denmark)	+		+					+					
Jensen (2017) (Denmark)		+											
Jensen (2013) (Denmark)		+											
Rasanen and Summala (2000) (Denmark)				+									
Sakshaug *et al.* (2010) (Sweden)									+				
Silvano et al. (2016) (Sweden)													+
Silvano *et al.* (2015) Sweden							+						
Arnold *et al.* (2013) U.S.									+				
Harkey and Carter (2006) U.S.	+												
The study in this thesis	+		+				+	+			+	+	

2.7 Chapter Conclusions

This chapter began by comprehensively reviewing a wide range of research on safety at roundabouts including vulnerable road users and cyclists in particular. From the general perspective of safety, converting priority or signalised junctions to roundabouts to improve safety by reducing the number of crashes for vehicles was considered. In addition, roundabouts increase the capacity of the intersection giving environmental benefits such as lower emissions due to less queuing. Therefore, converting junctions to roundabouts is a great benefit for vehicle safety and the environment. This has led to the wide range of applications of roundabouts with numbers increasing every day.

However, the performance of safety of roundabouts is not the same for all road users. Former studies present evidence that roundabouts do not improve safety for vulnerable users. Converting priority and signalised junctions to roundabouts has been shown to increase the number of crashes and casualty severity for vulnerable users and particularly cyclists. Therefore, former studies suggest that roundabouts are dangerous for cyclists compared to vehicle drivers. The main factors governing higher numbers of crash and severity for cyclists were higher speed and speed related geometric design parameters of a roundabout. An obvious awareness problem between driver and cyclist existed. Previous studies did not consider a wide range of variables, such as geometry, traffic, sociodemographic, environmental and behaviour related contributory factors, in one model mainly due to data availability. More specifically, the impact of these variables at tangential design style roundabouts is still not clear. A complete study of cyclist casualty severity reduction at roundabouts has not been carried out. With respect to the literature review, several research gaps were identified, namely cyclist casualty severity analysis, influence of geometric design parameters and driver/rider behaviour on casualty severity and investigating the consistency of casualty modelling including different countries. It is important to address these gaps because cycling is increasing every year in response to local government policy. Gaining a much deeper understanding of the impact of a wide range of variables on cyclist casualty severity at roundabouts is what the research presented in this thesis aims to achieve.

After carrying out a critical review of literature of statistical methods applied by previous research, it was evident that this study ensured compliance with several analytical methods. The structure of the data, assumptions and limitations of selected model needs careful investigation. The aim was not only to fill the gap in cyclist safety related studies but also to conduct a novel analytical approach to casualty severity analysis. It was observed that a Logistic Regression model, which contained several variables, was not conducted for determining the impact of each variable on cyclist casualty severity at roundabouts. In addition, statistical methods (i.e. descriptive statistics, test of normality, correlation analysis methods, dimension reduction and reliability) were not carried out to observe relationships between variables and their individual impacts on casualty severity. It was observed that the interpretation of safety issues for cyclists was very shallow with regard to advice to policy makers and design engineers and therefore, based on the critical review of literature, a methodological framework was developed to address this research.

Chapter 3 Research Methodology

3.1 Introduction

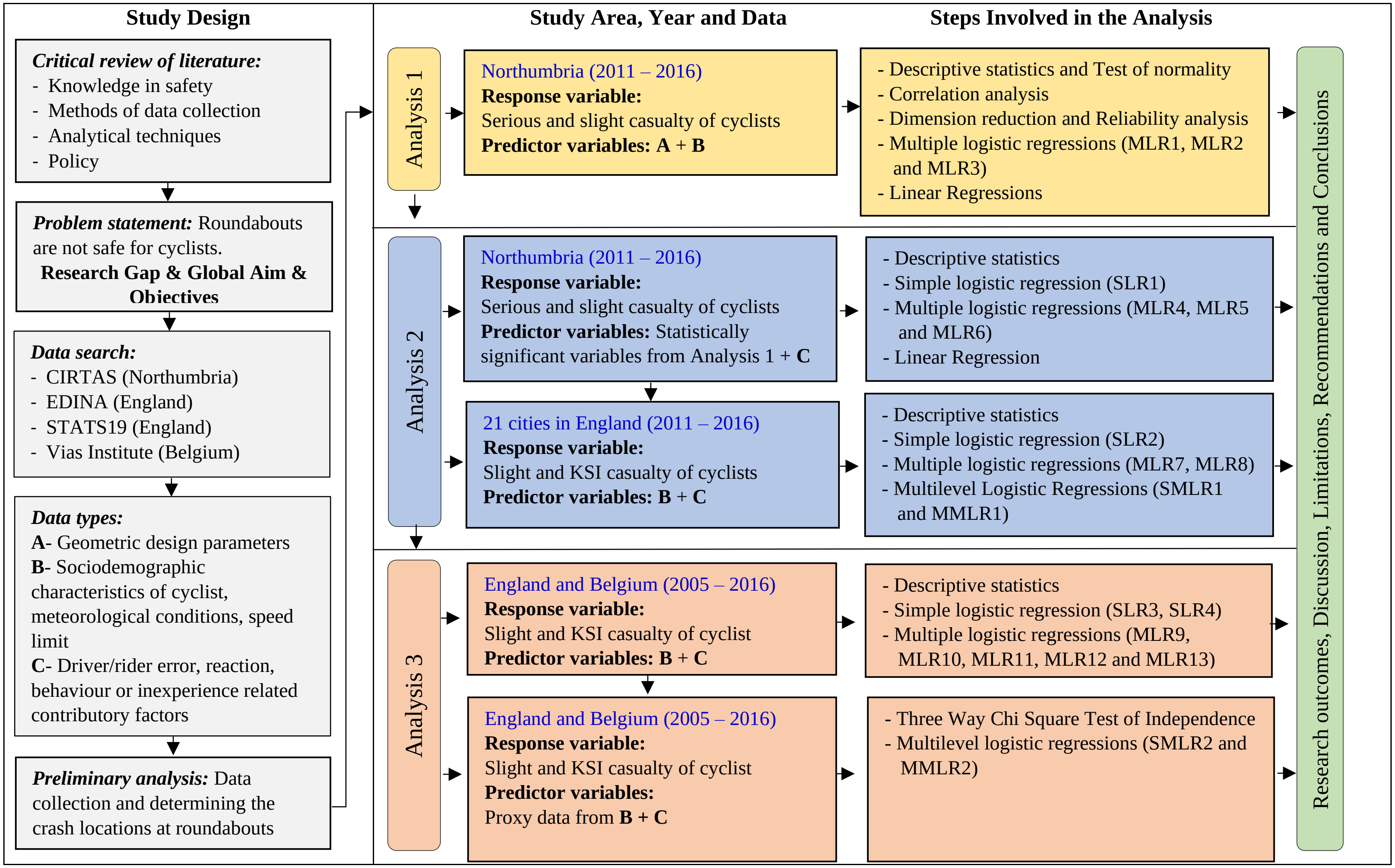
With attention to the research gaps that were identified in Chapter 2, the methodology was designed to ascertain to what extent do geometric design parameters, traffic characteristic, sociodemographic characteristics of cyclists, meteorological conditions and driver/rider behaviour related contributory factors influence cyclist casualty severity.

The methodological framework for the study was developed based on a critique of methods adopted in previous research. However, the combined approach of several different statistical methods required to meet the specific objectives is unique. First in Section 3.2 the total methodological framework will be elaborated upon. This is followed by a description of each three steps in study design in Section 3.3. Section 3.4, Section 3.5 and Section 3.6 respectively describe in detail Analysis 1, Analysis 2 and Analysis 3. Finally, the chapter is concluded in Section 3.7.

3.2 Methodological Framework

The methodological framework presented in Figure 3.1 outlines how the objectives of the study were achieved. These are elaborated upon in more detail below.

Study design started with a detailed critical review of the literature in order to understand the safety issues at roundabouts and to investigate the issues relating to cyclist safety at roundabouts to identify the research gap and formulate the global aim and objectives. A data search was carried out in order to establish the data types available and those needed to be collected to address the objectives. Before starting the main analysis steps, a preliminary analysis, such as determining the crash locations and data collection, was conducted to set the data to be ready for developing the models.



Notes:
SLR: Simple Logistic Regression
SMLR: Simple Multilevel Logistic Regression
MLR: Multiple Logistic Regression
MMLR: Multiple Multilevel Logistic Regression

Figure 3.1- Methodological Framework of This Study

The global aim and objectives of this study was achieved in three main steps of analysis (Analysis 1, 2, and 3) with reference to Figure 3.1.

- ***Analysis 1*** *was conducted to investigate the impact of geometric design parameters, as well as sociodemographic characteristics of cyclists, meteorological conditions, speed limit and traffic flow profile, on cyclist casualty severity at roundabouts, considering Northumbria as a case study.*
- ***Analysis 2*** *was carried out to explore the impact of driver/rider error, reaction, behaviour and inexperience related contributory factors on cyclist casualty severity. The analysis was conducted in two sections: a case study in Northumbria and an extended study in England.*
- ***Analysis 3*** *was designed to compare cyclist casualty severity England and Belgium in order to show the difference between two different design and cycling environments. This comparison was carried out to develop a richer understanding by introducing an international dimension.*

After carrying out statistical methods in Analysis 1, 2, and 3, the results were discussed, and limitations and recommendations were articulated based on the research outcomes. The details of each step are presented in the following sections.

3.3 Study Design

The study design consisted of the initial preparation process, critical review of existing literature, data search and making arrangements with relevant authorities (local authorities, Department for Transport for the UK, Vias Institute in Belgium) to receive the data. A critical review of literature began with a general approach to the safety at roundabouts for all road users. After identifying that roundabouts are not designed specifically for the safety of cyclists, previous studies of cyclist safety at roundabouts were critically reviewed as well as the strategic decision making process and policies (white papers and the reports published by local, regional and national authorities and institutes) related to the stated problem. Finally, the research gap, global aim and objectives for this study were determined.

The next step was understanding the data availability and collection needed to address the research gap and objectives. The data used in this study were categorised into three groups (**A**, **B** and **C**) (See Table 3.1). Data availability, collection and preliminary analysis for each step was provided in the following subsections.

Table 3-1 – Variables Considered in This Study

A **Geometric Design Parameters**	- Number of lanes on approach - Half width on approach - Entry path radius - Number of arms - Number of flare lanes on approach - Type of roundabout - Number of circulating lanes
B **Other Variables**	- Casualty gender - Casualty age - Lighting - Weather - Road surface condition - Speed limit - Traffic flow profile
C **Driver/Rider Error, Reaction, Behaviour and Inexperience**	- Junction overshoot - Junction restart (moving off at junction) - Poor turn or manoeuvre - Failed to signal or misleading signal - Failed to look properly - Failed to judge other person's path or speed - Passing too close to cyclists - Sudden braking - Swerved - Loss of control - Aggressive driving - Careless, reckless or in a hurry - Nervous, uncertain or panic - Driving too slow for conditions or slow vehicle - Learner or inexperienced driver/rider - Inexperience of driving on the left - Unfamiliar with model of vehicle

Data A was measured and B+C were made available by authorities.

3.3.1 *Data for Analysis 1*

Analysis 1 covered the area called Northumbria including the cities, namely Newcastle upon Tyne, Gateshead, Sunderland, North Tyneside, South Tyneside and Northumberland. In Analysis 1, the variables belong to data A (number of lanes on approach, half width on approach, entry path radius, number of arms, number of flare lanes on approach, type of roundabout and number of circulating lanes) were measured from roundabout geometric layouts using EDINA and Data Library, University of Edinburgh. Data B (speed limit, sociodemographic characteristics of cyclists and meteorological conditions) and cyclist casualty severity were obtained from CIRTAS database (developed based on Stats19 by Gateshead Council). In addition, limited number of traffic flow profiles for Newcastle upon Tyne and Gateshead were obtained from previous academic research carried out by the Transport Operations Research Group (TORG), Newcastle University.

3.3.2 *Data for Analysis 2*

As mentioned earlier, Analysis 2 was divided into two stages. The first stage was a case study in Northumbria analysing the impact of statistically significant variables from Analysis 1 and the variables in group C (driver/rider behaviour related contributory factors which were obtained CIRTAS) on cyclist casualty severity at roundabouts. It is worth noting here that the contributory factors in CIRTAS do not indicate that they are related to cyclists or driver. Therefore, the impact of behaviour on casualty severity was carried out regardless the road user type.

The second stage of the analysis was conducted to check the consistency of the results including the need to increase the number of samples considering several cities in England. 21 cities, namely Bedford, Blackpool, Brighton & Hove, City of Derby, County Durham, Darlington, Kingston upon Hull, Middlesbrough, Milton Keynes, Portsmouth, Southend, Stoke on Trent, Stockton on Tees, Warrington, York, Newcastle upon Tyne, Gateshead, Sunderland, South Tyneside, North Tyneside and Northumberland were selected. A further aim of the second stage was to determine the influence of regional variance in the regression analysis (i.e. comparing the cities and investigating the regional impact on casualty severity). The city selection was carried out based on initially determining the cities which included open access cyclist casualty data and then random selection from the determined ones. All local authorities in England were considered in this process (See Chapter 4).

For the second stage of the analysis, Group B data was available since it was open access; however, driver/rider behaviour related contributory factors were not available for the area covered. Therefore, the research team contacted the Department for Transport to receive the data for Group C and it was obtained for the years from 2011 to 2016. Group A was not included in the second stage of Analysis 2 since the explanation of casualty (including driving direction) was not provided by DfT because of a confidentiality agreement with the third parties.

3.3.3 *Data for Analysis 3*

Analysis 3 was designed to investigate the international dimension of the findings, comparing England with another country; hence, a comparative study between England (Northumbria) and

Belgium was carried out. The reason why Belgium was selected for this comparative analysis was based on the availability of data via collaborative links with Vias Institute located in Brussels, Belgium. A visit was made to the Vias Institute in April 2018 where the initial findings in England were presented. Several meetings with the academic and research staff there were found to be very beneficial which formed the basis for securing data from Belgium. Since the number of casualties occurred between 2011 to 2016 in Belgium was not adequate to make a good comparison, the data for cyclist casualty at give way roundabouts between 2005 and 2016 in Belgium was requested. A comparative analysis was conducted comparing England (Northumberland) and Belgium.

Since the location of the casualties that occurred was not available for the Belgian data, only group B and C data were considered in Analysis 3. The analysis was divided into two sections: i) individual analysis for each region (England and Belgium), ii) a comparative analysis with proxy data from England and Belgium together.

3.4 Analysis 1 – Investigating the Impact of Geometric Design of Roundabout on Cyclist Casualty Severity: A Case Study of Northumbria

This analysis was conducted to investigate the impact of variables, which were geometric design parameters, speed limit, traffic flow profile, sociodemographic characteristics of cyclists and meteorological conditions, on cyclist casualty severity of crashes that occurred at roundabouts. The data description has been given in the previous section.

The direction of the driver and rider was considered to measure the relevant geometric design parameters (i.e. number of lanes on approach, half width on approach, entry path radius and number of flare lanes on approach). The description of the crash that enabled the direction of driver/rider to be defined was available in the dataset obtained from CIRTAS, Gateshead Council records. Therefore, Northumbria (Newcastle upon Tyne, Gateshead, Sunderland, North Tyneside, South Tyneside and Northumberland which were located in North East of England) was selected as the case study area based on data availability. Traffic flow profile data was available for only Newcastle upon Tyne and Gateshead; hence the impact of traffic flow profile only was investigated for sub-sets of the data.

The analytical steps, which were applied in Analysis 1, are as follow:

Step 1. Descriptive statistics and Test of normality	The aim was understanding the structure of the data, applying descriptive statistical tests and determining the normality (from Kolmogorov-Smirnov and Shapiro-Wilk tests) using SPSS. The type of data and results of normality analysis guided to determine frequency of each variable regarding the cyclist casualty severity guided the decision as to the most appropriate correlation test in the following step.
Step 2. Correlation analysis	Used to determine the relationship between individual variables which were geometric design parameters, speed limit, sociodemographic characteristics of cyclists and meteorological conditions by using SPSS. Spearman's rho, Point Biserial and Cramer's V correlation analysis was applied.Statistically significance in correlation results were investigated in the step of dimension reduction and interpreted the outcomes of logistic regression were interpreted.
Step 3. Dimension reduction	The aim was to explore the relationship between variables, reducing the number of variables grouping them into factors and finally calculating the factor predictors by using SPSS. The rotation method was Promax with Kaiser Normalisation. The explained total variance and scree plot of eigenvalues were calculated. The resulting factors were used as predictors and applied in binary logistic regression.
Step 4. Reliability analysis	This aimed to investigate the reliability between variables which were grouped into the same factors determined by dimension reduction. Internal consistency was assessed by calculating Cronbach's Alpha values. This analysis was carried out in SPSS.
Step 5. Multiple logistic regression	The aim was to investigate the impact of variables individually and determining the influence of the calculated factors on cyclist casualty severity occurence. Multiple Logistic Regressions were carried out in STATA. Predictive margins were calculated in order to interpret the results and develop recommendations for policy makers.
Step 6. Linear regression	The statistically significant variables, which were continious values, in logistic regression models needed to be interpreted in more detail to develop suggestions for policy makers and design engineers. The upper safe limit of values was detemined by conducting Linear Regression. Response variables were continuous values and the predictor was casualty severity (slight and serious).

Figure 3.2 – Analytical Steps in Analysis 1

Descriptive Statistics and Test of Normality

Visualising the large number of samples or measures is very difficult, therefore descriptive statistics are essential to understand the data before considering any in-depth analysis. Descriptive statistics are useful when determining statistics including mean, median, standard deviation, minimum and maximum of the samples of interval predictors and number of records for categorical variables. Therefore, the analysis started with descriptive statistics to describe the basic features of the data and continued with test of normality.

The test of normality, comparing the sample distribution with a normal is defined by Field (2009), and the importance of checking for normality to eliminate error in further statistical procedures was endorsed by (Ghasemi and Zahediasl, 2012). The reason for this requirement is that normal distribution is the main assumption for many statistical applications such as a normality test which is applied to be able to understand the data structures as the results of the normality test will help to decide the correct correlation test. Normality can be determined by significance tests or plots (Altman and Bland, 1995; Field, 2009). There are several types of normality tests: Kilmogorov-Smirnov, Shapiro-Wilk, Lilliefors, Anderson-Darling, Cramer-von Mises, D'Agostino skewness, D'Agostino-Pearson omnibus, Anscombe-Glynn kurtosis and Jarque-Bera (Ghasemi and Zahediasl, 2012). There is a common statement in the literature that the Shapiro-Wilk test gives the best result with the highest power compared to other tests (Razali and Wah, 2011; Yap and Sim, 2011; Ghasemi and Zahediasl, 2012). In the Shapiro-Wilk test, the null hypothesis is that variable is normally distributed. If the result of the p-value is less than 0.05 then it is statistically significant at 95% confidence level, the null hypothesis is rejected. In this study, the test of normality was checked by using SPSS and the results were given for both Kolmogorov-Smirnov and Shapiro-Wilk.

Correlation Analysis with Spearman's rho, Point Biserial and Cramer's V Tests

There are several types of correlation analysis applications, which determine the strength of statistical relationship between individual variables, with specific assumptions for each of them (See Chapter 2). The most appropriate correlation method should be identified based on the data characteristics (i.e. continuous, categorical, normally and non-normally distributed). Regarding the results of descriptive statistics and test of normality, the suitable non-parametric correlation test was selected. These were Spearman's rho, Point Biserial and Cramer's V correlation analysis and were subsequently applied.

As recommended (Field, 2009) in the literature, Spearman's rho correlation coefficient was calculated when the data was not normally distributed. The formula for the Spearman's correlation coefficient is as follow (MEI, 2007):

$$R = 1 - \frac{6 \sum {d_i}^2}{(N^3 - N)}$$

Where:

R = the coefficient of Spearman correlation (rho),

d_i = the difference in rank between paired values of X and Y,

N = the sample size of *X* and *Y* in the selected sample.

Point Biserial correlation is used for determining the relationship between variables when one is dichotomous and the other one is continuous. In this correlation, the dichotomous variable should be discrete. The formula for the Point Biserial correlation coefficient is as follow (NCSS, 2011):

$$R_{pb} = \left(\frac{\bar{Y}_1 - \bar{Y}_0}{S_Y}\right)\sqrt{\frac{np_0(1 - p_0)}{n - 1}}$$

Where:

n = the sample size of *X* and *Y* in the selected sample

$$\bar{Y} = \frac{\sum_{k=1}^{n} Y_k}{n}$$

$$S_Y = \sqrt{\frac{\sum_{k=1}^{n}(Y_k - \bar{Y})^2}{n - 1}}$$

$$p_1 = \frac{\sum_{k=1}^{n} X_k}{n}$$

$$p_0 = 1 - p_1$$

Finally, when both variables are nominal Cramer's V correlation was conducted in order to establish any relationship. The formula for the Cramer's V correlation coefficient is as follow:

$$V = \sqrt{\frac{x^2}{Nt}}$$

Where:
N = the sample size of *X* and *Y* in the selected sample
t = Minimum (r-1, c-1), r is number of rows, c is number of columns

Dimension Reduction with Principal Component Analysis (PCA)

The next step of the analysis is dimension reduction using PCA. PCA is a technique to reduce the number of variables into a number of dimensions (groups) often referred to as principal components preparing the data for further analysis using for example MLR and visualising the variables. In this study, dimension reduction was carried out to group the geometric design parameters, also called principle components. In order to understand the reliability of the PCA, goodness of fit measures such as the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) and Bartlett's test of sphericity were suggested by previous literature (Hotelling, 1933; Bartlett, 1950; Field, 2009; Hair *et al.*, 2010). The KMO value ranges from 0 to 1, and the higher the value is close to 1 indicates that the dimension reduction is suitable for the data. When the KMO value is less than 0.5, the result of dimension reduction is not valid. In addition, Bartlett's test determines the statistical significance of the correlations in the correlation matrix which is the first step of the process of dimension reduction. The next step was determining the explained total variance and scree plot of eigenvalues. This step was the key criteria to determine how many dimensions or principal components were created in the PCA.

The rotation method was Promax with Kaiser Normalisation, which was recommended (Tabachnick and Fidell, 2013) to apply if the created factors were statistically significantly correlated. Tabachnick and Fidell (2013) suggested that when the component correlation matrix shows that the correlation between variables is greater than 0.32 or smaller than -0.32, there is a statistically significant correlation between the created factors. In this case, oblique rotation with

Promax with Kaiser Normalisation should be applied for rotating the axis. In this study, the most appropriate rotation method was selected when analysing the data in Chapter 5.

Reliability Analysis with Cronbach's Alpha Value

Following dimension reduction, a reliability analysis was conducted to investigate the internal consistency between variables which were allocated to the same factor by PCA. As the state of art review recommended (Cronbach, 1951), Cronbach's Alpha should be calculated to determine the internal consistency within grouped variables. A Cronbach's Alpha with a range between 0.7 and 0.9 means that there is an acceptable internal consistency between observed variables (Tavakol and Dennick, 2011). If the value is between 0.6 and 0.7, the internal consistency is questionable. However, the Cronbach's Alpha below 0.5 is not acceptable since it shows that there is very low observed internal consistency in the grouped variables. On the other hand, a high Cronbach's Alpha, above 0.9, should also be questioned because some items may be redundant, which means that the variables measure the same item but in a different way. Therefore, a Cronbach's has maximum limit of 0.9 for a reliable test of internal consistency (Streiner, 2003). The formula for Cronbach's Alpha is as follow:

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}}$$

Where:

N = the number of items

$\bar{c}$ = the average inter item covariance of each of the variables

$\bar{v}$ = the average variance of each of the variables

Logistic Regression Analysis

Logistic regression was applied to investigate the impact of the variables on cyclist casualty severity at roundabouts. As reported in the critical review of analytical methods, Logistic Regression included one response and one predictor variable. However, with more than one predictor Multiple Logistic Regression (MLR) was used. In this study, the response variable was cyclist casualty severity (slight and serious) and the predictors were geometric design parameters, sociodemographic characteristics of cyclists, meteorological conditions, speed limit and traffic flow profile.

The formula of MLR is as follow:

$$logit(p) = \log(Odds) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

Where:

$X_1, X_2 \ldots X_n$: Predictor variables

β_0 : Coefficient of the unknowns

$\beta_1, \beta_2 \ldots \beta_n$: Coefficients of the predictor variables

In MLR, the minimum limit of 10 events per variable (EPV) was suggested by the early research (Peduzzi *et al*., 1996). Although this limitation was a "*rule of thumb*" in the literature, this study aimed to develop the most reliable model by adhering to the recommendation of 10 EPV. With regard to the literature, there are two options to reduce the number of variables in MLR if the number of observations is not adequate to obtain a reliable full model. The first recommended option is relaxing the p-values criterion (Sperandei, 2013). In relaxing the p-values method, initially both Simple Logistic Regression (SLR) and Multiple Logistic Regression (MLR) should be applied. Accordingly, SLR for each predictor (independent) variable were estimated individually before modelling all predictor variables together in the MLR. Then, the results of both SLRs and the MLR were compared in terms of the statistical significance of the predictor variables before identifying the set of variables taken in a further full model of MLR. However, SLR is not recommended by (Wang *et al.*, 2017) for conducting variables which are continuous values. Therefore, it was decided to examine the second recommended (Suhr, 2005) option called Principal Component Analysis. This allows to group the predictor variables into several components before applying the MLR to deal with the limitation of 10 EPV. Regarding the limitations of MLR, the analysis was applied in three steps (MLR1, MLR2 and MLR3) considering the EPV limitation.

The probabilities of serious casualty severity occurrence, expressed as predictive margins, were calculated in order to gain a deeper understanding. The base value was selected as slight casualty therefore the outcome of the odds ratio gave the probability of serious casualty occurrence relative to slight.

Linear Regression

When the statistically significant predictor is a continuous variable, interpreting the result to provide suggestions for policy makers and design engineers is difficult because the predictive margins can be calculated for any value. Therefore, Linear Regression was applied to interpret the influence of continuous variables on cyclist casualty severity. The formula of Linear Regression is as follows (Schneider *et al.*, 2010):

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

Where:

Y = Response variable

β_0 = Coefficient of the unknowns

$X_1, X_2 \ldots X_n$ = Predictor variables

$\beta_1, \beta_2 \ldots \beta_n$ = Coefficients of the predictor variables

The results of Analysis 1 encouraged further research (Analysis 2) which included driver/rider behaviour related contributory factors in order to gain a richer understanding of the wider influences for better interpretation of the impacts of geometric design on cyclist casualty severity.

3.5 Analysis 2 – Investigating the Impact of Driver/Rider Behaviour on Cyclist Casualty Severity: A Case Study of Northumbria and Expanded Study across England

Analysis 2 (divided into two stages) was conducted in order to illustrate the impact of driver/rider error, reaction, behaviour and inexperience related contributory factors. The stages of the analysis are given in Figure (3.3).

Stage 1. A case study of Northumbria	*Descriptive Statistics:* Understanding the structure of the data. This analysis was conducted in SPSS. *Relaxing P-value Criterion* Applying SLR and MLR together and selecting the statistically significant variables at 90 and 95% confidence level. STATA was used to conduct logistic regressions. *Multiple Logistic Regression* A final full model of MLR including selected variables. Calculating the predictive margins. *Linear Regression* Reversed model to interpret the impact of statistically significant variables which are continuous values. STATA was used to conduct Linear Regression.
Stage 2. Expanded study around England	*Descriptive Statistics:* Understanding the structure of the data. This analysis was conducted in SPSS. *Relaxing P-value Criterion* Applying SLR and MLR together and selecting the statistically significant variables at 90 and 95% confidence level. STATA was used to conduct logistic regressions. *Multiple Logistic Regression* A final full model of MLR including selected variables. Calculating the redictive margins. *Multilevel Logistic Regression* Detemining the regional residual on the model by calculating the estimated variance.

Figure 3.3 – Stages in Conducting Analysis of Driver/Rider Behaviour

Stage 1 – A Case Study of Northumbria

In stage one, a case study covering the Northumbria area was considered to obtain an initial understanding. The casualty severity data used was similar to Analysis 1, but only group C data was included. Therefore, the dependent variable in the regession model was binary (serious and slight) cyclist casualty severity. This was because of no record of fatal casualty in the the given period (2011-2016) in Northumbria. Only statistically significant variables resulting from Analysis 1 and group C data were applied in the analysis.

Initially, descriptive statistics was carried out to gain similar understanding on the structure of the driver/rider related contributory factors, as in Analysis 1. This was followed by Logistic Regressions. The limitation for 10 EPV (Peduzzi *et al.*, 1996), was still valid; therefore as recommended by (Sperandei, 2013) the relaxed p-value criterion was applied. SLR1 and MLR4 was applied to identify the statistically significant variables at 90% and 95% confidence level to derieve a final reliable model MLR5 . The equations of the model were given in the previous section.

Stage 2 – Expanded Analysis Around England, Including 21 Cities

Regarding the result, the analysis needed to be carried out with increased numbers of observations and the consistency of the results should be examined by regional influence on the model. In additon, the influence of regional variance on model should be determined. Therefore, it was decided to carry out a second stage of Analysis 2 by expanding the study area around England. The data was collected for 15 cities in England (between 2011 and 2016) and merged with the data of Northumbria. Similar to the first stage, descriptive statistics and relaxed p-value method (i.e. variable selection from SLR2 and MLR7) was conducted. Finally a full model MLR8 was developed.

In the final analysis in Stage 2, Multilevel Logistic Regression was conducted to determine the influence of regional variance on cyclist casualty severity. Both Simple Multilevel Logistic Regression (SMLR1) and Multiple Multilevel Logistic Regression (MMLR1) were applied because the aim was to investigate the regional variance on a simple model including each individual variables and a full model. The formula of the full model of Multilevel Logistic Regression is as follow (Steele, 2009):

$$log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 X_{ij} + u_j$$

$$u_j \sim N(0, \sigma_u^2)$$

Where:

β_0 = the log-odds that y = 1 when x = 0 and u = 0

β_1 = the effect on log-odds of one unit increase in x for individuals in same group

u_j = is the effect of being in group j on the log-odds that y = 1

σ_u^2 = is the level 2 variance (residual)

3.6 Analysis 3 – Comparative Study of England with Belgium

The aim of Analysis 3 was to investigate the impact of different roundabout design structure, sociodemographic characteristics of road user and environments in England compared to Belgium. Analysis 3 is also divided into two stages: in the first stage, an individual model for each of England and Belgium was developed and in the second stage a comparison was made between the two countries including proxy data (See Figure 3.4).

Stage 1 – Individual Analysis for England and Belgium

A descriptive analysis for England and Belgium individually was conducted in order to determine the frequency of the variables for each of slight and serious severity. This was followed by regression analysis based on relaxing the p-value criterion. SLR and MLR were applied to select the statistically significant variables at 90% and 95% confidence level and finally a full MLR was conducted.

Stage 2 – Comparative Analysis between England and Belgium

The proxy data, which corresponded to the same variables for both countries, was applied in this stage of the analysis. The analysis started with three-way chi squared test of independence in order to gain deeper understanding of the data. This test is used for observing the relationship between two categorical variables based on an idea of comparing the observed frequencies in certain categories to expected frequencies in those categories by chance (Field, 2009). The equation of the chi-square is given following:

$$\chi^2 = \sum \frac{(Observed_{ij} - Model_{ij})^2}{Model_{ij}}$$

Where:
i: Represents the rows
j: Represents the columns in the probability table.

This was followed by Multilevel Logistic Regressions based on relaxing p-value criterion. Simple and Multiple Multilevel Logistic Regressions were applied to determine the estimated effect of covariates between two countries. It was expected that the outcome of this stage would help to understand the differences between two main design types and validate the results of previous stages at an international level.

Stage 1. Individual analysis for both countries	*Descriptive Statistics:* Understanding the structure of the data of England and Belgium *Relaxing P-value Criterion* Individually applying SLR and MLR together and selecting the statistically significant variables at 90 and 95% confidence level. *Multiple Logistic Regression* A final full model of MLR including selected variables for England and Belgium individually. Calculating the predictive margins.
Stage 2. Comparative analysis with merged data	*Three-way chi square test of indepencende:* Undertanding the relationship between variables. *Multilevel Logistic Regression* Detemining the regional influence on the model by calculating the estimated variance.

Figure 3.4 – Steps in Analysis 3 Comparative Study of England with Belgium

3.7 Chapter Conclusions

This chapter has presented the methodological framework, method of data collection and stages of conducted analysis in order to address the open questions which are given as follows:

- *What are the descriptive statistics of variables (geometry of a roundabout, traffic characteristics, socio-demography, meteorological and driver behaviour related contributory factors) in regard to cyclist casualty severity at roundabouts?*
- *What is the relationship between variables?*
- *Do these variables have an impact on casualty severity?*
- *Are any variables a proxy of each other's?*
- *Is the outcome of analysis for England consistent?*
- *Is the method of study applicable in another country?*
- *What is the difference between the impact of the variables in England (tangential design) and Belgium (radial design)?*

The study was divided into three stages: analysis of geometric design impact (Analysis 1), contributory factor analysis (Analysis 2) and comparative analysis (Analysis 3). Each stage was also divided into several steps with regard to the aims of the analysis. The appropriate applicable methods for each step were determined. These methods were descriptive statistics, test of normality, correlation analysis, dimension reduction, reliability test, logistic regression and multilevel logistic regression.

After all analyses mentioned above were successfully conducted, the process and outcomes were presented in the following chapters. In Chapter 4, the details of data collection and coding were explained. The analysis and results were presented in Chapter 5, 6 and 7. Finally, the outcomes were discussed in Chapter 8.

Chapter 4 Collection and Preparation of Data for the Analysis

4.1 Introduction

The approach taken in the multiphase analysis adopted in this research was presented step by step in Chapter 3. The research methods, including descriptive statistics, test of normality, correlation analysis, dimension reduction, reliability analysis, Simple Logistic Regression (SLR), Multiple Logistic Regression (MLR) and Multilevel Logistic Regression, were identified as appropriate methods to establish those variables that influence cyclist casualty severity occurring at roundabouts.

This chapter first describes the data and how it was collected. The study area was selected and the results of the initial descriptive statistics for cyclist casualty records for each of the three steps of analysis, Analysis 1, Analysis 2 and Analysis 3, are presented. Section 4.2 explains the data collection methods of the geometric design parameters, traffic characteristics, sociodemographic characteristics of cyclists and environmental conditions. Section 4.3 introduces the data collection for driver/rider error, reaction, behaviour and inexperience related contributory factors. Section 4.4 gives the details of the data collection for comparative analysis, followed by the chapter conclusions in Section 4.5.

4.2 Data Collection for Analysis 1: A Case Study of Northumbria

4.2.1 Data Collection and Description

The first stage of the data collection was to investigate the availability of data relevant to the research aim and objectives. It was found that cyclist casualty data was available easily from the STATS19 data bases recorded by police authorities in the UK. Given the convenience of physically visiting roundabouts to gain an appreciation of the interaction of cyclist and drivers at roundabouts, a local data set was explored as a first step. The STATS19 casualty records were available in the web-based software developed by Gateshead Council. The data includes the variables (cyclist casualty severity, location of occurred casualty, direction of driver/rider, speed limit, sociodemographic characteristics of cyclist and environmental conditions) and the Northumbria region which consists of six divisions, namely Newcastle upon Tyne, Gateshead, Sunderland, North Tyneside, South Tyneside and Northumberland, was considered as a case study in the Analysis 1(See Figure 4.1 (a) and (b)).

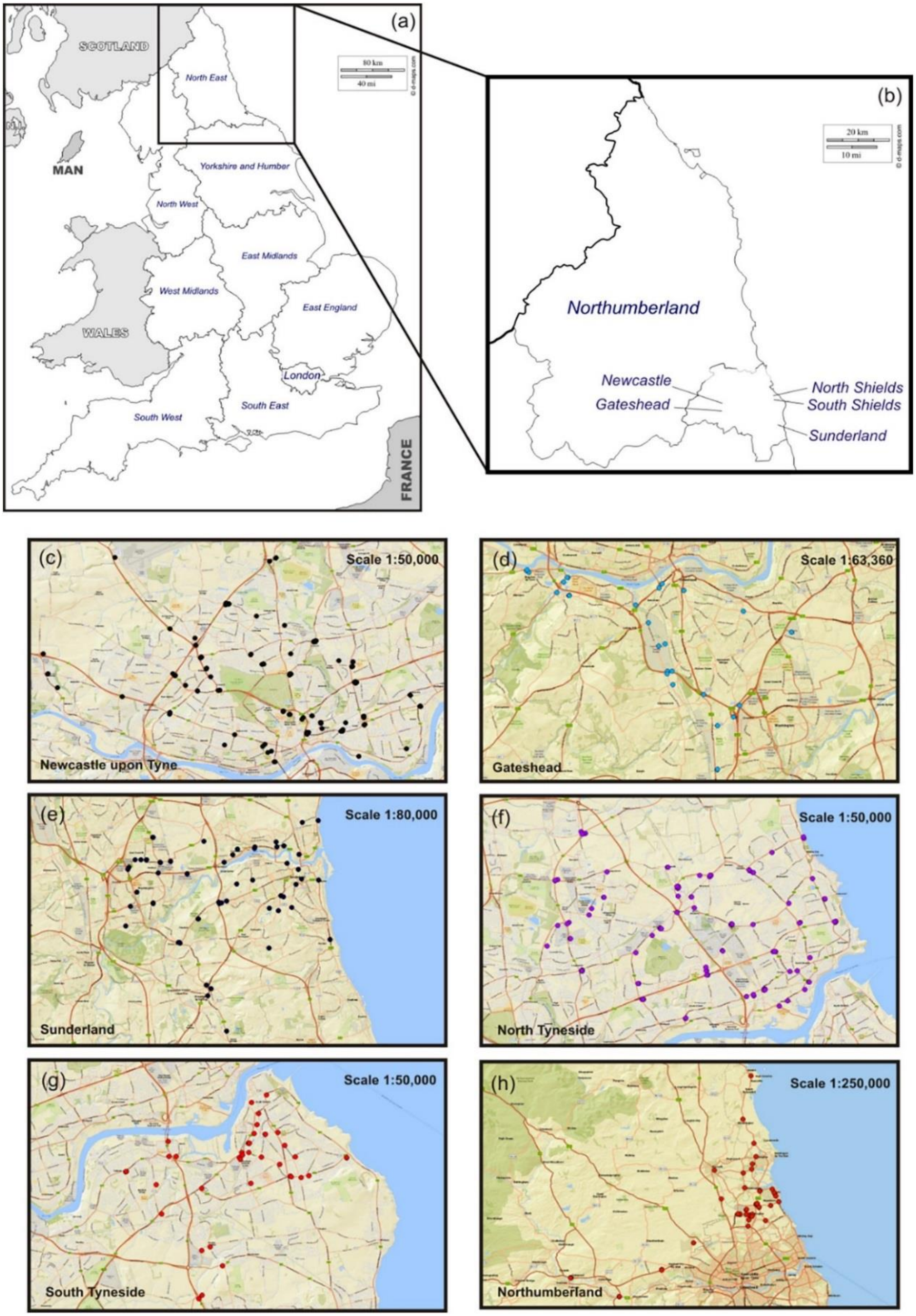


Figure 4.1 – Visualisation of Casualty Locations (map a: edited from http://d-maps.com/m/europa/uk/angleterre/ angleterre21) (map b: edited from http://d-maps.com/m/europa/uk/englandne/englandne12) (map c, d, e, f, g and h: generated in ArcMap)

In accordance with the announcement by the European Commission (2010) in their European Road Safety Policy Orientation for 2011-2020 publication, some changes to the road safety programmes have been implemented in the UK in 2010. In addition, the STATS19 data collection form was reviewed in 2011; therefore, the data recorded during 2011-2016 was considered in this study. Each data record consisted of the coordinates where the casualty occurred. The coordinates were plotted on the maps of the cities (See Figure 4.1-c/d/e/f/g/h). The dots on the map represent the locations of cyclist casualties in each city, respectively Newcastle, Gateshead, Sunderland, North Tyneside, South Tyneside and Northumberland.

Slight cyclist casualties (N = 370) and serious cyclist casualties (N = 69) were recorded at 209 roundabouts during the study period 2011 – 2016. There were no fatal casualties recorded at any roundabout in the given period. As shown in Figure 4.2, there has been a gradual increase in the number of casualties from 2011 to 2015. The 2016 data was not included in Figure 4.2 as it represents only the first 5 months of the year. The increase in trend for both slight and serious casualties involving cyclists is of concern and a there is a need to understand the reason for this rise. This could reflect the increase in cyclist flow; nevertheless, it remains the underlying concern of the safety of cyclists at roundabouts.

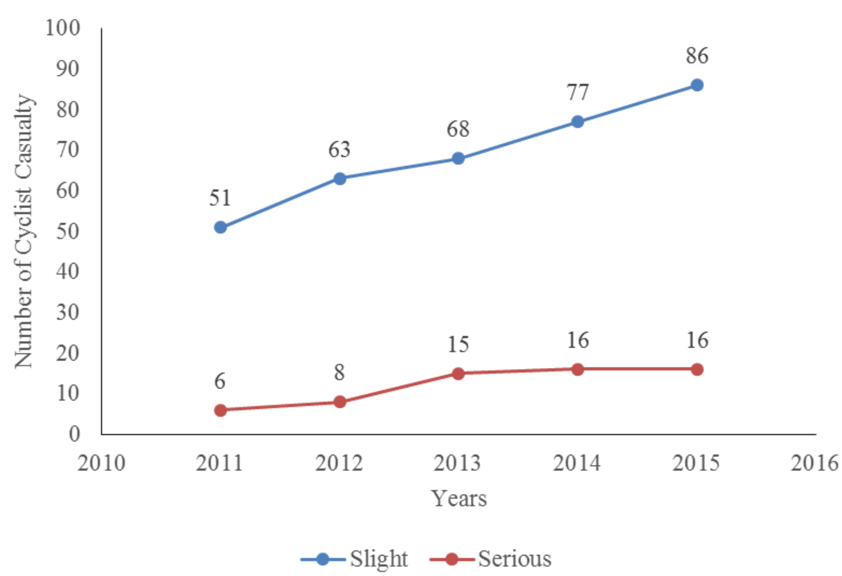


Figure 4.2- Number of Cyclist Casualties for Each Year between 2011 and 2015

With respect to number of casualty severity, the highest number recorded was in Newcastle upon Tyne with 125 slight and 20 serious. This was followed by North Tyneside with 92 slight and 14 serious cyclist casualties. The lowest record of casualty severity was in Gateshead with 23 slight and 4 serious (See Figure 4.3). The cyclist flow or cyclist miles travelled for given areas was not available; therefore, interpretation of these figures is difficult.

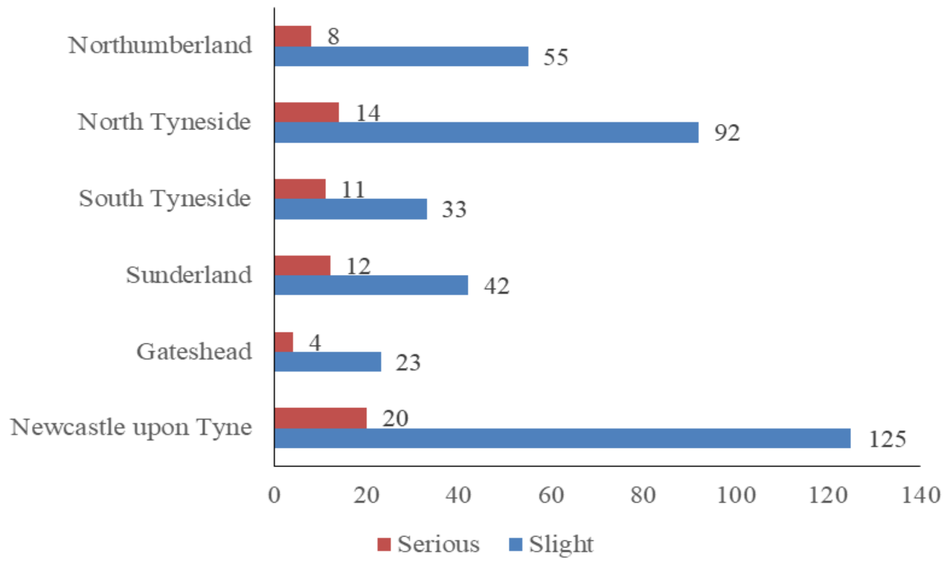


Figure 4.3 – Number of Casualties between 2011 and 2016 for Each City

The casualty records in the database consist of not only casualty severity but also crash reference number, the time the crash occurred, whether the junction was a signal control or give away roundabout, gender and age of cyclist, vehicle type (pedal cycle/car/bus/heavy goods vehicle), speed limit, lighting level, weather and road surface conditions. These types of data were made ready for analysis by defining a coding system. Each of the geometric design parameters were measured consistent with the procedures outlined in the Design Manual for Roads and Bridges DMRB (DfT, 2007), as described more in detail in the next subsection. Traffic flow data only was available for two out of six districts in Northumberland (i.e. Newcastle upon Tyne and Gateshead). Therefore, the impact of the traffic flow profile was investigated in a separate regression model, considering only these two cities. In summary, the variables used in this analysis are shown in Table 4.1. The following sections provide the details of measuring the geometric design parameters in detail and data coding (sociodemographic characteristics of cyclist, environmental conditions and traffic related parameters).

Table 4-1 – Data used for Analysis 1

Geometric design parameters of roundabouts	Number of lanes on approach
	Half width on approach
	Entry path radius
	Number of arms
	Number of flare lanes on approach
	Type of roundabout
	Number of circulating lanes
Sociodemographic characteristics of cyclist	Casualty gender
	Casualty age
Meteorological conditions	Lighting
	Weather
	Road surface condition
Traffic related parameters	Speed limit
	Traffic flow profile

4.2.2 Measuring the Geometric Design Parameters on Roundabout

The majority of the crashes occur at entry and approach locations at roundabouts. Therefore, the geometric design parameters which related to the approach, entry and circulating locations, as well as the size of roundabout were considered in this study. More specifically, the geometric design parameters considered were the number of lanes on approach, half width on approach, entry path radius, number of arms, number of flare lanes on approach, type of roundabout and number of circulating lanes, as shown in Figure 4.4. Determining the approach and entry locations was carried out based on the entry directions of the vehicles which were explained in the crash records. However, the vehicles' path after the entry was not always very clear in the explanatory column of the crash records. This is a main limitation for considering the exit related geometric design parameters (i.e. exit width and exit kerb radius). Therefore, exit related parameters were not considered in the analysis.

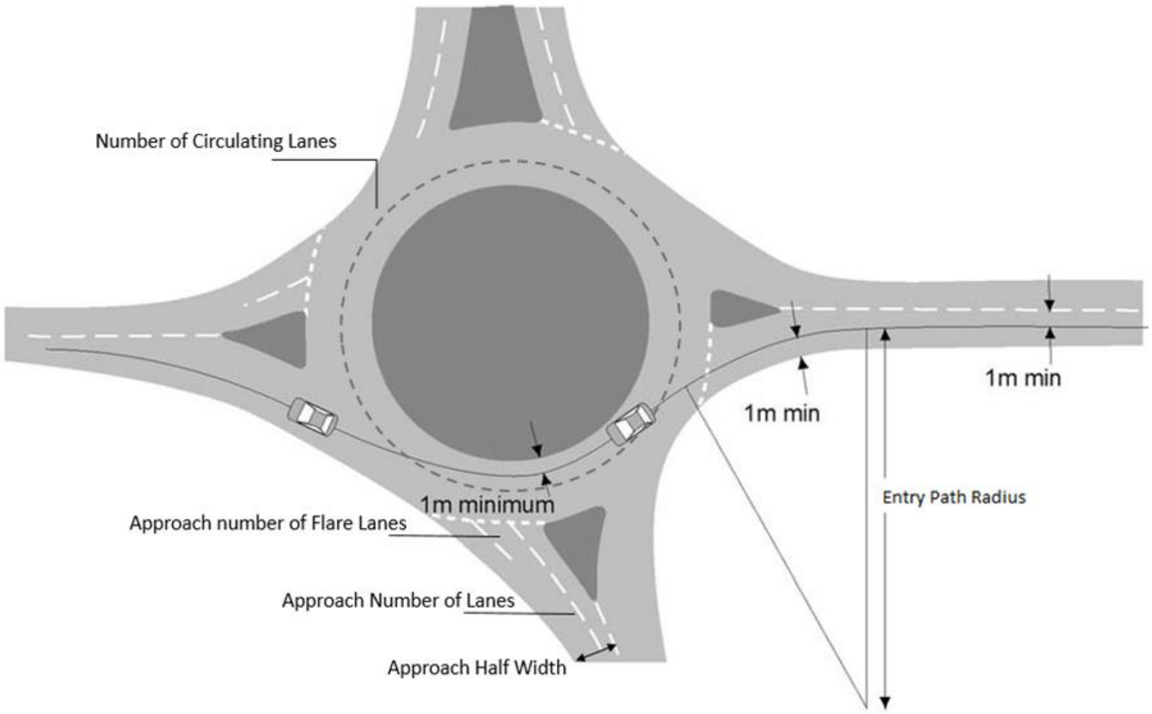


Figure 4.4 – Geometric Design Parameters on a Four Arms Roundabout

The geometric parameters, namely number of lanes on approach, number of arms, type of roundabout and number of circulating lanes, were specified based on Google Earth images. The geometric design parameters were measured and collected for each roundabout that was involved in cyclists' crashes during the study period 2011-16. There was a possibility that a change in the geometric design had occurred during these six years; therefore, the layout of the roundabout for each cyclist casualty also considered the year of the casualty occurrence. The historic imagery option in Google Earth was used to identify changes in geometric parameters, as appropriate. The main difficulty encountered during this process was that the image data for the years of 2011, 2013 and 2014 were not available. Therefore, for these years, either the previous or later years were considered in order to identify the geometric design change that occurred. This approach was necessary because, having approached the local authorities, it was found that such records were not available. Also, the majority of roundabouts were constructed largely before the study period.

When measuring the half width on approach and entry path radius a degree of engineering judgement was required, consistent with normal practice. The maps of roundabout layouts for Northumbria were downloaded from EDINA and Data Library created by the University of

Edinburgh by special licence to Newcastle University. Edina is known to be one of the most accurate databases measuring distance to accuracy of centimetres and widely adopted in the former research for the scientifically published works. The research in this thesis measured distances to one centimetre; however, distances were rounded to one metre in making recommendations. The measurement of these two parameters was carried out on the downloaded AutoCAD roundabout layout drawing. The criteria for each measurement was followed as given in DMRB (DfT, 2007).

By following the description in the design manual, the distance of half width on approach (v) in units of metres was measured as the shortest distance between the median line and the nearside edge of the road. There was no information given on how to locate the position of measurement of the half width of an approach road. This was located on a stretch of road of constant width in advance of the point where the road begins to splay.

Entry path radius is the radius of the kerb of the deflection along the shortest path of driving from approach up to the point of circulating. This deflection should be measured from a start point with a minimum of 1 metre away from the offside of the road until an end point at a minimum of 1 metre away from the edge of the central island. *The entry path radius cannot exceed 100 metres in all types of roundabouts.* Measuring the entry path radius again relied on engineering judgement (DfT, 2007) which introduces a limitation; however, given that the researcher, Akgun, carried out all the measurements and maintained consistency in the measuring method this was not considered and issue.

The geometric design parameters namely entry width, entry angle, average effective flare length and entry kerb radius also were measured using the given explanations in the design manual (DfT, 2007) and were applied in the analysis. However, they did not give any statistical significance in the models. Moreover, they increased the multicollinearity and reduced the statistical significance of the regression. Therefore, they needed to be excluded from the analysis.

All the design parameters were measured in units of metres and were recorded in a bespoke casualty dataset in Excel format. The next step was to transfer the nominal variables (sociodemographic characteristics of cyclists, environmental condition and traffic flow profile) into numeric variables.

4.2.3 Coding the Nominal and the Ordinal Variables into Numeric Values

STATS19 casualty records have a specific nominal classification in the sense that casualty severity, sociodemographic characteristics of cyclist, environmental conditions and driver behaviour related contributory factors were not recorded as numbers. Therefore, these have been coded as numeric figures to facilitate dimension reduction, reliability testing and fitting of regression models. The coding applied for each variable in STATS19 and the representative numeric values are given in Table 4.2.

Table 4-2 – Numeric Coding for the Variables from Stats19 Records of England

Variables	Stats19 records	Coding
Cyclist casualty severity	Fatal	2
	Serious	1
	Slight	0
Casualty gender	Male	1
	Female	0
Lighting	Daylight-street lights present) Daylight-no street lighting Daylight-street lighting unknown	(Daylight) 1
	Darkness-street lights present and lit Darkness-street lights present but unlit Darkness-no street lighting Darkness-street lighting unknown	(Darkness) 0
Weather	Fine with high winds Fine without high winds	(Fine) 0
	Raining without high winds Snowing without high winds Raining with high winds Snowing with high winds	(Special) 1
	Fog or mist – if hazard	2
	Other	3
	Unknown	4
Road surface condition	Dry	0
	Wet / Damp	1
	Snow / Frost / Ice	2
	Flood	3

With respect to lighting level, daylight and darkness were represented as a group regardless of the existence of streetlight. A similar approach was applied for weather conditions, regardless of the existence of wind. The reason for this application was to limit the number of variables influencing casualty severity. This will be investigated further in the descriptive statistics at each step of analysis.

As mentioned earlier, the traffic flow profile data was only available for Newcastle upon Tyne and Gateshead districts from an earlier study carried out by Newcastle University (Goodman *et al.*, 2014). Therefore, an analysis was carried out only to investigate the impact of specific traffic flow periods on cyclist casualty severity for Newcastle upon Tyne and Gateshead given "the time of crash" was available from the STATS19 data base. The report (Goodman *et al.*, 2014) provided the traffic counts for different peak times (See Table 4.3). Therefore, the traffic flow periods used in this study were defined AM peak from 7 am to 10 am, PM peak from 4 pm to 7 pm and all other times as the Inter-peak (See Table 4.4).

Given the times of the day when cyclists normally choose to travel not too late in the evening, there seemed a justification to separate the day and overnight inter peaks. In this thesis, the peak timetable was connected to the time when cyclist casualty occurred, and a new variable called "traffic flow profile" was created. Traffic flow profile was used in the analysis as a numeric variable (evening inter peak=0, day inter peak=1 and peak=2).

Table 4-3 – Peak Times for Newcastle upon Tyne and Gateshead (Goodman, 2014)

Start Hour	End Hour	TPM Network
00:00:00	01:00:00	Inter-Peak
01:00:00	02:00:00	Inter-Peak
02:00:00	03:00:00	Inter-Peak
03:00:00	04:00:00	Inter-Peak
04:00:00	05:00:00	Inter-Peak
05:00:00	06:00:00	Inter-Peak
06:00:00	07:00:00	Inter-Peak
07:00:00	08:00:00	AM-Peak
08:00:00	09:00:00	AM-Peak
09:00:00	10:00:00	AM-Peak
10:00:00	11:00:00	Inter-Peak
11:00:00	12:00:00	Inter-Peak
12:00:00	13:00:00	Inter-Peak
13:00:00	14:00:00	Inter-Peak
14:00:00	15:00:00	Inter-Peak
15:00:00	16:00:00	Inter-Peak
16:00:00	17:00:00	PM-Peak
17:00:00	18:00:00	PM-Peak
18:00:00	19:00:00	PM-Peak
19:00:00	20:00:00	Inter-Peak
20:00:00	21:00:00	Inter-Peak
21:00:00	22:00:00	Inter-Peak
22:00:00	23:00:00	Inter-Peak
23:00:00	00:00:00	Inter-Peak

Table 4-4 – Traffic Flow Periods for Newcastle upon Tyne and Gateshead

Start Hour	End Hour	Peak
19:00:00	07:00:00	Overnight Inter-Peak
07:00:00	10:00:00	Peak
10:00:00	16:00:00	Day Inter-Peak
16:00:00	19:00:00	Peak

4.3 Data Collection for Analysis 2: Northumbria and Several Cities in England

In Analysis 2, it was decided to investigate the influence of road users' behaviour on the casualty severity of cyclist crashes that occurred at roundabouts. This type of data is separate from open access records in STATS19 due to the fact that they are more personalised referred to contributory factors. There were several groups of contributory factors in the dataset; namely road environment, vehicle defects, injudicious action, driver/rider error or reaction, impairment or distraction, behaviour or inexperience, vision affected by, pedestrian only and special codes. As the prime aim of this study was to observe the road users' behaviour, the most relevant factors were selected, namely driver/rider error, reaction, behaviour and inexperience related contributory factors. One of the limitations of the STATS19 is that the contributory factors are not assigned specifically to the cyclist or the driver. Therefore, the contributory factors were recorded as driver/rider. Analysis 2 was conducted in two stages: firstly, a case study of Northumbria and secondly an extended study across England. The following sections provide the details of the data used for both stages.

4.3.1 Data Collection for the Case Study of Northumbria

The case study for Northumbria aimed to gain an initial understanding of the impact of road users' behaviour before conducting a country wide extended analysis. The Northumbria analysis was a continuation of Analysis 1; increase in the number of casualty observation was the same (370 slight and 69 serious casualties occurred during 2011 – 2016). For this data analysis, only the statistically significant variables from Analysis 1 were included along with behaviour related contributory factors. Finally, the case study of Northumbria, including the variables, namely speed limit, number of lanes on approach, entry path radius and driver/rider error, reaction, behaviour and inexperience related contributory factors, was conducted.

Data collection for speed limit, number of lanes and entry path radius was given in previous sections. Data of *Driver/ Rider Error or Reaction* and *Behaviour or Inexperience* were obtained from CIRTAS database of Gateshead Council for Northumbria (See Table 4.5). The inclusion of contributory factors was achieved using a binary number coded (either 0 or 1) in STATS19 casualty records, for instance if the factor existed for the casualty, it was coded as 1; on the other hand, if it was absent, it was coded as 0. All other coding remained the same. The contributory factors were merged with the main casualty data sheet using the crash reference number which is unique for each individual casualty.

Table 4-5 – Data for Driver/Rider Behaviour Related Contributory Factors

Driver/rider behaviour related contributory factors	
Driver/rider error or reaction	**Driver/rider behaviour and inexperience**
Junction overshoot	Aggressive driving
Junction restart (moving off at junction)	Careless, reckless or in a hurry
Poor turn or manoeuvre	Nervous, uncertain or panic
Failed to signal or misleading signal	Driving too slow for conditions or slow vehicle
Failed to look properly	Learner or inexperienced driver/rider
Failed to judge other person's path or speed	Inexperience of driving on the left
Passing too close to cyclists	Unfamiliar with model of vehicle
Sudden braking	
Swerved	
Loss of control	

4.3.2 Data Collection and Description for Extended Study across England

In order to increase the statistics and improve the confidence in the results of Analysis 1, the case study of Northumbria was extended across England and the presence of regional variance between areas was investigated through scrutiny of regression models. The CIRTAS database was only available for the North East of England. Casualty data for all England was publicly available (STATS19); however, contributory factors were not available in the STATS19 database due to security concerns. Therefore, the UK DfT was contacted in order to obtain the contributory factors for cyclist casualties at roundabouts for the cities/regions that were the focus of this study. This produced the data processing effort of the data processing unit and the data was delivered. Accordingly, several cities were selected, and the contributory factors related data were requested and permission to access granted. Initially, three sets of data (crash, casualty and vehicle) were downloaded from open access platform Road Safety Data for the UK. In these three datasets, the mutual identification was "*Crash Reference*" and, therefore, it was used to match and merge all the relevant variables in one dataset. The second identification was "*Local Authority*", which was unique to each area in England.

Whilst grouping the cyclist casualties with due reference to their *Local Authority* codes, it was realised that some local authorities' crash reference numbers were not identical for 2015 and 2016. Therefore, it was concluded that the database was not completed for all local authorities. This limitation was dealt with by selecting those local authorities which had a full set of casualty severity records and data was downloaded in August 2017. The data used in this study included Blackpool, Greater Manchester, Cheshire, Durham, York, West Yorkshire, South Yorkshire, Humberside, Cleveland, West Midlands, Stoke-on-Trent, Herefordshire, Derby, Nottingham, Leicester, Peterborough, Bedfordshire, Southend-on-Sea, Berkshire, Milton Keynes, Southampton, Medway, Brighton and Hove, Plymouth, Cornwall, Avon, Wiltshire and Bournemouth. These 28 local authorities covered 80 cities and counties.

The actual casualty data of a range of characteristics used in the study was selected by eliminating several cities on a *simple random selection* method. In this method, each city had equal chance or probability of being chosen, thus eliminating biased sampling. 15 cities/local authorities (Bedford, Blackpool, Brighton & Hove, City of Derby, County Durham, Darlington, Kingston upon Hull, Middlesbrough, Milton Keynes, Portsmouth, Southend-on-Sea, Stoke-on-Trent, Stockton-on-Tees, Warrington and York) were randomly selected to put forward to the DfT for the request for contributory factors.

The data for contributory factors received from the DfT was merged with casualty data from STATS19 (speed limit, sociodemographic characteristics of cyclist and meteorological conditions). Finally, the data for all 15 cities were merged with casualty records for Northumbria (See Figure 4.5). This analysis was carried out later in this research at a time when the contributory factor analysis was also available for crashes that occurred in 2017. Therefore, the data for the latter 6 months of 2016 for Northumbria also was compiled. This increased the number of data records to 1680 cyclist casualties (1394 slight, 284 serious and 2 fatal) in total that occurred at roundabouts between 2011 and 2016 at given 21 cities/counties. The two fatal casualties occurred in Northumberland and Kingston upon Hull.

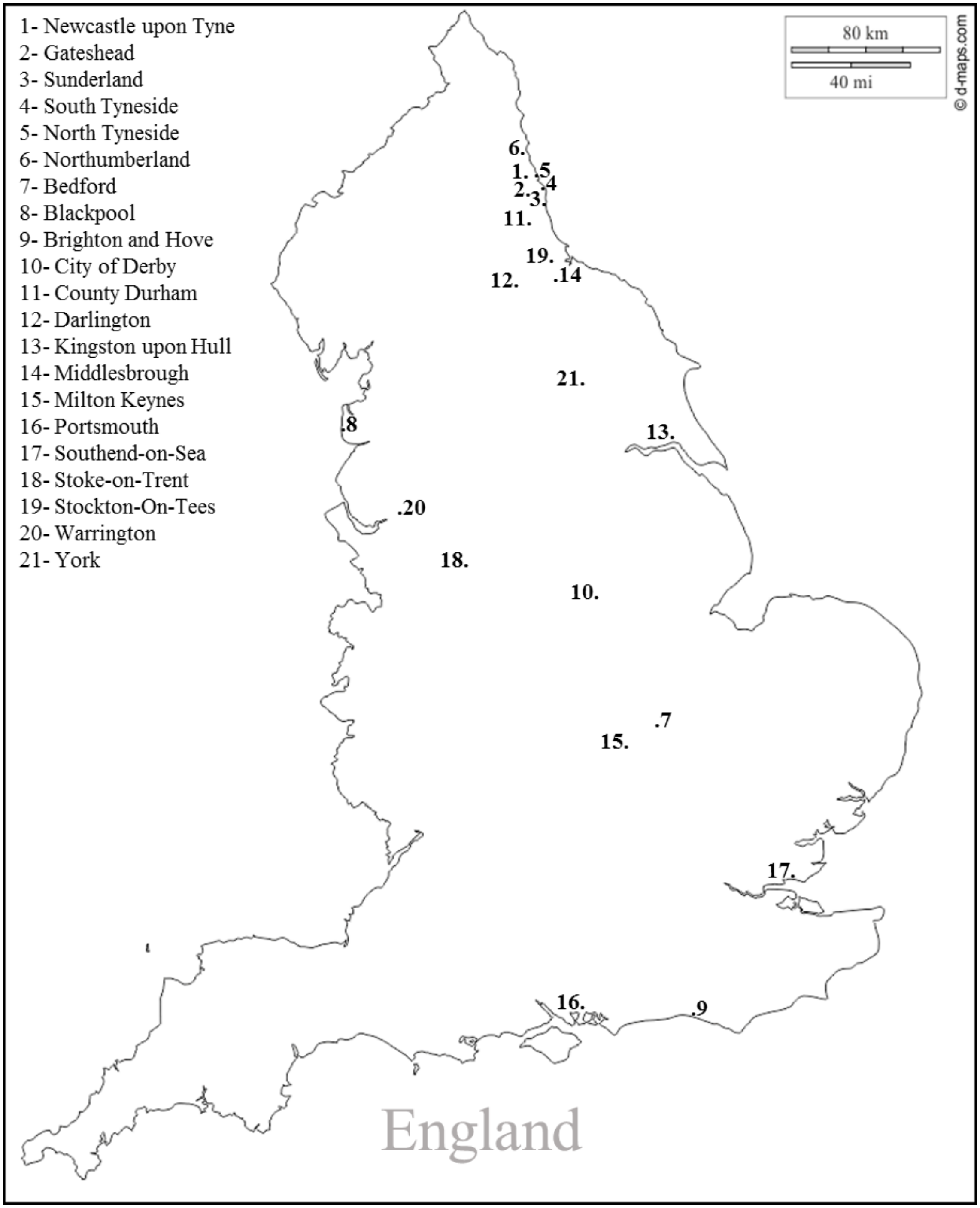


Figure 4.5- The Location of 21 Selected Cities in England

The original intention was to also include the geometric design parameters which emerged as statistically significant in Analysis 1; however, due to data protection rules, neither the open access data nor contributory factors included sufficient details of the crash to be able to identify the direction the driver was travelling and make it possible to determine the design parameters.

Also, the data for 15 cities did not include actual age of the cyclist, but an age band had been given in the dataset instead (See Table 4.6). Similar coding was applied in the extended analysis. Therefore, the actual age for the Northumbria data was allocated to an age band in order to maintain consistency in the dataset.

Table 4-6- Age Band in England

Age Band	Age Between
1	0-5
2	6-10
3	11-15
4	16-20
5	21-25
6	26-35
7	36-45
8	46-55
9	56-65
10	66-75
11	76-101

4.4 Data Collection for Analysis 3: Comparative Study between England and Belgium

In order to gain a better understanding of the differences in the safety performance between tangential and radial roundabouts and test the validity of the UK results in an international context, an additional analysis was conducted to compare cyclist crash casualty data and associated characteristics available in England with Continental Europe. Through close collaboration with Professor Stijn Daniels at the VIAS Institute in Belgium, the availability of a similar cyclist casualty dataset was identified. A visit to the VIAS institute in Belgium in April 2018 was organised to meet and have discussions with Professor Daniels and his colleagues. A brief presentation about this PhD study was shared and followed by a meeting during which the data relating to cyclist casualty in Belgium required for the comparative study was made clear. Accordingly, a confidentiality and IPR, Intellectual Property Rights, agreement was signed, and the relevant data was received in May 2018.

4.4.1 Data Description

The dataset received had cyclist casualties at give way roundabouts covering all areas in Belgium. More specifically, the variables included casualty severity, sociodemographic information of cyclist casualty, meteorological conditions and speed limit, as well as driver behaviour related contributory factors consistent with those shown in Table 4.7. Contributory factors had been recorded as binary digit numbers based on their occurrence in the crash. Therefore, a change to the original coding was not required. In addition, contributory factors were recorded for cyclists and vehicles, separately. This separate dataset had an advantage because the behaviour at roundabouts for either or both users whether cyclist and driver could be explored. Contributory factor records relating to crossing red lights were removed because only give way roundabouts are studied in this thesis.

Table 4-7- Data of Belgium

Contributory factors (for cyclists and vehicles separately)	Crossing the red light
	Non respect of the priority
	Cross the white line continues
	Junction overshoot
	Performs in extremis an avoidance manoeuvre
	Illegal place on the roadway
	Loss of control of the vehicle
	No respect for the distance between users
	Fall
Sociodemographic characteristics of cyclist	Casualty gender
	Casualty age
Meteorological Conditions	Lighting
	Weather
	Road surface condition
Traffic related parameters	Speed limit

The comparative analysis was carried out between Belgium and Northumbria. This decision was made because of the data availability. In earlier study data was collected for Northumbria between 2011 and 2016. Given that data was available in Belgium from 2005 to 2016, to keep the consistency between these two different datasets (in terms of number of casualties occurred during the same time period) the data for the period 2005-2010 was obtained for Northumbria. The contributory factors for the Belgium dataset were slightly different to England. Therefore, the factors were studied carefully and matched before carrying out the analysis. How this was achieved is explained in the following section and termed reconciled data.

4.4.2 Reconciling Data for Comparative Analysis

As mentioned earlier, speed limit, sociodemographic characteristics of cyclist, meteorological conditions and behaviour related contributory factors were considered in the comparative analysis of England with Belgium in terms of cyclists' casualties at roundabouts. Speed limit was given in units of kilometres per hour in the Belgium data. Therefore, this was converted to miles per hour. The differences in the contributory factors were reconciled with each other in data for both England and Belgium and given in Table 4.8. Only four variables were considered to be mutual: *Junction overshoot*, *Poor turn or manoeuvre*, *Passing too close to cyclists* and *Loss of control*. The data descriptor and numeric code assigned to the cyclist casualty severity, cyclist gender, lighting, weather and road surface condition are given in Table 4.9.

Table 4-8 – The Proxy of Contributory Factors of England and Belgium

Contributory factors for England		**Contributory factors for Belgium**
Junction overshoot	Proxy	Junction overshoot
Poor turn or manoeuvre		Performs in extremis an avoidance manoeuvre
Passing too close to cyclists		No respect for the distance between users
Loss of control		Loss of control of the vehicle
Junction restart (moving off at junction)		Non respect of the priority
Failed to judge other person's path or speed		Cross the white line continues
Failed to signal or misleading signal		Illegal place on the roadway
Failed to look properly		Fall
Sudden braking		
Swerved		
Aggressive driving		
Careless, reckless or in a hurry		
Nervous, uncertain or panic		
Driving too slow for conditions or slow vehicle		
Learner or inexperienced driver/rider		
Inexperience of driving on the left		
Unfamiliar with model of vehicle		

Table 4-9 – Data Coding for England and Belgium

Variables	England		Belgium	
	STATS19 records	**Coding**	**Police records**	**Coding**
Cyclist casualty severity	Fatal	2	Fatal	2
	Serious	1	Serious	1
	Slight	0	Slight or Unharmed	0
Casualty gender	Male	1	Male	1
	Female	0	Female	0
Lighting	Daylight-street lights present)	(Daylight)	Day	1
	Daylight-no street lighting	1	Dawn-Dusk	(Darkness)
	Daylight-street lighting unknown		Night-public lighting on	0
	Darkness-street lights present and lit		Night-without public lighting	
	Darkness-street lights present but unlit	(Darkness)	Unknown	2
	Darkness-no street lighting	0		
	Darkness-street lighting unknown			
Weather	Fine without high winds	(Fine)	Normal	0
	Fine with high winds	0	Rain	
	Raining without high winds		Strong wind, gust	(Special)
	Snowing without high winds	(Special)	Snowfall	1
	Raining with high winds	1	Hail	
	Snowing with high winds			
	Fog or mist – if hazard	2	Fog	2
	Other	3	Other	3
	Unknown	4	Unknown	4
Road surface condition	Dry	0	Dry	
	Wet / Damp	1	Clean	0
	Snow / Frost / Ice	2	Dirty (sand, gravel, leaves...)	
	Unknown	3	Wet	1
			Ice / snow	2
			Unknown	3

4.5 Chapter Conclusions

This chapter focused on the stages in data preparation including measurement and coding protocols. This study was divided into three steps: ***Analysis 1, Analysis 2*** and ***Analysis 3***. Geometric design variables, speed limit, traffic flow profile, sociodemographic characteristics of cyclists and meteorological conditions were obtained (from CIRTAS, Gateshead Council) to carry out the ***Analysis 1*** to investigate the impact of variables on cyclist casualty severity. The study area was selected as Northumbria (which included the six districts, namely Newcastle upon Tyne, Gateshead, Sunderland, South Tyneside, North Tyneside and Northumberland) which is in North East of England. The maps of roundabout layouts for Northumbria were downloaded from EDINA and Data Library, created and maintained by the University of Edinburgh, and the geometric design parameters on AutoCAD drawings were measured. The measurements were supported by Google Earth images. The variables (casualty severity, cyclist gender, lighting, weather and road surface conditions) were in either nominal or ordinal structure therefore they were coded into numeric values.

Following ***Analysis 1***, driver/rider behaviour contributory factors were included in the regression analysis. Therefore, in the first part of ***Analysis 2***, the contributory factors (driver/rider error or reaction and driver/rider behaviour and inexperience) were obtained from CIRTAS for the period 2011- 2016. The study area was then extended across England including the variables speed limit, sociodemographic characteristics of cyclists, environmental conditions and driver/rider behaviour related contributory factors. 15 local authorities (Bedford, Blackpool, Brighton & Hove, City of Derby, County Durham, Darlington, Kingston upon Hull, Middlesbrough, Milton Keynes, Portsmouth, Southend, Stoke on Trent, Stockton on Tees, Warrington and York) were selected for Analysis 2. Finally, a comparative analysis between Belgium and England was carried in ***Analysis 3***. Through an academic network, VIAS Institute in Belgium were contacted and similar datasets for the period 2005 to 2016 were obtained. The Northumbria datasets were extended back to 2005 and a careful comparison of the attributes assigned to the two datasets were carefully scrutinised to ensure similarity in the variables used in the final analysis. Chapter 4 has dealt in detail with the data requirements for each of the Analysis Stages 1, 2 and 3, the results of which are presented in chapters 5, 6 and 7, respectively.

Chapter 5 Investigating the Impact of Geometric Design, Traffic Network, Sociodemographic Characteristics of Cyclist and Meteorological Conditions on Cyclist Casualty Severity: A Case Study of Northumbria

5.1 Introduction

The previous chapter set out the data requirements for each of the three stages of the analysis. The aim of this chapter is to report the results of Analysis 1, the application of Logistic Regression methods to investigate the impact of *geometric design, as well as traffic characteristics, the sociodemographic of cyclists and environmental conditions* on cyclist casualty severity of crashes that occurred at roundabouts. However, before carrying out logistic regression models, several basic analyses and statistical tests will be carried out.

Initially, this chapter presents descriptive statistics and a test for normality in the distribution of each variable in Section 5.2. Following this initial exploration of the structure of data, the correlation analysis is carried out in Section 5.3 to determine statistical independence between variables. Previous research (Peduzzi *et al.,* 1996) recommended a minimum of 10 events per variable (EPV) to reduce the error in MLR. Accordingly, in order to reduce the number of the geometric design variables, Principal Component Analysis (PCA), as recommended by Suhr (2005), will be carried out in Section 5.4 along with reliability analysis to confirm the internal consistency between the variables grouped under the same component in PCA. After meeting the 10 EPV condition, Section 5.5 presents the results of several Logistic Regression models derived to identify those variables that have impact on cyclist casualty severity. Finally, the preliminary analysis results are concluded in Section 5.6. All analysis in this chapter was carried out using IBM SPSS and STATA statistical packages.

5.2 Steps in the Analysis in Chapter 5

For clarity an overview of the steps in the analysis presented in this chapter and how they relate to the wider analytical are illustrated in Figure 5.1. These will be dealt with systematically in this section.

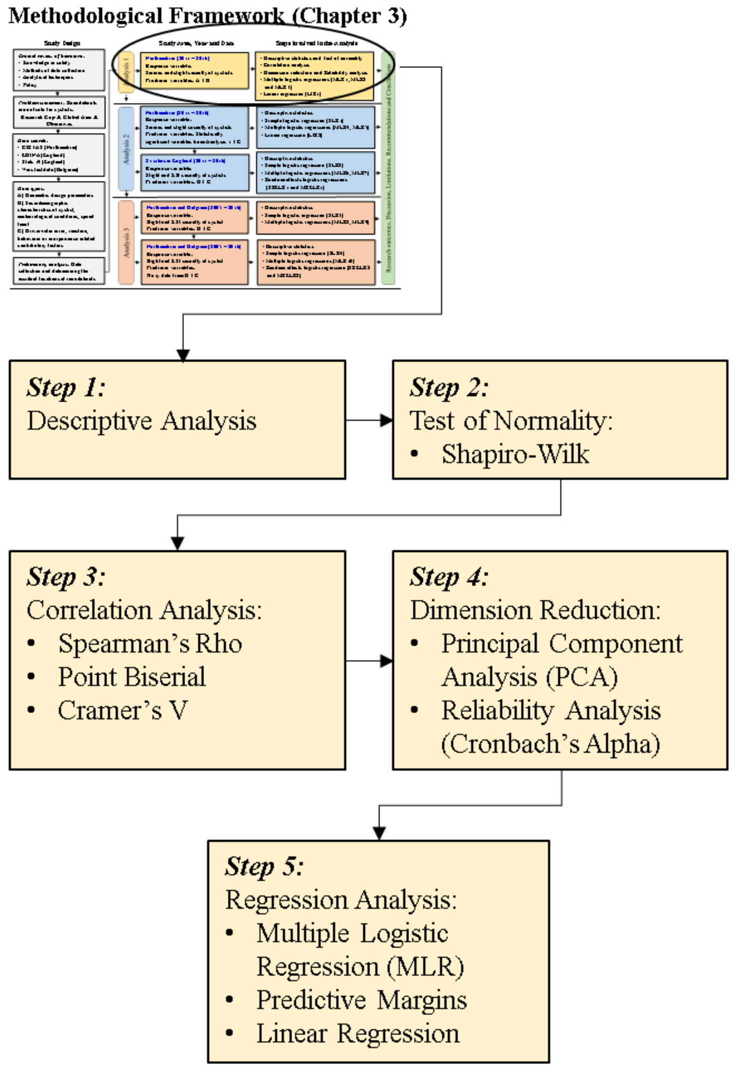


Figure 5.1 – Steps Involved in Chapter 5

Descriptive Analysis of Geometric Design Parameters and Speed Limit

The count, mean, standard deviation, minimum and maximum of all geometric design parameters and speed limit for slight separately from serious casualties are presented in Table 5.1. Slight casualty occurrence gradually decreases when the number of lanes on approach increases from one to three. For serious casualties, the higher number was recorded for roundabouts with two lanes on approach and this was followed by one and three lanes, respectively. Both slight and serious casualties occurred more at roundabouts with four arms compared to three, five and six arms. Roundabouts with two compared with one, three and four flare lanes on approach had more slight and serious casualty records. The serious and slight casualty records were observed to be significantly higher at normal roundabouts (recorded as type 1) compared to mini or grade separated roundabouts. The number of slight casualties decreased sharply when the number on circulating lanes increased from one to three. However, the trend was different for serious casualties as a very similar number of records were recorded for roundabouts with one and two circulating lanes. The highest number of serious casualties was observed to occur at the roundabouts with 30 mph speed limit followed by 60 mph and 40 mph.

Table 5-1 – Descriptive Statistics for Geometric Design Parameters and Speed Limit given the Number of Casualties in Parenthesis

Variable (abbreviation)	**Slight**	**Serious**
Number of lanes on approach (1; 2; 3)	1 (274); 2 (90); 3 (6)	1 (32); 2 (36); 3 (1)
Half width on approach (metre)	Min. (3); Max. (11.37); Mean (5.15); S.D. (1.79)	Min. (3); Max. (8.78); Mean (5.81); S.D. (1.66)
Entry path radius (metre)	Min. (19.23); Max. (99.83); Mean (64.36); S.D. (20.58)	Min. (23.77); Max. (99.98); Mean (80.74); S.D. (20.35)
Number of arms (3; 4; 5; 6)	3 (60); 4 (245); 5 (53); 6 (12)	3 (13); 4 (41); 5 (12); 6 (4)
Number of flare lanes on approach (1; 2; 3)	1 (168); 2 (180); 3 (22); 4 (0)	1 (21); 2 (41); 3 (6); 4 (1)
Type of roundabout (mini=1; normal=2; grade separated=3)	1 (17); 2 (301); 3 (52)	1 (6); 2 (53); 3 (10)
Number of circulating lanes (1; 2; 3)	1 (237); 2 (127); 3 (6)	1 (33); 2 (35); 3 (1)
Speed limit (20; 30; 40; 50; 60; 70) (mph)	20 (3); 30 (280); 40 (33); 50 (12); 60 (33); 70 (9)	20 (2); 30 (43); 40 (8); 50 (1); 60 (9); 70 (6)

Test of Normality of the Geometric Design Parameters and Speed Limit

Since geometric design parameters and speed limit are continuous data, some statistical procedures should be applied to eliminate the errors (Ghasemi and Zahediasl, 2012) by applying incorrect statistical test. Therefore, the data was tested for normality by comparing sample distributions with the normal (Field, 2009) using statistical tests or plots (Altman and Bland, 1995; Field, 2009).

Table 5.2 provides Shapiro-Wilk test for all geometric design parameters with p-value=0.00 at 95% confidence level. Therefore, we rejected the null hypothesis because there was statistically significant evidence that the geometric design parameters were not normally distributed.

Table 5-2 – Test for Normality for Geometric Design Parameters and Speed Limit

	Shapiro-Wilk		
Variables	**Statistic**	**df**	**Significance**
Number of lanes on approach	0.61	439	0.00
Half width on approach	0.89	439	0.00
Entry path radius	0.96	439	0.00
Number of arms	0.78	439	0.00
Number of flare lanes on approach	0.75	439	0.00
Type of roundabout	0.59	439	0.00
Number of circulating lanes	0.66	439	0.00
Speed limit	0.59	439	0.00

df: Degree of Freedom

The non-normal distributions for half width on approach and entry path radius were plotted as histograms (See Figure 5.2). The histogram for half width on approach was right skewed as a result of gathered peaks (most common values) around 4 metres. With respect to entry path radius, the skewness of non-normal distribution was on the left side. In addition, the normal and detrended Q-Q plots also illustrated the right skewness for half width on approach and left skewness for approach half width.

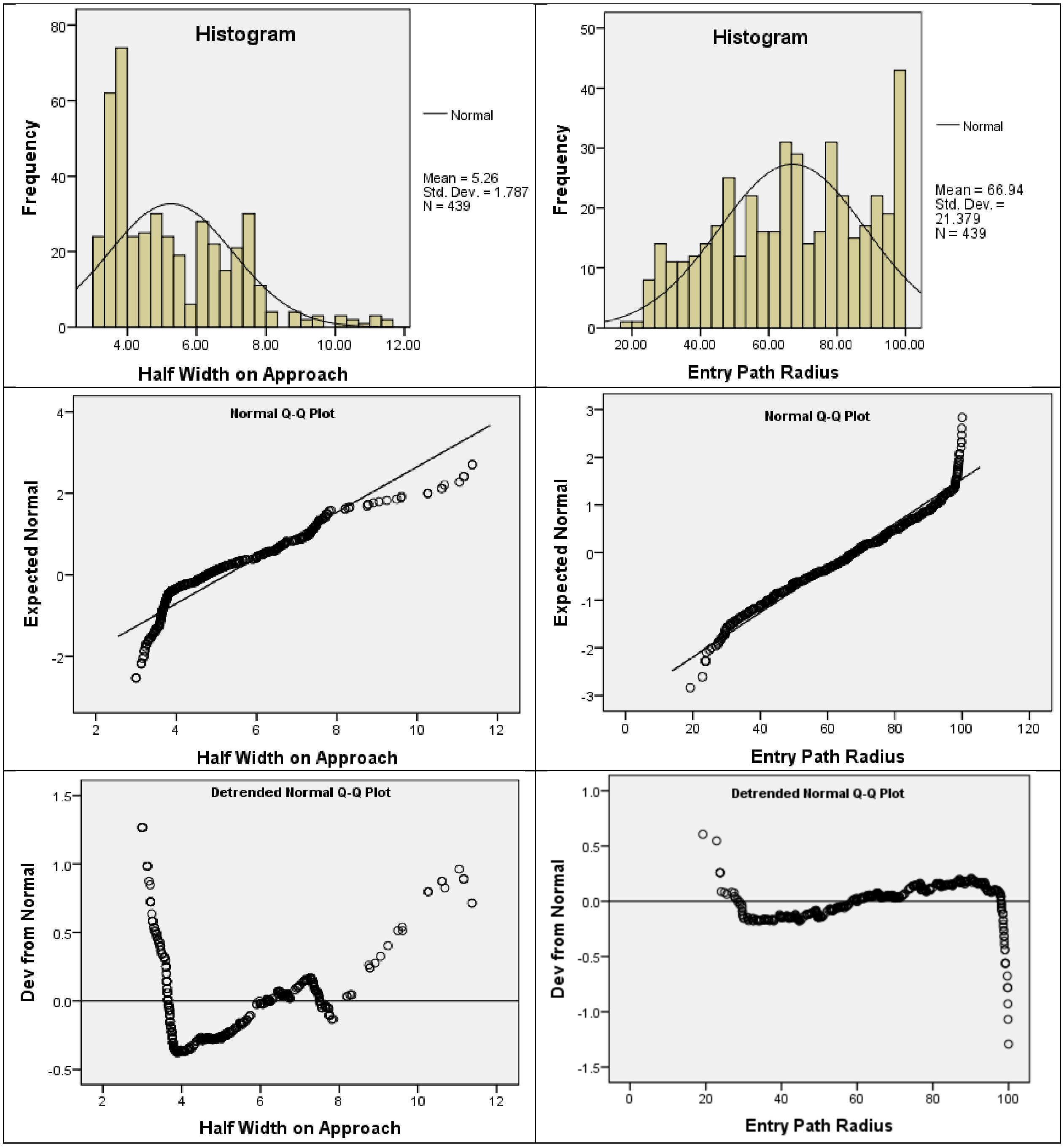


Figure 5.2 – Test of Normality for Half Width on Approach and Entry Path Radius with Histograms and Q-Q Plots

Descriptive Statistics for Sociodemographic Characteristics of Cyclists and Test for Normality of Cyclist Age

The sociodemographic characteristics data considered in this study were gender and age of the cyclist casualty. Both the number of slight (325) and serious (59) casualties were higher for men than women (45 slight and 10 serious), with this difference being around 7 times more for slight casualties and 6 times more for serious. The median, standard deviation, and the range for age of cyclist casualty data were 39.00, 14.17 and 5-75 respectively.

The normality test carried out using SPSS rejected the null hypothesis of normality (p-value= 0.01) with 95% confidence level for Shapiro-Wilk test (See Figure 5.3). The skewness on the histogram was not very clear, although in addition to the mode at 40 years a distinct mode, albeit small, seemed to emerge at 74 years. Thus, the Normal QQ plot for cyclist age variable clearly shows a right skew.

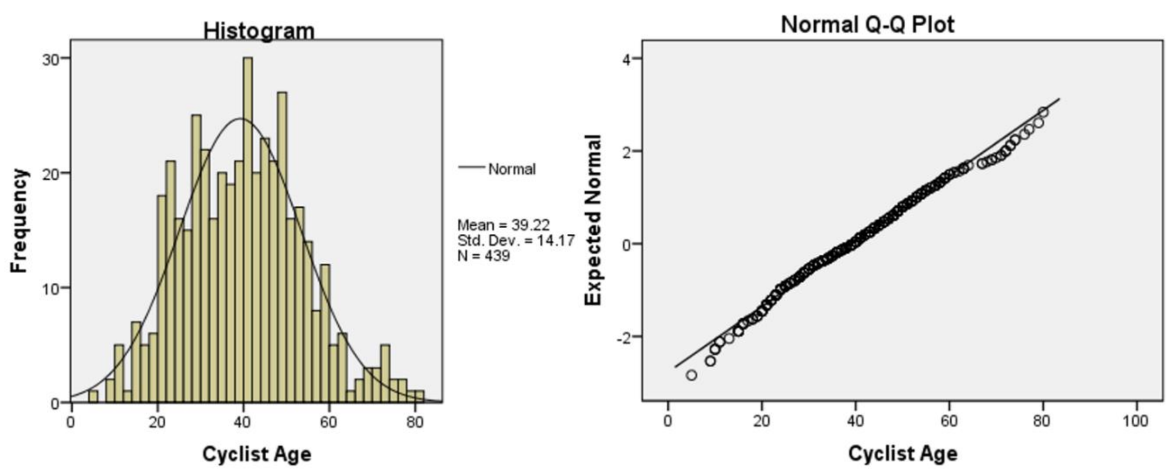


Figure 5.3- Histogram and Normal QQ Plot for Cyclist Age

5.2.3 Descriptive Statistics for Environmental Conditions

The environment (lighting, weather and road surface) were considered as part of meteorological related variables in the casualty dataset (See Table 5.3). The number of casualties that tended to occur at the daylight condition was 371, while at darkness it was 98. With regard to serious casualties, the number of occurrences at daylight was 3.6 times higher than darkness. The number of casualties recorded at fine weather was 373, with 312 casualties occurring when the road surface condition was dry. These figures reflect higher cyclist flows during daylight and fine weather when road conditions are generally dry.

Table 5-3 – Data Description for Environmental Conditions

Variable	Descriptive Statistics
Lighting (Slight)	Darkness (83); Daylight (287)
Lighting (Serious)	Darkness (15); Daylight (54)
Weather (Slight)	Fine (313); Rain (45); Other (12)
Weather (Serious)	Fine (60); Rain (9); Other (0)
Road surface (Slight)	Dry (261); Wet (100); Ice (9)
Road surface (Serious)	Dry (51); Wet (18); Ice (0)

5.3 Correlation Analysis

In order to determine the relationship between variables, correlation analysis was carried out. Spearman's rho was used to investigate the correlation between geometric design parameters and speed limit because they were not found to be normally distributed (See Table 5.4). Given that the number of lanes on approach, arms, flare lanes, circulating lanes and speed limit and also quantitative variables are in discrete structure, Spearman's rho correlation was applied. The formula for the Spearman's correlation coefficient is as follows (MEI, 2007):

$$R = 1 - \frac{6 \sum {d_i}^2}{(N^3 - N)}$$

Where:

R = the coefficient of Spearman correlation (rho),

d_i = the difference in rank between paired values of X and Y,

N = the sample size of *X* and *Y* in the selected sample.

Many of the roundabout geometric design parameters were found to be statistically significant at the 95% confidence level. However, many of the parameters were below the level of *rho* = 0.70 which was the level recommended by Mukaka (2012). The only exception was the number of lanes and half width on approach which can be considered as having had a relatively strong positive relationship with *rho* = 0.76.

Table 5-4 – Spearman's Rho Correlation Results of Speed Limit and Geometric Design Parameters

Variable	1	2	3	4	5	6	7	8
1. Speed limit	1.00							
2. Number of lanes on approach	0.33**	1.00						
3. Half width on approach	0.23**	0.76**	1.00					
4. Entry path radius	−0.09	0.19**	0.23**	1.00				
5. Number of arms	0.17**	0.17**	0.18**	−0.03	1.00			
6. Number of flare lanes on approach	0.30**	0.61*	0.58**	0.21**	0.08	1.00		
7. Type of roundabout	0.20**	0.26**	0.34**	0.00	0.41**	0.26**	1.00	
8. Number of circulating lanes	0.36**	0.42**	0.37**	0.07	0.35**	0.46**	0.41**	1.00

* Statistically significant at 95% confidence level.

**Statistically significant at 99% confidence level.

Sociodemographic characteristics of cyclists (i.e., cyclist gender and age) belonged to the 2nd group. Cyclist gender can be considered as a nominal variable and cyclist age was a continuous value. Cyclist age was normally distributed for each category of gender with p-values Shapiro-Wilk 0.07 (male) and 0.08 (female). Therefore, a point biserial correlation test was considered as the most appropriate one to determine the relationship between them (Kornbrot, 2005). The formula for the Point Biserial correlation coefficient is as follows (NCSS, 2011):

$$R_{pb} = \left(\frac{\bar{Y}_1 - \bar{Y}_0}{S_Y}\right)\sqrt{\frac{Np_0(1-p_0)}{N-1}}$$

Where:
N = the sample size of *X* and *Y* in the selected sample

$$\bar{Y} = \frac{\sum_{k=1}^{N} Y_k}{N}$$

$$S_Y = \sqrt{\frac{\sum_{k=1}^{N}(Y_k - \bar{Y})^2}{N-1}}$$

$$p_1 = \frac{\sum_{k=1}^{N} X_k}{N}$$

$$p_0 = 1 - p_1$$

The results of the analysis showed a statistically significant relationship (R*pb*=0.11) at a 99% statistical confidence level between the two variables but the strength of the correlation at a value of 0.11 is below the recommended level of 0.7.

Environmental variables such as lighting, weather and road surface condition were considered as nominal variables; therefore, Cramer's V correlation analysis was applied as suggested by Field (2009). The formula for the Cramer's V correlation coefficient is as follow:

$$V = \sqrt{\frac{x^2}{Nt}}$$

Where:

V= Cramer's V coefficient

N = the sample size of *X* and *Y* in the selected sample

t = Minimum (r-1, c-1), r is number of rows, c is number of columns

The result of Cramer's V for lighting, weather and road surface condition were statistically significant, but the strength of the relationships between them were below the recommended level of 0.7 (See Table 5.5).

Table 5-5 – Cramer's V Correlation Results of Meteorological Variables

Variable	**1**	**2**	**3**
1. Lighting	1.00		
2. Weather	0.24**	1.00	
3. Road surface condition	0.23**	0.46**	1.00

* Statistically significant at 95% confidence level.

**Statistically significant at 99% confidence level.

5.4 Dimension Reduction

Dimension reduction is a technique to explore dependency within a set of variables, which can be reduced into smaller sets of components (groups) that collectively better explain the variance of the original variables. In this study, Principal Component Analysis (PCA) was carried out to reduce the number of variables by grouping the geometric design parameters into relevant components (See Figure 5.4).

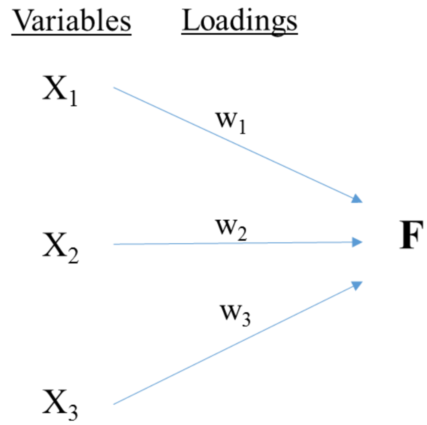


Figure 5.4- The Method of Creating Components (F) in PCA Based on the Loadings (w_1, w_2 and w_3) for Each Variable (X_1, X_2 and X_3)

It should be noted that large values of a specific variable tend to have respectively dominating effects on the PCA compared to smaller values. Therefore, to eliminate this effect to some extent and make the absolute value better describe the relative variation, a log transformation (changing the value of a variable by calculating logarithmic) is made. For instance, for the variables which have considerably large difference in scales, such as number of lanes from 1 to 3 and entry path radius from 20m to 100m, this log transformation is required to reduce the influence of extreme values and outliers. The importance of this transformation is illustrated by the PCA on the geometric design parameters first without and logarithmic transformation (Appendix C) where Cronbach's Alpha value for variable number of lanes on approach, half width, number of flare lanes and entry path radius prior to transformation was 0.07. This is greatly below the recommended value of 0.7, but despite this, these variables were grouped in the PCA and created the component labelled Approach Capacity indicated by the results of KMO and Bartlett's tests of sphericity. The low Cronbach's Alpha level observed was caused by the large values of entry path radius compared to the other variables in that component, including number of lanes on approach, half width and number of flare lanes. Therefore, a log transformation of all the geometric design parameters was conducted and the PCA process repeated. Furthermore, it should be noted that the grouping variables into components by PCA is not affected by log transformation and the internal consistency between variables is achieved.

5.4.1 Principal Component Analysis for Geometric Design Parameters

The results of the log transformed variables in PCA showed that the assumption of sampling adequacy presented by Kaiser-Meyer-Olkin (KMO) was met (KMO value = 0.73). Please note that the KMO value ranges from 0 to 1, and so a closer value to 1 means that the dimension reduction is suitable for the data. The statistical significance of the correlations in the correlation matrix, which is the first step process of the dimension reduction, was determined using the Bartlett's Test of Sphericity and showed that at a statistical significance ($p<0.05$) at a 95% confidence level the variables were unrelated (See Table 5.6). Therefore, it was appropriate for the variables to be used in dimension reduction (Hotelling, 1933; Bartlett, 1950; Field, 2009; Hair *et al.,* 2010).

Table 5-6 – KMO and Bartlett's Test

Kaiser – Meyer – Olkin Measure of Sampling Adequacy		0.73
Bartlett's Test of Sphericity	Approximate Chi-Square	1044
	Degree of freedom	21
	Statistical significance	0.00

Statistically significance at 95% confidence level

The next step was to establish how much of total variance is systematically explained by a given number of components or groups of geometric design variables. This is achieved by scree plot of eigenvalues. Table 5.7 provides the eigenvalues and the cumulative figures of eigenvalues that explain the total variance of the parameters. As can be seen from Table 5.7, the first two components respectively explain 42% and 20% of the variance by 2.97 and 1.37 eigenvalues, both of which were above the suggested limit of eigenvalue with 1.00. This result infers that the given variables were grouped in two main components and together explained 62% of the total variance. Eigenvalues for each component were illustrated in Figure 5.5, and it was seen that the values for the first and second components were above the recommended 1.00.

Table 5-7 – Total Variance Explained by Geometric Design Parameters

Component	Initial Eigenvalues			Extraction Sums of Squared Loading			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.97	42.38	42.38	2.97	42.38	42.38	2.76
2	1.37	19.60	61.98	1.37	19.60	61.98	2.06
3	0.89	12.73	74.71				
4	0.63	9.00	83.71				
5	0.58	8.22	91.94				
6	0.39	5.52	97.46				
7	0.18	2.54	100.00				

Extraction Method: Principal Component Analysis
[a] When components are correlated, sums of squared loadings cannot be added to obtain a total variance

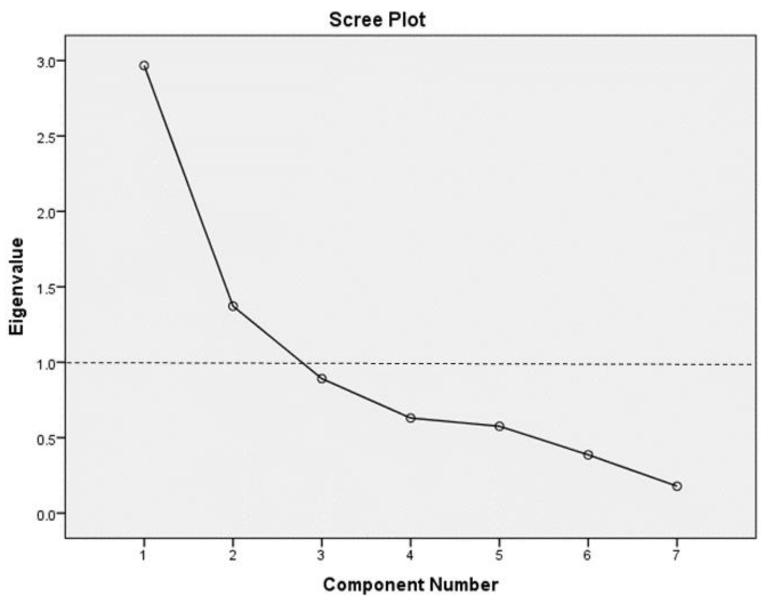


Figure 5.5 – Scree Plot of PCA with the Number of Selected Components above the Eigenvalue 1.0

Tabachnick and Fidell (2013) suggested that when the component correlation matrix shows that the coefficient is greater than 0.32 or smaller than -0.32, then oblique rotation Promax with Kaiser Normalisation should be used to generate the regression coefficients (loadings) of the variables. If the correlation value is between the recommended values, orthogonal rotation (Varimax) should be applied (Tabachnick and Fidell, 2013). In this study, the component correlation matrix gave a correlation value between Component 1 and Component 2 of 0.34. This was slightly above the recommended value of 0.32 for the use of orthogonal rotation of Varimax. Therefore, Promax with

Kaiser Normalisation was applied and the components for each of the variable was determined by the loadings in the pattern matrix.

Given the correlation value of 0.34 being close to 0.32, the orthogonal rotation with Varimax was also applied, but only to gain a deeper understanding of the difference between the two types of rotation (See Appendix D). The result of the orthogonal rotation was similar to the oblique rotation and interestingly created the same components called Approach Capacity and Size of Roundabout. Component 1 labelled "*Approach Capacity*" has variables associated with the impact of number of lanes on approach, half width, number of flare lanes and entry path radius with loadings 0.87, 0.85, 0.82 and 0.51, respectively (See Table 5.8). Component 2 labelled "*Size of Roundabout*" is associated with number of arms, type of roundabout and number of circulating lanes with loadings 0.85, 0.77 and 0.58, respectively. There is no overlap of the loadings observed in the pattern matrix as all variables were in their separate components. The descriptive statistics for the components *Approach Capacity* and *Size of Roundabout* is given in Table 5.9.

Table 5-8 – Pattern Matrix Loadings

Variables	Component 1 Approach Capacity	Component 2 Size of roundabout
Number of lanes on approach	**0.87**	0.05
Half with on approach	**0.85**	0.05
Number of flare lanes on approach	**0.82**	0.02
Entry path radius	**0.51**	-0.36
Number of arms	-0.21	**0.85**
Type of roundabout	0.04	**0.77**
Number of circulating lanes	0.33	**0.58**

Extraction Method: Principal Component Analysis
Rotation Method: Promax with Kaiser Normalisation
[a] Rotation converged in 3 iterations

Table 5-9 – Descriptive Statistics for Approach Capacity and Size of Roundabout

Variable	**Range**	**Minimum**	**Maximum**	**Standard Deviation**
Approach Capacity	4.52	-1.39	3.13	1.00
Size of Roundabout	5.68	-2.44	3.24	1.00

5.4.2 Reliability Analysis with Cronbach's Alpha

The internal consistency between variables grouped into the components (Approach Capacity and Size of Roundabout) was investigated using Cronbach's Alpha calculated for variables in each component (Cronbach, 1951). The formula for Cronbach's Alpha is as follow:

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}}$$

Where:
N = the number of items
$\bar{c}$ = the average inter item covariance of each variables
$\bar{v}$ = the average variance of each variables

Cronbach's Alpha ranges from 0 to 1, and a value between 0.7 and 0.9 indicates that the internal consistency between observed variables is acceptable (Tavakol and Dennick, 2011). If the value of Cronbach's Alpha is between 0.6 and 0.7, the internal consistency is questionable and below 0.5 is a very low internal consistency and therefore not acceptable. On the other hand, a high Cronbach's Alpha, above 0.9, needs to be scrutinised because items in the group might be redundant, which means that the variables measure the same item but in a different way (Streiner, 2003).

The reliability analysis of the raw data was carried out in SPSS. The value of Cronbach's Alpha for the number of arms, type of roundabout and number of circulating lanes was 0.63 which was acceptable but questionable. On the other hand, the Cronbach's Alpha value for number of lanes on approach, half width, number of flare lanes and entry path radius was 0.07. This is way below the recommended value with 0.7. However, these variables were perfectly grouped in PCA and created the component called *Approach Capacity* with statistically significantly acceptable results of KMO and Bartlett's tests. This inconsistency was due to the log transformation of values of the variable needed to harmonise the difference in scale between variables.

The Cronbach's Alpha values for the log transformed variables showed a high internal consistency between the variables grouped under *Approach Capacity* (Alpha value= 0.73) and under *Size of Roundabout* (Alpha value= 0.60) (See Table 5.10). Whilst the alpha value for *Approach Capacity* related variables was acceptable as recommended by Tavakol and Dennick (2011), the value for variables related to *Size of Roundabout* was questionable. However, the two components (*Approach Capacity* and *Size of Roundabout*) were considered sufficiently robust to be taken forward for further analysis including MLR.

Table 5-10 – Reliability Analysis with Cronbach's Alpha

Variables	Components	Cronbach's Alpha with original data	Cronbach's Alpha with log transformed data
Number of lanes on approach Half width on approach Number of flare lanes on approach Entry path radius	Approach Capacity	0.07	0.73
Number of arms Type of roundabout Number of circulating lanes	Size of Roundabout	0.63	0.60

5.5 Multiple Logistic Regression Model

MLR is a predictive analysis that includes several predictors and one dichotomous (binary) dependent variable in a full model to investigate the collective impact of predictor variables on the dependent variable. In this study, the dependent variable of the MLR was cyclist casualty severity, which was coded as a binary variable (i.e. slight=0 and serious=1). By having slight casualty as the base level of the model, the MLR allows the impact of the predictors (i.e. geometric design parameters, speed limit, sociodemographic characteristics of cyclists, weather conditions, approach capacity, size of roundabout and traffic flow profile) on the increase of probability of serious casualties to be investigated. The statistical significance at 95% confidence level (p-value equal to or less than 0.05) was considered to identify the statistically significant variables. The variables with statistical significance at 90% confidence level also was considered for further investigation considering the possibility that the sample size may not be sufficient to achieve the required statistical significance. The formula of MLR is as follow:

$$logit(p) = \log(Odds) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

Where:

$X_1, X_2 \ldots X_n$: Predictor variables

β_0: Coefficient of the unknowns

$\beta_1, \beta_2 \ldots \beta_n$: Coefficients of the predictor variables

p = Probability of Serious Casualty

1-*p* = Probability of Slight Casualty

5.5.1 Multiple Logistic Regression Model for Northumbria

In this section, the MLR was conducted in three steps: i) MLR1 considering variables speed limit, sociodemographic characteristics of cyclists and meteorological conditions; ii) MLR2 with the geometric design parameters; and finally iii) MLR3 with all variables, namely approach capacity, size of roundabout, speed limit, sociodemographic characteristics of cyclists and environmental conditions. The results of these models now will be presented in turn.

Multiple Logistic Regression 1 (MLR1)

The first MLR1, which considered only speed limit, sociodemographic characteristics and environmental conditions, produced the coefficients of speed limit, cyclist gender, cyclist age, lighting, and weather and road surface condition (0.02, -0.27, 0.01, -0.22, -0.22, and -0.22, respectively) defined the coefficients in the equation:

$$Logit\ (p) = ln\frac{p}{1-p}$$
$$= -2.49 + 0.02 * Speed\ limit - 0.27 * Cyclist\ gender + 0.01 * Cyclist\ age - 0.22 * Daylight - 0.22 * Weather - 0.22 * Road\ surface\ condition$$

The dependent outcome of logistic regression was in Logit (p) form (i.e., natural logarithm of probability of serious casualty occurrence (p) divided by the probability of slight casualty occurrence, (1-p)), therefore, the predicted probabilities were calculated based on the exponential version of the Logit (p).

$$p = \frac{exp^{logit(p)}}{1 + exp^{logit(p)}}$$

Where p is the predicted probability (predictive margins).

With regard to the given equation of MLR1, any required probability can be calculated based on selected values of variables. For example, the predictive margins (probability of serious casualty occurrence) for each speed limit was calculated in this case for females with average age of casualty 39.22 years during daylight with fine weather and dry road surface and presented in Table 5.11.

Table 5-11 – Predictive Margins for Each Speed Limits

Speed Limit	Predictive Margins	p-value	95% Confidence Interval	
			Lower	Upper
20	0.14	0.01	0.04	0.24
30	0.17	0.00	0.07	0.28
40	0.21	0.00	0.09	0.33
50	0.25	0.00	0.11	0.40
60	0.30	0.00	0.12	0.48
70	0.35	0.00	0.13	0.51

Whilst the equation of MLR1 suggests that all variables influence the probability of serious relative to slight casualty, it is important to establish how significant these variables are. Table 5.12 presents the statistical significance results for the MLR1 model. The only variable that has a statistically significant influence on cyclist casualty severity at roundabouts at a 95% confidence level is speed limit. The positive coefficient (0.02) indicates that a higher speed limit increases the probability of serious relative to slight casualty severity at roundabouts. This increase could be interpreted in a meaningful way by using the odds ratio of the speed limit and predicted margins probabilities. The odds ratio (1.02) suggests that serious casualty occurrence was 2% more likely than slight casualty for a one unit (10mph) increase of speed limit at roundabouts. The upper and lower intervals for the odds ratios at 95% confidence were 1.00 and 1.05 respectively, and these suggest 95% confidence interval of statistical significance of the result.

In Linear Regression, R^2 value shows the significance of the model. If the R2 value is greater than 0.8, the regression is statistically significant compared to a null model (Bakar and Tahir, 2009). However, R^2 value is not representative for statistical significance of the Logistic Regression; therefore, it is recommended that statistically significant P-value (at 95% confidence level) should be considered while interpreting Logistic Regression model (UCLA, 2019). The result of MLR1 showed that P-value of the model is 0.24 which suggested that the model is not statistically significant.

This result was interesting because it is expected that other variables may influence the casualty severity. For example, it is expected that people drive slower in darkness and are more cautious in rain. This result is likely to be due to insufficient sample data; hence the analysis should be extended and applied across England. In addition, given that little of the variation is explained by speed, this result is suggesting that behaviour related contributory factors may have a role on probability of casualty severity and should be considered.

Table 5-12 – The Results of MLR1

Casualty	**Coefficient**	**P ≥|z|**	**Odds Ratio**	**95% Confidence Interval for Odds Ratio P-value= 0.24**	
				Lower	**Upper**
Speed limit	**0.02**	**0.02****	**1.02****	**1.00**	**1.05**
Cyclist gender	-0.27	0.48	0.76	0.36	1.62
Cyclist age	0.01	0.24	1.01	0.99	1.03
Lighting	-0.22	0.52	0.81	0.42	1.56
Weather	-0.22	0.56	0.80	0.39	1.68
Road surface condition	-0.22	0.49	0.81	0.44	1.49
Constants	-2.49	0.00	0.08	0.02	0.29

**Statistically significant at 95% confidence level.

In order to gain a deeper understanding of the impact of speed limit on casualty severity, the predictive margins of speed limit in MLR1 were calculated from a Simple Logistic Regression including only speed limit (Table 5.13). The predictive margins for each speed limit (i.e. 20, 30, 40, 50, 60 and 70) were determined and presented in Figure 5.6. The P-value dropped to 0.03, which showed that the simple logistic regression model including only speed limit is statistically significant, after excluding the sociodemographic and environmental conditions from MLR1. This suggested that either these variables were not related to cyclist casualty severity at roundabouts or number of observations was not representative for the model. This needed a further investigation with greater number of cyclist casualty records.

Table 5-13 – Simple Logistic Regression Including Only Speed Limit

Casualty	**Coefficient**	**P ≥|z|**	**Odds Ratio**	**95% Confidence Interval for Odds Ratio P-value= 0.03**	
				Lower	**Upper**
Speed limit	**0.02**	**0.02****	**1.02****	**1.00**	**1.04**
Constant	-2.52	0.00	0.08	0.04	0.18

**Statistically significant at 95% confidence level.

$$Logit\ (p) = ln\frac{p}{1-p} = -2.51 + 0.02 * Speed\ limit$$

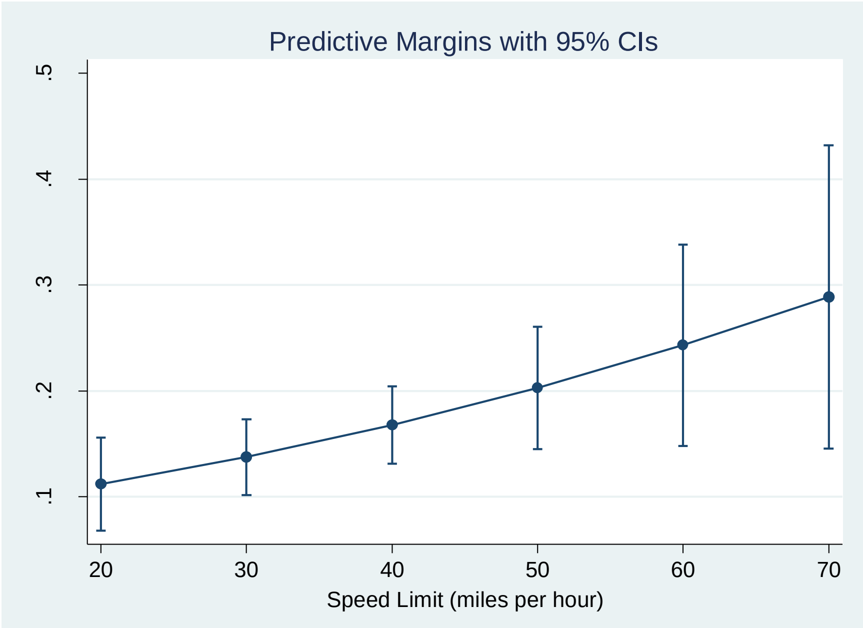


Figure 5.6- Predictive Margins for Speed Limit

Multiple Logistic Regression 2 (MLR2)

The impact of geometric design parameters on cyclist casualty severity at roundabouts was explored in MLR2. The coefficient values of each variable lead to the equation in Logit (p) form:

$$Logit\ (p) = ln\frac{p}{1-p}$$
$$= -5.07 + 1.60 * Number\ of\ lanes\ on\ approach - 0.22 * Half\ width\ on\ approach$$
$$+ 0.04 * Entry\ Path\ Radius + 0.11 * Number\ of\ arms - 0.32$$
$$* Number\ of\ flare\ lanes\ on\ approach - 0.65 * Type\ of\ roundabout + 0.53$$
$$* Number\ of\ circulating\ lanes$$

This equation can be used to calculate the probability of serious casualty occurrence for any contribution of values combination for each predictor where the predicted probability (predictive margin) is given as follow:

$$p = \frac{exp^{logit(p)}}{1 + exp^{logit(p)}}$$

Table 5.14 shows that the *number of lanes on approach* and *entry path radius* reached the expected statistical significance level of 95% with p-values of 0.01 and 0.00, respectively. The P-value of the model was 0.00. The signs of the coefficients are important because a positive coefficient, as for the number of lanes on approach along with a high level of odds ratio of 4.97 suggests that for one more lane on the approach arm increases the occurrence of serious casualty of cyclists relative to slight at roundabouts to around 5 times. The sign of the entry path radius also indicates that a higher entry path radius increases the severity of the cyclist casualties at roundabouts with odds ratio 1.04. This value suggests that a serious casualty is 1.04 times more likely than slight casualty occurrence for cyclists for one unit (1 metre) increase of entry path radius at roundabouts. This result suggests that the number of lanes on approach was a dominant variable in the MLR2 model. The Spearman's rho correlation matrix suggested that there was a statistically significant positive high correlation between number of lanes on approach and half width on approach. However, half width on approach did not show any statistical significance in the MLR2.

Table 5-14 – The Results of MLR2

Casualty	**Coefficient**	**P ≥|z|**	**Odds Ratio**	**95% Confidence Interval for Odds Ratio P-value= 0.00**	
				Lower	**Upper**
Number of lanes on approach	**1.60**	**0.01****	**4.97****	**1.55**	**15.91**
Half width on approach	-0.22	0.18	0.80	0.58	1.11
Entry path radius	**0.04**	**0.00****	**1.04****	**1.03**	**1.06**
Number of arms	0.11	0.64	1.11	0.71	1.73
Number of flare lanes on approach	-0.32	0.32	0.72	0.38	1.36
Type of roundabout	-0.65	0.07	0.52	0.26	1.06
Number of circulating lanes	0.53	0.12	1.69	0.88	3.26
Constants	-5.07	0.00	0.00	0.00	0.03

**Statistically significant at 95% confidence level.

Considering the results so far, the statistically significant influence of number of lanes on approach and entry path radius were consistent and dominant. Therefore, it was decided to develop a predictive model which includes these variables together. Table 5.15 shows that a unit increase in number of lanes on approach or entry path radius increases the probability of serious casualty. The coefficient of entry path radius remained the same but the number of lanes on approach is reduced which shows that the model is sensitive to this variable.

Table 5-15 – Multiple Logistic Regression

Casualty	**Coefficient**	**P ≥|z|**	**Odds Ratio**	**95% Confidence Interval for Odds Ratio P-value= 0.00**	
				Lower	**Upper**
Number of lanes on approach	**0.77**	**0.00****	**2.16**	**1.31**	**3.56**
Entry path radius	**0.04**	**0.00****	**1.04**	**1.02**	**1.06**
Constant	-5.63	0.00	0.00	0.00	0.01

**Statistically significant at 95% confidence level.

$$\begin{aligned} Logit\ (p) &= ln\frac{p}{1-p} \\ &= -5.63 + 0.77 * Number\ of\ lanes\ on\ approach + 0.04 * Entry\ path\ radius \end{aligned}$$

The further step was to predict the margins considering each additional lane and over the range between the maximum and minimum value of entry path radius. The predictive margins for serious casualty occurrence were calculated from 19 to 99 metre for every 1 metre increase in entry path radius for each of 1, 2 and 3 lanes. The calculated predictive margins were statistically significant at 95% confidence level. In addition, the lower and upper 95% confidence interval of margins were very close to each other for each case (See Appendix E). In order to gain deeper understanding of this increased impact, the margins were plotted, and an equation for the margins was created. Figure 5.7 illustrates the logarithmic relationship between the probability of serious casualty occurrence and 1 metre increase in entry path radius for each additional lane on approach. It is seen that there is an overlapping for 95% confidence intervals except lane 1.

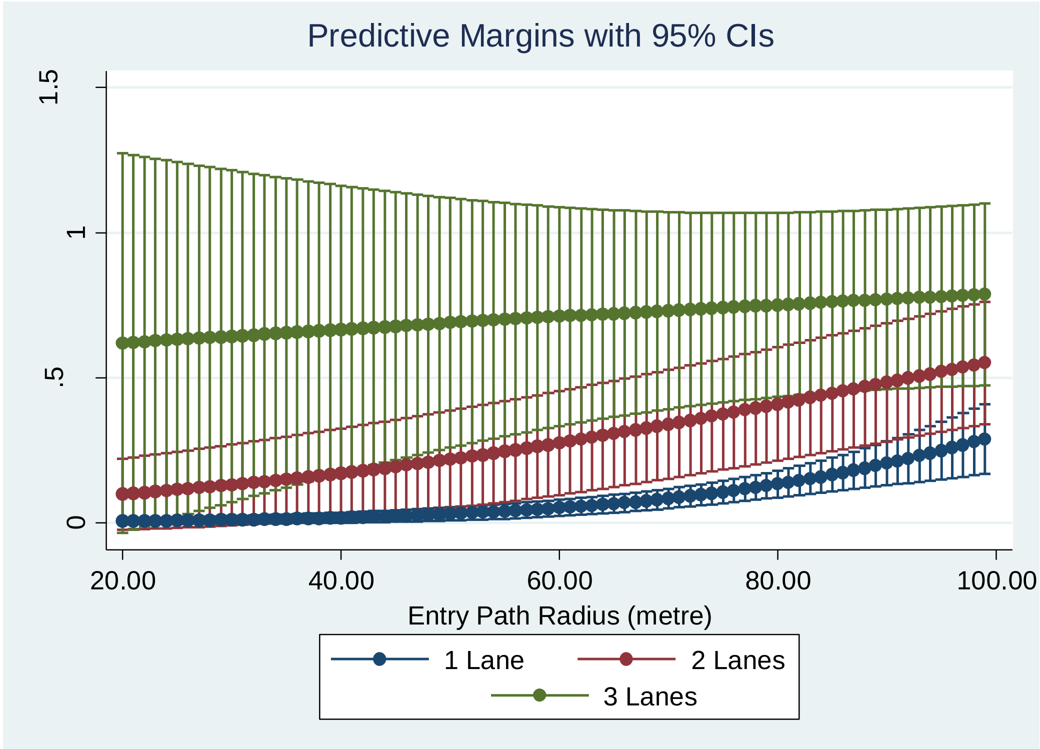


Figure 5.7- Plot of Predictive Margins for Each 1 Metre Increase in Entry Path Radius and Number of Lanes on Approach

Linear Regression 1

The impact of entry path radius on cyclist casualty severity was estimated based on logistic regression modelling because the dependent variable was binary. By switching round the dependent variable (entry path radius) and the independent predictor (casualty severity) and applying the Linear Regression model given by the formula (Schneider *et al.*, 2010):

$$Y = \beta_0 + \beta_1 X_1$$

Where:

Y = Response variable

β_0 = Coefficient of the unknowns

X_1 =Predictor variable

β_1 = Coefficient of the predictor variable

The resulting equation of Linear Regression is as follows:

$$Y = 64.36 + 16.39 * Entry\ Path\ Radius$$

Linear Regression shows that when casualty severity increased from 0 to 1 (slight to serious) the roundabout was more likely to have higher entry path radius (See Table 5.16). The predictive margin for serious casualty (severity=1) was 80.75. This result suggests that if entry path radius was more than 80 metre, the casualty severity was more likely to be serious.

Table 5-16 – Linear Regression Model

Casualty	Coefficient	P ≥|z|	95% Confidence Interval		Predictive Margins
			Lower	Upper	
Casualty Severity					
0	Base				64.36
1	16.39	0.00**	11.09	21.68	80.75
Constants	64.36	0.00	62.26	66.46	

**Statistically significant at 95% confidence level.

Multiple Logistic Regression 3 (MLR3)

MLR3 considered speed limit, sociodemographic characteristics of cyclists, meteorological conditions and the geometric design related components (approach capacity and size of roundabout) derived from PCA. The model gives the coefficients for each variable resulting in the following equation:

$$
\begin{aligned}
Logit\ (p) \quad &= ln\frac{p}{1-p} \\
&= -2.14 + 0.02 * Speed\ limit - 0.34 * Cyclist\ gender + 0.01 * Cyclist\ age - 0.15 \\
&\quad * Daylight - 0.12 * Weather - 0.43 * Road\ surface\ condition + 0.62 \\
&\quad * Approach\ capacity - 0.29 * Size\ of\ roundabout
\end{aligned}
$$

Where the predicted probability (predictive margin):

$$p = \frac{exp^{logit(p)}}{1 + exp^{logit(p)}}$$

The statistical significance of MLR3, final merged model, was P-value=0.00 at 95 % confidence level. With reference for the statistics presented in Table 5.17, it is clear that only approach capacity was statistically significant at 95% confidence level. The result suggested that a serious compared with slight casualty for cyclists was 86% more likely (odd ratio=1.86) for one unit increase in approach capacity.

Table 5-17 – The Results of MLR3

Casualty	**Coefficient**	**P ≥|z|**	**Odds Ratio**	**95% Confidence Interval for Odds Ratio P-value= 0.00**	
				Lower	**Upper**
Speed limit	0.02	0.22	1.02	0.99	1.04
Gender of casualty	-0.34	0.38	0.71	0.33	1.53
Age of casualty	0.01	0.38	1.01	0.99	1.03
Lighting	-0.15	0.67	0.86	0.43	1.71
Weather	-0.12	0.76	0.89	0.41	1.91
Road surface condition	-0.43	0.21	0.65	0.33	1.28
Approach capacity	**0.62**	**0.00****	**1.86****	**1.40**	**2.47**
Size of roundabout	-0.29	0.06	0.75	0.55	1.01
Constants	-2.14	0.00	0.12	0.03	0.45

**Statistically significant at 95% confidence level.

Given that only one variable emerged as statistically significant, Simple Logistic Regression was applied including only Approach Capacity in order to determine the predictive margins (See Table 5.18). Considering the minimum and maximum value of Approach Capacity (0.5 to 3.0), the predictive margins were calculated for 47 cases with 0.1 increase in the approach capacity. The predictive margins were the 95% confidence intervals of the probability of serious relative to slight cyclist casualty as a function of the statistically significant variable approach capacity (See Appendix F). The relationship between probability of cyclist serious casualty occurrence at roundabouts and Approach Capacity was found to be logarithmic (See Figure 5.8).

Table 5-18 – Predictive Margins of Approach Capacity

Casualty	**Coefficient**	**P ≥|z|**	**Odds Ratio**	**95% Confidence Interval for Odds Ratio P-value= 0.00**	
				Lower	**Upper**
Approach capacity	**0.55**	**0.00****	**1.74**	**1.36**	**2.22**
Constant	-1.78	0.00	0.18	0.13	0.22

**Statistically significant at 95% confidence level.

$$Logit\ (p) = ln\frac{p}{1-p} = -1.78 + 0.55 * Approach\ capacity$$

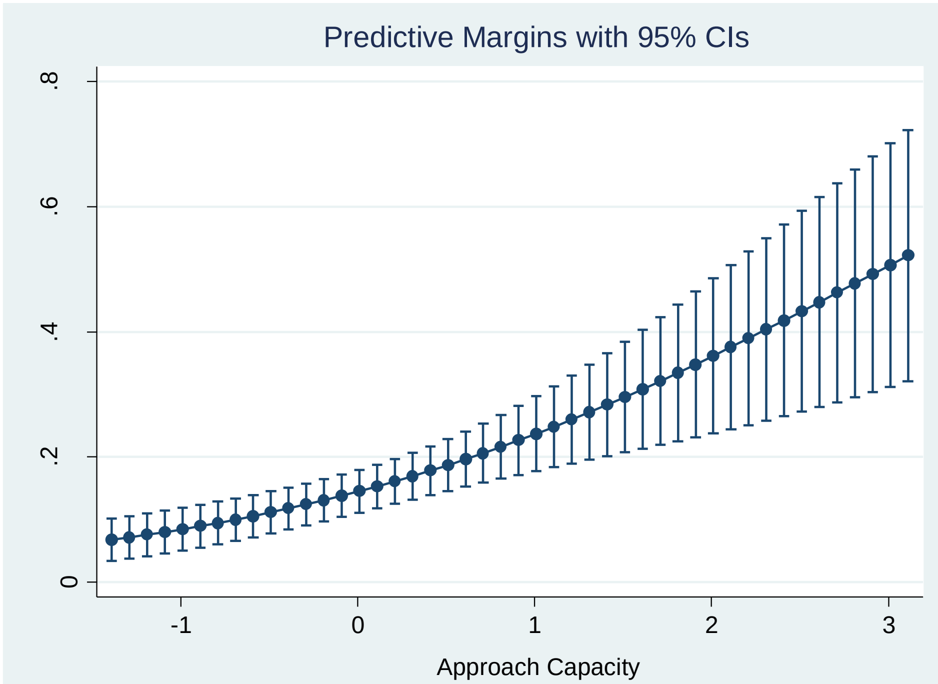


Figure 5.8- Plot of Predictive Probabilities for Each 0.1 Increase in Approach Capacity

Linear Regression

Since Approach Capacity was a continuous variable, it was possible to switch axes in the Linear Regression to create the relationship between Approach Capacity as the dependent variable and casualty severity as the independent predictor. The result is given in Table 5.19 and shows the casualty severity was more likely to be serious rather than slight at roundabouts with an approach capacity higher the 0.6.

Table 5-19 – Linear Regression Model

Casualty	Coefficient	P ≥|z|	95% Confidence Interval		Predictive Margins
			Lower	Upper	
Casualty Severity					
0	Base				-0.09
1	0.60	0.00**	0.35	0.85	0.60
Constants	-0.09	0.07	-0.19	0.01	

**Statistically significant at 95% confidence level.

5.5.2 Traffic Flow Periods Impact

As explained in Chapter 4, the data for traffic flow periods (peak, inter peak, off peak) were available only for Newcastle upon Tyne and Gateshead. Therefore, the analysis with 172 number of observations was conducted to investigate the impact of traffic flow periods on cyclist casualty severity at roundabouts. Descriptive statistics showed that the casualty occurrence gradually increased for slight severity with increase in traffic flow period. However, this increase followed a difference pattern for serious severity occurrence since the number of serious cyclist casualties for both inter peak and peak times were equal (See Figure 5.9). This may be due to the increase in cyclist vehicle kilometres during the day compared to the evening and overnight off-peak period.

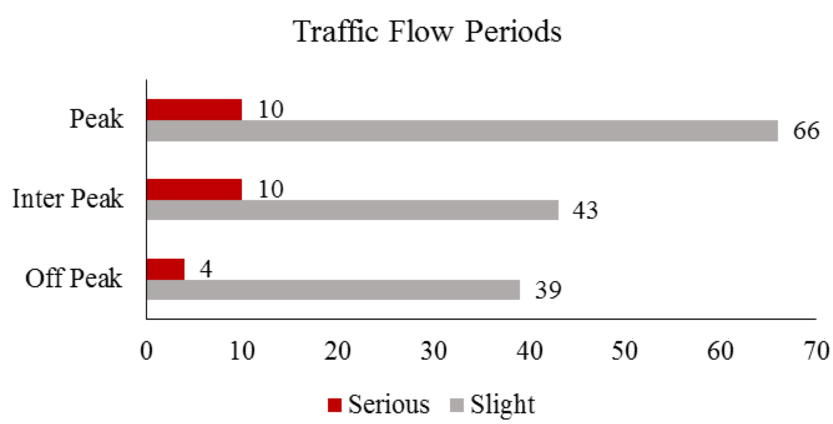


Figure 5.9- The Bar Chart Illustration of Traffic Flow Profile

As a result of the EPV limitation, the MLR was conducted for traffic flow periods along with the significant variables previously observed in the former models (MLR1 and MLR2). Approach capacity was not considered because this variable was derived as a component of PCA of geometric design parameters and the number of lanes on approach and entry path radius have already been included in the analysis by virtue of their statistical significance from MLR analysis.

The results given in Table 5.20 (with P-value= 0.07 at 95% confidence level) suggests that traffic flow periods did not show any statistically significant influence on casualty severity for cyclist at roundabouts. Interestingly, the statistically significant variables from previous models (i.e. number of lanes on approach and speed limit) were not statistically significant in this model. However, entry path radius did emerge at 95% confidence level of statistical significance with a consistent value of odds ratio by 1.04. This result suggested that the impact of entry path radius dominated other variables in the model.

Table 5-20 – Multiple Logistic Regression Model

Variables	Coefficient	P ≥|z|	Odds Ratio	95% Confidence Interval for Odds Ratio P-value= 0.07	
				Lower	**Upper**
Entry path radius	**0.04**	**0.02****	**1.04**	**1.01**	**1.07**
Number of lanes on approach	0.43	0.33	1.54	0.64	3.67
Speed limit	-0.02	0.59	0.98	0.93	1.04
Traffic flow period	0.21	0.46	1.23	0.71	2.16
Constants	-5.25	0.00	0.01	0.00	0.15

**Statistically significant at 95% confidence level.

5.6 Chapter Conclusions

In this chapter, the impact of the variables geometric design parameters, speed limit, sociodemographic characteristics of cyclists and meteorological conditions on cyclist casualty severity at roundabouts was investigated. The correlation analysis showed that there is a strong relationship exists between number of lanes on approach and half width on approach. The relationship between other variables were found uncorrelated.

PCA was used to group the geometric design parameters and two components, namely *Approach Capacity* (derived from number of lanes on approach, half width, number of flare lanes and entry path radius) and *Size of Roundabouts* (created from number of arms, type of roundabout and number of circulating lanes) were identified. The reliability analysis was conducted to investigate the internal consistency between variables within the same group. When the original data was conducted to the analysis, the Cronbach's Alpha for *Size of Roundabouts* was 0.63, which was questionable because it was slightly lower than recommended value of 0.7. On the other hand, the Cronbach's Alpha for *Approach Capacity* was 0.07. This value was far below the recommended value of 0.7, however these values were perfectly grouped in PCA with KMO by 0.73 and Bartlett's test by 0.00. The reason of the initial low Cronbach's Alpha was the absence of the normalisation of the data. When the normalised data of the variables conducted were included in reliability analysis, the Cronbach's Alpha for *Approach Capacity* increased to 0.74. The result of the reliability analysis suggested that the outcomes of the PCA could be used to further the analysis using MLRs.

MLR1 suggested that higher speed limit reduces the safety for cyclists by increasing the casualty severity with odds ratio 1.02. However, this increase is not linear for each increase in speed limit based on predictive margins. Sociodemographic characteristics of cyclists and meteorological conditions did not show any statistically significant impact on severity. MLR2 provided a statistically significant impact of number of lanes on approach and entry path radius with odd ratios 4.97 and 1.04, respectively. Serious casualty of cyclists was more likely by one lane increase on approach arm. A unit increase in entry path radius decreases the safety for cyclists. Linear Regression suggested that if the entry path radius was more than 80 metres, the casualty severity was more likely to be serious. Although, considering the outcomes of correlation analysis number of lanes on approach and half width on approach were statistically significantly correlated, half width on approach did not show impact on casualty severity. MLR3 suggested that higher Approach Capacity increases the probability of serious casualty occurrence by 86% (odds ratio 1.86). Linear Regression suggested that if the Approach Capacity was more than 0.50, the casualty severity was more likely to be serious. With respect to traffic flow profile, there was no observed impact on casualty severity of cyclists at roundabouts.

Finally, the results of MLR3 from which approach capacity emerged as statistically significant was consistent with MLR2 showing that number of lanes on approach and entry path radius were statistically significant and descriptive statistics endorsing that geometric design parameters do influence roundabout safety for cyclists. Speed limit is a proxy of number of lanes on approach and entry path radius and can be used in further studies as a representative of influence of wider approach capacity in situations of lack of knowledge of geometric design parameters in a model. Whilst MLR2 gave more insight into the relative contribution of each geometric design variable to a roundabout safety the changes in the weights when fewer variables were included in the regression model suggests that a greater number of casualty records is required. Therefore, in the next chapter the Northumbria data is increased by extending the study including 21 cities across England.

Chapter 6 Analysis 2: Investigating the Influence of Driver/Rider Error, Reaction, Behaviour and Inexperience on Cyclist Casualty Severity

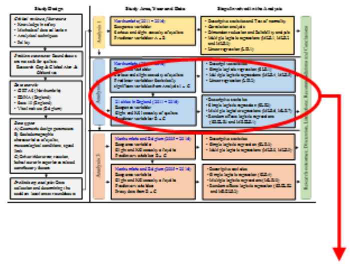
6.1 Introduction

In Chapter 5, descriptive statistics, test of normality, correlation analysis, dimension reduction, reliability analysis and Multiple Logistic Regressions were conducted. The results suggest that wider approach capacity statistically significantly increases the probability of serious cyclist casualty at give way roundabouts with mixed traffic. The results also provide a deeper knowledge that one unit increase in speed limit, number of lanes on approach and entry path radius increase the probability of serious casualty. In addition, speed limit is a proxy of entry geometric design parameters. The results of Chapter 5, encourage extending the analysis including more numbers of data records to address three objectives: i) investigating the driver/rider behaviour related; ii) improving the statistical significance and reducing the influence of sensitivity of the prediction model; iii) exploring the influence of regional variance to observe the consistency of the created model.

Therefore, this chapter aims to investigate the impact of the driver/rider behaviour related predictors on cyclist casualty severity through a case study for Northumbria and increasing the number of casualty records by means of an extended analysis in England (covering areas namely Newcastle upon Tyne, Gateshead, Sunderland, North Tyneside, South Tyneside, Northumberland, Bedford, Blackpool, Brighton & Hove, City of Derby, County Durham, Darlington, Kingston upon Hull, Middlesbrough, Milton Keynes, Portsmouth, Southend on Sea, Stoke on Trent, Stockton on Tees, Warrington and York)

Section 6.2 presents the case study of Northumbria starting with relaxing p-value criterion in order to select the statistically significant variables at 90% confidence level. A full model is shown including the selected variables and finally linear regression is presented to establish whether any recommendations for road design engineers and policy makers emerge from the analysis. Section 6.3 includes an extended study across 21 cities in England and is provided in order to explore the consistency of the developed models. Relaxing p-value criterion and developing a full model including selected variables is presented. In addition, the section includes a Multilevel Logistic Regression analysis to determine the regional variance in the model. Finally, the chapter is concluded in Section 6.4. The steps of this chapter are shown in Figure 6.1.

Methodological Framework (Chapter 3)



Step 1 Case Study of Northumbria	**Step 2** 21 Cities in England
Data (2011-2016):	
• Statistically significant variables from Analysis 1	• Speed limit, Sociodemographic characteristics of cyclist and Environmental conditions
Driver/rider behaviour related contributory factors.	
Analytical Methods:	
• Descriptive statistics • Relaxing p-value criterion (with Simple and Multiple Logistic Regressions) • Predictive margins	
Linear Regression	Multilevel Logistic Regression (null , simple and full models)

Figure 6.1 – Steps Involved in Analysis reported in Chapter 6

6.2 Analysis of Driver/Rider Error, Reaction, Behaviour and Inexperience: A Case Study of Northumbria

The case study area (Northumbria) and the number of cyclist casualties were the same as in Analysis 1, as explained in Chapter 5. With respect to Analysis 1, it was already identified that only a few variables, namely speed limit, number of lanes on approach and entry path radius, emerged as having statistically significant impacts on cyclist casualty severity. Therefore, in terms of geometric design variables, only statistically significant variables, except approach capacity, were used in the analysis in this section. Approach capacity was not considered since it was calculated from the statistical relationship between the variables, namely number of lanes on approach, half width on approach, number of flare lanes on approach and entry path radius by the PCA, and this would lead to the redundancy in the regression model. In summary, the case study for Northumbria was considered including only statistically significant predictor variables as well as driver/rider behaviour related contributory factors.

As explained in Analysis 1 (in Chapter 5), 439 cyclist casualties (370 slight and 69 serious) occurred at 209 roundabouts in Northumbria between 2011 and 2016. The statistically significant variables of speed limit, number of lanes on approach and entry path radius were combined with the contributory factors, namely junction overshoot, junction restart, poor turn or manoeuvre, failed to signal or misleading signal, failed to look properly, failed to judge other person's path or speed, passing too close to cyclist, sudden braking, swerved, loss of control, aggressive driving, careless, reckless or in a hurry, nervous, uncertain or panic, driving too slow for condition or slow vehicle, learner or inexperienced driver/rider, inexperience of driving on left, unfamiliar with model of vehicle. The fact that these contributory factors were not specifically linked to the cyclist or the driver as individuals meant that they were unable to be differentiated between road users involved in the casualties.

6.2.1 Descriptive Statistics of Statistically Significant Variables from Analysis 1

The descriptive statistics for speed limit, number of lanes on approach and entry path radius (See Section 5.2) are briefly summarised in Table 6.1. Speed limit, number of lanes on approach and entry path radius are not normally distributed.

Table 6-1- Descriptive Statistics for Speed Limit, Number of Lanes on Approach and Entry Path Radius

Variable (abbreviation)	**Severity of Casualty**	**Descriptive Statistics**
Speed limit (20; 30; 40; 50; 60; 70)	Slight	20 (3); 30 (280); 40 (33); 50 (12); 60 (33); 70 (9)
	Serious	20 (2); 30 (43); 40 (8); 50 (1); 60 (9); 70 (6)
Number of lanes on approach (1; 2; 3)	Slight	1 (274); 2 (90); 3 (6)
	Serious	1 (32); 2 (36); 3 (1)
Entry path radius (metre)	Slight	Min.= 19.23; Max.= 99.83; Median= 65.63
	Serious	Min.= 23.77; Max= 99.98; Median= 90.37

6.2.2 Descriptive Statistics of Driver/Rider Error or Reaction Related Contributory Factors

Table 6.2 provides the descriptive statistics in terms of the frequency of mention of contributory factors related to driver/rider error or reaction separately for slight and serious casualty records for cyclists at roundabouts. By far *failed to look properly* was the highest contributory factor noted in 78.6% of slight and in 84.1% serious casualty severity records.

Table 6-2 – Descriptive Statistics for Driver/Rider Error or Reaction Related Contributory Factors

Variable	Slight			Serious		
	Yes	No	%	Yes	No	%
Junction overshoot	16	354	4.3	6	63	8.7
Junction restart	10	360	2.7	4	65	5.8
Poor turn or manoeuvre	45	325	12.2	6	63	8.7
Failed to signal or misleading signal	9	361	2.4	2	67	2.9
Failed to look properly	291	79	78.6	58	11	84.1
Failed to judge other person's path or speed	91	279	24.6	10	59	14.5
Passing too close to cyclist	49	321	13.2	9	60	13.0
Sudden braking	7	363	1.9	6	63	8.7
Swerved	0	370	0.0	2	67	0.0
Loss of control	1	369	0.3	2	67	2.9

Regarding the slight casualties, *failed to judge other person's path or speed*, *passing too close to cyclist* and *poor turn or manoeuvre* contributed moderately with 24.6%, 13.2% and 12.2%, respectively. On the other hand, a similar trend of contribution was observed for serious casualties for *failed to judge other person's path or speed*, *passing too close to cyclist* and *poor turn or manoeuvre* with 14.5%, 13.0% and 8.7%, respectively.

However, the other of the variables (*loss of control, sudden braking, failed to signal or misleading signal, junction restart* and *junction overshoot*) were found to have a fewer number of records. There has been no record of swerved in slight casualty occurrence and it should not be included in further models as a predictive variable. In addition, the variables, which have few contributions in either slight or serious casualties or both together, should be interpreted regarding the 95% confidence interval with regard to the results of regression models.

6.2.3 Descriptive Statistics for Driver/Rider Behaviour or Inexperience

The results of descriptive analysis are presented in Table 6.3. Only one variable, namely *careless, reckless or in a hurry,* had a moderate contribution of 23.0% and 20.3% for *slight* and *serious* casualties, respectively. On the other hand, the variables *driving too slow for condition or slow vehicle, learner or inexperienced driver/rider*, *inexperience of driving on left,* and *unfamiliar with model of vehicle* did not have any record, therefore these variables should be excluded in further parametric models. The absence of these variables in this data set does not mean they would have no statistically significant influence in a larger data set. The impact of *aggressive driving* and *nervous, uncertain or panic* should be examined regarding the 95% confidence interval in regression models.

Table 6-3 – Descriptive Statistics of Driver/Rider Behaviour or Inexperience

Variable	**Slight**			**Serious**		
	Yes	**No**	**%**	**Yes**	**No**	**%**
Aggressive driving	2	368	0.5	1	68	1.4
Careless, reckless or in a hurry	85	285	23.0	14	55	20.3
Nervous, uncertain or panic	4	366	1.1	1	68	1.4
Driving too slow for condition or slow vehicle	0	370	0.0	0	69	0.0
Learner or inexperienced driver/rider	2	368	0.5	0	69	0.0
Inexperience of driving on left	0	370	0.0	0	69	0.0
Unfamiliar with model of vehicle	0	370	0.0	0	69	0.0

6.2.4 Simple and Multiple Logistic Regression Models

The response variable in the Logistic Regression models is casualty severity (slight and serious). The predictors were the variables that emerged as statistically significant in Chapter 5 (speed limit, number of lanes on approach and entry path radius) and driver/rider behaviour related contributory factors (junction overshoot, junction restart, poor turn or manoeuvre, failed to signal or misleading signal, failed to look properly, failed to judge other person’s path or speed, passing too close to cyclist, sudden braking, loss of control, aggressive driving, careless, reckless or in a hurry and nervous, uncertain or panic).

Simple Logistic Regression (SLR1) used to test each of the independent variables in isolation was followed by Multiple Logistic Regression (MLR4) using all variables. The results are presented in Table 6.4 for SLR1 and MLR4. The variables speed limit, number of lanes on approach, entry path radius, sudden braking and loss of control emerged as statistically significant at 95% confidence level from both SLR1 and MLR4 with odds ratios 1.02, 2.64, 1.04, 4.94 and 11.01, respectively. The coefficients for SLR1 were the same (speed limit, entry path radius), higher (number of lanes on approach), or lower (sudden braking, loss of control) compared to MLR4, but the same five variables emerged as being statistically significant at 95% confidence level. The results from the two approaches are consistent differing only in the magnitude of the coefficients and the odds ratios.

Striking results include for a unit change in sudden braking is 5 and 7 times increase in probability of serious compared to slight casualty severity for SLR1 and MLR4, respectively. The similar odds ratios for loss of control are 11 and 18 times. These behavioural related variables are far more influential than the geometrical related variables of number of lanes on approach (3 and 2 times for SLR1 and MLR4), speed limit 1.02 and entry path radius 1.04 for both SLR1 and MLR4. The previous results (See Chapter 5) showed that the influence of speed limit diminished when geometrical parameters were included in the model. However, MLR4 showed that these three variables were statistically significant together in the same model. This situation suggests that the influence of speed limit was supported by contributory factors and the model is sensitive. This is emerging in a further full model including a higher number of observations.

The descriptive statistics analysis suggested that the number of observations for each predictor might influence the outcome of regression models evident for the variables *sudden braking* and *loss of control* given the wide range in the 95% confidence intervals. Conversely, the 95% confidence interval values for the other variables (*speed limit, number of lanes on approach* and *entry path radius*) were narrow.

The number of Events per Variable (EPV) were highlighted in Chapter 2 as a limitation in the logistic regression modelling and Peduzzi *et al.* (1996) recommended that 10 EPV should be considered to avoid biased results. Given that in this step of the analysis the number of serious casualties was 69, a maximum of seven variables could be considered to develop a reliable model. Therefore, only the statistically significant variables (*speed limit*, *number of lanes on approach, entry path radius, failed to look properly, failed to judge other person's path or speed, sudden braking* and *loss of control*) at 90% confidence level or better were selected by using relaxing p value (SLR1 and MLR4) to create MLR5 (See Table 6.5).

Table 6-4- Relaxing P-value Criteria by Simple Logistic Regression (SLR1) and Multiple Logistic Regression (MLR4) for Variable Selection

Variable name	**SLR1**					**MLR4**				
	Coefficient	P- Value	Odds ratio	95% confidence interval for odds ratio Lower	Upper	Coefficient	P- Value	Odds ratio	95% confidence interval for odds ratio Lower	Upper
Speed limit	**0.02**	**0.02****	**1.02**	**1.00**	**1.04**	**0.02**	**0.05****	**1.02**	**1.00**	**1.05**
Number of lanes on approach	**0.97**	**0.00****	**2.64**	**1.65**	**4.22**	**0.63**	**0.03****	**1.89**	**1.07**	**3.31**
Entry path radius	**0.04**	**0.00****	**1.04**	**1.03**	**1.06**	**0.04**	**0.00****	**1.04**	**1.03**	**1.06**
Junction overshoot	0.75	0.13	2.11	0.79	5.59	0.62	0.33	1.85	0.53	6.44
Junction restart	0.79	0.19	2.22	0.67	7.27	1.02	0.14	2.79	0.72	10.75
Poor turn or manoeuvre	-0.37	0.41	0.69	0.28	1.68	-0.62	0.28	0.54	0.17	1.67
Failed to signal or misleading signal	0.18	0.82	1.19	0.25	5.66	0.01	0.99	1.01	0.15	6.65
Failed to look properly	0.36	0.31	1.43	0.72	2.86	0.76	0.08*	2.13	0.92	4.97
Failed to judge other person's path or speed	-0.65	0.07*	0.52	0.25	1.06	-0.73	0.09*	0.48	0.21	1.11
Passing too close to cyclist	-0.02	0.96	0.98	0.46	2.11	0.43	0.35	1.53	0.62	3.77
Sudden braking	**1.60**	**0.01****	**4.94**	**1.61**	**15.18**	**1.90**	**0.01****	**6.71**	**1.71**	**26.37**
Loss of control	**2.40**	**0.05****	**11.01**	**0.98**	**123.20**	**2.90**	**0.04****	**18.11**	**1.09**	**300.29**
Aggressive driving	1.00	0.42	2.71	0.24	30.26	0.21	0.12	9.10	0.55	150.68
Careless, reckless or in a hurry	-0.16	0.63	0.85	0.45	1.61	-0.27	0.49	0.76	0.36	1.63
Nervous, uncertain or panic	0.30	0.79	1.35	0.15	12.22	1.25	0.33	3.48	0.28	43.09
Constant	---	---	---	---	---	-7.25	0.00	0.00	0.00	0.00

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

Table 6-5 – Multiple Logistic Regression (MLR5) Including Statistically Significant Variables

Variable name	**MLR5 (P-value= 0.00)**				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**	
				Lower	**Upper**
Speed limit	**0.02**	**0.05****	**1.02**	**1.00**	**1.05**
Number of lanes on approach	**0.67**	**0.02****	**1.95**	**1.12**	**3.37**
Entry path radius	**0.04**	**0.00****	**1.04**	**1.03**	**1.06**
Failed to look properly	0.71	0.08*	2.05	0.92	4.59
Failed to judge other person's path or speed	-0.62	0.12	0.54	0.25	1.18
Sudden braking	**1.97**	**0.00****	**7.14**	**1.96**	**26.06**
Loss of control	2.25	0.09*	9.44	0.72	123.26
Constant	-7.04	0.00	0.00	0.00	0.01

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

The outcome of MLR5 (with P-value= 0.00 at 95% confidence level) showed that *speed limit, number of lanes on approach, entry path radius* and *sudden braking* emerged at 95% statistical confidence, whilst loss of control was at 90%. The model equation shows that for each one-unit increase of *speed limit* (10 mph) a serious is 2% more likely to occur than slight casualty (odds ratio 1.02). In addition, for each additional lane on approach and each unit increase in *entry path radius* the likelihood of serious is 95% and 4% more likely than slight casualty, with odds ratios of 1.95 and 1.04 respectively. By far the biggest contribution is *sudden braking* with 7 times more likely to be a serious than slight casualty. However, it should be pointed out that the 95% confidence interval was very wide because the number of records was very low.

A further step eliminates the variables which were not statistically significant at 95% confidence level and observes the sensitivity by creating a full model. MLR6 (with P-value=0.00) given in Table 6.6 shows that statistically significant predictors at 95% confidence level were number of lanes on approach, entry path radius and sudden braking. In addition, the coefficients of these three predictors did not change remarkably compared to MLR5. However, the influence of speed limit declined. This result suggests that failed to look properly, failed to judge other person's path and loss of control supported the influence of speed limit. However, these contributory factors were not influential directly in the model. This result suggests that the model is still sensitive, and it emerges that a further analysis including more casualty records should be conducted.

Table 6-6 – Multiple Logistic Regression 6 (MLR6)

Variable name	**MLR6 (P-value= 0.00)**				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**	
				Lower	**Upper**
Speed limit	0.02	0.09*	1.02	1.00	1.04
Number of lanes on approach	**0.69**	**0.01****	**2.00**	**1.16**	**3.44**
Entry path radius	**0.04**	**0.00****	**1.04**	**1.03**	**1.06**
Sudden braking	**1.92**	**0.00****	**6.85**	**1.92**	**24.49**
Constant	-6.44	0.00	0.00	0.00	0.01

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

The equation of MLR6 is as follow:

$$Logit\ (p) = ln\frac{p}{1-p}$$
$$= -6.44 + 0.02 * Speed\ limit + 0.69 * Number\ of\ lanes\ on\ approach$$
$$+ 0.04 * Entry\ path\ radius + 1.92 * Sudden\ braking$$

Where the predicted probability (predictive margin):

$$p = \frac{exp^{logit(p)}}{1 + exp^{logit(p)}}$$

p = Probability of Serious Casualty

1-*p* = Probability of Slight Casualty

Finally, the model was reversed to determine the limit of safe value for entry path radius. In the reversed model, the response variable was entry path radius and the predictive variable was casualty severity; therefore, Liner Regression was conducted. The predictive margin of Linear Regression is given in Table 6.7 which shows that if the entry path radius was equal or more than 80 metres, the casualty severity was more likely to be *serious* than *slight*. On the other hand, regarding the entry path radius with equal or less than 64 metres, the casualty severity was more likely to be slight than serious.

Table 6-7 – Linear Regression

Casualty	Coefficient	P ≥|z|	95% Confidence Interval		Predictive Margins
			Lower	Upper	
Casualty Severity					
0	Base				64.36
1	16.39	0.00**	11.09	21.68	80.75
Constants	64.36	0.00	62.26	66.46	

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

The results of the case study of Northumbria suggested that the variables of speed limit, number of lanes on approach and entry path radius remained strong in their influence on casualty severity. In addition, sudden braking as a contributory factor showed a statistically significant impact on severity. However, under further scrutiny other variables emerged as statistically significant even though at a lower confidence. In addition, the coefficients and odds ratios were unstable. This endorses the finding of Chapter 5 that more crash records are needed, extending the study to include roundabouts beyond Northumbria to embrace areas across England, the transferability of the analytical approach and to identify similarities and differences in casualty risk. The following section reports the results of the extended analysis of driver/rider behaviour influence on cyclist casualty severity in crashes at roundabouts across England.

6.3 Extended Analysis of Driver/Rider Behaviour: A Case Study England

The extended analysis was carried out across England including the areas namely *Newcastle upon Tyne, Gateshead, Sunderland, North Tyneside, South Tyneside, Northumberland, Bedford, Blackpool, Brighton & Hove, City of Derby, County Durham, Darlington, Kingston upon Hull, Middlesbrough, Milton Keynes, Portsmouth, Southend, Stoke on Trent, Stockton on Tees, Warrington* and *York*. The casualty data was collected for between 2011 and 2016. 1394 *slight*, 284 *serious* and 2 *fatal* cyclist casualties occurred. Since the number of *fatal* casualties is low at 2, these were merged with the serious casualties and named as *Killed and Seriously Injured (KSI)* to meet the constraints of Logistic Regression. Therefore, the casualty severity variable consisted of 1394 *slight* and 286 *KSI* casualty.

The distribution of number of casualties for each year between 2011 and 2016 for slight and KSI crashes is shown in Figure 6.2. The number of slight casualties gradually increased from 2011 until 2014 when they then fall sharply. A similar trend was not observed for KSI casualties: whilst they increased up until 2015, there was a fall in 2016.

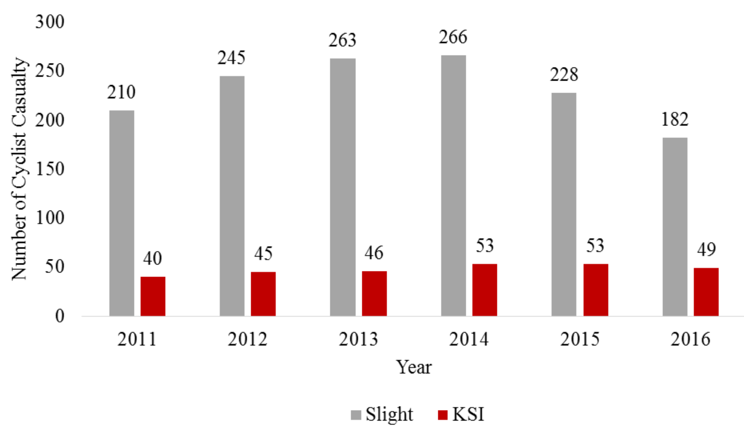


Figure 6.2 – Number of Slight and KSI Casualty between 2011 and 2016

Cyclist count in each region for the given period 2011-2016 was made available in DfT (2018) (See Figure 6.3). The number of casualties varies substantially for the areas considered due to many reasons: population densities, number of cyclists and roundabouts, traffic flows, road type and kilometre length of road, topography and weather. In addition, the variation across regions is huge for numbers of both number of roundabouts and cycle counts. Therefore, in order to gain an appreciation of the influence of number of roundabouts and cyclist count the slight and KSI were normalised and the results are presented in Figure 6.4 and 6.5, respectively. It was observed that not only cyclist counts but also cycling mileage might have an influence; however, this data was not available and is a limitation in this study.

The total number of roundabouts in each city/region is unknown. Therefore, the total number of roundabouts at which a casualty occurred is used to explore whether roundabouts are less safe in one city/region compared to another. A particular observation is that Kingston upon Hull has approximately 3 times more serious casualties than Brighton and Hove. However, the serious casualties per roundabout where a crash occurred is lower. This suggests that particular roundabouts in Brighton and Hove are more dangerous than in Kingston upon Hull. Regarding the normalised number of slight and KSI casualties by cyclist counts, Milton Keynes has a remarkably higher value than any other city/region. This suggests that Milton Keynes, given such low number of cyclists, emerges as the least safe.

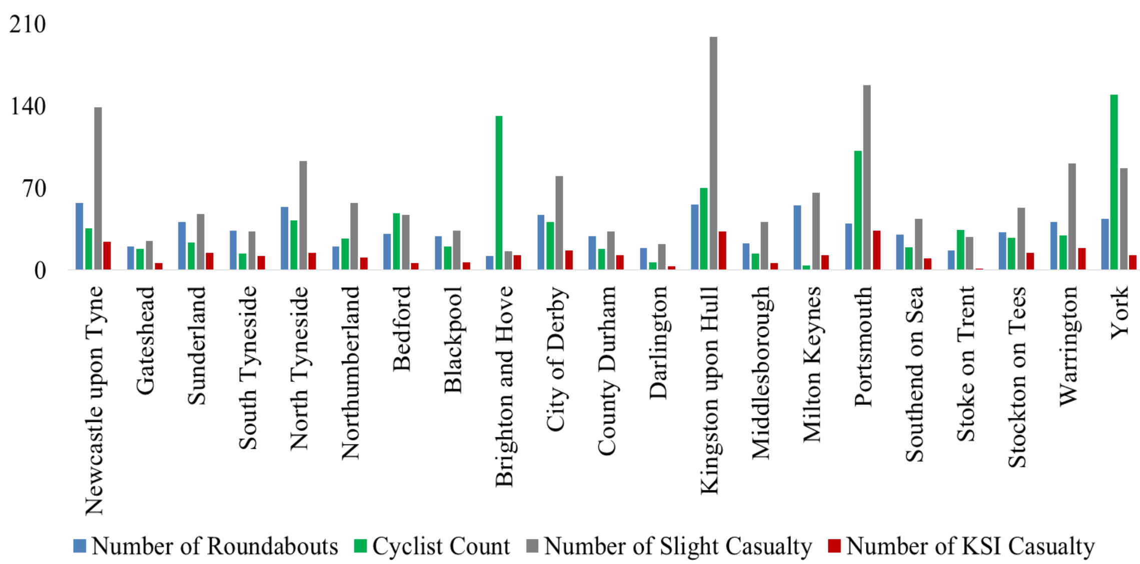


Figure 6.3 – Number of Roundabouts at Which a Cyclist Casualty Occurred, Cyclist Count (per 1000) Slight and KSI Casualty in Each Local Authority

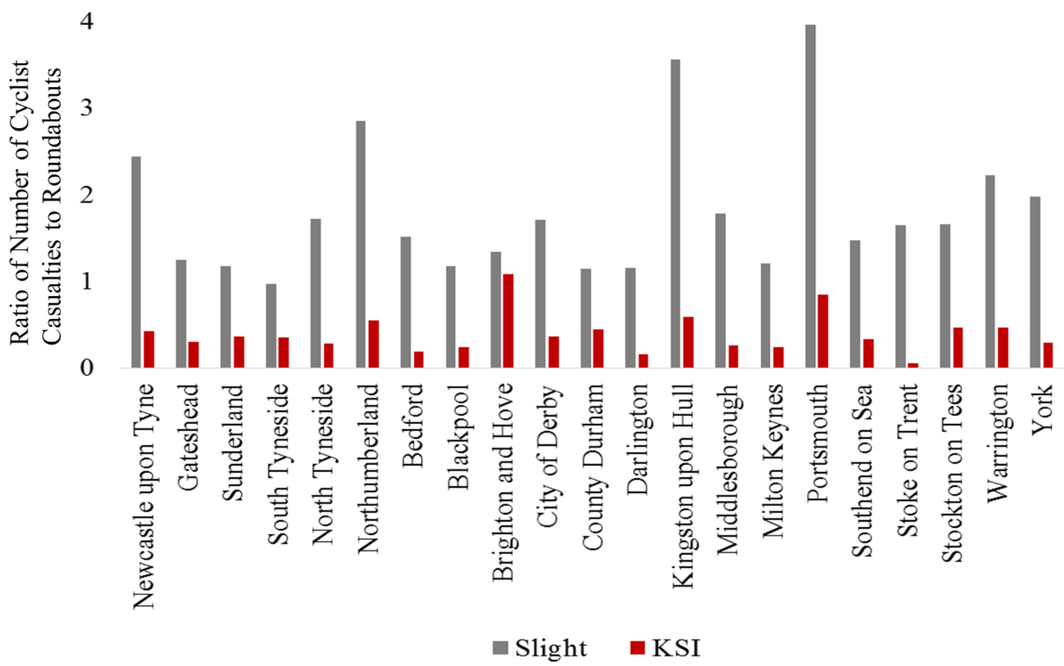


Figure 6.4 – Ratio of Number of Cyclist Casualties to Roundabouts for each Authority

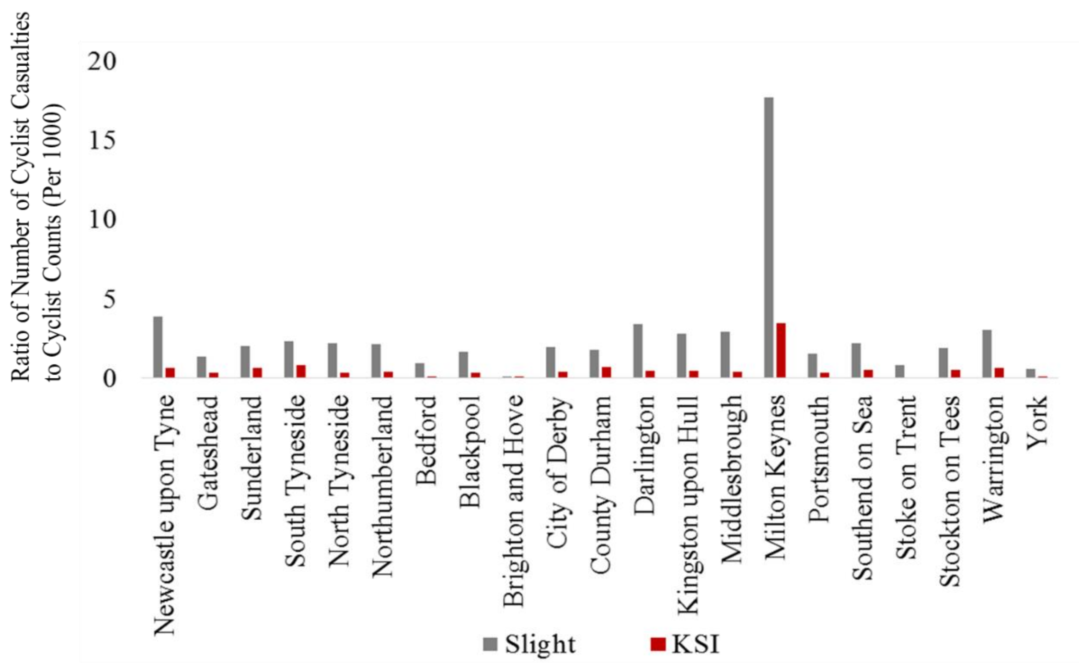


Figure 6.5 – Ratio of Number of Cyclist Casualties to Cyclist Counts per 1000 for each Authority

6.3.1 Descriptive Statistics of Speed Limit, Sociodemographic Characteristics of Cyclist and Meteorological Conditions, Driver/Rider Behaviour Related Contributory Factors

The casualty data for across England did not include the crash details (the direction of driving) (See Chapter 4); therefore, geometric design parameters were unable to be included in the analysis. However, results (See Chapter 5) showed that speed limit can be used as a proxy for geometric parameters for entry and approach. Therefore, the variables considered in the analysis were speed limit, sociodemographic characteristics of cyclist, meteorological conditions and driver/rider error, reaction, behaviour and inexperience.

The descriptive statistics were derived for all the variables, namely *speed limit*, *cyclist age group*, *cyclist gender*, *lighting*, *weather*, *road surface condition*, *junction overshoot*, *junction restart*, *poor turn or manoeuvre*, *failed to signal or misleading signal*, *failed to look properly*, *failed to judge other person's path or speed*, *passing too close to cyclist*, *sudden braking*, *swerved*, *loss of control*, *aggressive driving*, *careless, reckless or in a hurry*, *nervous, uncertain or panic*, *driving too slow for condition or slow vehicle*, *learner or inexperienced driver/rider*, *inexperience of driving on left* and *unfamiliar with model of vehicle*.

Speed Limit

Table 6.8 provides the summary statistics for speed limit. The highest number of records for both slight and KSI casualty were recorded when the speed limit was 30 mph with 1102 and 205, respectively. This showed that 72% of the KSI casualties occurred at roundabouts with a 30 mph speed limit. This was followed by 40 mph and 60 mph with 11% and 10 % KSI casualties, respectively. This result in part reflects the fact that there are likely to be fewer cyclists riding on roads with higher speed limits.

Table 6-8 – Number of Slight and KSI Casualty at Each Speed Limit

Severity of Casualty	Speed Limit (mph)					
	20	30	40	50	60	70
Number of Slight	20	1102	132	26	84	30
Number of KSI	6	205	32	2	29	12

Sociodemographic Characteristics of Cyclist

With regard to cyclist gender, 60 female and 226 male KSI casualties were recorded, with a similar trend for slight casualty occurrence with 335 female and 1039 male of all cyclist casualties, 62% were slight and male. Since the gender split was not known, parametric analysis will be only based on the number of observations. Regarding the cyclist age group, it is seen that the number of slight casualties sharply increases with age until age group 6 (26-35 years) then gradually decreases to age group 8 (46-55 years) when the number falls dramatically for age groups to 11 (76-101 years) (See Figure 6.6). Repeatedly, the cyclist volume for each age group was not included in the dataset, therefore it was not possible to interpret the sharp decrease after age group 8. With respect to KSI casualties, the gradual increase continued until age group 8 and a decrease occurred beyond. In summary, majority of the casualties were recorded between age groups 6 and 8.

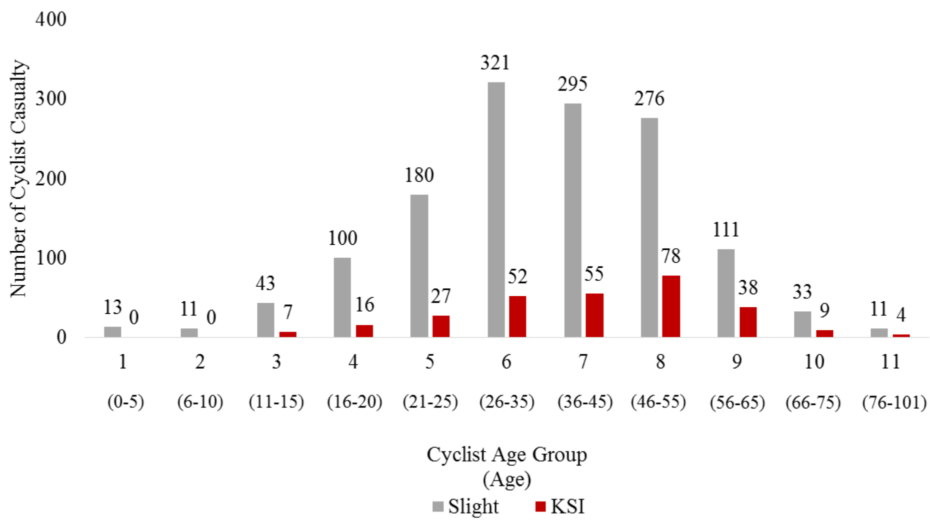


Figure 6.6 – Descriptive Statistics for Cyclist Age Group

Lighting, Weather and Road Surface Condition

Regarding the lighting level, most of the casualties (1039 slight and 213 KSI) occurred in daylight time. The number of casualties that occurred in darkness was 355 for slight and 73 for KSI. A high number of slight and KSI casualties occurred when the weather was fine with 1126 and 246, respectively (See Figure 6.7), followed by rain or snow (special) with 222 slight and 35 KSI. 21 casualties occurred in unknown weather conditions.

As can be seen from Figure 6.7, the majority of casualties occurred when the road surface condition was either dry or wet/damp. 927 slight and 208 KSI casualties occurred in dry road surface conditions, while 432 slight and 73 KSI casualties recorded when the road was wet or damp.

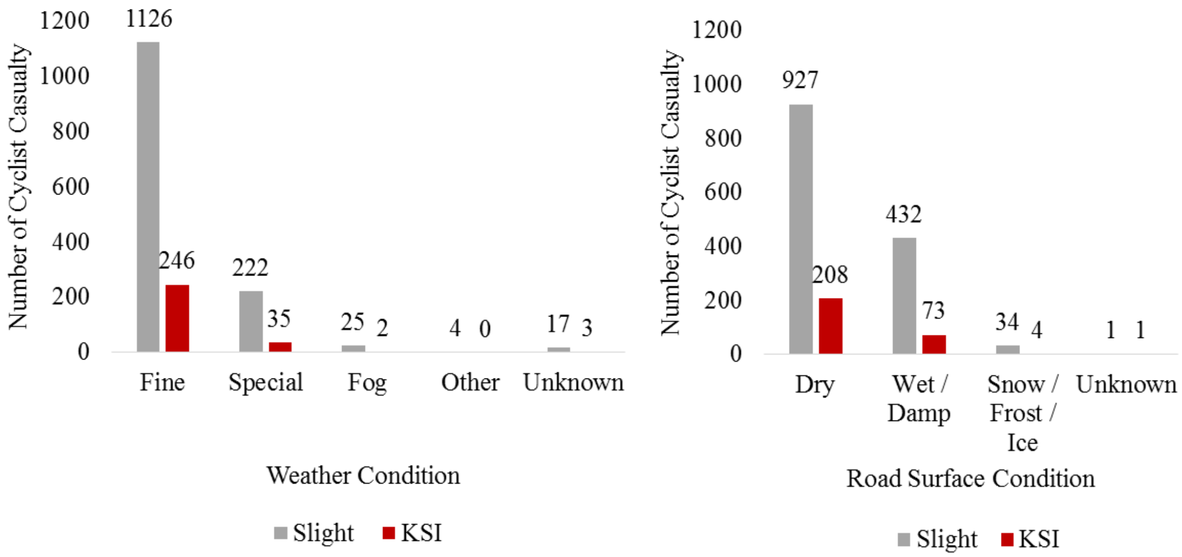


Figure 6.7 – Descriptive Statistics of Weather and Road Surface Conditions

Driver/Rider Error or Reaction Related Contributory Factors

Driver/rider behaviour related contributory factors were divided into two groups. The first group called driver/rider error or reaction is presented in Table 6.9. Failed to look was most relevant for both slight and KSI casualties by 912 and 233, respectively. In other words, failed to look involved 68% of the cyclist casualties that occurred at roundabouts. This was followed by failed to judge other person's path or speed, poor turn or manoeuvre and passing too close to cyclist. There was no KSI casualty record for swerved; therefore, this variable was excluded from regression analysis.

Table 6-9 – Number of Slight and KSI Casualties for each Driver/Rider Error or Reaction Related Variables

Variable	Slight			KSI		
	Yes	No	%	Yes	No	%
Junction overshoot	39	1355	2.8	11	275	3.8
Junction restart	34	1360	2.4	15	271	5.2
Poor turn or manoeuvre	88	1306	6.3	28	258	9.8
Failed to signal or misleading signal	27	1367	1.9	10	276	3.5
Failed to look properly	912	482	65.4	233	53	81.5
Failed to judge other person's path or speed	283	1111	20.3	90	196	31.5
Passing too close to cyclist	99	1295	7.1	16	270	5.6
Sudden braking	12	1382	0.9	24	262	8.4
Swerved	2	1392	0.1	0	286	0.0
Loss of control	9	1385	0.6	4	282	1.4

Driver/Rider Behaviour or Inexperience Related Contributory Factors

The second group of contributory factors (driver/rider behaviour or inexperience), shown in Table 6.10, were not often included in cyclist casualties occurring at roundabouts, with *careless, reckless or in hurry* (214 *slight* and 42 *KSI*) only worthy of note. *Inexperience of driving on left* was excluded from predictive variables in further analysis because KSI casualties were absent.

The other variables (*aggressive driving, nervous, uncertain or panic, driving too slow for condition or slow vehicle, learner or inexperienced driver/rider* and *unfamiliar with model of vehicle*) also had very low casualty records. These variables were included in the regression analysis, but care should be taken with respect to 95% confidence interval values when interpreting the results.

Previous logistic regression analysis (Peduzzi et al., 1996) recommended 10 EPV are considered necessary to obtain a reliable model. In this step 2 of the analysis, the number of EPV (KSI casualty) was 286; therefore, a full model including up to 28 variables in the same model was permitted. However, including many variables which were not statistically significant in a full model might reduce its statistical power overall. Therefore, the method of relaxation p-value criterion when conducting univariate and multivariate logistic regression, as recommended (Sperandei, 2013), was adopted. The analysis, namely Simple Logistic Regression (SLR), Multiple

Logistic Regression (MLR) and finally the full model of MLR including selected variables, is applied and the results are presented in the following subsection.

Table 6-10 – Number of Slight and KSI Casualties for each Driver/Rider Behaviour or Inexperience Related Variables

Variable	**Slight**			**Serious**		
	Yes	**No**	**%**	**Yes**	**No**	**%**
Aggressive driving	7	1387	0.5	4	282	1.4
Careless, reckless or in a hurry	214	1180	15.4	42	244	14.7
Nervous, uncertain or panic	10	1384	0.7	1	285	0.3
Driving too slow for condition or slow vehicle	5	1389	0.4	1	285	0.3
Learner or inexperienced driver/rider	9	1385	0.6	2	284	0.7
Inexperience of driving on left	1	1393	0.1	0	286	0.0
Unfamiliar with model of vehicle	1	1393	0.1	1	285	0.3

6.3.2 Simple and Multiple Logistic Regression Models

The results obtained from SLR2 and MLR7 are summarised in Table 6.11. For SLR2, seven variables (*speed limit, cyclist age group, junction restart, poor turn or manoeuvre, failed to look properly, failed to judge other person's path or speed* and *sudden braking*) were statistically significant at 95% level of confidence with odds ratios of 1.02, 1.20, 2.20, 1.60, 2.34, 1.83 and 10.50, respectively. The coefficients for each variable were positive and this indicated that one unit increase in value of each variable reduces the cyclist safety with the probability of KSI casualty occurrence increasing by 2%, 20%, two times, 60%, two times, 83% and 11 times respectively for *speed limit*, *cyclist age group*, *junction restart, poor turn or manoeuvre, failed to look properly*, *failed to judge other person's path or speed* and *sudden braking*.

The multiple modelling, MLR7 produced similar results to SLR2 with the same variables, namely speed limit, cyclist age group, junction restart, poor turn or manoeuvre, failed to look properly, failed to judge other person's path or speed and sudden braking, emerging as statistically significant at 95% confidence level. Interestingly, the odds ratios of the statistically significant variables in MLR7 were very close to the odds ratio values in SLR2.

Table 6-11- Simple Logistic Regression (SLR2) and Multiple Logistic Regression (MLR7)

Variable name	**SLR2**					**MLR7**				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**		**Coefficient**	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**	
				Lower	**Upper**				**Lower**	**Upper**
Speed limit	**0.02**	**0.00****	**1.02**	**1.01**	**1.03**	**0.02**	**0.01****	**1.02**	**1.00**	**1.03**
Cyclist gender	-0.24	0.13	0.78	0.57	1.08	-0.32	0.06*	0.73	0.52	1.02
Cyclist age group	**0.18**	**0.00****	**1.20**	**1.11**	**1.29**	**0.19**	**0.00****	**1.21**	**1.11**	**1.32**
Lighting	0.00	0.99	1.00	0.75	1.34	-0.24	0.15	0.79	0.57	1.09
Weather	-0.23	0.08*	0.80	0.62	1.02	-0.17	0.28	0.84	0.62	1.15
Road surface condition	-0.25	0.06*	0.78	0.61	1.01	-0.29	0.07*	0.75	0.54	1.03
Junction overshoot	0.33	0.35	1.38	0.70	2.74	-0.05	0.90	0.95	0.44	2.06
Junction restart	**0.79**	**0.01****	**2.20**	**1.18**	**4.10**	**0.67**	**0.04****	**1.96**	**1.03**	**3.75**
Poor turn or manoeuvre	**0.47**	**0.04****	**1.60**	**1.03**	**2.50**	**0.45**	**0.06****	**1.56**	**1.00**	**2.50**
Failed to signal or misleading signal	0.60	0.11	1.83	0.87	3.81	0.50	0.22	1.65	0.74	3.69
Failed to look properly	**0.85**	**0.00****	**2.34**	**1.70**	**3.21**	**0.89**	**0.00****	**2.44**	**1.74**	**3.44**
Failed to judge other person’s path or speed	**0.60**	**0.00****	**1.83**	**1.38**	**2.42**	**0.52**	**0.00****	**1.69**	**1.24**	**2.29**
Passing too close to cyclist	-0.26	0.35	0.77	0.45	1.33	**-0.61**	**0.05****	**0.54**	**0.30**	**1.00**
Sudden braking	**2.35**	**0.00****	**10.50**	**5.19**	**21.26**	**2.53**	**0.00****	**12.67**	**5.92**	**27.14**
Loss of control	0.78	0.20	2.17	0.66	7.11	0.59	0.39	1.81	0.47	6.94
Aggressive driving	1.03	0.10*	2.80	0.81	9.62	1.12	0.10*	3.05	0.82	11.34
Careless, reckless or in a hurry	-0.06	0.76	0.94	0.66	1.35	-0.07	0.72	0.93	0.64	1.37
Nervous, uncertain or panic	-0.73	0.50	0.48	0.06	3.79	-0.53	0.62	0.59	0.07	4.80
Driving too slow for condition or slow vehicle	-0.03	0.98	0.97	0.11	8.34	0.30	0.80	1.35	0.14	13.33
Learner or inexperienced driver/rider	0.08	0.92	1.08	0.23	5.02	0.00	0.99	1.00	0.19	5.17
Unfamiliar with model of vehicle	1.58	0.26	4.87	0.30	78.04	2.12	0.15	8.30	0.46	149.93
Constant						-3.85	0.00	0.02	0.01	0.05

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

This result suggests that these variables (speed limit, cyclist age group, junction restart, poor turn or manoeuvre, failed to look properly, failed to judge other person's path or speed and sudden braking) have a dominant impact on cyclist casualty severity in crashes occurring at roundabouts. The larger number of data records is providing stability and consistency in the results. However, an additional variable, namely passing too close to cyclist, emerged as being statistically significance at 95% confidence level in MLR7. The p-value of the model was 0.00, which indicated that the MLR7 was statistically significantly better compared to a null model.

Relaxing the statistical significance to a 90% confidence level for both SLR and MLR, the final full model of MLR8 was reworked (See Table 6.12). Again, speed limit, cyclist age group, junction restart, failed to look properly, failed to judge other person's path or speed and sudden braking remained statistically significant at 95% confidence level; however, the statistical significance of poor turn or manoeuvre dropped in statistical significance to 90% confidence level.

A 10 mph increase in speed limit was 2% more likely to be a KSI casualty with odds ratio 1.02. The odds ratio of cyclist age group by 1.20 suggests that every unit increase in age group increase the probability of KSI casualty occurrence by 20% compared to slight casualty. The cyclist casualties which included junction restart were almost twice more likely to be KSI than slight casualties. Failed to look properly and failed to judge other person's path or speed increase the probability of KSI casualty occurrence by odds ratios 2.41 and 1.70, respectively. The influence of sudden braking on cyclist casualty was similar to the previous analysis (MLR7) with a very high odds ratio of 12.83, suggesting a KSI compared to slight was almost 13 times more likely although with a wide range of 95% confidence interval. Although the confidence interval was very wide, the lower value was still over 6 times, suggesting that the outcome for sudden braking is statistically significant and reliable.

Table 6-12 – Multiple Logistic Regression 8 (MLR8)

Variable name	**MLR8 (P-value= 0.00)**				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio** **Lower**	**Upper**
Speed limit	**0.02**	**0.01****	**1.02**	**1.00**	**1.03**
Cyclist gender	-0.31	0.07*	0.74	0.53	1.03
Cyclist age group	**0.18**	**0.00****	**1.20**	**1.11**	**1.30**
Weather	-0.18	0.26	0.84	0.62	1.14
Road surface condition	-0.26	0.11	0.77	0.57	1.06
Junction restart	**0.67**	**0.04****	**1.94**	**1.02**	**3.72**
Poor turn or manoeuvre	0.42	0.07*	1.53	0.96	2.45
Failed to look properly	**0.88**	**0.00****	**2.41**	**1.71**	**3.38**
Failed to judge other person's path or speed	**0.53**	**0.00****	**1.70**	**1.26**	**2.30**
Passing too close to cyclist	-0.55	0.06*	0.57	0.32	1.03
Sudden braking	**2.55**	**0.00****	**12.83**	**6.05**	**27.20**
Aggressive driving	1.06	0.11	2.88	0.78	10.58
Constant	-3.93	0.00**	0.02	0.01	0.04

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

The equation of the full model was created based on the coefficients of each of the variables given in MLR8 (with P-value= 0.00) and this is given as follows:

$$Logit\ (p) = ln\frac{p}{1-p}$$
$$= -3.93 + 0.01 * SL - 0.31 * CG + 0.18 * CA - 0.18 * W - 0.26 * RS + 0.67 * JR$$
$$+ 0.42 * PTM + 0.88 * FL + 0.53 * FJ - 0.55 * PC + 2.55 * SB + 1.06 * AD$$

Where the predicted probability (predictive margin):

$$p = \frac{exp^{logit(p)}}{1 + exp^{logit(p)}}$$

(SL) Speed limit; (CG) Cyclist gender; (CA) Cyclist age group; (W) Weather; (RS) Road surface condition; (JR) Junction restart; (PTM) Poor turn or manoeuvre; (FL) Failed to look properly; (FJ) Failed to judge other person's path or speed; (PC) Passing too close to cyclist; (SB) Sudden braking; (AD) Aggressive driving.

6.3.3 Multilevel Logistic Regression Model

When the data is nested in groups, the response variables (casualty severity) which are in the same groups are more likely to function in a similar way and different from other groups (Sommet and Morselli, 2017). In this situation, the nested cluster impact in the model occurs and Multilevel Logistic Regression is the recommended statistical method to apply and compare the results with normal Logistic Regression outcomes (Sommet and Morselli, 2017). While normal Logistic Regression creates a full model, it does not explore the potential for a nested cluster influence and thus to estimate the effect of covariates at a regional level (Li *et al.*, 2011).

Previous studies did not consider this type of regression model in their analysis probably because they did not conduct a study which has a large amount of nested data in groups. However, this study brought together data from cities/areas across England which potentially form 21 groups that may exhibit similarities and/or differences. Therefore, to further investigate these differences and similarities it was decided to conduct Multilevel Logistic Regression analysis and compare the outcomes with normal Logistic Regression Analysis presented in the previous section. In this analysis, the first level contains predictor variables and the second level the 21 cities/areas.

Initially, a null model of Multilevel Logistic Regression which includes only cities/areas and cyclist casualty severity was applied. The aim of obtaining the results from the null model was to determine the regional impact based only on casualty severity and excluding any explanatory external influences. The equation of null model of Multilevel Logistic Regression is given as follow:

$$log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + u_j$$

Where β_0 is overall intercept and u_j is level 2 residual (regional effect/regional residual).

The likelihood ratio statistic for testing the null hypothesis is given in Table 6.13 and shows that the cities/areas variance of level 2 was not statistically significant at 95% confidence level (p-value=0.34); the estimated regional effect parameters u_j= 0.02 and intercept β_0= -1.57. In other words, there was strong evidence that there was no statistically significant difference in the cyclist casualty severity ratio between cities/areas. The plot of the estimated residuals for 21 local authorities is shown in Figure 6.8.

Table 6-13 – Null Model of Multilevel Logistic Regression

	Coefficient	P-Value	Estimated Regional Effects Parameters
Constant of Model	-1.57	0.00	
Residual of Local Authorities			0.02
			p-value = 0.34

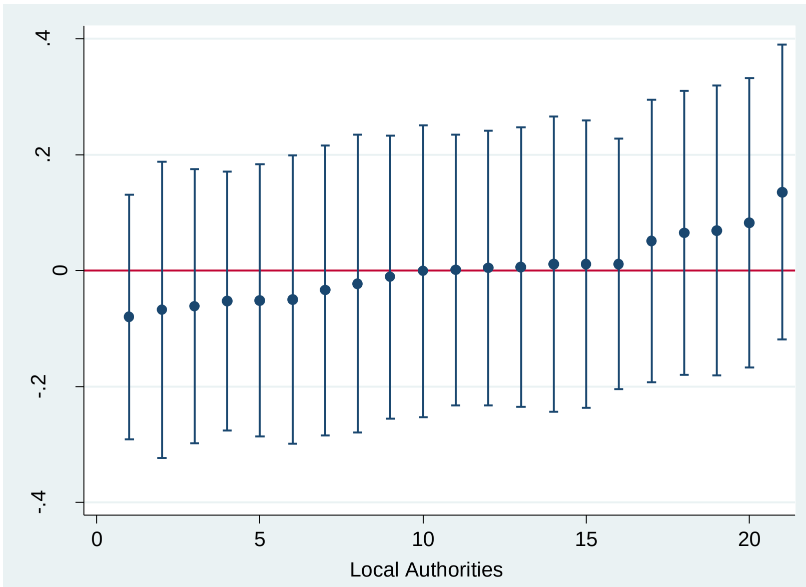


Figure 6.8 – Estimated Residuals (u_j) for each 21 Local Authority

After confirming the regional similarity in the casualty severity ratio between considered cities/areas across England, Simple Multilevel Logistic Regression (SMLR1) and Multiple Multilevel Logistic Regression (MMLR1) was conducted to investigate the regional variance of influence of variables on casualty severity (See Table 6.14). SMLR1 was carried out to investigate the individual impact of the variables with regard to different cities. The equation of likelihood ratio statistic including one predictor variable (Steele, 2010) is as follows:

$$Y = log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 X_{ij} + u_j$$

Where:
β_0 = the log-odds that y = 1 when x = 0 and u = 0
β_1 = the effect on log-odds of one unit increase in predictor x for individuals in same group
u_j = is the level 2 (region) variance

Table 6-14- Simple Multilevel Logistic Regression (SMLR1) and Multiple Multilevel Logistic Regression (MMLR1)

Variables	**SMLR1**				**MMLR1 (P-value= 0.00)**		
	Coefficient	**P-Value**	**Estimated Residual of Cities/Areas**	**P-Value of Residual**	**Coefficient**	**P-Value**	**Estimated Residual of Cities/Areas**
Speed limit	0.02	0.00	0.01	0.44	0.02	0.01	
Cyclist gender	-0.26	0.12	0.03	0.28	-0.31	0.07	
Cyclist age group	0.18	0.00	0.03	0.26	0.19	0.00	
Weather	-0.24	0.07	0.03	0.23	-0.19	0.24	
Road surface condition	-0.25	0.06	0.02	0.28	-0.25	0.13	
Junction restart	0.80	0.01	0.02	0.32	0.67	0.05	
Poor turn or manoeuvre	0.48	0.04	0.02	0.30	0.47	0.05	
Failed to look properly	0.89	0.00	0.06	0.13	0.93	0.00	
Failed to judge other person's path or speed	0.61	0.00	0.03	0.28	0.55	0.00	
Passing too close to cyclist	-0.26	0.35	0.02	0.34	-0.52	0.09	
Sudden braking	2.35	0.00	0.00	1.00	2.56	0.00	
Aggressive driving	1.02	0.11	0.01	0.37	1.09	0.10	
Constant	---	---	---	---	-4.06	0.00	0.07
					P-Value = 0.10		

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

The results in SMLR1 shows that seven coefficients have statistically significant impact on casualty severity, but none of the residuals (the influence of different cities/areas) are statistically significant at 95 confidence level with p-values of residuals way higher than 0.05. This result suggested that the generalised model including the selected cities in England was consistent with the not statistically significant regional residual for the influence of individual variables on casualty severity.

Also, MMLR1 was conducted to investigate the regional impact on the full model. The result of MMLR1 suggested that there was evidence that the model had a not statistically significant residual of cities/regions with p-value 0.10 and estimated regional effect parameters 0.07. This result is expected because selected cities in England were designed based on the same engineering design manual (DfT, 2007) and casualty severity data is collected using standard protocol by the police and local authorities (Stats19). However, a regional residual would be observed when the selected areas have a very different environment, such as countries. Therefore, the next chapter will focus on comparative analysis between two regions, namely Northumbria (England) and Belgium, to investigate the consistency of the model in international approach.

The results of variable coefficients from Multiple Logistic Regression and Multilevel Multiple Logistic Regression are given in Table 6.15 for a comparison. It can be seen that the coefficients of predictors in both models are similar. The results show both similarities and differences, but on the whole in the investigating within error of prediction entire consistency is observed (in the magnitude and direction/positive/negative, increase/decrease) and very close or identical values emerge for all coefficients, adding credibility to the analytical approach and leading to outputs useful to local authority engineers. This suggests that the study succeeded in reaching a final stable consistent model for cyclist casualty severity analysis for England.

Table 6-15 – Difference of Coefficients between Multiple Logistic Regression and Multilevel Logistic Regression

Predictor variable	Multiple Logistic Regression (MLR8)	Multilevel Multiple Logistic Regression (MMLR1)
Speed limit	0.02	0.02
Cyclist gender	-0.31	-0.31
Cyclist age group	0.18	0.19
Weather	-0.18	-0.19
Road surface condition	-0.26	-0.25
Junction restart	0.67	0.67
Poor turn or manoeuvre	0.43	0.47
Failed to look properly	0.88	0.93
Failed to judge other person's path or speed	0.53	0.55
Passing too close to cyclist	-0.55	-0.52
Sudden braking	2.55	2.56
Aggressive driving	1.06	1.09
Constant intercept	-3.93	-4.06

6.4 Chapter Conclusions

This chapter has investigated the impact of driver/rider behaviour related contributory factors on cyclist casualty severity in two sections: a case study of Northumbria and an extended analysis across England. The case study for Northumbria included speed limit, number of lanes on approach, entry path radius and driver/rider behaviour related contributory factors.

The analysis started with a detailed descriptive statistic and continued with regression modelling. Relaxed p-value criterion was used in the selection of statistically significant variables resulting from the SLR1 and MLR4 to determine the predictors for MLR5. MLR5 was based on the selected variables which were speed limit, number of lanes on approach, entry path radius and sudden braking statistically significant at 95% confidence level. For each 10 mph increase in speed limit a serious compared to slight casualty was 1.02 times more likely. In addition, for each additional lane on approach and higher unit increases in entry path radius increase the likelihood of a serious rather than slight casualty with odds ratio 1.95 and 1.04, respectively. The contribution of sudden braking to the serious casualty was over 7 times more likely than slight. The model for entry path radius was reversed and Linear Regression was carried out to investigate the upper safe limit of entry path radius. The result suggested that the casualty severity was more likely to be serious than slight at a roundabout when there was more than 80 metres entry path radius

The relative contributions of number of lanes on approach, entry path radius and speed limit on sudden braking were observed with applying simple and multiple logistic regressions. It was found that there was no statistical significance in the regression models. This was not expected because in the earlier results in this thesis it was suggested that entry geometric design parameters were associated with speed limit at the roundabouts. The reason may be due to the few records of sudden braking and this needs a further investigation with greater number of observations. In addition, the relative contributions of considered geometric design variables on all recorded contributory factors were investigated and the results showed that there was no statistical significance. This suggests that the impacts on contributory factors are related on other influences rather than geometry of the roundabout. This suggestion needs a further investigation in future studies.

The analysis continued with increased number of data records by extending the analysis to include 21 cities/areas across England. The analysis procedure for Northumbria was repeated for the extended data with the descriptive statistical analysis followed by several regression models (SLR1, MLR7 and MLR8). In MLR8, speed limit, cyclist age group, junction restart, failed to look properly, failed to judge other person's path or speed and sudden braking remained as statistically significant at 95 confidence level. 10 mph increase in speed limit was 1.02 times more likely to be a KSI casualty. The odds ratio of cyclist age group by 1.20 suggested that every unit increase in age group increased the probability of KSI casualty occurrence by 20% compared to slight casualty. The cyclist casualties which included junction restart was 1.94 times more likely KSI casualties than slight. Failed to look properly and failed to judge other person's path or speed increase the probability of KSI casualty occurrence by odds ratios of 2.41 and 1.70, respectively. The influence of sudden braking on cyclist casualty was with a very high odds ratio of 12.83 and a wide range of 95% confidence interval.

When exploring differences across 21 cities, the null or base model of Multilevel Logistic Regression suggested that there was no statistically significant difference between cities/areas regarding cyclist casualty severity, given the p-value of 0.34 for the residual and estimated regional effect parameter (u_j) of 0.02. In addition, the simple and full models (SMLR1 and MMLR1) showed that there was no statistically significant evidence of difference between regions for any of the variables considered. This result suggested that the developed model of Northumbria was consistent with the randomly selected regions in England. In addition, the Chapter 6 analysis confirms the need for data sets that are typically more than 1000 number of records to gain statistical confidence 95% in the model results and that the modelling approach adopted in this study was suitable; the need for local highway engineers to design roundabouts using a standard protocol leads to consistency in crash risk and, in general, cyclist/driver behaviour is not different in cities/areas across England. The standard method (STATS19) used by the police record data has allowed error in data used in the analysis to be minimised.

The consistency of results emerging from Chapters 5 and 6 led to the question of whether, by applying a similar analysis approach to a different roundabout design and driver/rider behaviour, differences in cyclist casualty severity could be investigated. Therefore, the next chapter carries out a comparative analysis of data sets available from England and compared to Belgium.

Chapter 7 Comparison Analysis between North East of England (Northumbria) and Belgium

7.1 Introduction

The impact of sociodemographic characteristics of cyclists, weather conditions and speed limits on cyclist casualty severity was investigated in Analysis 1 (See Chapter 5). This was followed by Analysis 2 (See Chapter 6) which identified which driver/rider behaviour related contributory factors influence cyclist casualty severity. The results in Analysis 2 showed that there is no statistically significant evidence of differences in the influence of variables on casualty severity between the 21 randomly selected cities/areas in England. This suggests that there is consistency in the roundabout design and driver/cyclist behaviour between cities, which was expected because the highway design process is governed by standards across the country. As well as this, the data records are reported in a standard format (STATS19) by the police and the geometric design parameters were measured systematically and consistently by the researcher. Finally, the results supported the appropriateness of the modelling approach.

This chapter reports Analysis 3, which included Belgium, where a different policy drives their design of roundabouts by adopting the radial geometric layout that potentially causes differences in driver/cyclist behaviour, in order to truly demonstrate the consistency of the results from the developed modelling approach. This offers the opportunity to compare data from two countries with a typically different roundabout design, namely tangential (England) and radial (Belgium).

As mentioned earlier in Chapter 4, the study area representative of England that was selected was Northumbria and was compared with Belgium regarding the data availability. Consistent with the analysis of the 21 cities data in England, Simple and Multiple Multilevel Logistic Regression was used to explore differences between Belgium and Northumbria (used as a representative of England). The analysis was carried out in three steps (See Figure 7.1). Identical analysis was first performed on the Northumbria data (step 1) and second on the Belgium data (step 2). This analysis included descriptive statistics, test of normality, relaxing p-value applied to SLR and MLR and finally, developing the full MLR model based on statistically significant variables. The final analysis (step 3) applied the generalised model, having the three-way chi square test of independence. The null model of Multilevel Logistic Regression was derived before carrying out SMLR and MMLR following the same procedure of the 21 cities in England (See Section 6.4).

The analysis of the Northumbria data is presented in Section 7.2 and is followed by the analysis of Belgium reported in section 7.3 separately. The regional influence on the relationship between the considered variables and the generalised model across both countries is presented in Section 7.4. Finally, the chapter is concluded in Section 7.5.

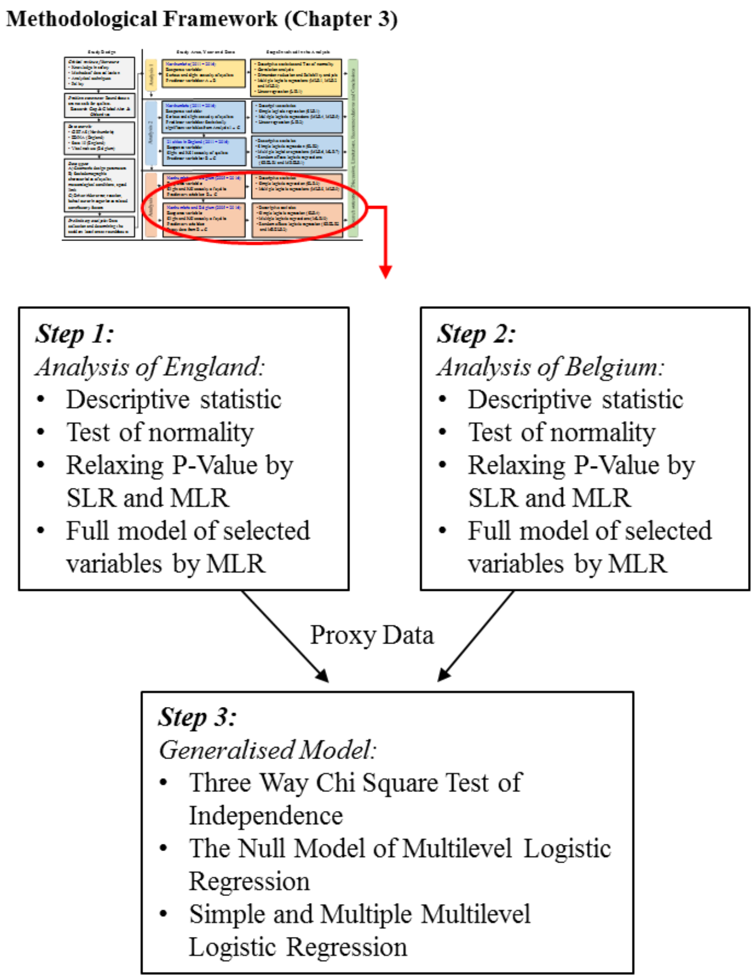


Figure 7.1- Applied Steps in Chapter 7

7.2 Analysis of Northumbria

The analysis starts with descriptive statistics which were carried out in order to gain a better understanding for the data before interpreting the results of further regression analysis. Frequency results showed that 729 Slight, 133 serious and 2 fatal casualties occurred in Northumbria between 2005 and 2016. Since the number of fatal casualties were 2, to retain binary data for casualty severity serious and fatal casualty records were combines as Killed and Seriously Injured (*KSI*). The distribution of slight and KSI casualties per year in Northumbria are shown in Figure 7.2.

The number of slight casualties fluctuated year and year exhibiting a significant fall at 2011 followed by a steady increase before dropping again in 2016. However, what was striking in this data was the highest number of KSI casualties were recorded in the recent four years from 2013 to 2016.

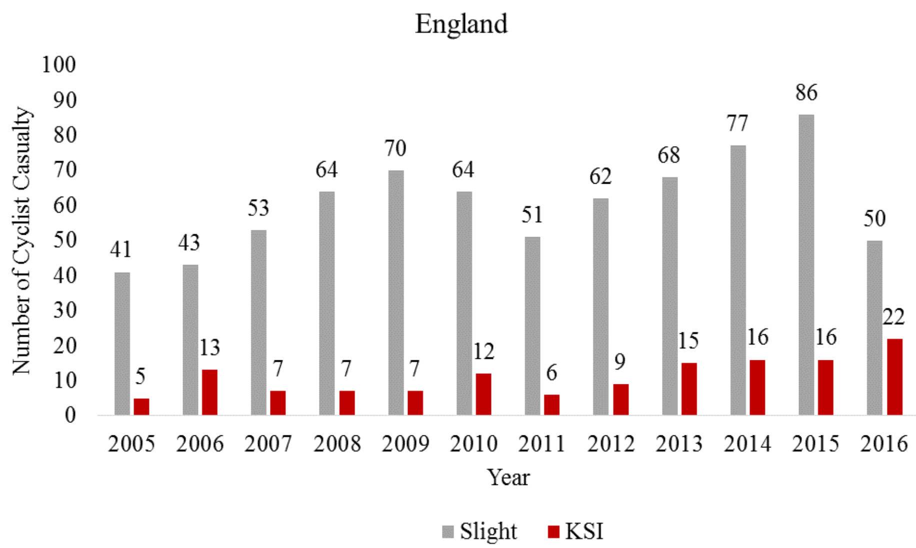


Figure 7.2 - Number of Slight and KSI Cyclist Casualties per Year between 2005 and 2016 Occurred at Northumbria

It is known that substantial funding of facilities (such as safer streets, cycle routes and secure parking) encourages people to cycle (DfT, 2014). Therefore, cyclist count data was sourced from Northumbria and plotted for the period 2005 to 2016 in Figure 7.3. As anticipated since 2005, there has been a systematic increase in cyclist counts. Therefore, when the number of KSI casualties per year were normalised using the cyclist counts, as can be seen in the Figure 7.4, the number of KSI casualties per 10,000 has fluctuated over time but increased from 3 per 10,000 cyclists in 2005 to 8 per 10,000 cyclists in 2016. It can be concluded that whilst promoting cycling as a more sustainable mode also has health benefits, it is important to maintain a safe network for this more vulnerable mode of travel. This finding endorses the importance of an in-depth study of cyclist casualty severity at roundabouts to understand how to reduce crashes.

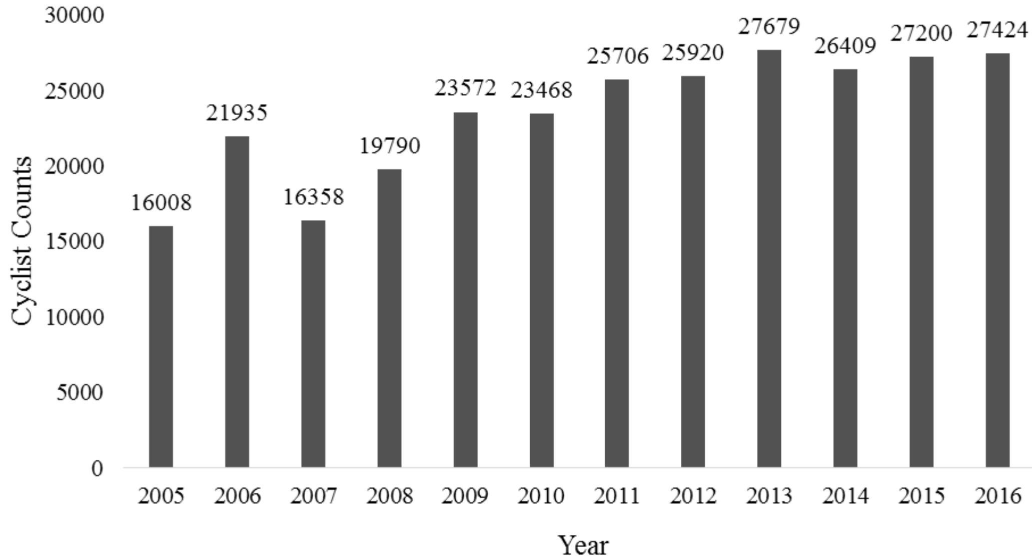


Figure 7.3 - Traffic Counts for Cyclists in Northumbria between 2005 and 2016

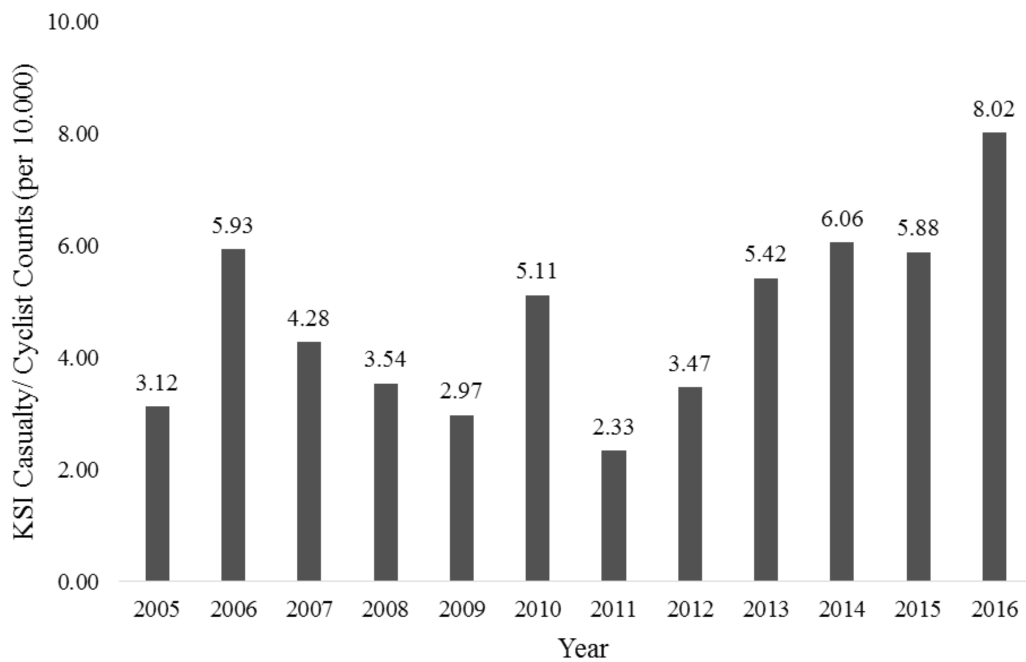


Figure 7.4 – Ratio of Number of KSI Casualty to Cyclist Counts per 10.000 for Given Period between 2005 and 2016

7.2.1 Descriptive Statistics

This section describes the descriptive statistics of speed limit, cyclist age, cyclist gender, lighting, weather, road surface condition and driver/rider error and reaction (junction overshoot, junction restart, poor turn or manoeuvre, failed to signal or misleading signal, failed to look properly, failed to judge other person's path or speed, passing too close to cyclist, sudden braking, swerved, loss of control) and driver/rider behaviour and inexperience (aggressive driving, careless, reckless or in a hurry, nervous, uncertain or panic, driving too slow for condition or slow vehicle learner or inexperienced driver/rider inexperience of driving on left unfamiliar with model of vehicle).

Speed Limit

The number of Slight and KSI casualties for each speed limit at roundabouts is given in Figure 7.5. Majority of the cyclist casualties occurred at roundabouts with 30 mph speed limit with 555 Slight and 92 KSI and it was followed by 40 mph and 60 mph.

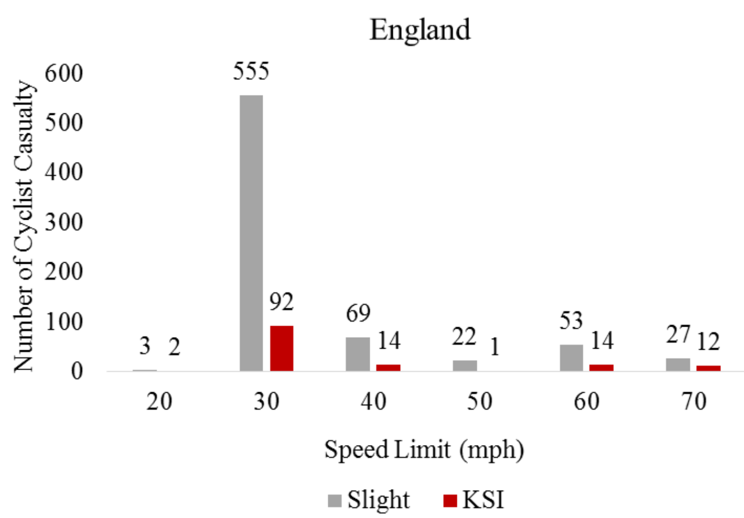


Figure 7.5 - Frequency Analysis for Speed Limit for Data of Northumbria

In the UK, speed limit is given in miles per hour and is generally 30 mph for urban areas, 60 mph for single lane carriageways, and 70 mph for dual carriageways and motorways. For some specific situations, for instance in an urban area near a school, speed limits are reduced to 20 mph, and for suburban ring and radial roads in towns and cities 40 mph and 50 mph roads are prevalent. Local Authorities are responsible for setting speed limits (DfT, 2019) and many roundabouts located in urban areas are restricted to a speed limit 30 mph.

Sociodemographic Characteristics of Cyclists

The descriptive statistics showed that the mean cyclist age was 37.8 years with 95% confidence interval 36.8 lower and 38.8 upper bound. The minimum age was 4 and the maximum age was 83. In the Shapiro-Wilk test for normality determined in SPSS the null hypothesis of normality (states that the data is normally distributed) was rejected (p-value= 0.00) with 95% confidence level. The skewness on the histogram given in Figure 7.6 was not pronounced. Therefore, Normal Q-Q plot suggested that the cyclist age data was right skewed showing a tendency for cyclist casualties to be older (tail to the right side of the histogram). This is reflecting the fact that whilst there are fewer cyclists from the younger group of the population there is a higher number of 20-25 years old.

Regarding the cyclist gender, there were more male (653 slight and 116 KSI) casualties than female (76 slight and 19 KSI) involved in crashes between year 2005 and 2016 in Northumbria.

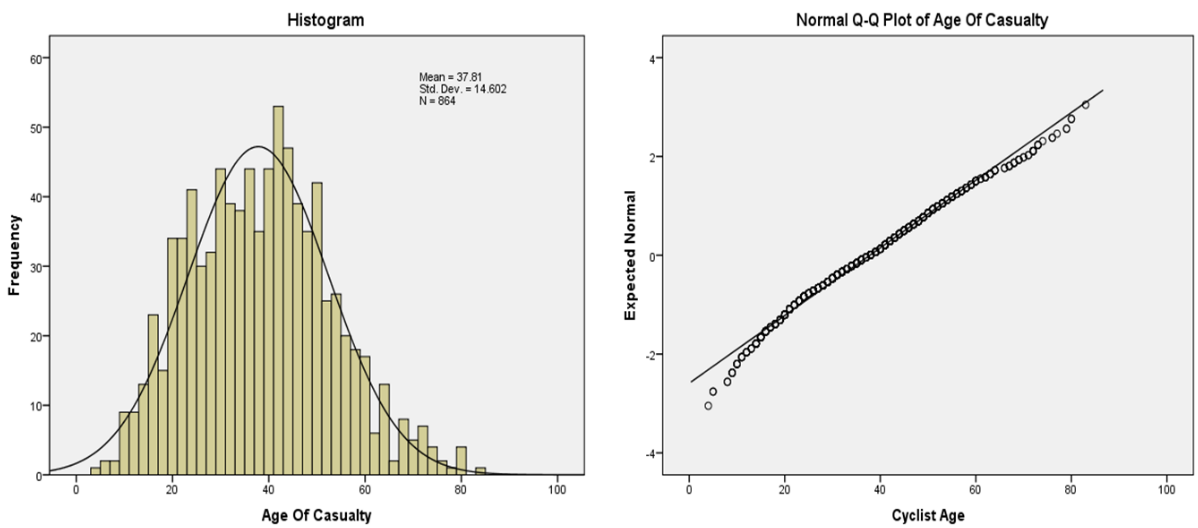


Figure 7.6 – Histogram and Normal Q-Q Plot of Non-Normally Distributed data of Cyclist Age

Lighting, Weather and Road Surface Condition

Meteorological conditions were considered in three groups: lighting, weather and road surface conditions. Weather and road surface condition were related to each other, for instance if the weather was rainy the road surface was wet. Therefore, descriptive statistics were presented for lighting separately and for weather and road surface condition together on the figures. Descriptive statistics of lighting showed that majority of the both slight and KSI casualties occurred during the daylight condition by 559 and 107, respectively. Number of slight and KSI casualties for darkness were 28 and 107, respectively. These statistics reflect that fewer cyclists use roads during darkness.

Approximately 84% of the cyclist casualties occurred when the weather was fine without rain or low visibility. 488 slight and 99 serious casualties occurred when the road surface condition was dry (See Figure 7.7). In order to carry a statistical comparison of slight and KSI casualty occurrence for lighting, weather and road surface condition, the cyclist flow at each specific environment should be known. This was not possible and therefore represents a limitation of this study.

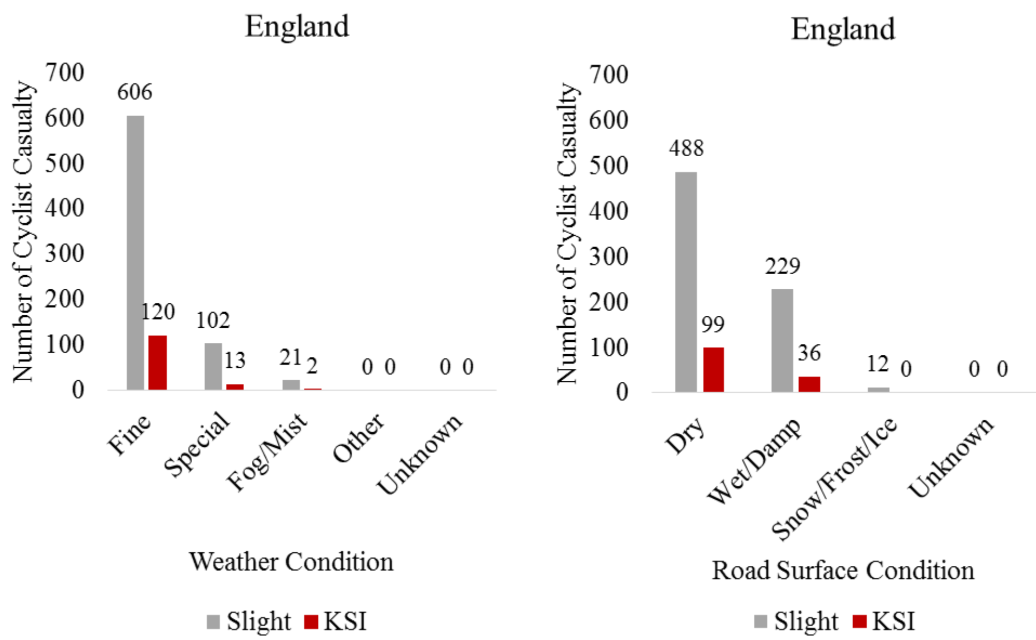


Figure 7.7 - Descriptive Statistics for Weather and Road Surface Condition

Driver/Rider Contributory Factors

The available contributory factors data were recorded without specific reference to the road user i.e. whether driver or rider. In the absence of this differentiation the contributory factors were appropriately assigned driver/rider. Therefore, the two groups of contributories were named *driver/rider error and reaction* and *driver/rider behaviour and inexperience* (See Table 7.1).

Regarding error and reaction related factors, failed to look properly was by far the higher for both slight and KSI casualties by 517 and 91, respectively. This suggested that 67% of KSI casualties occurred when driver/rider's failed to look properly. This was followed by failed to judge other person's path, passing too close to cyclist and poor turn or manoeuvre.

Table 7-1- Frequency Analysis for Driver/Rider Error, Reaction, Behaviour and Inexperience Related Contributory Factors for Northumbria

	Variable Name	**Slight**		**KSI**	
		Yes	**No**	**Yes**	**No**
Driver/Rider Error and Reaction	Junction overshoot	33	696	8	127
	Junction restart	26	703	7	128
	Poor turn or manoeuvre	74	655	10	125
	Failed to signal or misleading signal	15	714	2	133
	Failed to look properly	517	212	91	44
	Failed to judge other person's path or speed	159	570	17	118
	Passing too close to cyclist	83	646	13	122
	Sudden braking	12	717	6	129
	Swerved	3	726	0	135
	Loss of control	9	720	3	132
Driver/Rider Behaviour and Inexperience	Aggressive driving	8	721	1	134
	Careless, reckless or in a hurry	154	575	28	107
	Nervous, uncertain or panic	6	723	1	134
	Driving too slow for condition or slow vehicle	63	729	0	135
	Learner or inexperienced driver/rider	3	726	0	135
	Inexperience of driving on left	1	728	0	135
	Unfamiliar with model of vehicle	63	729	0	135

With respect to the variables related to driver/rider behaviour and inexperience, the highest casualty record was observed for careless, reckless or in a hurry with 154 Slight and 28 KSI. There has been no KSI casualty record for: swerved, driving too slow for condition or slow vehicle learner or inexperienced driver/rider, inexperience of driving on left, unfamiliar with model of vehicle; therefore, these variables should be excluded from the further regression analysis. In addition, the variables with a very low Slight and KSI records should be interpreted carefully in further regression analysis based on with 95% confidence interval because the low amount of observation gives biased results.

7.2.2 Regression Modelling to Determine the Influence of Considered Variables on Cyclist Casualty Severity

In the Northumbria data, the number of KSI casualties was 135; 13 variables could be applied into the regression model. However, the number of variables in the dataset was 18 and this does not meet with the suggested number of events per variable (Peduzzi *et al.*, 1996). This limitation led to applying relaxing p-value criterion, which was a variable selection method by comparing the results of Simple Logistic Regression (SLR) and a full model of Multiple Logistic Regression (MLR) and determining statistically significant predictors at 90 and 95% confidence level (Sperandei, 2013).

As seen in Table 7.2, the predictors were applied in SLR3 and MLR9 for Northumbria data to select the statistically significant variables that were at least 90% confidence level. Speed limit, cyclist age, failed to judge other person's path or speed and sudden braking were statistically significant at 95% confidence level in both SLR2 and MLR9 models. While weather and road surface condition were statistically significant at 90% confidence level in SLR3, their significance disappeared in the MLR9 because other predictors in the model dominated the influence of weather and road surface condition. This situation was opposite way for the influence of cyclist gender. While cyclist gender was statistically significant at 90% confidence level in the MLR9 model, it did not emerge as statistically significant in SLR3.

In summary, the two statistical approaches exhibited a degree of instability, so the variables selected for further modelling were speed limit, cyclist age, cyclist gender, weather, road surface condition, failed to judge other person's path or speed and sudden braking.

Table 7-2 - Simple Logistic Regression (SLR3) and Multiple Logistic Regression (MLR9) for Northumbria Data

Variable name	**SLR3**					**MLR9**				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**		**Coefficient**	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**	
				Lower	**Upper**				**Lower**	**Upper**
Speed limit	0.02	**0.02****	1.02	1.00	1.03	0.02	**0.05****	1.02	1.00	1.03
Cyclist age	0.02	**0.00****	1.02	1.01	1.03	0.02	**0.00****	1.02	1.01	1.04
Cyclist gender	-0.34	0.22	0.71	0.41	1.22	-0.54	**0.06***	0.59	0.33	1.03
Lighting	0.15	0.51	1.16	0.74	1.82	-0.19	0.45	0.83	0.51	1.35
Weather	-0.41	**0.10***	0.66	0.41	1.07	-0.20	0.47	0.82	0.47	1.41
Road surface condition	-0.34	**0.09***	0.71	0.48	1.05	-0.38	0.11	0.69	0.43	1.09
Junction overshoot	0.28	0.48	1.33	0.60	2.94	0.35	0.41	1.42	0.62	3.28
Junction restart	0.39	0.37	1.47	0.63	3.48	0.48	0.29	1.62	0.67	3.95
Poor turn or manoeuvre	-0.35	0.33	0.71	0.36	1.41	-0.44	0.23	0.64	0.31	1.32
Failed to signal or misleading signal	-0.33	0.67	0.72	0.16	3.17	-0.18	0.82	0.84	0.18	3.79
Failed to look properly	-0.16	0.41	0.85	0.57	1.23	-0.15	0.47	0.86	0.57	1.30
Failed to judge other person's path or speed	-0.66	**0.02****	0.52	0.30	0.88	-0.65	**0.02****	0.52	0.30	0.91
Passing too close to cyclist	-0.19	0.55	0.93	0.45	1.54	-0.22	0.51	0.80	0.41	1.54
Sudden braking	1.02	**0.05****	2.78	1.02	7.53	1.21	**0.03****	3.36	1.17	9.67
Loss of control	0.60	0.38	1.82	0.49	6.80	0.62	0.38	1.87	0.46	7.57
Aggressive driving	-0.40	0.71	0.67	0.08	5.42	-0.80	0.48	0.45	0.05	4.10
Careless, reckless or in a hurry	-0.02	0.92	0.98	0.62	1.54	0.00	0.99	1.00	0.63	1.61
Nervous, uncertain or panic	-0.11	0.92	0.90	0.11	7.53	-0.22	0.85	0.81	0.09	7.48
Constant						-2.10	0.00	0.12	0.05	0.32

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

When the selected variables were processed in MLR10 (with P-value=0.00) (See Table 7.3) and it was identified that higher speed limit increased the probability of KSI compared to slight casualty by 1% (odds ratio of 1.01) lower than the 2% in models SLR3 and MLR9. Older cyclists in one unit increase in age were 2% more likely to be involved in KSI compared to slight casualty (odds ratio 1.02). However, failed to judge other person’s path or speed was statistically significant at 95% confidence level, the coefficient value of the predictor was negative since the casualties were more likely to be slight for this reported contributory factor. The final statistically significant predictor at 95% confidence level was sudden braking with odds ratio 3.02, which suggests that a slight casualty was three times more probable than a KSI. Given that the descriptive statistics revealed that the number of observations for sudden braking was very low the 95% confidence intervals had a wide range from 1.09 to 8.38, suggesting that slight compared to KSI could be similar or over eight times more probable. Earlier, it was suggested that identifying the influence of geometric design parameters on sudden braking should be investigated in a further study including higher number of casualty records (See Section 6.4). This suggestion was applied, and it was observed that there was no statistically significant relationship between geometry and sudden braking. This suggests that the impact on sudden braking is related to other influences rather than geometry of the roundabout.

Table 7-3 - Multiple Logistic Regression (MLR10) Including Selected Predictors

Variable name	**MLR10 (P-value= 0.00)**				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio** **Lower**	**Upper**
Speed limit	0.02	0.05**	1.01	1.00	1.03
Cyclist age	0.02	0.00**	1.02	1.01	1.03
Cyclist gender	-0.50	0.07*	0.60	0.35	1.06
Weather	-0.18	0.51	0.84	0.49	1.43
Road surface condition	-0.31	0.18	0.74	0.47	1.15
Failed to judge other person’s path or speed	-0.67	0.02**	0.51	0.30	0.88
Sudden braking	1.11	0.03**	3.02	1.09	8.38
Constant	-2.30	0.00	0.10	0.04	0.23

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

Finally, a further model MLR11 (with P-value=0.00) which considered only statistically significant variables (See Table 7.4), provides a robust model with all the variables statistically significant at the 95% confidence level. The coefficients in MLR11 compared to MLR10 are of a similar magnitude. This suggests that the prediction model including speed limit, cyclist age, failed to judge other person's path or speed and sudden braking is the stable reliable final model.

Table 7-4 - Multiple Logistic Regression (MLR11)

Variable name	MLR10 (P-value= 0.00)				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**	
				Lower	**Upper**
Speed limit	**0.01**	**0.05****	**1.01**	**1.00**	**1.03**
Cyclist age	**0.02**	**0.00****	**1.02**	**1.00**	**1.03**
Failed to judge other person's path or speed	**-0.64**	**0.02****	**0.53**	**0.31**	**0.91**
Sudden braking	**1.07**	**0.04****	**2.93**	**1.06**	**8.06**
Constant	-2.82	0.00**	0.06	0.03	0.13

**Statistically significantly at 95% confidence level

The equation for predictive purposes was derived based on the coefficients of each variable given in MLR11, and is as follows:

$$Logit\ (p) = ln\frac{p}{1-p}$$
$$= -2.82 + 0.01 * Speed\ limit + 0.02 * Cyclist\ age - 0.64$$
$$* Failed\ to\ judge\ other\ person's\ path\ or\ speed + 1.07 * Sudden\ braking$$

Where p is probability of KSI and 1-p is probability of slight casualty.

Linear Regression for Determining the Relationship between Cyclist Age and Casualty Severity

As mentioned above, the model was reversed, and linear regression was carried out to determine the safe cyclist age limit for policy makers. The reason why Linear Regression was applied was that cyclist age became the dependent variable and casualty severity was the predictor in the model. The result of Linear Regression showed that cyclists who were over the age of 41 were more likely to be in KSI casualty than Slight (See Table 7.5).

Table 7-5- Linear Regression Model

Casualty	Coefficient	P ≥|z|	95% Confidence Interval		Predictive Margins
			Lower	Upper	
Casualty Severity					
0	Base				37.17
1	4.05	0.00**	1.36	6.70	41.23
Constants	37.18	0.00	36.12	38.23	

**Statistically significantly at 95% confidence level

7.3 Analysis of Belgium

Belgium data between 2005 and 2016 included 924 casualties (8 unharmed, 855 Slight, 60 serious and 1 fatal). Due to the low number of unharmed crashes and fatal casualties, it was not possible to apply ordinal logistic regression. Therefore, slight and unharmed data was combined as slight. Serious and fatal were merged as Killed or Seriously Injured (KSI). The descriptive statistics in Figure 7.8 showed that Slight records were fluctuating throughout the study period and KSI was highest in 2012.

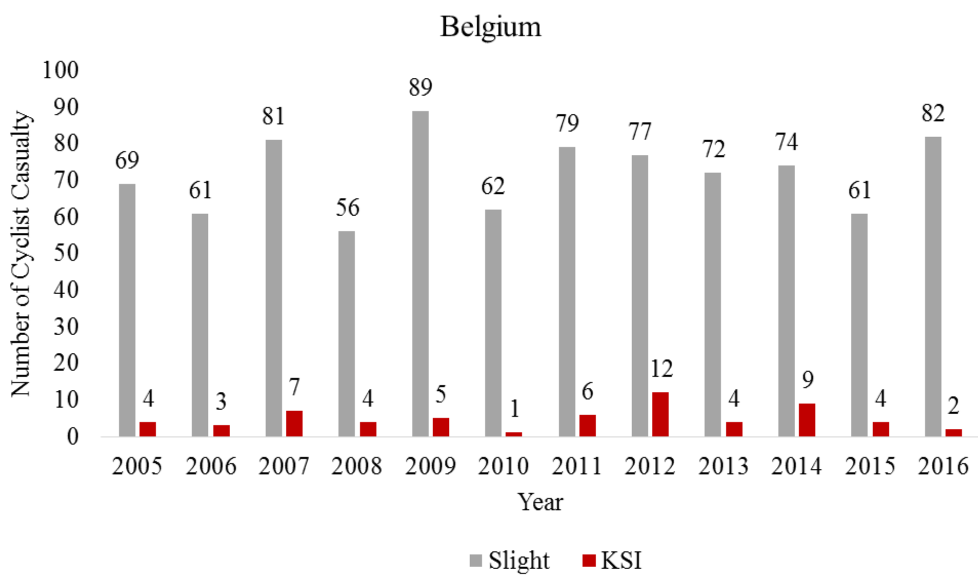


Figure 7.8- Number of Slight and KSI Cyclist Casualties per Year between 2005 and 2016 Occurred at Belgium Roundabouts

7.3.1 Descriptive Statistics of Speed Limit, Sociodemographic Descriptors of Cyclists, Meteorological Conditions and Behaviour Related Contributory Factors

Descriptive statistics for Belgium were produced for the variables, namely speed limit, sociodemographic characteristics (cyclist age and gender), environmental conditions (lighting, weather and road surface) and driver/rider behaviour related contributory factors. The behaviour contributory factors, unlike data from England, were recorded for driver and rider individually, so therefore the descriptive statistics were applied separately.

Speed Limit

Speed limit is given in kilometres per hour (kph) in Belgium (See Chapter 4). Figure 7.9 illustrates the number of slight and KSI casualties for each speed limits in kph. The converted values to miles per hour (mph) was also given below in parenthesis. The highest number by far of slight (722) and KSI (45) casualty severity was recorded at roundabouts with 50 kph speed limit. This was followed by 30 kph, 70 kph and 90 kph, respectively. Given that other data such as number of roundabouts for each speed band was not known; therefore, in absence of normalisation further investigation was not possible.

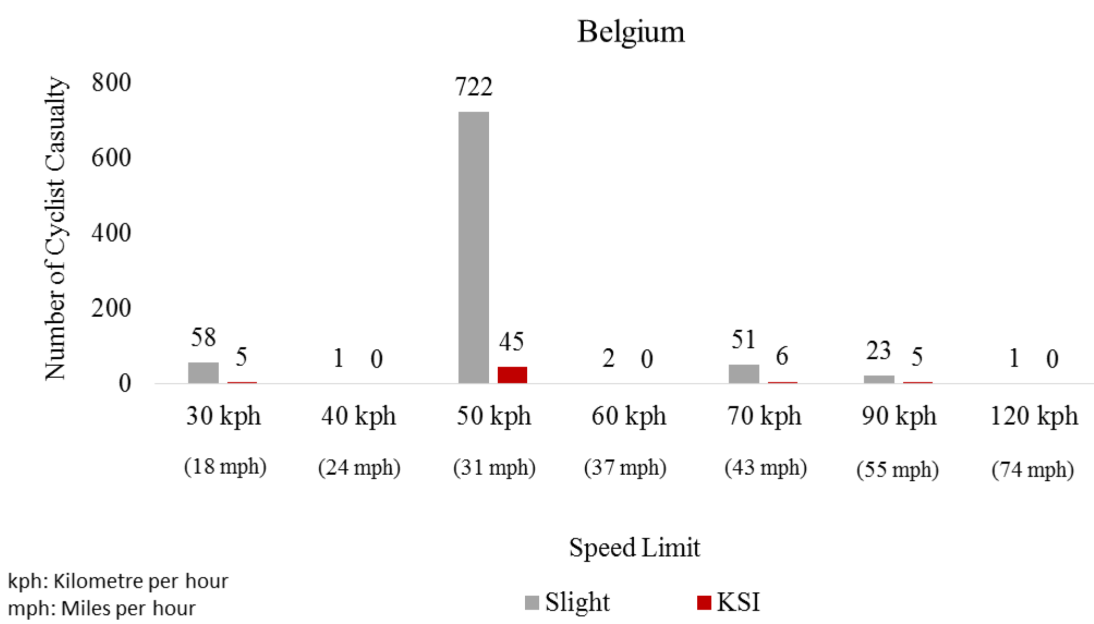


Figure 7.9 - Descriptive Statistics of Speed Limit of Belgium Data

Sociodemographic Characteristics of Cyclists

Descriptive statistics showed that the mean cyclist age was 39.2 years with 95% confidence interval range with lower 37.9 and upper 40.5 bounds. The Shapiro-Wilk test for normality using SPSS stated that the p-value at 95% confidence level was 0.00. Therefore, the null hypothesis of normality (data is normally distributed) was rejected. The skewness on the histogram given in Figure 7.10 was much more prominent with similar features observed in the Normal Q-Q plot, suggesting that the cyclist casualty age was a right skewed data with the younger age ≤25 years dominating.

More males than females were recorded with 556 slight and 42 KSI, and 307 slight and 19 KSI casualties for female cyclists. Cycling volume at mixed traffic roundabouts according to gender was not known, therefore further investigation was limited. The large difference between the number of male and female casualties in part, may be due to fewer cycle kilometres travelled by females.

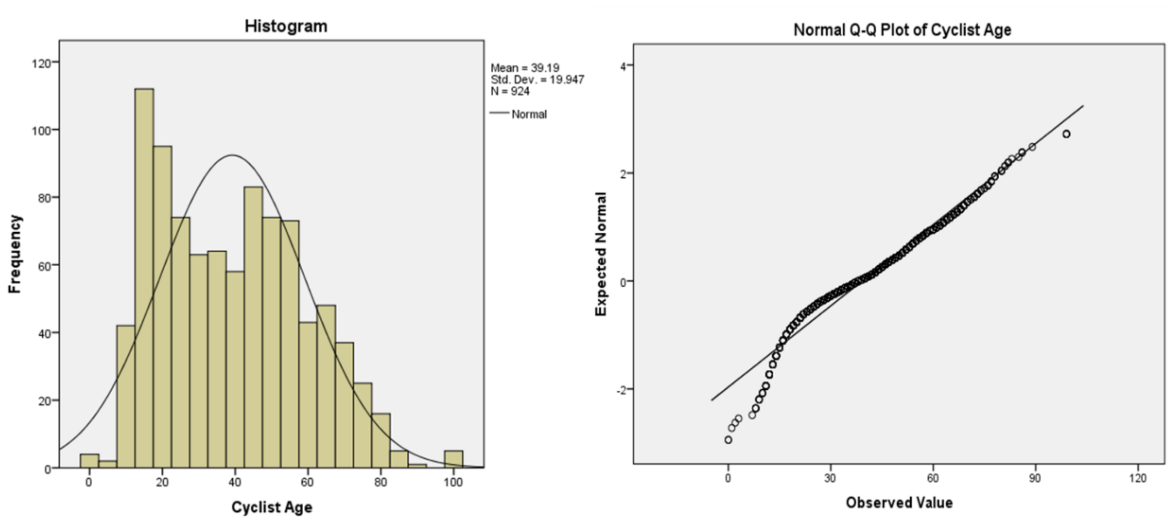


Figure 7.10 - Test of Normality with Histogram and Normal Q-Q Plot of Cyclist Age for Belgium Data

Lighting, Weather and Road Surface Condition

The majority of the number of slight (756) and KSI (57) casualties recorded occurred when the lighting condition was in daylight. The casualty records for darkness was significantly less than daylight with 103 slight and 4 KSI. 4 slight casualties were recorded as unknown in the lighting level category demonstrating the difficulties of drawing conclusions when dealing with low occurrence. Descriptive statistics for weather and road surface conditions were illustrated next to each other in order to reveal any relationship since if the weather is fine, the road surface is dry. The Figure 7.11 shows that 740 slight and 53 KSI casualties occurred when the weather condition was fine. Whilst 583 slight and 40 KSI casualties were recorded when the road surface was dry.

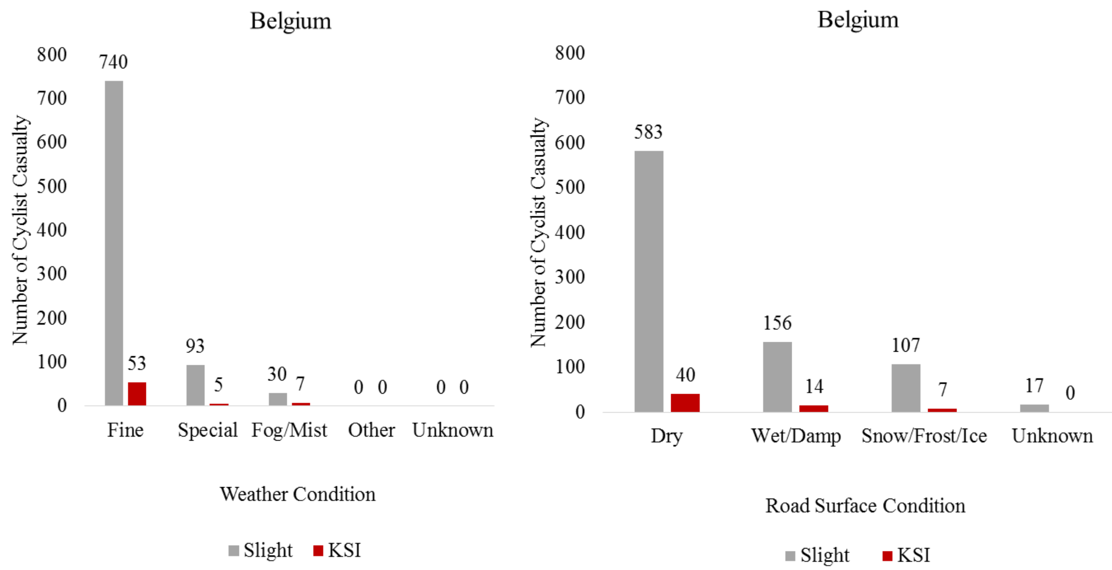


Figure 7.11 - Descriptive Statistics of Weather (Left) and Road Surface (Right) Conditions for Belgium Data

Driver and Rider Behaviour Related Contributory Factors

In contrast to England data behaviour related contributory factors in the Belgian data was separate for cyclists and vehicle drivers. Table 7.6 showed that driver's non respect of the priority was the leading contributory factor with 282 Slight and 19 KSI. Indeed, the other driver related factors had either very few or no recorded casualties and cyclist's behaviour in particular fall had high amount of records with 117 slight and 7 KSI. This was followed by illegal place of the roadway and non-respect of the priority. Few numbers of observation increase the 95% confidence interval in logistic regression models. Therefore, the variables with no records should be excluded in the further regression analysis and few records should be interpreted carefully with respect to the interval of 95% confidence.

Table 7-6 - Frequency Test for Cyclist and Driver Behaviour Related Contributory Factors for Belgium

Variable	Cyclist		Driver	
	Slight	KSI	Slight	KSI
Non respect of the priority	0(827) 1(36)	0(55) 1(6)	0(579) 1(284)	0(42) 1(19)
Cross the white line continues	0(862) 1(1)	0(61) 1(0)	0(863) 1(0)	0(61) 1(0)
Performs in extremis an avoidance manoeuvre	0(854) 1(9)	0(61) 1(0)	0(861) 1(2)	0(60) 1(1)
No respect for the distance between users	0(856) 1(7)	0(60) 1(1)	0(841) 1(22)	0(59) 1(2)
Loss of control of the vehicle	0(857) 1(6)	0(61) 1(0)	0(860) 1(3)	0(61) 1(0)
Junction overshoot	0(860) 1(3)	0(60) 1(1)	0(855) 1(8)	0(61) 1(0)
Illegal place on the roadway	0(807) 1(56)	0(55) 1(6)	0(857) 1(6)	0(60) 1(1)
Fall	0(746) 1(117)	0(54) 1(7)	0(859) 1(4)	0(60) 1(1)

7.3.2 Regression Modelling to Determine the Influence of Considered Variables on Cyclist Casualty Severity

The relaxation p-value criterion was applied at 90% and 95% confidence level in selecting statistically significant variables in either SLR4 or MLR12 for Belgium data (See Table 7.7). Cyclist age and cyclist's non respect of the priority was statistically significant at 95% confidence level in both SLR4 and MLR12. Speed limit was statistically significant at 90% confidence level but only in SLR4. For the driver related contributory factors, performs in extremis an avoidance manoeuvre showed statistical significance at 90% confidence level in MLR12. In summary, the four selected variables, namely cyclist age, speed limit, cyclist's non respect of the priority and vehicle's performs in extremis an avoidance manoeuvre were applied in the final MLR model for Belgium.

Table 7-7 - Simple Logistic Regression (SLR4) and Multiple Logistic Regression (MLR12) for Belgium Data

Variable name	**SLR4**					**MLR12 (P-value= 0.02)**				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**		**Coefficient**	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio**	
				Lower	**Upper**				**Lower**	**Upper**
Cyclist age	0.02	**0.00****	1.02	1.01	1.04	0.02	**0.00****	1.02	1.01	1.04
Speed limit (kph)	0.02	**0.06***	1.02	1.00	1.04	0.02	0.21	1.02	0.99	1.04
Cyclist gender	0.20	0.49	1.22	0.70	2.14	0.15	0.61	1.16	0.65	2.10
Weather	0.01	0.96	1.01	0.58	1.76	0.10	0.73	1.11	0.62	1.98
Road surface condition	-0.05	0.78	0.95	0.68	1.34	-0.01	0.97	0.99	0.69	1.43
Lighting	0.54	0.26	1.71	0.67	4.35	0.64	0.22	1.90	0.69	5.29
Cyclist’s non respect of the priority	0.92	**0.05****	2.51	1.01	6.20	1.13	**0.02****	3.10	1.19	8.08
Cyclist’s junction overshoot	1.56	0.18	4.78	0.49	46.63	1.70	0.15	5.47	0.54	55.14
Cyclist’s illegal place on the roadway	0.45	0.32	1.57	0.65	3.81	0.72	0.13	2.06	0.81	5.22
Cyclist’s no respect for the distance between users	0.71	0.51	2.04	0.25	16.84	0.89	0.43	2.42	0.27	21.99
Cyclist’s fall	-0.19	0.65	0.83	0.37	1.86	-0.26	0.56	0.77	0.33	1.82
Driver’s non respect of the priority	-0.08	0.78	0.92	0.53	1.61	0.09	0.75	1.10	0.61	1.98
Vehicle’s performs in extremis an avoidance manoeuvre	1.97	0.11	7.18	0.64	80.26	2.30	**0.09***	10.00	0.72	139.53
Vehicle’s illegal place on the roadway	0.87	0.43	2.38	0.28	20.09	1.25	0.26	3.49	0.39	31.14
Vehicle’s no respect for the distance between users	0.26	0.73	1.30	0.30	5.64	0.41	0.59	1.51	0.33	6.85
Vehicle’s fall	1.28	0.26	3.58	0.39	32.53	1.44	0.21	4.24	0.43	41.38
Constant						-5.42	0.00	0.00	0.00	0.02

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

The MLR13 (with P-value=0.00 at 95% confidence level), including selected variables, suggested that a unit increase of cyclist age increased the probability of KSI casualty occurrence with odds ratio of 1.02 (See Table 7.8). What was interesting in this result was that only cyclist related predictor was statistically significant at 95% confidence level with 2.71. Interestingly, the influence of speed limit on casualty severity was no longer statistically significant (p-value 0.14) in MLR13, and instead the model was dominated by cyclist age and cyclist's non respect of the priority.

Table 7-8 - Multiple Logistic Regression (MLR13) with Selected Variables

Variable name	**MLR13 (P-value= 0.00)**				
	Coefficient	**P- Value**	**Odds ratio**	**95% confidence interval for odds ratio Lower**	**Upper**
Cyclist age	**0.02**	**0.00****	**1.02**	**1.01**	**1.04**
Speed limit	0.02	0.14	1.02	0.99	1.04
Cyclist's non respect of the priority	**1.00**	**0.04****	**2.71**	**1.07**	**6.87**
Vehicle's performs in extremis an avoidance manoeuvre	1.62	0.20	5.06	0.42	60.70
Constant	-4.64	0.00	0.01	0.00	0.04

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

Predictive Margins for Cyclist Age and Cyclist's non respect of the priority

The statistically significant variables were applied in a multiple logistic regression analysis resulting in the following:

$$Logit\ (p) = ln\frac{p}{1-p} = -3.75 + 0.02 * Cyclist\ age + 1.04 * Cyclist's\ non\ respect\ of\ the\ priority$$

Where the predicted probability (predictive margin):

$$p = \frac{exp^{logit(p)}}{1 + exp^{logit(p)}}$$

The predictive margins for cyclist age were calculated and the logarithmic relationship is illustrated in Figure 7.12. An older cyclist was more likely to suffer a KSI, and the 95% confidence interval range was substantially higher particularly beyond age 50 years.

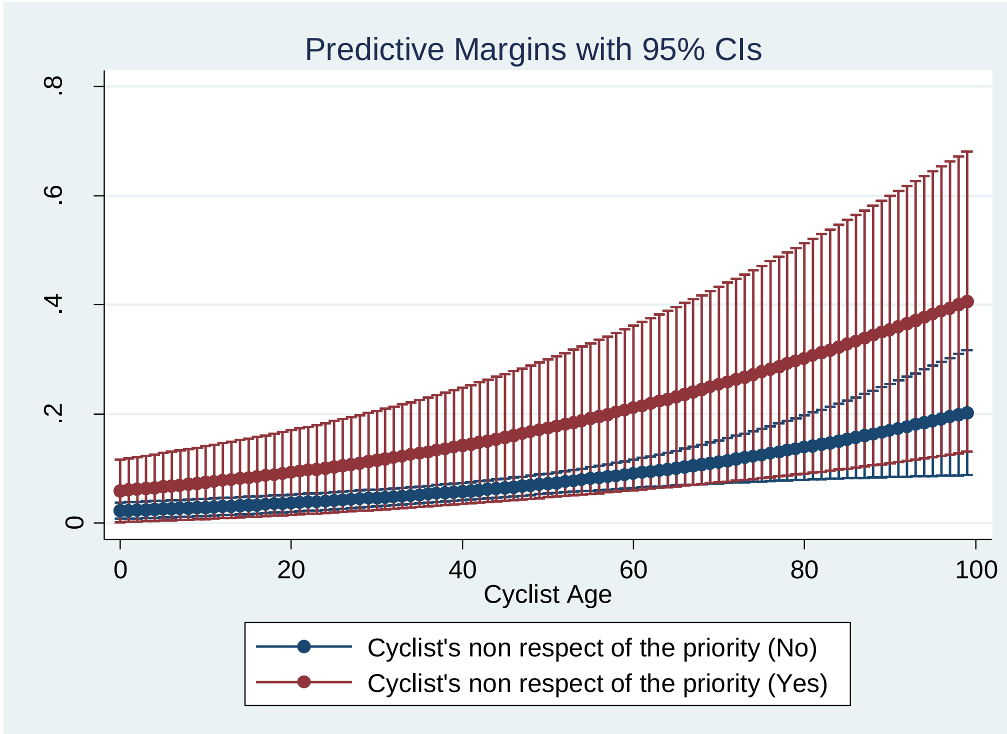


Figure 7.12 - Predictive Margins with 95% Confidence Interval for Cyclist Age

Linear Regression

The influence of cyclist age on casualty severity was investigated further by fitting a Linear Regression with the independent and dependent variables switched. Linear Regression showed that cyclists who are over age 48 were more likely to be involved in KSI compared to slight casualty (See Table 7.9).

Table 7-9- Linear Regression Model

Casualty	Coefficient	P ≥|z|	95% Confidence Interval		Predictive Margins
			Lower	Upper	
Casualty Severity					
0	Base				38.55
1	9.61	0.00**	4.46	14.77	48.16
Constants	38.55	0.00	37.22	39.87	

* Statistically significantly at 90% confidence level
**Statistically significantly at 95% confidence level

The current results of Northumbria and Belgium were obtained from individual analysis. It was aimed to conduct a further analysis in order to show the relative contribution of regional influence for response and predictor variables in the regression model. Therefore, further stage of the study was a comparison analysis between Northumbria and Belgium with proxy variables.

7.4 Exploring the Regional Influence Regarding Country Base: A Study of England (Northumbria) and Belgium with Proxy Data

Given that the data recording protocol used by the police in the UK is different from Belgium data, descriptors were needed to be associated with each other in order to standardise the data across the two countries before carrying out a comparative analysis. In this section the Northumbria data is used as representative of the UK because the required variables across 2005-2016 were not available for other cities.

The resulting ten variables were *cyclist age, cyclist gender, speed limit, lighting, weather, road surface condition, junction overshoot, poor turn or manoeuvre, passing too close cyclist and loss of control*. The UK and Belgium have a different design approach and environment; therefore, the initial stage of the regional influence analysis was to understand the association between cyclist casualty severity, considered variables and countries. The three-way chi square test of independence was deemed appropriate and details are given in the next sub-section.

7.4.1 Three Way Chi Square Test of Independence

The literature review of analytical methods identified the *chi-square test of independence* (χ^2) as a method to explore the statistical relationship between variables. The equation of χ^2 is given as follow:

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where "*O*" is the observed and "*E*" the expected value of the variable, c is degrees of freedom.

There are two assumptions of χ^2: the variables should be ordinal or nominal and must consist of two or more categorical values. The data used in this study met both χ^2 assumptions. The null hypothesis in the χ^2 test is that there is statistically significant evidence of independence between variables. If the p-value is less than 0.05 (statistically significance at 95% confidence level), the null hypothesis should be rejected. This means that there is statistically significant evidence of dependence between variables. χ^2 calculates the dependency between two variables; however, it is possible to include a third variable in the analysis to serve as a control level. For example, in this study casualty severity was considered along with the ten variables and the two countries as a three-way analysis shown diagrammatically in Figure 7.13.

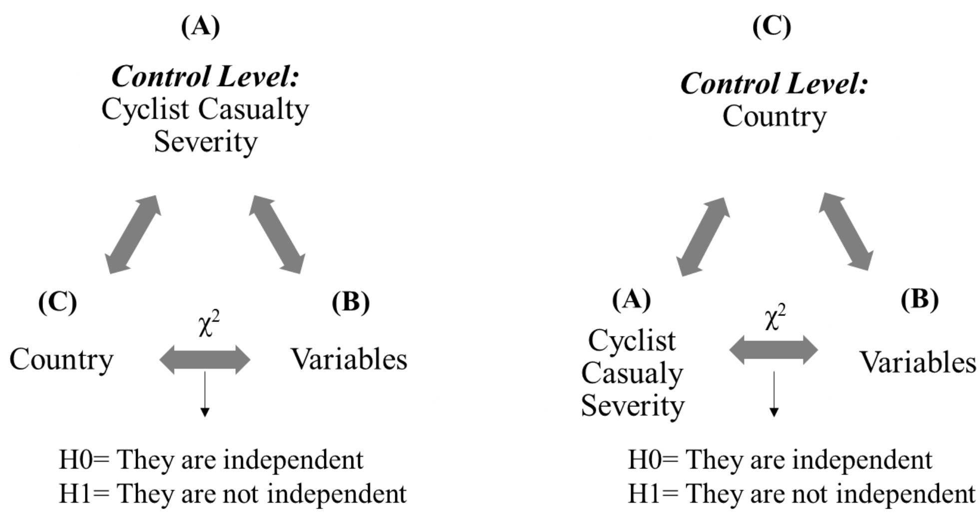


Figure 7.13 - Three Way Chi Square (χ^2) Test of Independence

Initially, the aim is to explore the relationship between country (C) and variables (B) among cyclist casualty severity (A) in order to gain a deep understanding on the regional influence on associated data. It was expected that the results would lead to an appropriate regression model based on the aggregated data. The three-way chi square test of independence was applied and the results presented in Table 7.10 at 95% confidence level. A statistically significant relationship was found between variables (i.e. cyclist age group, cyclist gender, speed limit, lighting, road surface condition, junction overshoot, poor turn or manoeuvre and passing too close to cyclists) and country among slight casualties. This suggested that the proportion of number of observations of given variables for Northumbria and Belgium were statistically significantly different among slight casualty severity.

Table 7-10- Three Way Chi Square Test of Independence Based on Casualty Severity

Chi-square test of independence between country and variables		Variables									
		Cyclist age group	**Cyclist gender**	**Speed limit**	**Lighting**	**Weather**	**Road surface condition**	**Junction overshoot**	**Poor turn or manoeuvre**	**Passing too close to cyclist**	**Loss of control**
Slight Casualty	Pearson χ^2	101.03	136.80	122.83	38.96	4.09	104.57	232.86	61.60	50.09	0.13
	P-value	**0.00****	**0.00****	**0.00****	**0.00****	0.13	**0.00****	**0.00****	**0.00****	**0.00****	0.72
KSI Casualty	Pearson χ^2	29.26	7.84	11.59	6.19	2.06	16.08	34.35	2.64	2.40	1.38
	P-value	**0.00****	**0.01****	**0.04****	**0.01****	0.36	**0.00****	**0.00****	0.10	0.12	0.24
Total Casualty	Pearson χ^2	115.11	146.29	138.24	43.80	4.08	119.51	268.26	62.36	51.69	0.66
	P-value	**0.00****	**0.00****	**0.00****	**0.00****	0.13	**0.00****	**0.00****	**0.00****	**0.00****	0.42

**Statistically significant at 95% confidence level

However, the null hypothesis was not rejected for weather (χ^2=4.09 with p-value=0.13) and loss of control (χ^2=0.13 with p-value=0.72); the variables were independent from country among slight casualty severity.

Regarding KSI casualties, the null hypothesis was rejected for six variables, namely cyclist age group, cyclist gender, speed limit, lighting, road surface condition and junction overshoot, which suggested that these variables were dependent to country with χ^2 values 29.26, 7.84, 11.59, 6.19, 16.08 and 34.35 (p-values$\leq$0.05), respectively. On the other hand, weather, poor turn or manoeuvre, passing too close to cyclist and loss of control were independent from regional difference. Comparing the χ^2 values among casualty severity, more variables were independent in KSI than Slight. Regarding the total casualty severity, statistically significantly independent variables to country were weather (χ^2= 4.08 with p-value=0.13) and loss of control (χ^2= 0.66 with p-value=0.42).

At this point, the relationship between country and variables among casualty severity was investigated. The next step explored the three-way chi square test of independence between casualty severity and variables among country (See Table 7.11). The results suggested that there was a statistically significant relationship between speed limit and casualty severity among Northumbria (χ^2= 13.74 with p-value=0.02). On the other hand, it is seen that the relationship between cyclist age group and casualty severity was statistically significant among Belgium (χ^2= 24.37 with p-value=0.01). Regarding the total data as a control level, both the cyclist age group and speed limit were statistically significantly dependent on casualty severity (See Appendix H).

The outcomes of chi square test of independence determine the statistical interdependency of three variables in order to gain a deep understanding of data. However, chi square test of independence cannot provide a detailed explanation on to what extent this relationship and applying a regression method is required to create a prediction model. Therefore, it was decided that further analysis should be Multilevel Logistic Regression model.

Table 7-11 -Three Way Chi Square Test of Independence Based on Country

Chi-square test of independence between severity and variables		Variables									
		Cyclist age group	**Cyclist gender**	**Speed limit**	**Lighting**	**Weather**	**Road surface condition**	**Junction overshoot**	**Poor turn or manoeuvre**	**Passing too close cyclist**	**Loss of control**
Northumbria	Pearson χ^2	9.88	1.55	13.74	0.43	2.92	3.74	0.49	0.98	0.36	0.81
	P-value	0.45	0.21	0.02**	0.51	0.23	0.15	0.48	0.32	0.55	0.37
Belgium	Pearson χ^2	24.37	0.49	7.82	1.92	0.70	2.01	0.25	0.06	0.12	0.64
	P-value	0.01**	0.48	0.10	0.38	0.71	0.57	0.62	0.80	0.73	0.42
Total Casualty	Pearson χ^2	20.85	2.11	27.58	0.59	1.90	6.34	3.00	0.03	0.31	0.24
	P-value	0.02**	0.15	0.00**	0.75	0.39	0.10	0.08	0.87	0.58	0.62

**Statistically significant at 95% confidence level

7.4.2 Multilevel Logistic Regression Model

In Chapter 4, it was mentioned that associated variables were identified based on Northumbria and Belgium datasets and merged with including the regional identity as an additional variable. It was aimed to estimate the effect of covariates at regional level by applying *Multilevel Logistic Regression* model by recommended Li *et al.* (2011).

Null Model of Multilevel Logistic Regression

Initially a null model of Multilevel Logistic Regression was conducted to investigate the country variance for only cyclist casualty severity (See Table 7.12). The null model of Multilevel Logistic Regression suggested that there was a statistically significant (p-value = 0.00) difference between the two countries with estimated country effects parameter by 0.22.

Table 7-12- The Null Model of Multilevel Logistic Regression

	Coefficient	**P-Value**	**Estimated Regional Effects Parameter**
Constant of Model	-2.16	0.00**	
Variance of Countries			0.22
P-value of regional residual = 0.00			

**Statistically significant at 95% confidence level

Simple and Multiple Multilevel Logistic Regressions

The aim of carrying out the Simple Multilevel Logistic Regression (SMLR) on each variable separately was to investigate the country influence with respect to each predictor individually. Multiple Multilevel Logistic Regression (MMLR) was conducted to explore the influence of regional variance on the full model.

The results in Table 7.13 show that the estimated country variance of multiple multilevel regression model was 0.25 (p-value=0.00). This suggested that there was statistically significant evidence of difference between the two countries with regard to the generalised model and more specifically for variables speed limit and cyclist age.

Table 7-13 - Simple and Multiple Multilevel Logistic Regression

Variables	**SMLR2**				**MMLR2**		
	Coefficient	**P-Value**	**Estimated Variance of Region**	**P-Value of Residual**	**Coefficient**	**P-Value**	**Estimated Variance of Region**
Speed limit	0.02	0.00	0.17	0.00	0.02	0.01	
Cyclist age group	0.18	0.00	0.24	0.00	0.18	0.00	
Cyclist gender	-0.05	0.82	0.23	0.00	-0.13	0.52	
Lighting	0.22	0.28	0.23	0.00	0.01	0.96	
Weather	-0.25	0.19	0.22	0.00	-0.15	0.47	
Road surface condition	-0.19	0.17	0.21	0.00	-0.15	0.32	
Junction overshoot	0.15	0.52	0.24	0.00	0.17	0.47	
Poor turn or manoeuvre	-0.28	0.40	0.23	0.00	-0.33	0.33	
Passing too close to other vehicle	-0.12	0.69	0.23	0.00	-0.18	0.56	
Loss of control	0.24	0.71	0.22	0.00	0.50	0.44	
Constant					-3.86	0.00	0.25
					P-value of regional residual = 0.00		

Statistically significance at 95% confidence level

In addition, a consistent result emerged from the SMLR2 which showed that the statistically significant variance for countries was observed for all individual variables taken separately with p-value≤0.05. However, the p-value of the coefficient was only statistically significant for speed limit and cyclist age, interestingly for speed limit and cyclist age the coefficients were the same from SMLR2 and MMLR2.

7.5 Chapter Conclusions

In this chapter, the results of Analysis 3, which explored differences in cyclist casualty severity at roundabout in England (Northumbria) and Belgium. Data was available for the period 2005-2016 for the two countries which adopt different roundabout designs (tangential in England and radial in Belgium). Initially, descriptive statistics, test of normality and logistic regression models were applied to both countries individually since the data recording method was slightly different. The aim was to investigate how the predictors influence cyclist casualty severity for each country.

Regarding the analysis of Northumbria, relaxing p-value criterion suggested that the selected variables of speed limit, cyclist age, cyclist gender, weather, road surface condition, failed to judge other person's path or speed and sudden braking should be applied to develop a full model (MLR10). MLR10 suggested that speed limit, cyclist age, failed to judge other person's path or speed and sudden braking were statistically significant at 95% confidence level with odds ratios 1.01, 1.02, 0.51 and 3.02, respectively. One unit (10mph) higher in speed limit increased the casualty severity; the probability of a KSI casualty was more likely than Slight in case of subjects of a higher age as compared to lower age; sudden braking had a negative influence of casualty severity by reducing the safety for cyclists.

With respect to the results of Belgium, one unit increase of cyclist age were more likely to be in KSI casualty by 1.02 odds ratio. Interestingly, none of the vehicle driver related contributory factors were statistically significant. One cyclist behaviour related predictor, cyclist's non respect of the priority, showed an influence on casualty severity. The contribution of cyclist's non respect of the priority in a crash caused being more likely to be a KSI casualty by 2.71 odds ratio.

The three-way χ^2 demonstrated that in both Northumbria and Belgium a dependent relation exists between cyclist age group, cyclist gender, speed limit, lighting, road surface condition and junction overshoot and both slight and KSI casualties. In addition, poor turn or manoeuvre and passing too close to cyclists was statistically significant for slight but not KSI, with weather also not being statistically significant for either slight or KSI casualty. The important result from this chi square test was that except for junction overshoot, all non-behavioural variables such as cyclist age group, cyclist gender, lighting, road surface condition and more importantly speed were all found to be statistically different between Northumbria and Belgium for both slight and KSI casualty severity.

The second three-way χ^2 showed that of all the variables in Northumbria speed limit merged as the single most statistically significant variable and for Belgium it was cyclist age group. This result suggests that the geometry of international roundabout may be reducing speed effectively. This was investigated further in the Multilevel Logistic Regression analysis.

The third part of the chapter continued with conducting a comparison analysis based on the associated data. The analysis started with a three-way chi-square test of independence and was followed by Multilevel Logistic Regression. SMLR2 and MMLR2 was applied to identify the regional variance influence on individual predictors and a full model, respectively. Both SMLR2 and MMLR2 showed that there was statistically significant regional variance between. Although the proxy data was applied, each region had its own characteristics regarding the influence of predictors on casualty severity.

The previous analysis (See Section 6.3.3) showed that the regional residual between cities located in England is not statistically significant. This is expected due to the consistency in design protocol and that the casualty record collection method is similar for each city/region in England. On the other hand, the residual of country is statistically significant for associated data of Northumbria and Belgium. This suggests that two countries have own independency in the developed model. It is difficult to carry out a comparative study between regions with different design approach, environment and data collection methodology.

Chapter 8 Discussions, Limitations, Recommendations and Conclusions

8.1 Introduction

This thesis has presented a detailed literature review which was carried out to identify the following research questions:

1. What are the relative contributions of geometric parameters of give way roundabouts with mixed traffic, speed limit and traffic flow profile to cyclist casualty severity?
2. What are the relative contributions of sociodemographic characteristics of cyclist and environmental conditions to cyclist casualty severity?
3. What is the relative contribution of driver/rider behaviour related contributory factors to cyclist casualty severity?
4. What is the consistency of the results between tangential and radial roundabouts based on a comparative analysis?
5. What is the appropriate statistical method to analyse the safety impact of variables on cyclist casualty severity?

The aims and objectives were developed to deliver research that addressed the research questions. This led to the formulation of a novel methodological approach to investigate the variables that influence cyclist casualty severity at give way roundabouts with mixed traffic. The novelty in the analysis included combination of statistical techniques that were applied to identify consistency within and between regions in England and two countries in the predictor variables (relating to the cyclist/driver characteristics, environmental conditions and geometric design parameters) on cyclist casualty severity.

Section 8.2 first discusses the results emerging from the analytical steps descriptive statistics, normality test, correlation analysis, dimension reduction, reliability analysis, three-way chi square test of independence, Simple, Multiple and Multilevel Logistic Regressions and Linear Regression which were used to address the research gap.

This is followed in Section 8.3 with the main findings, and in Section 8.4 the secondary findings. Next in Section 8.5 the limitations are detailed to ensure the outcomes are considered within the constraints posed by, for example, data availability. This research has identified useful insights which have policy implications and recommendations which are elaborated upon in Section 8.6. Finally, ideas for further research are given in Section 8.7.

8.2 Discussion

Cyclist safety at roundabouts has been studied by many research teams (Harkey and Carter, 2006; Brabander and Vereeck, 2007; Daniels *et al.,* 2010; Arnold et al., 2010; Daniels *et al.,* 2011; Polders *et al.,* 2014); however, the general concern has related to assessing that roundabouts are safe for vehicle drivers not for cyclists. More detailed studies (Davies *et al.,* 1997; Rasanen and Summala, 2000; Lawton *et al.,* 2003; Hels and Bekkevold, 2007; Moller and Hels, 2008; Daniels *et al.,* 2008; Daniels *et al.*, 2009; Sakshaug *et al.,* 2010; Jurisich *et al.,* 2011; Jensen, 2013; Silvano *et al.,* 2015; Silvano *et al.*, 2016; Jensen, 2017) have attempted to discuss possible safety issues for cyclists, but commonly they recognise the need for comprehensive studies which also include geometric design parameters, as well as sociodemographic characteristics and behaviour related contributory factors in order to better understand their specific influence on cyclist safety at roundabouts. Therefore, the research in this study aimed to address the recommendations from previous research. In achieving this goal original methodology in collection of data statistical analysis were formulated based on a critique of traditional methods applied casualty severity analysis research.

Influence of Speed Limit and Geometric Design of Roundabout

The studies carried out by Daniels et al., 2010 and Polders et al., 2014 suffered from several important limitations such as lack of knowledge of speed limits and geometric design parameters. Speed limit was considered in a study carried out by Jensen 2017, however this research only focused on exploring whether converting signalised junctions to roundabouts changed the junction safety record for cyclists. Therefore, comparing the results found in this thesis is not directly comparable given the different methodological approach and research question. However, the key message that speed limit influences cyclist casualty severity at roundabouts emerging from previous studies is consistent with the findings in this research.

The study (Jensen, 2017) found that converting an intersection to a roundabout with a 70 km/h or more speed limit increased safety for cyclists by reducing both the number of casualties and their severity. This suggested that a roundabout is a good choice as junction type at higher speed limits but a poor choice as low speed limit. The reason for this given by Jensen (2017) was that the speed variance among motor vehicles is much lower at roundabouts compared to signalized or non-signalized intersections. Jensen continues by suggesting that a roundabout is rather “robust” in relation to speed limits. In this thesis, the influence speed limit on casualty severity was statistically significant in all three independent analysis steps Analysis 1, 2 and 3. A higher speed limit increased the probability of serious over slight casualty at give way roundabouts with mixed traffic. A comparative safety analysis of roundabouts with priority or signalised junctions was out of the scope of this thesis.

The drive through curve, which is related to the entry path radius, is known to have significant influence on the number of cyclist crashes occurring at roundabouts (Hels and Orozova-Bekkevold, 2007). The result in this thesis study demonstrated that a higher entry path radius also increases the probability of serious over slight casualties of cyclists. Entry path radius is known to influence the speed of a vehicle on approach and at entry locations (DfT, 2007). And this research result is consistent with the finding that higher speed along with higher entry path radius are statistically significant contributors to increasing cyclist casualty severity at roundabouts.

Furtado (2004) pointed out that determining one size for the safest balanced design is difficult to achieve as there are several design solutions applied in different regions such as tangential in the UK and radial in Europe. A main design feature of the UK design roundabouts (tangential) is allowing a higher approach capacity having a wider approach and entry geometry. Whilst delivering an increased level of service and less delay to traffic they increase the number of cyclist crashes due to the higher entry speed (Davies *et al.*, 1997; Lawton *et al.*, 2003). The study reported in this thesis has deepened the knowledge and understanding that higher approach capacity also increases the severity of cyclist casualties. For example, an additional number of lanes increases the probability of serious over slight casualty by 95%. This suggests that tighter geometry should be applied at the approach arm of roundabouts to improve the safety for cyclists. However, because tangential roundabouts are designed for increasing the capacity for motor vehicles, this suggests that a tighter geometry does not meet with the basic principal of roundabout design in the UK. Therefore, the balance between wider geometry at tangential roundabouts to increase capacity and cyclist safety needs further attention.

Approach capacity and entry geometric predictors are major influential parameters in cyclist casualty severity analysis, but such data is not readily available from any source. Also, crash records available for research do not always locate where in the roundabout the crash occurred to be allowed to measure the geometric design parameters using Google Earth. This data was only available for the casualty records in Northumbria. Absence of these variables is a limitation in this study, however, the bespoke methodology allowed the statistically significant design variables of number of lanes on approach and entry path radius to be identified as the major variables influencing casualty severity. In addition, the results in this study suggested that speed limit is a proxy for entry geometrical parameters. Therefore, in extending the study across the UK and transferring the methodological approach to Belgium speed limits could be equivalent to entry geometric design variables. By considering only speed limit as a proxy of entry geometry the influence of environment and driver/cyclist behaviour could be explored in the development of predictive models.

Influence of Sociodemographic and Environmental Conditions

The study carried out by Daniels *et al.* (2010) in Belgium found that the number of severe casualties increases for the higher age groups for all road user types at roundabouts. The study in this research included data from England and Belgium. The results showed that older cyclists are more likely to be involved in KSI compared to slight casualties. This result is consistent with the previous study (Daniels *et al.*, 2010) and is expected given slower reaction and lower physical ability as people age. Whilst there was no evidence that cyclist gender had impact on casualty severity both in a previous study (Daniels *et al.*, 2010) and the research in this thesis, Evan (2004) suggested that the probability of killed crashes with the same impact for females was higher than males; however, this study did not specifically focus on roundabouts.

Weather and road surface conditions were expected to have an impact on cyclist casualty severity because rain, ice, and snow reduce traction on the road surface for both cyclists and drivers and therefore may increase the serious or slight or both casualties due to the potential loss of control. However, weather and road surface condition did not show a statistically significant impact on cyclist casualty severity at roundabouts. Previous research by Daniels *et al.* (2010) indicated that the probability of serious casualty severity for all types of road user was higher at night at roundabouts. The results of the research reported in this thesis illustrated that lighting conditions did not have any statistical significance on cyclist casualty. The reason for the difference in the results here with those of Daniels *et al*. (2010) may be due to there being different levels of cyclist flow in the respective regions. The cyclist flow at the time of the crash was not known in the study reported and therefore could not be considered.

Influence of Driver/Rider Behaviour Contributory Factors

Considering the influence of behaviour related contributory factors along with environmental and geometric design variables on the analysis of cyclist casualty severity occurring at roundabouts is unique, and so therefore comparative discussion is limited to the outcomes of the behaviour related contributory factors. Previous studies (Rasanen and Summala, 2000; Silvano *et al.,* 2015) in investigating driver's and cyclist's perception were reviewed to gain awareness of previous work and find synergy with the results presented in this thesis. These studies focused on driver's or cyclist's yielding behaviour to each other and were often based on a very limited number of video recording observations (See Section 2.4).

Rasanen and Summala (2000) suggest that the main contributory factor to number of crashes is that drivers are not looking properly to the right side from where cyclists appear unexpectedly. This is consistent with the result in this thesis because the variable *failed to look properly* was recorded in 65% of slight and 82% of KSI cyclist casualties in England analysis. The analysis in this thesis went further and quantified the influence of failed to look properly on the casualty severity as two and half times.

The literature also suggests that driver's speed decreases yielding behaviour towards cyclists (Rasanen and Summala, 2000) and driver's behaviour has a strong impact on cyclist position at the roundabout (Silvano *et al.*, 2015). The study in this thesis agrees that there is a strong relationship between driver/rider behaviour and speed which significantly influences cyclist casualty severity. However, the analysis of Belgium casualty records revealed that was not driver's but cyclist's non respect of the priority (as a proxy of yielding) that was a statistically significant contributory factor in casualty severity. This result suggests that cyclist behaviour should be considered and more attention in research focused on behaviour related causes of crashes rather than focusing on driver behaviour which has been the case in the past.

8.3 Main Findings

Table 8.1 provides an overview of the main outcomes from Analysis 1, 2 and 3 conducted in this thesis. The key messages from these outcomes will be used to answer the research questions of this study. With reference to Table 8.1 the results are now presented for each research question in turn.

1. *What are the relative contributions of geometric parameters of give way roundabouts with mixed traffic, speed limit and traffic flow profile to cyclist casualty severity?*

   Higher speed limit reduces the cyclist safety at roundabouts and this result was consistent in all following applied models. For geometric design parameters, one unit increase in the number of lanes on approach and entry path radius increased the probability of casualty severity. The reverse linear regression suggested that entry path radius should not exceed 81 metre to reduce the casualty severity. This was consistent in both Analysis 1 and 2. These three variables (speed limit, number of lanes on approach an entry path radius) emerged the importance of capacity on approach arm of roundabout on cyclist safety. This was endorsed by the statistically significant impact of approach capacity (as a derived variable from PCA) on casualty severity. The overall result of the analysis of the geometric design parameters suggests that cyclist safety at give way roundabouts in mixed traffic is compromised by approach capacity. Also, by far the number of lanes with five times increase in probability of a serious compared to a slight cyclist casualty is the most influential geometrical variable. Second is entry path radius with 4% and third speed limit at 2% increase in cyclist casualty severity.

2. *What are the relative contributions of sociodemographic characteristics of cyclist and environmental conditions to cyclist casualty severity?*

   For each increase in age of one-year cyclists were 2% more likely to be involved in a more serious casualty crash compared to slight but gender did not show any influence on cyclist safety. When the cyclist age was higher than 41, the casualty was more likely to be severe. The environmental predictors, lighting level, weather and road surface condition were not influential on cyclist casualty severity. However, this result is likely to be unreliable as cycling is a dry weather mode choice and in the absence of cyclist flow data correction for this could not be made.

3. *What is the relative contribution of driver/rider behaviour related contributory factors to cyclist casualty severity?*

   The influence of sudden braking was consistent in both stages of Analysis 2 (England) and the first stage of Analysis 3 (Northumbria). Sudden braking was more likely to contribute in severe casualties than slight. Clearly this behavioural factor is emerging as significant within larger data sets (across England and across more years in the Northumbria analysis).The other contributor variables in severe casualties were failed to judge other person's path or speed, junction restart and failed to look properly but their influence was not consistent in the complete study. In summary, it was seen that the variables, which had an impact on cyclist casualty severity, were speed related predictors in the analysis for England. Surprisingly, the analysis of Belgium showed that cyclist's non respect of the priority to drivers increased the probability of casualty severity.

4. *What is the consistency of the results between tangential and radial roundabouts based on a comparative analysis?*

   Analysis 3 was conducted to identify the consistency between tangential and radial roundabouts based on a comparative analysis. Higher speed limit increased the probability of casualty severity at tangential roundabouts (England), most likely because this type of design allows drivers to enter the roundabout with higher speed to increase the junction capacity with the wider approach geometry. On the other hand, speed limit was not a statistically significant variable in the model for the radial design (Belgium), which is designed to reduce the speed of the driver given its narrow entry geometry. Therefore, as expected the impact of speed limit on casualty severity was not consistent between tangential and radial designs. The results showed that older cyclists were more likely to be involved in KSI rather than slight casualty crashes and this was consistent for both countries.

   An important result emerging from the chi square test was that all non-behavioural variables such as cyclist age group and gender, lighting, road surface condition and more importantly speed limit were all found to be statistically different between Northumbria (representative of England) and Belgium for both slight and KSI casualty severity. On the other hand of the three behavioural variables, only junction over-shoot emerged as having statistically significantly different influence in casualty severity in the two countries. This suggests that driver/cyclist interaction and behaviour in cities in the two countries is generally similar whilst speed limit, sociodemographic characteristics of cyclists and environmental conditions are specific for each country. Therefore, further research specifically into driver-rider behaviour influence of cyclist casualty severity in different countries would be useful. However, considering other variables in particular geometric design, in the same model is limited due to the different protocols applied in recording details of accidents in the different countries.

5. *What is the appropriate statistical method to analyse the safety impact of variables on cyclist casualty severity?*

Logistic regression was found to be the most suitable method to investigate the impact of influencing variables on casualty severity. However, this type of model was found to be very sensitive and the results prone to be bias when the number of observations was low. The results in this thesis showed that the statistical significances of some variables changed when the study area was extended or casualty records over a larger period were considered. As expected, this suggests that the reliability of the results improves as the number of observations increases.

This research increased the number of observations by expanding the study area to include more cities, local authority areas and two counties. However, the nested grouped data cannot be applied in Logistic Regression with one level because a regional residual occurs. Therefore, Multilevel Logistic Regression was used so that the regional residual could be included in the model. The Multilevel Logistic Regression analysis in this study showed that regional residual for 21 cities nested data across England was not statistically significant. This was expected because the local authorities in England are required by the government to use the same design manual to maintain consistency across the UK highway networks. Additionally, the police use a standard protocol for recording casualty data. However, in the comparative analysis of England with Belgium the regional residual was highly statistically significant. This was because of the difference of design and data recording in two different countries, with tangential for England and radial for Belgium, as well as different data recording protocol.

Table 8-1- Summary of the Main Findings of This Thesis

Descriptive Statistic: In Analysis 1, 2 and 3, the dataset is mixed with continuous and categorical variables.

Normality Test: In Analysis 1, 2 and3, variables are not normally distributed.

Analysis 1: Statistically significantly correlated variables were number of lanes on approach and half width on approach, and PCA created approach capacity and size of roundabout

	Variable Groups	**Analysis1 (Ch5)**	**Analysis 2 (Ch6)**		**Analysis 3 (Ch7)**	
		MLR Northumbria (2011-2016)	***MLR Northumbria (2011-2016)***	***MLR England (2011-2016)***	***MLR Northumbria (2005-2016)***	***MLR Belgium (2005-2016)***
Statistically Significant Variables at 95% Confidence Level	Geometric Design Parameters and Speed Limit	• Speed limit (O.R.= 1.02) → • Number of lanes on approach (O.R.= 4.97) → • Entry path radius (O.R.= 1.04) → • Approach capacity (O.R.= 1.86)	• Speed limit (O.R.= 1.02) → • Number of lanes on approach (O.R.= 1.95) • Entry path radius (O.R.= 1.04)	• Speed limit (O.R.= 1.02) →	• Speed limit (O.R.= 1.01)	
	Sociodemographic Characteristics of Cyclist			• Cyclist age group (O.R.= 1.20) →	• Cyclist age (O.R.= 1.02) →	• Cyclist age (O.R.= 1.02)
	Driver and Cyclist Behaviour Related Contributory Factors		• Sudden braking (O.R.= 7.14) →	• Sudden braking (O.R.= 12.83) → • Failed to judge other person's path or speed (O.R.= 1.70) → • Junction restart (O.R.= 1.94) • Failed to look properly (O.R.= 2.41)	• Sudden braking (O.R.= 3.02) • Failed to judge other person's path or speed (O.R.= 0.51)	• Cyclist's non respect of the priority (O.R.= 2.71)
			Multilevel Logistic Regression: City residual for England was not statistically significant (0.07 with p-value 0.10) →		***Multilevel Logistic Regression with associated data:*** Country residual for England and Belgium was statistically significant (0.25 with p-value 0.00)	
					Three-way chi square test of independence: 1-Statistically significant relationship was found between variables (cyclist age group, cyclist gender, speed limit, lighting level, road surface condition and junction overshoot) and countries among both slight and KSI casualties. 2- Statistically significant relationship was found between speed limit and casualty severity in Northumbria and cyclist age group and casualty severity for Belgium.	

PCA: Principal Component Analysis
MLR: Multiple Logistic Regression
O.R.: Odds ratio

8.4 Secondary Findings

Along with the main findings above, other interesting results emerged from the researched proposed in this thesis. These findings are as follows:

1. The originality of the research conducted in this thesis was identified by reviewing a wide number of former studies available in the literature which also revealed that the attention on cyclist safety was not in a sufficient level. This is quite alarming given that local authorities are investing in schemes to promote sustainable transport and cycling is increasing year on year.
2. The data used in this thesis is mixed in the sense that it includes continuous and categorical variables and none of the variables are normally distributed. This led to the need to apply non-parametric analytical techniques in the analysis.
3. Principal component was used to reduce the number of dimensions by grouping together correlated variables. Two components were identified namely approach capacity and size of roundabout. Although the variables in the same group, except number of lanes and half width on approach, were not statistically significantly correlated in correlation analysis they were statistically significantly grouped with high Cronbach's alpha. Furthermore, the PCA component approach capacity emerged as statistically significant in the probability prediction model.

8.5 Limitations of the Study

In general, a limitation of cyclist safety studies is that crash data records are uncompleted. Hels and Bekkevold (2007) found that 75% of the casualty records reported in hospital were not recorded in police crash reports. Most studies rely on police records as the data source for analysis. The study in this thesis also used police records in England (STATS19) and Belgium (VIAS) and the author is well aware of the limitation of unreported cyclist casualties.

STATS19 crash database, which is recorded by local processing authorities through co-operation between (police and local councils), is available for public consumption by permission for the UK DfT. The accumulated data is widely used in research and in designing road safety measures, and so therefore, the data should be collected in a high accuracy (DfT, 2011). STATS21 describes how the local processing authority to increase the accuracy of recording and creating database by applying checking the validity protocols error (TS, 2013). Despite the enormous attention on accuracy in the data, there is a limitation in recording contributory factors. Contributory factors started to be recorded since year 2005 and are to a certain extent based on a police officer's subjective assessment of whether they believe the specific factor has contributed to the crash or not (Rolison *et al.*, 2018). Moreover, contributory factors for England data was not assigned individually to the cyclist or driver. The factor represents either cyclist or driver or both in the records. This presents a limitation into interpret action of the behavioural impact on casualty severity in the England analysis.

Measuring entry path radius is another limitation in the dataset. A standard way to measure the entry path radius is not described in specific terms in the DMRB (DfT, 2007). As described earlier in Section 4.2.2, entry path radius, the shortest path when entering a roundabout, is measured based on engineering judgement with due consideration of the minimum distance from edges of the roads. Whilst this may lead to error these when kept to minimum as all measurements were carried out by Akgun to maintain consistency.

Traffic flow is an important parameter because at higher cyclist and driver flow the number of potential encounters increases. This study cyclist and driver flow at roundabouts could not be considered as they were not available at the location where the casualty was recorded. The only available data related to traffic flow was aggregated over periods of the day (peak, inter peak and off peak) and only for Newcastle and Gateshead, with the number of observations being limited to only the strategic highway network. This data would have improved the prediction models.

8.6 Policy Implications of the Study

This study offers scientific evidence-based recommendations for policy makers in the future. Roundabout geometric design is planned based on two main requirements: reducing the delay to traffic and improving the safety for all road users; however, a compromise is needed between these two, particularly at roundabouts with high level of cycle flows.

In order to reduce cyclist casualty severity, the study reported in this thesis recommends that:

1. Highway engineers should not create a give way roundabout layout with mixed traffic with an entry path radius exceeding 81 metres. It is acknowledged that there are several existing roundabouts and most of them do not meet this recommendation. Changing the geometric design of existing roundabouts might not be an economical solution (Montella, 2011) and this raises doubt as to the practicalities of modifying existing layouts. Therefore, reducing speed limit or pavement treatments on the approach to roundabouts to decrease the driver/rider speed is recommended when entry path radius exceeds 81 metres.

2. Tighter approach and entry geometries to reduce speed are accommodated at roundabouts involved in cyclist schemes.

3. Adopt a radial designing in the design of new give way roundabouts with mixed traffic. It is acknowledged that in some situations, a tangential design with wider approach and entry geometry may be essential to increase the capacity of roundabout and such tangential design results in minimal speed reduction (Parkin, 2018). Therefore, in these situations, efforts should be made to divert cyclists onto alternative routes, and when it is not possible, safety for cyclists can be improved by applying pavement markings on approach for speed reduction and channelization of cycle paths should be considered.

4. Geometric parameters and speed are objective measures which can be controlled in the design if consistent with policy. However, human behaviour is subjective and beyond the design process. The results of this study suggest that driver/cyclist contributory factors failing to reduce speed or look properly, non-respect of giving priority to other road users and failing to judge other's path or speed all play a part in increasing the severity of cyclist casualty. Therefore, measures should be taken by policy makers to try to improve driver's awareness of other more vulnerable users by introducing yielding road marks/signs on the roadside posters and include cycle awareness in the driver training and formal practical. However, the results presented in this thesis suggest that it is not driver's but cyclist's failure of yielding that increase the casualty severity. Given that the proportion of cycling is increasing in traffic every year, it is recommended that ideally formal training for a

cycling licence should be required before being able to cycle in mixed traffic. However, the author of this study is aware that this also may present a barrier for encouraging cycling. Therefore, it is suggested that a basic safety and rules of the road booklet should be introduced with a basic training course made available free of charge for cyclists.

8.7 Recommendations for Future Research

Whilst this study has successfully contributed new knowledge, the results suggest several recommendations for future research:

- Cyclist and motor vehicle flow needs to be considered in further research because this will help to enhance the interpretation of the descriptive statistics and regression models.
- Tangential roundabouts should have a tighter geometry at approach and entry locations; however, they would have less capacity for motor vehicles and congestion will increase. Further research should be conducted to identify the optimum balance of approach capacity and safety for cyclists as well as given due consideration of environmental issues. Novel roundabout designs which introduce a degree of segregation of cyclist flows should be explored.
- Comparative analysis should be carried out to include several countries to increase the statistical outcome of the regional residual. This will provide a more reliable generic model which may include additional variables. However, bearing in mind limitations in data availability, statistical techniques such as Monte Carlo simulation can be applied to generate the required data.
- Given this, research has recommended that a cycling training, and ideally requirement of a cycling licence, handbook should be introduced and acknowledged. However, this is a barrier to cycling, and so a further study should be conducted to investigate the level of support for cycling training in the nature of content and format of a handbook, the level of acceptance of licence and the extent to which the introduction of such regulation would detect the public from choosing cycling as a travel mode.

Appendices

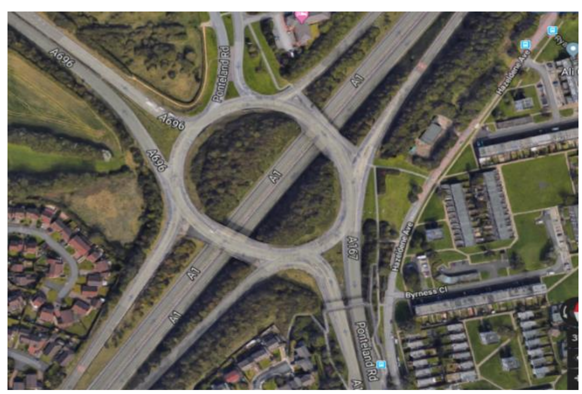
Appendix A

Types of Roundabouts

Normal roundabout: Normal roundabout has minimum 4 m central island diameter with single or multiple approach and entry lanes. Additionally, it usually has flared entry and exit in order to increase the entry and exit capacity of roundabout. Number of legs can be three or more however, UK roundabout design guideline claims that if the number is higher than four, the probability of having higher circulatory speed increases which may result in safety problems. In this case, signalisation is recommended to be used at roundabouts (DfT, 2007).

Mini roundabout: Mini roundabout has 1 to 4 m road marking centre of island which should be painted as unavoidable by drivers (DfT, 2007).

Grade separated roundabout: Grade separated roundabout is mainly used at motorway junctions. Coming lanes are located with a level of road. This type of roundabout can also be used in order to link the flyovers, underpasses and multiple level intersections. Usually, the speed limit is higher at these roundabouts (DfT, 2007).



Double roundabout: Double roundabout (also called dumb-bell) consists of two roundabouts which are linked together with a lane of junction. These two roundabouts might be normal, compact or mini roundabouts. Double roundabout is used for improving segregated junctions, joining two parallel roads, reducing the circulating flow and increasing the capacity (DfT, 2007).

Ring junction (Magic roundabout): Magic roundabout is applied when the number of arms are higher than four. There is a big central island and mini roundabouts around located at entry/exit of each arms. The main aim for this design is maintain the traffic movement at central island through clockwise and at mini roundabouts through anti clockwise for gain higher capacity. This type of roundabout only was applied in the UK (Tollazzi, 2015).

Roundabout with a transitional central island (Hamburger roundabout): There is a straight through carriageway which divides the central island in two splits. The central island should be minimum 60m (Tollazzi, 2015).

Roundabout with segregated right-hand turning lanes: There is a segregated lane at the approach arm which leads vehicles to turn right hand side. In this condition, traffic right hand turn is separated from the roundabout. This type of roundabout also is called ‘bypasses’ or ‘free-flow lanes’ or a ‘channelized turn lanes’ (Tollazzi, 2015).

Turbo roundabout: Some directed traffic flows are separated with multiple centres. There is a physical separated lane at the inner circle of the central island (Tollazzi, 2015).

Dog-bone roundabout: This roundabout has a similar method as double roundabout. The main difference is that central islands of two roundabouts are connected.

Appendix B

The statistical methods used in former studies of cyclist safety at roundabouts

Linear Regression Model:

Linear Regression analysis creates quantitative dependent variable's (Y) distribution based on one or more quantitative independent variables (x1, x2, x3… xn) (n= number of independent variables) (Fox, 1997). It is the basic statistical model of predictive analysis, which gives the relationship between one dependent variable and two or more independent variables. The basic formula of the model is Y = A + BX, where Y is estimated variables (or predicted variables), A is constant, B is regression coefficient and X is independent variables. The result of the estimated variable gives a linear line in plot (Statistics Solutions, 2016).

There are two types of linear regression: Simple Linear Regression and Multiple Linear Regression. The Simple Linear Regression is used for investigating the response of dependent variable Y, changes with the value of only one independent variable X.

Simple Linear Regression equation will take the following form (Olive, 2017):

$$Y_i = \beta_1 + \beta_2 X_i + e_i$$

Where:

Y_i: Quantitative dependent variable,
X_i: Independent variable
β_1: Constant of the unknowns
β_2: Constant of the independent variable
e_i: The error of the regression
Note: The random variables are Y_i and e_i; X_i is the known constant and unknown constants are β_1 and β_2.

When the number of independent variables is more than one, Multiple Linear Regression is applied. It is mainly used for forecasting an impact, trend forecasting and casual analysis. Dependent variable (interval or ratio) should be continuous, while independent variables (interval or ratio or dichotomous) might not be continuous. The data needs to be normally distributed. The linear relationship between variables is found (Statistics Solutions, 2016). In this case, the Multiple Linear Regression equation will take the following form:

$$Y_i = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + \ldots + \beta_{i+1} X_i + e_i$$

Where:

Y_i: Quantitative dependent variable
$X_i, X_2 \ldots X_i$: Independent variables
β_1: Constant of the unknowns
$\beta_2, \beta_3 \ldots \beta_{i+1}$: Constants of the independent variable
e_i: The error of the regression

Cyclist safety studies were one of the areas where Multiple Linear Regression is applied (Moller and Hels, 2008; Daniels *el al.*, 2009). Moller and Hels (2008) investigated response of the number of cyclist casualty changes with the cyclist and vehicle flow, cycling facilities and gender. Multiple Linear Regression was applied because the dependent variable, which is the number of cyclist casualty, was a quantitative variable. Daniels *et al.* (2009) carried a similar study and applied Multiple Linear Regression to measure the changes on number of cyclist casualties based on the independent variables, such as year of the casualty happened, built up area, road signal and barrier, cycling facility, road marking and number of lanes. Previous studies show that if the considered dependent variable is a quantitative value, such as number of crashes, Multiple Linear Regression is the suitable model.

<u>Logistic Regression Model:</u>

Logistic Regression is the version of linear regression model when the dependent variable, which is a stochastic event, is binary (dichotomous). It gives prediction results of one dependent binary variable from one or more independent variables (interval or ratio) (Statistics Solutions, 2016).

The equation for Logistic Regression with two independent variables will take the following form (Peng *et al.*, 2002):

$$Logit\ (Y) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2$$

$$\pi = \frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2}}{1 + e^{\alpha+\beta_1 X_1+\beta_2 X_2}}$$

Where:

Π: The probability of event

α: The constant of the unknowns

β_1 and β_2: Regression coefficients of the independent variables which are X_1 and X_2.

Former studies (Hels and Bekkevold, 2007; Daniels *et al.*, 2010; Silvano *et al.*, 2015) on cyclist safety analysis used Logistic Regression since their outcome variable was binary. For instance, Hels and Bekkevold (2007) applied Logistic Regression in order to find the probability of cyclist crash at roundabouts. Following, Daniels *et al.* (2010) investigated the probability of severity of the cyclist casualties at roundabouts based on Logistic Regression outcomes. The final application of Logistic Regression in research of cyclist safety at roundabouts was carried by Silvano *et al.* (2015) to determine the probability of conflict and yielding between vehicles and cyclists. It is clearly seen that Logistic Regression model is suitable for investigating the probability of casualty severity and yielding related research.

<u>Regression to the Mean</u>:

The regression to the mean effect is likely to occur when it was decided to construct a roundabout since increase of crash is more important entity than among others for constructing a roundabout at a specific location (Daniels et al., 2009). Correcting the regression to the mean is critical when there is a relationship between crash history of the entity and the reason why its safety is estimated (Hauer et al., 2009).

It is a method used for effectiveness index calculation of selected group of roundabouts. Effectiveness index reflects the treatment evaluation of the odds-ratio that gives the statistical result of presence or absence of casualties. Effectiveness index is commonly used in regression models in order to determine the risk profiles of the variables which are used in the multivariable model (Harrell, 2001).

This method is mainly used in before and after or comparison studies and Meta-analysis is applied to generate the results (Studies by Brabander & Vereeck, 2007; Daniels et al., 2008; Daniels et al., 2009). Meta-analysis is known as analysing of primary and secondary analysis (Glass, 1976). According to Glass (1976) primary analysis is the application of statistical method on original data and secondary analysis is the reanalysis of the data in order to answer the new questions in the research.

<u>Pearson's Chi-Square</u>:

Pearson's Chi-Square is used for observing the relationship between two categorical variables based on an idea of comparing the observed frequencies in certain categories to expected frequencies in those categories by chance (Field, 2003).The equation of the Pearson's chi-square is given following:

$$X^2 = \sum \frac{(Observed_{ij} - Model_{ij})^2}{Model_{ij}}$$

Where:
i: Represents the rows
j: Represents the columns in the probability table.

Polders *et al.* (2015) applied Pearson's chi-square in order to observe the impact of independent variables, such as roundabout segments, weather, light condition, cycling facilities, and number of lanes, on distribution of cyclist and moped crashes. This test is applicable for investigating the impact of each categorical variables individually on cyclist casualties, however observing the impact of several variables on the casualties should be carried by regression modelling.

<u>Association Rules:</u>

The only example of using association rules was applied in the study by Montella (2011). In this study (Montella, 2011), the interdependencies between contributory factors and the relationship between these contributory factors and different crash types were observed. Relative frequency of the number of crash contributory factors is given in the results. This method is a statistical analysis for identification of sets of factors which exist in a given case (Montella, 2011).

Empirical Bayes:

Empirical Bayes (EB) model is mainly used in before and after studies (Persaud et al., 2001; Daniels et al., 2009) in order to observe the level of crash number change after converting the junction from one type to another or adding a safety facility. In detail, EB is used in road safety estimation for increasing the precision of estimated results and correcting the biased regression to the mean (Hauer et al., 2009; Daniels et al., 2009). Precision becomes essentially important if the usual estimate is unreliable to apply (Hauer et al., 2009). A sensible estimate is a mixture of two main clues of EB model for safety related research (Hauer et al., 2009): i) crashes count is not the only entity, ii) the knowledge in similar entities. Therefore, EB model is a combination of information contained in crashes counts and in knowing the safety of similar entities. The only application of EB model in cyclist safety at roundabouts was carried by Daniels et al. (2009) in order to investigate the impact of different types of cycling facilities based on a before and after study. This study (Daniels et al., 2009) observed the safety change regarding relationship between estimated value for the effectiveness per location and some known characteristics of the roundabout location.

Appendix C

Principal Component Analysis without logarithmic transformation

FILE='F: \Method and analysis\1 Preliminary analysis\Preliminary Analysis. sav'.
DATASET NAME DataSet1 WINDOW=FRONT. FACTOR
/VARIABLES Number of Lanes- Approach Half Width- Entry Path Radius- Arms- Flare Lane- Type of Roundabout- Circulating lanes
/MISSING LISTWISE
/PRINT UNIVARIATE INITIAL KMO EXTRACTION ROTATION
/FORMAT SORT
/PLOT EIGEN ROTATION
/CRITERIA MINEIGEN (1) ITERATE (25)
/EXTRACTION PC
/CRITERIA ITERATE (25)
/ROTATION PROMAX (4)
/METHOD=CORRELATION.

PCA
[DataSet1] F:\Method and analysis\1 Preliminary analysis\Preliminary Analysis.sav

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Number of Lanes	1.32	.500	439
Approach Half Width	5.25	1.787	439
Entry Path Radius	66.94	21.379	439
Arms	4.06	.675	439
Flare Lane	1.64	.611	439
Type of Roundabout	1.09	.431	439
Circulating Lanes	1.40	.522	439

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.729
Bartlett's Test of Sphericity Approx. Chi-Square	1044.5
Degrees of Freedom	21
Sig.	.000

Communalities

	Initial	Extraction
Number of Lanes	1.000	.789
Approach Half Width	1.000	.760
Entry Path Radius	1.000	.266
Arms	1.000	.644
Flare Lane	1.000	.685
Type of Roundabout	1.000	.615
Circulating Lanes	1.000	.579

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	2.967	42.379	42.379	2.967	42.379
2	1.372	19.599	61.979	1.372	19.599
3	.891	12.733	74.712		
4	.630	9.002	83.714		
5	.576	8.223	91.937		
6	.386	5.519	97.456		
7	.178	2.544	100.000		

Pattern Matrix[a]

	Component	
	1	2
Number of Lanes	.869	.052
Approach Half Width	.853	.052
Flare Lane	.822	.016
Entry Path Radius	.512	-.358
Arms	-.211	.849
Type of Roundabout	.038	.771
Circulating Lanes	.334	.580

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.[a]

a. Rotation converged in 3 iterations.

Component Correlation Matrix

Component	1	2
1	1.000	.340
2	.340	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

Principal Component Analysis with logarithmic transformation

VARIABLES Log lanes- Log half width- Log entry path radius- Log arms- Log flare lanes- Log type of roundabout – Log circulating lanes
/MISSING LISTWISE
/PRINT UNIVARIATE INITIAL SIG KMO EXTRACTION ROTATION
/PLOT EIGEN ROTATION
/CRITERIA MINEIGEN (1) ITERATE (25)
/EXTRACTION PC
/CRITERIA ITERATE (25)
/ROTATION PROMAX (4)
/SAVE REG (ALL)

/METHOD=CORRELATION.

PCA

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Log lanes	.0940	.14439	439
Log half width	.6982	.13798	439
Log entry path radius	1.7992	.15990	439
Log arms	.6023	.07136	439
Log flare lanes	.1833	.16556	439
Log type of round	.3101	.09521	439
Log circulating lanes	.1187	.15177	439

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.737
Bartlett's Test of Sphericity Approx. Chi-Square	998.9
Degrees of Freedom	21
Sig.	.000

Communalities

	Initial	Extraction
Log lanes	1.000	.770
Log half width	1.000	.760
Log entry path radius	1.000	.310
Log arms	1.000	.638
Log flare lanes	1.000	.676
Log type of round	1.000	.600
Log circulating lanes	1.000	.555

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	2.969	42.417	42.417	2.969	42.417
2	1.339	19.132	61.550	1.339	19.132
3	.885	12.645	74.194		
4	.616	8.801	82.995		
5	.598	8.536	91.531		
6	.398	5.681	97.212		
7	.195	2.788	100.000		

Component Correlation Matrix

Component	1	2
1	1.000	.316
2	.316	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

Pattern Matrix[a]

	Component	
	1	2
Log lanes	.872	.018
Log half width	.857	.044
Log entry path radius	.504	-.444
Log arms	-.134	.831
Log flare lanes	.820	.006
Log type of round	.102	.736
Log circulating lanes	.408	.507

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Appendix D

Orthogonal rotation with Varimax

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.729
Bartlett's Test of Sphericity	Approx. Chi-Square	1044.471
	df	21
	Sig.	.000

Communalities

	Initial	Extraction
Number of Lanes	1.000	.789
Approach Half Width	1.000	.760
Entry Path Radius	1.000	.266
Flare Lane	1.000	.685
Arms	1.000	.644
Type of Roundabout	1.000	.615
Circulating Lanes	1.000	.579

Extraction Method: Principal Component Analysis.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.967	42.379	42.379	2.967	42.379	42.379	2.525	36.075	36.075
2	1.372	19.599	61.979	1.372	19.599	61.979	1.813	25.904	61.979
3	.891	12.733	74.712						
4	.630	9.002	83.714						
5	.576	8.223	91.937						
6	.386	5.519	97.456						
7	.178	2.544	100.000						

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component	
	1	2
Number of Lanes	.845	-.273
Approach Half Width	.830	-.268
Entry Path Radius	.243	-.455
Flare Lane	.777	-.284
Arms	.359	.717
Type of Roundabout	.541	.568
Circulating Lanes	.692	.317

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

Rotated Component Matrix[a]

	Component	
	1	2
Number of Lanes	.863	.212

Approach Half Width	.846	.209
Entry Path Radius	.446	-.259
Flare Lane	.810	.167
Arms	-.072	.799
Type of Roundabout	.161	.768
Circulating Lanes	.422	.634

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Component Transformation Matrix

Component	1	2
1	.850	.526
2	-.526	.850

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Appendix E

Predictive Margins for entry path radius and number of lanes on approach

```
. do "C:\Users\B30517~1.038\AppData\Local\Temp\STD01000000.tmp"
. logit CasualtySeverity NumberofLanes EntryPathRadius
Iteration 0:   log likelihood = -190.94578
Iteration 1:   log likelihood =  -169.7388
Iteration 2:   log likelihood = -167.59707
Iteration 3:   log likelihood = -167.58853
Iteration 4:   log likelihood = -167.58853

Logistic regression                 Number of obs    =       439
LR chi2(2)        =      46.71
Prob > chi2       =     0.0000
Log likelihood = -167.58853                 Pseudo R2        =     0.1223
. margins, at(EntryPathRadius=(20(1)99) NumberofLanes=(1(1)3)) plot
Adjusted predictions                 Number of obs    =       439            Model VCE    : OIM
Expression   : Pr(CasualtySeverity), predict()
Delta-method
Margin   Std. Err.      z    P>z     [95% Conf. Interval]
1     .0167518       .007857     2.13   0.033     .0013524    .0321511
2     .0174103       .0080374     2.17   0.030     .0016573    .0331634
3     .0180944        .00822     2.20   0.028     .0019835    .0342052
4     .0188047   .0084046     2.24   0.025     .0023319    .0352775
5     .0195424   .0085913     2.27   0.023     .0027039     .036381
6     .0203085   .0087797     2.31   0.021     .0031005    .0375164
7     .0211039     .00897     2.35   0.019     .003523    .0386847
8     .0219298   .0091619     2.39   0.017     .0039728    .0398867
9     .0227872   .0093553     2.44   0.015     .0044512    .0411233
10    .0236774   .0095501     2.48   0.013     .0049594    .0423953
11    .0246014   .0097463     2.52   0.012     .0054991    .0437038
12    .0255606   .0099435     2.57   0.010     .0060717    .0450495
13    .0265562   .0101418     2.62   0.009     .0066787    .0464337
14    .0275894   .0103408     2.67   0.008     .0073217    .0478571
15    .0286617   .0105406     2.72   0.007     .0080024     .049321
16    .0297743    .010741     2.77   0.006     .0087224    .0508263
17    .0309288   .0109417     2.83   0.005     .0094834    .0523742
18    .0321266   .0111427     2.88   0.004     .0102872    .0539659
19    .0333691   .0113438     2.94   0.003     .0111357    .0556025
20     .034658   .0115448     3.00   0.003     .0120306    .0572853
21    .0359948   .0117456     3.06   0.002     .0129739    .0590157
22    .0373811    .011946     3.13   0.002     .0139674    .0607949
23    .0388188    .012146     3.20   0.001     .0150131    .0626244
24    .0403094   .0123453     3.27   0.001     .0161129    .0645058
25    .0418547    .012544     3.34   0.001     .0172689    .0664405
26    .0434566   .0127419     3.41   0.001     .0184829    .0684303
27    .0451169    .012939     3.49   0.000     .019757    .0704768
28    .0468375   .0131352     3.57   0.000     .021093    .0725821
29    .0486205   .0133306     3.65   0.000     .0224929     .074748
30    .0504677   .0135252     3.73   0.000     .0239587    .0769766
31    .0523812   .0137192     3.82   0.000     .0254921    .0792703
32    .0543631   .0139126     3.91   0.000     .0270949    .0816313
33    .0564155   .0141057     4.00   0.000     .0287688    .0840623
34    .0585407   .0142989     4.09   0.000     .0305154     .086566
```

35	.0607407	.0144924	4.19	0.000	.0323361	.0891453
36	.0630179	.0146868	4.29	0.000	.0342322	.0918035
37	.0653745	.0148826	4.39	0.000	.036205	.0945439
38	.0678128	.0150806	4.50	0.000	.0382554	.0973702
39	.0703353	.0152815	4.60	0.000	.0403841	.1002865
40	.0729442	.0154863	4.71	0.000	.0425916	.1032968
41	.0756421	.0156961	4.82	0.000	.0448782	.1064059
42	.0784312	.0159122	4.93	0.000	.0472439	.1096186
43	.0813142	.016136	5.04	0.000	.0496882	.1129402
44	.0842935	.0163691	5.15	0.000	.0522106	.1163763
45	.0873715	.0166133	5.26	0.000	.05481	.119933
46	.0905508	.0168706	5.37	0.000	.057485	.1236166
47	.0938339	.0171432	5.47	0.000	.0602338	.127434
48	.0972233	.0174334	5.58	0.000	.0630544	.1313923
49	.1007216	.0177438	5.68	0.000	.0659443	.1354989
50	.1043311	.0180771	5.77	0.000	.0689007	.1397616
51	.1080545	.0184361	5.86	0.000	.0719204	.1441886
52	.1118942	.0188238	5.94	0.000	.0750002	.1487882
53	.1158526	.0192433	6.02	0.000	.0781364	.1535687
54	.1199321	.0196977	6.09	0.000	.0813252	.1585389
55	.124135	.0201903	6.15	0.000	.0845628	.1637072
56	.1284638	.0207241	6.20	0.000	.0878454	.1690823
57	.1329206	.0213023	6.24	0.000	.0911689	.1746723
58	.1375077	.0219279	6.27	0.000	.0945298	.1804855
59	.142227	.0226038	6.29	0.000	.0979244	.1865297
60	.1470808	.0233328	6.30	0.000	.1013494	.1928121
61	.1520708	.0241173	6.31	0.000	.1048018	.1993398
62	.1571989	.0249597	6.30	0.000	.1082788	.2061189
63	.1624668	.025862	6.28	0.000	.1117782	.2131553
64	.167876	.026826	6.26	0.000	.1152981	.220454
65	.1734281	.0278531	6.23	0.000	.1188369	.2280192
66	.1791242	.0289447	6.19	0.000	.1223937	.2358548
67	.1849656	.0301015	6.14	0.000	.1259677	.2439635
68	.1909532	.0313242	6.10	0.000	.1295589	.2523474
69	.1970877	.032613	6.04	0.000	.1331673	.261008
70	.2033697	.033968	5.99	0.000	.1367937	.2699458
71	.2097997	.0353888	5.93	0.000	.1404389	.2791605
72	.2163777	.0368749	5.87	0.000	.1441042	.2886513
73	.2231038	.0384254	5.81	0.000	.1477914	.2984162
74	.2299776	.0400391	5.74	0.000	.1515024	.3084527
75	.2369985	.0417146	5.68	0.000	.1552395	.3187575
76	.2441659	.0434501	5.62	0.000	.1590052	.3293265
77	.2514785	.0452437	5.56	0.000	.1628024	.3401546
78	.2589351	.0470932	5.50	0.000	.1666341	.3512361
79	.2665341	.048996	5.44	0.000	.1705036	.3625646
80	.2742736	.0509494	5.38	0.000	.1744145	.3741326
81	.0355552	.0174336	2.04	0.041	.0013859	.0697245
82	.0369253	.0178227	2.07	0.038	.0019934	.0718572
83	.0383461	.0182154	2.11	0.035	.0026446	.0740476
84	.0398193	.0186113	2.14	0.032	.0033418	.0762968
85	.0413467	.0190102	2.17	0.030	.0040873	.078606
86	.04293	.0194118	2.21	0.027	.0048835	.0809765
87	.0445711	.0198157	2.25	0.024	.005733	.0834093
88	.046272	.0202217	2.29	0.022	.0066382	.0859057
89	.0480344	.0206292	2.33	0.020	.007602	.0884669
90	.0498605	.0210379	2.37	0.018	.008627	.0910941

91	.0517523	.0214474	2.41	0.016	.0097161	.0937885
92	.0537118	.0218573	2.46	0.014	.0108722	.0965513
93	.0557411	.0222671	2.50	0.012	.0120983	.0993838
94	.0578423	.0226764	2.55	0.011	.0133975	.1022872
95	.0600178	.0230846	2.60	0.009	.0147729	.1052627
96	.0622697	.0234912	2.65	0.008	.0162278	.1083116
97	.0646002	.0238958	2.70	0.007	.0177653	.1114351
98	.0670118	.0242978	2.76	0.006	.019389	.1146345
99	.0695066	.0246966	2.81	0.005	.0211022	.1179111
100	.0720872	.0250918	2.87	0.004	.0229082	.1212662
101	.0747559	.0254827	2.93	0.003	.0248107	.1247012
102	.0775152	.0258689	3.00	0.003	.0268131	.1282173
103	.0803674	.0262498	3.06	0.002	.0289188	.131816
104	.0833151	.0266247	3.13	0.002	.0311316	.1354987
105	.0863608	.0269934	3.20	0.001	.0334548	.1392668
106	.0895069	.0273551	3.27	0.001	.035892	.1431219
107	.0927561	.0277094	3.35	0.001	.0384465	.1470656
108	.0961107	.028056	3.43	0.001	.0411219	.1510994
109	.0995733	.0283943	3.51	0.000	.0439214	.1552252
110	.1031464	.0287241	3.59	0.000	.0468482	.1594446
111	.1068326	.029045	3.68	0.000	.0499054	.1637597
112	.1106342	.0293568	3.77	0.000	.0530959	.1681725
113	.1145538	.0296593	3.86	0.000	.0564225	.172685
114	.1185937	.0299526	3.96	0.000	.0598877	.1772996
115	.1227563	.0302365	4.06	0.000	.0634938	.1820187
116	.1270439	.0305113	4.16	0.000	.0672428	.186845
117	.1314589	.0307773	4.27	0.000	.0711365	.1917814
118	.1360034	.0310349	4.38	0.000	.0751761	.1968308
119	.1406796	.0312847	4.50	0.000	.0793628	.2019965
120	.1454895	.0315274	4.61	0.000	.0836969	.207282
121	.150435	.031764	4.74	0.000	.0881788	.2126913
122	.1555181	.0319956	4.86	0.000	.0928079	.2182283
123	.1607404	.0322236	4.99	0.000	.0975834	.2238974
124	.1661036	.0324495	5.12	0.000	.1025037	.2297034
125	.1716091	.0326751	5.25	0.000	.107567	.2356512
126	.1772583	.0329026	5.39	0.000	.1127704	.2417462
127	.1830524	.0331341	5.52	0.000	.1181107	.2479941
128	.1889924	.0333723	5.66	0.000	.123584	.2544008
129	.1950791	.0336197	5.80	0.000	.1291857	.2609725
130	.2013132	.0338795	5.94	0.000	.1349106	.2677157
131	.2076951	.0341547	6.08	0.000	.1407531	.2746371
132	.214225	.0344488	6.22	0.000	.1467067	.2817434
133	.220903	.0347651	6.35	0.000	.1527646	.2890414
134	.2277289	.0351073	6.49	0.000	.1589197	.296538
135	.2347021	.0354791	6.62	0.000	.1651644	.3042398
136	.241822	.0358839	6.74	0.000	.1714908	.3121531
137	.2490876	.0363254	6.86	0.000	.1778911	.320284
138	.2564976	.036807	6.97	0.000	.1843573	.3286379
139	.2640505	.0373318	7.07	0.000	.1908815	.3372196
140	.2717446	.037903	7.17	0.000	.1974562	.346033
141	.2795777	.0385229	7.26	0.000	.2040742	.3550812
142	.2875474	.0391939	7.34	0.000	.2107287	.3643661
143	.2956511	.0399178	7.41	0.000	.2174137	.3738886
144	.3038858	.0406958	7.47	0.000	.2241235	.383648
145	.3122481	.0415286	7.52	0.000	.2308536	.3936426
146	.3207345	.0424164	7.56	0.000	.2375998	.4038692

147	.3293411	.043359	7.60	0.000	.244359	.4143232
148	.3380637	.0443554	7.62	0.000	.2511289	.4249986
149	.3468979	.045404	7.64	0.000	.2579077	.4358881
150	.3558388	.046503	7.65	0.000	.2646947	.4469829
151	.3648814	.0476498	7.66	0.000	.2714896	.4582732
152	.3740204	.0488414	7.66	0.000	.2782929	.4697478
153	.3832501	.0500746	7.65	0.000	.2851058	.4813944
154	.3925648	.0513454	7.65	0.000	.2919296	.4932
155	.4019583	.0526499	7.63	0.000	.2987663	.5051502
156	.4114243	.0539836	7.62	0.000	.3056183	.5172303
157	.4209564	.055342	7.61	0.000	.3124881	.5294246
158	.4305477	.05672	7.59	0.000	.3193785	.5417169
159	.4401915	.0581129	7.57	0.000	.3262923	.5540906
160	.4498806	.0595154	7.56	0.000	.3332326	.5665285
161	.0738794	.0442967	1.67	0.095	-.0129407	.1606994
162	.0766089	.0453707	1.69	0.091	-.012316	.1655339
163	.0794307	.0464593	1.71	0.087	-.0116278	.1704892
164	.0823471	.047562	1.73	0.083	-.0108727	.1755669
165	.0853607	.0486784	1.75	0.080	-.0100472	.1807686
166	.0884739	.0498079	1.78	0.076	-.0091478	.1860957
167	.0916893	.05095	1.80	0.072	-.0081708	.1915495
168	.0950094	.0521039	1.82	0.068	-.0071124	.1971313
169	.0984367	.0532691	1.85	0.065	-.0059689	.2028422
170	.1019736	.0544448	1.87	0.061	-.0047362	.2086835
171	.1056228	.0556302	1.90	0.058	-.0034104	.214656
172	.1093867	.0568245	1.92	0.054	-.0019873	.2207607
173	.1132677	.0580269	1.95	0.051	-.0004629	.2269982
174	.1172682	.0592363	1.98	0.048	.0011672	.2333692
175	.1213907	.0604519	2.01	0.045	.0029071	.2398743
176	.1256375	.0616727	2.04	0.042	.0047613	.2465137
177	.1300109	.0628975	2.07	0.039	.0067341	.2532877
178	.1345131	.0641253	2.10	0.036	.0088298	.2601964
179	.1391463	.065355	2.13	0.033	.0110528	.2672397
180	.1439125	.0665853	2.16	0.031	.0134076	.2744174
181	.1488138	.0678151	2.19	0.028	.0158986	.281729
182	.153852	.0690431	2.23	0.026	.01853	.289174
183	.1590289	.070268	2.26	0.024	.0213061	.2967516
184	.1643461	.0714885	2.30	0.022	.0242313	.3044609
185	.1698053	.0727032	2.34	0.020	.0273097	.3123009
186	.1754077	.0739107	2.37	0.018	.0305453	.3202701
187	.1811546	.0751098	2.41	0.016	.0339421	.3283671
188	.1870471	.0762989	2.45	0.014	.037504	.3365902
189	.1930861	.0774767	2.49	0.013	.0412346	.3449376
190	.1992722	.0786417	2.53	0.011	.0451374	.3534071
191	.2056061	.0797925	2.58	0.010	.0492157	.3619964
192	.2120879	.0809276	2.62	0.009	.0534727	.3707031
193	.2187179	.0820457	2.67	0.008	.0579112	.3795245
194	.2254957	.0831453	2.71	0.007	.0625339	.3884576
195	.2324212	.0842251	2.76	0.006	.067343	.3974993
196	.2394935	.0852836	2.81	0.005	.0723407	.4066463
197	.2467119	.0863196	2.86	0.004	.0775286	.4158951
198	.2540752	.0873316	2.91	0.004	.0829083	.425242
199	.2615819	.0883185	2.96	0.003	.0884809	.4346829
200	.2692303	.0892789	3.02	0.003	.0942469	.4442137
201	.2770185	.0902117	3.07	0.002	.1002069	.4538301
202	.2849441	.0911156	3.13	0.002	.1063608	.4635275

203	.2930046	.0919897	3.19	0.001	.1127081	.4733011
204	.3011971	.0928328	3.24	0.001	.1192482	.483146
205	.3095183	.0936439	3.31	0.001	.1259797	.493057
206	.3179648	.0944221	3.37	0.001	.132901	.5030287
207	.3265329	.0951664	3.43	0.001	.1400101	.5130556
208	.3352183	.0958761	3.50	0.000	.1473046	.523132
209	.3440167	.0965504	3.56	0.000	.1547815	.533252
210	.3529235	.0971885	3.63	0.000	.1624376	.5434094
211	.3619336	.0977897	3.70	0.000	.1702693	.5535979
212	.3710419	.0983535	3.77	0.000	.1782725	.5638113
213	.3802428	.0988793	3.85	0.000	.1864428	.5740427
214	.3895305	.0993667	3.92	0.000	.1947754	.5842856
215	.3988991	.0998151	4.00	0.000	.2032652	.594533
216	.4083423	.1002241	4.07	0.000	.2119067	.6047779
217	.4178536	.1005934	4.15	0.000	.2206942	.6150131
218	.4274265	.1009227	4.24	0.000	.2296216	.6252313
219	.437054	.1012117	4.32	0.000	.2386827	.6354253
220	.4467291	.1014602	4.40	0.000	.2478708	.6455874
221	.4564448	.1016679	4.49	0.000	.2571793	.6557103
222	.4661937	.1018348	4.58	0.000	.2666012	.6657862
223	.4759686	.1019606	4.67	0.000	.2761294	.6758077
224	.4857618	.1020453	4.76	0.000	.2857566	.685767
225	.495566	.1020889	4.85	0.000	.2954755	.6956565
226	.5053737	.1020912	4.95	0.000	.3052787	.7054687
227	.5151772	.1020522	5.05	0.000	.3151585	.7151958
228	.524969	.101972	5.15	0.000	.3251075	.7248305
229	.5347416	.1018506	5.25	0.000	.335118	.7343652
230	.5444877	.1016882	5.35	0.000	.3451826	.7437928
231	.5541998	.1014847	5.46	0.000	.3552935	.7531061
232	.5638707	.1012403	5.57	0.000	.3654433	.7622981
233	.5734933	.1009553	5.68	0.000	.3756245	.771362
234	.5830605	.1006298	5.79	0.000	.3858298	.7802912
235	.5925656	.100264	5.91	0.000	.3960517	.7890795
236	.6020019	.0998584	6.03	0.000	.4062831	.7977208
237	.611363	.0994132	6.15	0.000	.4165168	.8062092
238	.6206427	.0989287	6.27	0.000	.4267459	.8145394
239	.6298349	.0984056	6.40	0.000	.4369636	.8227063
240	.638934	.0978441	6.53	0.000	.4471631	.830705

Variables that uniquely identify margins: Entry Path Radius Number of Lanes.
end of do-file
.

Appendix F

Predictive margins for approach capacity

```
. do "C:\Users\B30517~1.038\AppData\Local\Temp\STD01000000.tmp"
. logit Casualty Severity Approach Capacity

Iteration 0:   log likelihood = -190.94578
Iteration 1:   log likelihood = -181.69855
Iteration 2:   log likelihood = -181.21238
Iteration 3:   log likelihood = -181.21195
Iteration 4:   log likelihood = -181.21195

Logistic regression                    Number of obs     =        439
                                       LR chi2(1)        =      19.47
                                       Prob > chi2       =     0.0000
Log likelihood = -181.21195            Pseudo R2         =     0.0510

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |    .072304   .0173694     4.16   0.000     .0382606    .1063473
          2  |   .0760985   .0174569     4.36   0.000     .0418836    .1103134
          3  |   .0800749   .0175188     4.57   0.000     .0457386    .1144111
          4  |   .0842401   .0175554     4.80   0.000     .0498322    .1186481
          5  |   .0886012   .0175676     5.04   0.000     .0541692    .1230331
          6  |    .093165   .0175575     5.31   0.000      .058753    .1275771
          7  |   .0979387    .017528     5.59   0.000     .0635845    .1322929
          8  |   .1029292   .0174835     5.89   0.000     .0686622    .1371963
          9  |   .1081436   .0174301     6.20   0.000     .0739811     .142306
         10  |   .1135886   .0173759     6.54   0.000     .0795324    .1476447
         11  |   .1192711   .0173311     6.88   0.000     .0853028    .1532394
         12  |   .1251977   .0173085     7.23   0.000     .0912737    .1591218
         13  |    .131375   .0173236     7.58   0.000     .0974213    .1653286
         14  |    .137809   .0173945     7.92   0.000     .1037163    .1719016
         15  |   .1445056   .0175417     8.24   0.000     .1101245    .1788867
         16  |   .1514705   .0177874     8.52   0.000     .1166078    .1863332
         17  |   .1587089   .0181545     8.74   0.000     .1231266    .1942911
         18  |   .1662253   .0186655     8.91   0.000     .1296417     .202809
         19  |   .1740241   .0193405     9.00   0.000     .1361175    .2119308
         20  |   .1821089   .0201966     9.02   0.000     .1425243    .2216936
         21  |   .1904827   .0212466     8.97   0.000     .1488402    .2321252
         22  |   .1991478   .0224983     8.85   0.000      .155052    .2432436
         23  |   .2081057   .0239552     8.69   0.000     .1611543    .2550571
         24  |   .2173572   .0256168     8.48   0.000     .1671493    .2675651
         25  |   .2269021   .0274789     8.26   0.000     .1730445    .2807597
         26  |   .2367394   .0295351     8.02   0.000     .1788517    .2946271
         27  |    .246867   .0317769     7.77   0.000     .1845854    .3091486
         28  |   .2572818   .0341945     7.52   0.000     .1902618    .3243018
         29  |   .2679796    .036777     7.29   0.000      .195898    .3400613
         30  |   .2789552   .0395128     7.06   0.000     .2015116    .3563988
         31  |   .2902021   .0423893     6.85   0.000     .2071206    .3732837
```

32	.3017128	.0453937	6.65	0.000	.2127427	.3906828
33	.3134783	.0485123	6.46	0.000	.218396	.4085606
34	.3254888	.0517308	6.29	0.000	.2240982	.4268794
35	.3377331	.0550346	6.14	0.000	.2298672	.4455989
36	.3501988	.0584082	6.00	0.000	.2357209	.4646767
37	.3628725	.0618357	5.87	0.000	.2416768	.4840682
38	.3757397	.0653007	5.75	0.000	.2477527	.5037268
39	.3887848	.0687865	5.65	0.000	.2539658	.5236038
40	.4019911	.0722758	5.56	0.000	.2603331	.5436491
41	.4153411	.0757514	5.48	0.000	.2668712	.563811
42	.4288166	.0791956	5.41	0.000	.273596	.5840372
43	.4423985	.0825911	5.36	0.000	.2805228	.6042741
44	.4560671	.0859206	5.31	0.000	.2876659	.6244683
45	.4698021	.0891669	5.27	0.000	.2950382	.644566
46	.483583	.0923135	5.24	0.000	.3026518	.6645143

.

end of do-file

.

Appendix G

Predictive margins for cyclist age and cyclist’s non respect for the priority

. do "C:\Users\B30517~1.038\AppData\Local\Temp\STD01000000.tmp"

. logit casualty cyclist age cyclist’s non respect of the priority

```
Iteration 0:   log likelihood = -224.72877
Iteration 1:   log likelihood = -217.29345
Iteration 2:   log likelihood = -216.32512
Iteration 3:   log likelihood = -216.32308
Iteration 4:   log likelihood = -216.32308

Logistic regression                     Number of obs     =        924
                                        LR chi2(2)        =      16.81
                                        Prob > chi2       =     0.0002
Log likelihood = -216.32308             Pseudo R2         =     0.0374
```

. margin, at (cyclist age=(0(1)99) cyclist’s non respect of the priority =(0(1)1)) plot

```
Adjusted predictions          Number of observation     =        924          Model VCE    : OIM


------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |      .0229   .0076497     2.99   0.003      .007907    .0378931
          2  |   .0621425   .0306405     2.03   0.043     .0020882    .1221969
          3  |   .0234412   .0076904     3.05   0.002     .0083683     .038514
          4  |   .0635507   .0310873     2.04   0.041     .0026207    .1244807
          5  |   .0239948   .0077292     3.10   0.002     .0088458    .0391438
          6  |   .0649886   .0315408     2.06   0.039     .0031698    .1268073
          7  |   .0245612   .0077661     3.16   0.002     .0093399    .0397824
          8  |   .0664566   .0320011     2.08   0.038     .0037357    .1291776
          9  |   .0251406    .007801     3.22   0.001     .0098509    .0404302
         10  |   .0679555   .0324685     2.09   0.036     .0043184    .1315926
         11  |   .0257333   .0078338     3.28   0.001     .0103793    .0410872
         12  |   .0694856   .0329432     2.11   0.035     .0049182     .134053
         13  |   .0263395   .0078645     3.35   0.001     .0109254    .0417536
         14  |   .0710476   .0334253     2.13   0.034     .0055352    .1365599
         15  |   .0269597    .007893     3.42   0.001     .0114897    .0424297
         16  |   .0726419   .0339151     2.14   0.032     .0061695    .1391143
         17  |   .0275941   .0079193     3.48   0.000     .0120726    .0431156
         18  |   .0742691   .0344129     2.16   0.031     .0068211    .1417171
         19  |    .028243   .0079433     3.56   0.000     .0126744    .0438115
         20  |   .0759298   .0349187     2.17   0.030     .0074904    .1443693
         21  |   .0289066    .007965     3.63   0.000     .0132956    .0445176
         22  |   .0776245    .035433     2.19   0.028     .0081772    .1470719
         23  |   .0295854   .0079843     3.71   0.000     .0139365    .0452343
         24  |   .0793538   .0359559     2.21   0.027     .0088816    .1498261
         25  |   .0302796   .0080013     3.78   0.000     .0145975    .0459618
         26  |   .0811183   .0364877     2.22   0.026     .0096038    .1526328
         27  |   .0309896   .0080158     3.87   0.000     .0152789    .0467004
```

28	.0829184	.0370286	2.24	0.025	.0103437	.1554932
29	.0317157	.008028	3.95	0.000	.0159811	.0474504
30	.0847548	.037579	2.26	0.024	.0111014	.1584083
31	.0324583	.0080379	4.04	0.000	.0167044	.0482122
32	.0866281	.0381391	2.27	0.023	.0118768	.1613793
33	.0332176	.0080453	4.13	0.000	.0174491	.0489862
34	.0885387	.0387092	2.29	0.022	.0126701	.1644073
35	.0339941	.0080505	4.22	0.000	.0182155	.0497728
36	.0904873	.0392896	2.30	0.021	.013481	.1674935
37	.0347881	.0080534	4.32	0.000	.0190038	.0505724
38	.0924744	.0398807	2.32	0.020	.0143097	.1706391
39	.0355999	.0080541	4.42	0.000	.0198142	.0513856
40	.0945006	.0404827	2.33	0.020	.015156	.1738453
41	.03643	.0080527	4.52	0.000	.0206469	.052213
42	.0965665	.041096	2.35	0.019	.0160199	.1771132
43	.0372786	.0080494	4.63	0.000	.0215021	.0530552
44	.0986727	.0417209	2.37	0.018	.0169012	.1804442
45	.0381463	.0080443	4.74	0.000	.0223798	.0539128
46	.1008196	.0423578	2.38	0.017	.0178	.1838393
47	.0390333	.0080375	4.86	0.000	.02328	.0547866
48	.103008	.0430069	2.40	0.017	.018716	.1873
49	.0399401	.0080294	4.97	0.000	.0242028	.0556775
50	.1052383	.0436688	2.41	0.016	.0196491	.1908274
51	.0408671	.0080202	5.10	0.000	.0251479	.0565863
52	.107511	.0443436	2.42	0.015	.0205992	.1944229
53	.0418147	.0080101	5.22	0.000	.0261152	.0575141
54	.1098269	.0450318	2.44	0.015	.0215661	.1980877
55	.0427832	.0079995	5.35	0.000	.0271044	.058462
56	.1121863	.0457339	2.45	0.014	.0225496	.201823
57	.0437732	.0079889	5.48	0.000	.0281152	.0594311
58	.1145899	.04645	2.47	0.014	.0235496	.2056302
59	.0447849	.0079787	5.61	0.000	.0291471	.0604228
60	.1170382	.0471807	2.48	0.013	.0245658	.2095106
61	.045819	.0079693	5.75	0.000	.0301994	.0614386
62	.1195318	.0479262	2.49	0.013	.0255981	.2134655
63	.0468758	.0079615	5.89	0.000	.0312715	.06248
64	.1220711	.0486871	2.51	0.012	.0266462	.2174961
65	.0479557	.0079558	6.03	0.000	.0323625	.0635488
66	.1246568	.0494636	2.52	0.012	.0277099	.2216037
67	.0490592	.0079531	6.17	0.000	.0334714	.0646469
68	.1272893	.0502562	2.53	0.011	.028789	.2257896
69	.0501867	.007954	6.31	0.000	.0345972	.0657763
70	.1299691	.0510652	2.55	0.011	.0298832	.230055
71	.0513388	.0079595	6.45	0.000	.0357384	.0669392
72	.1326968	.051891	2.56	0.011	.0309923	.2344012
73	.0525158	.0079706	6.59	0.000	.0368937	.068138
74	.1354727	.052734	2.57	0.010	.032116	.2388295
75	.0537184	.0079883	6.72	0.000	.0380615	.0693752
76	.1382976	.0535945	2.58	0.010	.0332542	.2433409
77	.0549468	.0080138	6.86	0.000	.03924	.0706536
78	.1411716	.054473	2.59	0.010	.0344065	.2479368
79	.0562017	.0080483	6.98	0.000	.0404274	.071976
80	.1440955	.0553698	2.60	0.009	.0355727	.2526182
81	.0574835	.008093	7.10	0.000	.0416216	.0733454
82	.1470695	.0562851	2.61	0.009	.0367526	.2573863
83	.0587927	.0081492	7.21	0.000	.0428205	.0747649

84	.1500941	.0572195	2.62	0.009	.037946	.2622422
85	.0601298	.0082185	7.32	0.000	.0440219	.0762377
86	.1531698	.0581731	2.63	0.008	.0391526	.2671869
87	.0614953	.0083021	7.41	0.000	.0452235	.0777672
88	.1562969	.0591463	2.64	0.008	.0403723	.2722215
89	.0628898	.0084016	7.49	0.000	.046423	.0793566
90	.1594758	.0601394	2.65	0.008	.0416048	.2773469
91	.0643138	.0085183	7.55	0.000	.0476182	.0810093
92	.1627069	.0611527	2.66	0.008	.0428499	.282564
93	.0657677	.0086537	7.60	0.000	.0488067	.0827286
94	.1659906	.0621864	2.67	0.008	.0441075	.2878736
95	.0672521	.0088092	7.63	0.000	.0499865	.0845177
96	.1693271	.0632407	2.68	0.007	.0453775	.2932767
97	.0687675	.008986	7.65	0.000	.0511553	.0863798
98	.1727168	.064316	2.69	0.007	.0466597	.2987738
99	.0703146	.0091854	7.65	0.000	.0523114	.0883177
100	.1761599	.0654124	2.69	0.007	.047954	.3043658
101	.0718937	.0094087	7.64	0.000	.053453	.0903344
102	.1796568	.06653	2.70	0.007	.0492604	.3100532
103	.0735055	.0096568	7.61	0.000	.0545786	.0924325
104	.1832077	.0676691	2.71	0.007	.0505788	.3158366
105	.0751505	.0099307	7.57	0.000	.0556867	.0946144
106	.1868128	.0688296	2.71	0.007	.0519092	.3217164
107	.0768293	.0102313	7.51	0.000	.0567764	.0968823
108	.1904722	.0700118	2.72	0.007	.0532515	.3276929
109	.0785424	.0105593	7.44	0.000	.0578466	.0992383
110	.1941863	.0712157	2.73	0.006	.054606	.3337665
111	.0802904	.0109154	7.36	0.000	.0588966	.1016842
112	.197955	.0724414	2.73	0.006	.0559726	.3399374
113	.0820738	.0113001	7.26	0.000	.059926	.1042216
114	.2017786	.0736887	2.74	0.006	.0573514	.3462057
115	.0838932	.0117139	7.16	0.000	.0609343	.1068521
116	.2056571	.0749577	2.74	0.006	.0587427	.3525714
117	.0857492	.0121573	7.05	0.000	.0619213	.109577
118	.2095905	.0762483	2.75	0.006	.0601466	.3590345
119	.0876423	.0126305	6.94	0.000	.062887	.1123976
120	.213579	.0775604	2.75	0.006	.0615634	.3655946
121	.0895731	.0131338	6.82	0.000	.0638313	.1153149
122	.2176225	.0788939	2.76	0.006	.0629933	.3722516
123	.0915422	.0136675	6.70	0.000	.0647543	.11833
124	.2217209	.0802485	2.76	0.006	.0644368	.379005
125	.0935501	.0142318	6.57	0.000	.0656562	.1214439
126	.2258742	.081624	2.77	0.006	.0658942	.3858543
127	.0955974	.0148268	6.45	0.000	.0665373	.1246575
128	.2300824	.0830202	2.77	0.006	.0673658	.392799
129	.0976847	.0154528	6.32	0.000	.0673978	.1279716
130	.2343452	.0844367	2.78	0.006	.0688523	.3998381
131	.0998125	.0161098	6.20	0.000	.0682379	.1313872
132	.2386625	.0858732	2.78	0.005	.0703541	.4069709
133	.1019815	.016798	6.07	0.000	.0690579	.134905
134	.2430341	.0873293	2.78	0.005	.0718718	.4141964
135	.1041921	.0175176	5.95	0.000	.0698582	.138526
136	.2474597	.0888046	2.79	0.005	.073406	.4215135
137	.106445	.0182687	5.83	0.000	.070639	.1422509
138	.2519391	.0902984	2.79	0.005	.0749575	.4289208
139	.1087406	.0190514	5.71	0.000	.0714006	.1460806

140	.256472	.0918104	2.79	0.005	.0765269	.4364171
141	.1110796	.0198658	5.59	0.000	.0721434	.1500159
142	.261058	.09334	2.80	0.005	.078115	.4440009
143	.1134626	.0207121	5.48	0.000	.0728675	.1540576
144	.2656966	.0948864	2.80	0.005	.0797227	.4516705
145	.11589	.0215905	5.37	0.000	.0735733	.1582066
146	.2703875	.096449	2.80	0.005	.0813509	.4594242
147	.1183623	.0225011	5.26	0.000	.074261	.1624637
148	.2751302	.0980272	2.81	0.005	.0830004	.46726
149	.1208803	.023444	5.16	0.000	.0749309	.1668297
150	.2799242	.0996201	2.81	0.005	.0846724	.475176
151	.1234442	.0244194	5.06	0.000	.0755831	.1713054
152	.2847689	.1012269	2.81	0.005	.0863677	.4831701
153	.1260548	.0254275	4.96	0.000	.0762179	.1758917
154	.2896637	.1028469	2.82	0.005	.0880876	.4912398
155	.1287125	.0264683	4.86	0.000	.0768356	.1805893
156	.294608	.1044789	2.82	0.005	.0898331	.4993829
157	.1314177	.027542	4.77	0.000	.0774363	.1853991
158	.2996011	.1061222	2.82	0.005	.0916054	.5075968
159	.1341711	.0286489	4.68	0.000	.0780204	.1903219
160	.3046423	.1077757	2.83	0.005	.0934057	.5158788
161	.1369731	.0297889	4.60	0.000	.0785879	.1953583
162	.3097307	.1094385	2.83	0.005	.0952353	.5242262
163	.1398241	.0309623	4.52	0.000	.0791392	.200509
164	.3148657	.1111093	2.83	0.005	.0970955	.532636
165	.1427247	.0321691	4.44	0.000	.0796745	.2057749
166	.3200464	.1127872	2.84	0.005	.0989876	.5411051
167	.1456752	.0334094	4.36	0.000	.080194	.2111564
168	.3252718	.1144708	2.84	0.004	.1009131	.5496305
169	.1486762	.0346834	4.29	0.000	.080698	.2166543
170	.3305411	.1161592	2.85	0.004	.1028732	.5582089
171	.151728	.035991	4.22	0.000	.0811869	.2222691
172	.3358532	.117851	2.85	0.004	.1048695	.5668369
173	.154831	.0373324	4.15	0.000	.0816608	.2280012
174	.3412072	.1195449	2.85	0.004	.1069034	.5755109
175	.1579857	.0387076	4.08	0.000	.0821201	.2338513
176	.346602	.1212398	2.86	0.004	.1089764	.5842275
177	.1611924	.0401167	4.02	0.000	.0825652	.2398196
178	.3520365	.1229341	2.86	0.004	.1110901	.5929829
179	.1644515	.0415595	3.96	0.000	.0829964	.2459065
180	.3575096	.1246266	2.87	0.004	.1132459	.6017733
181	.1677633	.043036	3.90	0.000	.0834142	.2521124
182	.3630201	.126316	2.87	0.004	.1154453	.6105949
183	.1711281	.0445463	3.84	0.000	.083819	.2584373
184	.3685669	.1280007	2.88	0.004	.1176901	.6194437
185	.1745463	.0460902	3.79	0.000	.0842112	.2648814
186	.3741486	.1296794	2.89	0.004	.1199816	.6283155
187	.1780181	.0476675	3.73	0.000	.0845914	.2714447
188	.379764	.1313506	2.89	0.004	.1223215	.6372064
189	.1815437	.0492783	3.68	0.000	.0849601	.2781273
190	.3854117	.1330129	2.90	0.004	.1247113	.6461121
191	.1851235	.0509221	3.64	0.000	.0853179	.284929
192	.3910905	.1346647	2.90	0.004	.1271527	.6550284
193	.1887575	.052599	3.59	0.000	.0856654	.2918496
194	.396799	.1363045	2.91	0.004	.129647	.6639509
195	.1924461	.0543085	3.54	0.000	.0860034	.2988888

196	.4025357	.1379309	2.92	0.004	.132196	.6728754
197	.1961893	.0560504	3.50	0.000	.0863325	.3060461
198	.4082992	.1395424	2.93	0.003	.1348011	.6817973
199	.1999872	.0578244	3.46	0.001	.0866534	.313321
200	.414088	.1411374	2.93	0.003	.1374637	.6907123

Variables that uniquely identify margins: cyclist age cyclist's non respect of the priority

.

end of do-file

.

Appendix H

Three way-chi square test of independence

Country * Age Group * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)
Slight	Pearson Chi-Square	101.031[b]	10	.000
	Likelihood Ratio	105.184	10	.000
	Linear-by-Linear Association	.006	1	.936
	N of Valid Cases	1592		
KSI	Pearson Chi-Square	29.257[c]	9	.001
	Likelihood Ratio	30.144	9	.000
	Linear-by-Linear Association	5.605	1	.018
	N of Valid Cases	196		
Total	Pearson Chi-Square	115.113[a]	10	.000
	Likelihood Ratio	118.946	10	.000
	Linear-by-Linear Association	.018	1	.894
	N of Valid Cases	1788		

a. 2 cells (9.1%) have expected count less than 5. The minimum expected count is 3.87.

b. 2 cells (9.1%) have expected count less than 5. The minimum expected count is 3.66.

c. 8 cells (40.0%) have expected count less than 5. The minimum expected count is .62.

Country * Speed Limit mph * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)
Slight	Pearson Chi-Square	122.829[b]	5	.000
	Likelihood Ratio	148.293	5	.000
	Linear-by-Linear Association	94.156	1	.000
	N of Valid Cases	1587		
KSI	Pearson Chi-Square	11.586[c]	5	.041
	Likelihood Ratio	14.893	5	.011
	Linear-by-Linear Association	6.598	1	.010
	N of Valid Cases	196		
Total	Pearson Chi-Square	138.235[a]	5	.000
	Likelihood Ratio	166.832	5	.000
	Linear-by-Linear Association	106.584	1	.000
	N of Valid Cases	1783		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.15.

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.11.

c. 5 cells (41.7%) have expected count less than 5. The minimum expected count is .31.

Country * Gender * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Slight	Pearson Chi-Square	136.796[c]	1	.000		
	Continuity Correction[b]	135.423	1	.000		
	Likelihood Ratio	145.811	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	136.711	1	.000		
	N of Valid Cases	1592				
KSI	Pearson Chi-Square	7.837[d]	1	.005		
	Continuity Correction[b]	6.782	1	.009		
	Likelihood Ratio	7.408	1	.006		
	Fisher's Exact Test				.010	.005
	Linear-by-Linear Association	7.797	1	.005		
	N of Valid Cases	196				
Total	Pearson Chi-Square	146.290[a]	1	.000		
	Continuity Correction[b]	144.944	1	.000		
	Likelihood Ratio	153.460	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	146.208	1	.000		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 203.44.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 175.38.

d. 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.83.

Country * Weather * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)
Slight	Pearson Chi-Square	4.094[b]	2	.129
	Likelihood Ratio	4.082	2	.130
	Linear-by-Linear Association	.750	1	.387
	N of Valid Cases	1592		
KSI	Pearson Chi-Square	2.058[c]	2	.357
	Likelihood Ratio	1.880	2	.391
	Linear-by-Linear Association	.713	1	.398
	N of Valid Cases	196		
Total	Pearson Chi-Square	4.089[a]	2	.129
	Likelihood Ratio	4.096	2	.129
	Linear-by-Linear Association	.165	1	.684
	N of Valid Cases	1788		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 27.06.

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 23.35.

c. 2 cells (33.3%) have expected count less than 5. The minimum expected count is 1.56.

Country * Light Condition * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Slight	Pearson Chi-Square	38.953[b]	2	.000		
	Likelihood Ratio	40.489	2	.000		
	Linear-by-Linear Association	38.117	1	.000		
	N of Valid Cases	1592				
KSI	Pearson Chi-Square	6.187[c]	1	.013		
	Continuity Correction[d]	5.192	1	.023		
	Likelihood Ratio	7.094	1	.008		
	Fisher's Exact Test				.012	.008
	Linear-by-Linear Association	6.155	1	.013		
	N of Valid Cases	196				
Total	Pearson Chi-Square	43.797[a]	2	.000		
	Likelihood Ratio	45.735	2	.000		
	Linear-by-Linear Association	42.784	1	.000		
	N of Valid Cases	1788				

a. 2 cells (33.3%) have expected count less than 5. The minimum expected count is 1.93.

b. 2 cells (33.3%) have expected count less than 5. The minimum expected count is 1.83.

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.96.

d. Computed only for a 2x2 table

Country * Road Surface Condition * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)
Slight	Pearson Chi-Square	104.571[b]	3	.000
	Likelihood Ratio	121.799	3	.000
	Linear-by-Linear Association	17.102	1	.000
	N of Valid Cases	1592		
KSI	Pearson Chi-Square	16.076[c]	2	.000
	Likelihood Ratio	16.935	2	.000
	Linear-by-Linear Association	5.309	1	.021
	N of Valid Cases	196		
Total	Pearson Chi-Square	119.511[a]	3	.000
	Likelihood Ratio	138.961	3	.000
	Linear-by-Linear Association	23.240	1	.000
	N of Valid Cases	1788		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.21.

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 7.78.

c. 2 cells (33.3%) have expected count less than 5. The minimum expected count is 2.18.

Country * Overshoot * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Slight	Pearson Chi-Square	232.857[c]	1	.000		
	Continuity Correction[b]	230.998	1	.000		
	Likelihood Ratio	266.236	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	232.711	1	.000		
	N of Valid Cases	1592				
KSI	Pearson Chi-Square	34.348[d]	1	.000		
	Continuity Correction[b]	31.945	1	.000		
	Likelihood Ratio	31.956	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	34.172	1	.000		
	N of Valid Cases	196				
Total	Pearson Chi-Square	268.263[a]	1	.000		
	Continuity Correction[b]	266.366	1	.000		
	Likelihood Ratio	300.640	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	268.113	1	.000		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 182.17.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 157.98.

d. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.96.

Country * Poor Turn or Manoeuvre * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Slight	Pearson Chi-Square	61.604[c]	1	.000		
	Continuity Correction[b]	59.860	1	.000		
	Likelihood Ratio	66.874	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	61.565	1	.000		
	N of Valid Cases	1592				
KSI	Pearson Chi-Square	2.639[d]	1	.104		
	Continuity Correction[b]	1.662	1	.197		
	Likelihood Ratio	3.236	1	.072		
	Fisher's Exact Test				.178	.092
	Linear-by-Linear Association	2.625	1	.105		
	N of Valid Cases	196				
Total	Pearson Chi-Square	62.355[a]	1	.000		
	Continuity Correction[b]	60.708	1	.000		
	Likelihood Ratio	69.039	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	62.320	1	.000		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 46.39.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 38.92.

d. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 3.42.

Country * Passing Too Close to Cyclist * Severity Cross abulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Slight	Pearson Chi-Square	50.086[c]	1	.000		
	Continuity Correction[b]	48.662	1	.000		
	Likelihood Ratio	52.116	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	50.054	1	.000		
	N of Valid Cases	1592				
KSI	Pearson Chi-Square	2.398[d]	1	.122		
	Continuity Correction[b]	1.583	1	.208		
	Likelihood Ratio	2.765	1	.096		
	Fisher's Exact Test				.153	.100
	Linear-by-Linear Association	2.386	1	.122		
	N of Valid Cases	196				
Total	Pearson Chi-Square	51.691[a]	1	.000		
	Continuity Correction[b]	50.340	1	.000		
	Likelihood Ratio	54.702	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	51.662	1	.000		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 57.99.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 48.08.

d. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 4.67.

Country * Loss of Control * Severity Cross Tabulation Chi-Square Tests

Severity		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Slight	Pearson Chi-Square	.130[c]	1	.719		
	Continuity Correction[b]	.015	1	.902		
	Likelihood Ratio	.129	1	.719		
	Fisher's Exact Test				.814	.449
	Linear-by-Linear Association	.130	1	.719		
	N of Valid Cases	1592				
KSI	Pearson Chi-Square	1.377[d]	1	.241		
	Continuity Correction[b]	.297	1	.586		
	Likelihood Ratio	2.258	1	.133		
	Fisher's Exact Test				.554	.324
	Linear-by-Linear Association	1.370	1	.242		
	N of Valid Cases	196				
Total	Pearson Chi-Square	.662[a]	1	.416		
	Continuity Correction[b]	.353	1	.552		
	Likelihood Ratio	.663	1	.416		
	Fisher's Exact Test				.512	.276
	Linear-by-Linear Association	.662	1	.416		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.15.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.24.

d. 2 cells (50.0%) have expected count less than 5. The minimum expected count is .93.

Severity * Age Group * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)
England	Pearson Chi-Square	9.884[b]	10	.451
	Likelihood Ratio	12.179	10	.273
	Linear-by-Linear Association	7.341	1	.007
	N of Valid Cases	864		
Belgium	Pearson Chi-Square	24.368[c]	10	.007
	Likelihood Ratio	28.535	10	.001
	Linear-by-Linear Association	13.105	1	.000
	N of Valid Cases	924		
Total	Pearson Chi-Square	20.852[a]	10	.022
	Likelihood Ratio	24.333	10	.007
	Linear-by-Linear Association	17.295	1	.000
	N of Valid Cases	1788		

a. 2 cells (9.1%) have expected count less than 5. The minimum expected count is .88.

b. 5 cells (22.7%) have expected count less than 5. The minimum expected count is .47.

c. 5 cells (22.7%) have expected count less than 5. The minimum expected count is .33.

Severity * Gender * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
England	Pearson Chi-Square	1.550[c]	1	.213		
	Continuity Correction[b]	1.199	1	.273		
	Likelihood Ratio	1.460	1	.227		
	Fisher's Exact Test				.230	.138
	Linear-by-Linear Association	1.548	1	.213		
	N of Valid Cases	864				
Belgium	Pearson Chi-Square	.489[d]	1	.484		
	Continuity Correction[b]	.314	1	.575		
	Likelihood Ratio	.498	1	.481		
	Fisher's Exact Test				.580	.291
	Linear-by-Linear Association	.488	1	.485		
	N of Valid Cases	924				
Total	Pearson Chi-Square	2.114[a]	1	.146		
	Continuity Correction[b]	1.863	1	.172		
	Likelihood Ratio	2.198	1	.138		
	Fisher's Exact Test				.154	.084
	Linear-by-Linear Association	2.113	1	.146		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 46.15.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 14.84.

d. 0 cells (.0%) have expected count less than 5. The minimum expected count is 21.52.

Severity * Speed Limit mph * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)
England	Pearson Chi-Square	13.735[b]	5	.017
	Likelihood Ratio	12.648	5	.027
	Linear-by-Linear Association	5.533	1	.019
	N of Valid Cases	864		
Belgium	Pearson Chi-Square	7.824[c]	4	.098
	Likelihood Ratio	6.075	4	.194
	Linear-by-Linear Association	4.626	1	.031
	N of Valid Cases	919		
Total	Pearson Chi-Square	27.580[a]	5	.000
	Likelihood Ratio	22.368	5	.000
	Linear-by-Linear Association	21.217	1	.000
	N of Valid Cases	1783		

a. 2 cells (16.7%) have expected count less than 5. The minimum expected count is 2.53.

b. 3 cells (25.0%) have expected count less than 5. The minimum expected count is .78.

c. 5 cells (50.0%) have expected count less than 5. The minimum expected count is .07.

Severity * Light Condition * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
England	Pearson Chi-Square	.429[b]	1	.513		
	Continuity Correction[c]	.295	1	.587		
	Likelihood Ratio	.437	1	.509		
	Fisher's Exact Test				.578	.297
	Linear-by-Linear Association	.428	1	.513		
	N of Valid Cases	864				
Belgium	Pearson Chi-Square	1.927[d]	2	.382		
	Likelihood Ratio	2.434	2	.296		
	Linear-by-Linear Association	1.276	1	.259		
	N of Valid Cases	924				
Total	Pearson Chi-Square	.586[a]	2	.746		
	Likelihood Ratio	1.023	2	.600		
	Linear-by-Linear Association	.039	1	.843		
	N of Valid Cases	1788				

a. 2 cells (33.3%) have expected count less than 5. The minimum expected count is .44.

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 30.94.

c. Computed only for a 2x2 table

d. 2 cells (33.3%) have expected count less than 5. The minimum expected count is .26.

Severity * Weather * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)
England	Pearson Chi-Square	2.916[b]	2	.233
	Likelihood Ratio	3.186	2	.203
	Linear-by-Linear Association	2.848	1	.092
	N of Valid Cases	864		
Belgium	Pearson Chi-Square	.698[c]	2	.706
	Likelihood Ratio	.693	2	.707
	Linear-by-Linear Association	.002	1	.961
	N of Valid Cases	924		
Total	Pearson Chi-Square	1.897[a]	2	.387
	Likelihood Ratio	2.013	2	.365
	Linear-by-Linear Association	1.576	1	.209
	N of Valid Cases	1788		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.14.

b. 1 cells (16.7%) have expected count less than 5. The minimum expected count is 3.59.

c. 1 cells (16.7%) have expected count less than 5. The minimum expected count is 2.18.

Severity * Road Surface Condition * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)
England	Pearson Chi-Square	3.744[b]	2	.154
	Likelihood Ratio	5.617	2	.060
	Linear-by-Linear Association	2.936	1	.087
	N of Valid Cases	864		
Belgium	Pearson Chi-Square	2.010[c]	3	.570
	Likelihood Ratio	3.081	3	.379
	Linear-by-Linear Association	.078	1	.780
	N of Valid Cases	924		
Total	Pearson Chi-Square	6.335[a]	3	.096
	Likelihood Ratio	8.925	3	.030
	Linear-by-Linear Association	3.714	1	.054
	N of Valid Cases	1788		

a. 1 cells (12.5%) have expected count less than 5. The minimum expected count is 1.86.

b. 1 cells (16.7%) have expected count less than 5. The minimum expected count is 1.88.

c. 1 cells (12.5%) have expected count less than 5. The minimum expected count is 1.12.

Severity * Overshoot * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
England	Pearson Chi-Square	.493[c]	1	.482		
	Continuity Correction[b]	.232	1	.630		
	Likelihood Ratio	.465	1	.495		
	Fisher's Exact Test				.507	.303
	Linear-by-Linear Association	.493	1	.483		
	N of Valid Cases	864				
Belgium	Pearson Chi-Square	.251[d]	1	.617		
	Continuity Correction[b]	.132	1	.717		
	Likelihood Ratio	.248	1	.618		
	Fisher's Exact Test				.680	.355
	Linear-by-Linear Association	.250	1	.617		
	N of Valid Cases	924				
Total	Pearson Chi-Square	2.996[a]	1	.083		
	Continuity Correction[b]	2.683	1	.101		
	Likelihood Ratio	3.161	1	.075		
	Fisher's Exact Test				.095	.048
	Linear-by-Linear Association	2.994	1	.084		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 41.33.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.41.

d. 0 cells (.0%) have expected count less than 5. The minimum expected count is 22.18.

Severity * Poor Turn or Manoeuvre * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
England	Pearson Chi-Square	.977[c]	1	.323		
	Continuity Correction[b]	.689	1	.406		
	Likelihood Ratio	1.041	1	.307		
	Fisher's Exact Test				.428	.206
	Linear-by-Linear Association	.976	1	.323		
	N of Valid Cases	864				
Belgium	Pearson Chi-Square	.059[d]	1	.808		
	Continuity Correction[b]	.000	1	1.000		
	Likelihood Ratio	.055	1	.815		
	Fisher's Exact Test				.562	.562
	Linear-by-Linear Association	.059	1	.808		
	N of Valid Cases	924				
Total	Pearson Chi-Square	.026[a]	1	.873		
	Continuity Correction[b]	.000	1	1.000		
	Likelihood Ratio	.025	1	.874		
	Fisher's Exact Test				.866	.487
	Linear-by-Linear Association	.026	1	.873		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.52.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 13.13.

d. 1 cells (25.0%) have expected count less than 5. The minimum expected count is .79.

Severity * Passing Too Close to Cyclist * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
England	Pearson Chi-Square	.356[c]	1	.551		
	Continuity Correction[b]	.200	1	.655		
	Likelihood Ratio	.368	1	.544		
	Fisher's Exact Test				.655	.335
	Linear-by-Linear Association	.355	1	.551		
	N of Valid Cases	864				
Belgium	Pearson Chi-Square	.120[d]	1	.729		
	Continuity Correction[b]	.000	1	1.000		
	Likelihood Ratio	.111	1	.739		
	Fisher's Exact Test				.669	.479
	Linear-by-Linear Association	.120	1	.729		
	N of Valid Cases	924				
Total	Pearson Chi-Square	.312[a]	1	.577		
	Continuity Correction[b]	.166	1	.684		
	Likelihood Ratio	.301	1	.583		
	Fisher's Exact Test				.546	.332
	Linear-by-Linear Association	.312	1	.577		
	N of Valid Cases	1788				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 13.15.

b. Computed only for a 2x2 table

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 15.00.

d. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 1.58.

Severity * Loss of Control * Country Cross Tabulation Chi-Square Tests

Country		Value	Degrees of Freedom	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
England	Pearson Chi-Square	.811[c]	1	.368		
	Continuity Correction[b]	.250	1	.617		
	Likelihood Ratio	.711	1	.399		
	Fisher's Exact Test				.414	.285
	Linear-by-Linear Association	.810	1	.368		
	N of Valid Cases	864				
Belgium	Pearson Chi-Square	.642[d]	1	.423		
	Continuity Correction[b]	.016	1	.899		
	Likelihood Ratio	1.236	1	.266		
	Fisher's Exact Test				1.000	.539
	Linear-by-Linear Association	.642	1	.423		
	N of Valid Cases	924				
Total	Pearson Chi-Square	.241[a]	1	.624		
	Continuity Correction[b]	.019	1	.889		
	Likelihood Ratio	.222	1	.637		
	Fisher's Exact Test				.495	.409
	Linear-by-Linear Association	.240	1	.624		
	N of Valid Cases	1788				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 2.30.

b. Computed only for a 2x2 table

c. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 1.88.

d. 1 cells (25.0%) have expected count less than 5. The minimum expected count is .59.

References

AASHTO (1999) *A Guide for the Development of Bicycle Facilities*. United States of America: American Association of State Highway and Transportation Officials.

Agresti, A. (2007) *An introduction to categorical data analysis*. 2nd edn. Haboken, New Jersey: John Wiley & Sons, Inc.

Altman, D. and Bland, J. (1995) 'Statistics notes: the normal distribution'.

Arnold, L.S., Flannery, A., Ledbetter, L., Bills, T., Jones, M.G., Ragland, D.R. and Spautz, L. (2010) 'Identifying Factors that Determine Bicyclist and Pedestrian: Involved Collision Rates and Bicyclist and Pedestrian Demand at Multi-Lane Roundabouts' UC Berkeley Safe Transportation Research & Education Center, I.o.T.S. University of California, Berkeley.

Austroads (2009) 'Guide to Road Design Part 4B Roundabouts', *Australian Design Standards*.

Bakar, N. M. and tahir, I.M. (20019) Applying Multiple Linear Regression and Neural Network to Predict Bank Performance. International Business Research. Volume 2 (4)

Bartlett, M.S. (1950) 'Test of significance in factor analysis', *British Journal of Statistical Psychology*, 3(2), pp. 77-85.

Bastos Silva, A. and Seco, A. (2005) 'Trajectory Deflection Influence on the Performance of Roundabouts', *European Transport Conference (ETC)*. Strasbourg, France. Association for European Transport.

Bastos Silva, A., Seco, A.J.M. and Silva, J.P.C. (2006) 'Characterization of Trajectories Adopted at Roundabout Crossings', *European Transport Conference (ETC)*. Strasbourg, France. Association for European Transport.

Berry, W. (1993) *Understanding Regression Assumptions*. Thousand Oaks, CA, US: Sage Publications, Inc. Sage university papers series.

Brown, M. (1995) *The Design of Roundabouts: State of the Art Review*. London: Transport Research Laboratory.

Bruce, W., Rodegerdts, L., Scarborough, W., Kittelson, W., Troutbeck, R., Brilon, W., Bondzio, L., Courage, K., Kyte, M., Mason, J., Flannery, A., Myers, E., Bunker, J. and Jacquemart, G. (2000) *Roundabouts: An Informational Guide* (FHWA-RD-00-067). US Department of Transport: Federal Highway Administration AASHTO.

CEGB (2016) 'Cycling Embassy of Great Britain', *http://www.cycling-embassy.org.uk/wiki/roundabouts*.

Chard, B., Thomson, R. and Bargh, A. (2009) *Signal Controlled Roundabout Methodology and Its Introduction to NZ at Welcome Bay, Maungatapu and Brookfield Roundabouts in Tauranga North Island* NZ Transport Agency Fulton Hogan Ltd and Beca.

Cronbach, L.J. (1951) 'Coefficient Alpha and the Internal Structure of Tests', *Psychometrika*, 16(3), pp. 297-334.

Dalgaard, P. (2008) *Introductory Statistics with R*. Second edn. NY USA: Springer Science Business Media, LLC.

Daniels, S., Brijs, T., Nuyts, E. and Wets, G. (2009) 'Injury Crashes with Bicyclists at Roundabouts: Influence of Some Location Characteristics and the Design of Cycle Facilities', *Journal of Safety Research*, 40(2), pp. 141-148.

Daniels, S., Brijs, T., Nuyts, E. and Wets, G. (2010) 'Externality of Risk and Crash Severity at Roundabouts', *Accident Analysis and Prevention*, 42(6), pp. 1966-73.

Daniels, S., Brijs, T., Nuyts, E. and Wets, G. (2011) 'Extended Prediction Models for Crashes at Roundabouts', *Safety Science*, 49(2), pp. 198-207.

Daniels, S., Nuyts, E. and Wets, G. (2008) 'The Effects of Roundabouts on Traffic Safety for Bicyclists: An Observational Study', *Accident Analysis and Prevention*, 40(2), pp. 518-526.

Davies, D.G., Taylor, M.C., Ryley, T.J. and Halliday, M.E. (1997) *Cyclists at roundabouts — the effects of 'Continental' design on predicted safety and capacity*. Tansport Research Laboratory.

De Brabander, B. and Vereeck, L. (2007) 'Safety Effects of Roundabouts in Flanders: Signal Type, speed Limits and Vulnerable Road Users', *Accident Analysis and Prevention*, 39(3), pp. 591-599.

DfT (2007) 'DfT TD 16-07 Geometric Design of Roundabouts', *Department for Transport*.

DfT (2008) *Local Transport Note 2/08 Cycle Infrastructure Design*. London: Department for Transport.

DfT (2014) Cycling Delivery Plan. Department for Transport. Crown

DRIVINGED (2018) *The Official Highway Code*. Available at: https://www.highwaycodeuk.co.uk/roundabouts.html#.

Efron, B. (2013) *Empirical Bayes Modeling, Computation and Accuracy*. Stanford University, USA.

Elvik, R. (2003) 'Effects on Road Safety of Converting Intersections to Roundabouts: Review of Evidence from Non-US Studies', *Transportation Research Record: Journal of the Transportation Research Board*, (1847), pp. 1-10.

Field, A. (2009) *Discovering Statistics Using SPSS*. 3rd edn. SAGE Publications Ltd.

Fromme, V. (2010) *Roundabout*. Colorado, USA.

Furtado, G. (2004) 'Accommodating Vulnerable Road Users in Roundabout Design', *Annual Conference of the Transportation*. Canada, Quebec City.

Ghasemi, A. and Zahediasl, A. (2012) 'Normality Tests for Statistical Analysis: A Guide for Non-Statisticians', *International Journal of Endocrinol Metabolism*, 10(2), pp. 486-489.

Goodman, P., Galatioto, F., Namdeo, A. and Bell, M.C. (2014) 'Vehicle Emissions and Air Quality Modelling - Newcastle/Gateshead Low Emission Zone Feasibility Study Version 1.2'.

Gross, F., Lyon, C., Persaud, B., Srinivasan, R. (2013) 'Safety Effectiveness of Converting Signalized Intersections to Roundabouts', *Accident Analysis Prevention*, 50, pp. 234–241.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010) *Multivariate Data Analysis: A Global Perspective*. 7 edn. London: Pearson Education.

Harkey, D. and Carter, D. (2006) 'Observational Analysis of Pedestrian, Bicyclist, and Motorist Behaviors at Roundabouts in the United States', *Transportation Research Record Journal of the Transportation Research Board 1982*, pp. 155-165.

Harrell, F. (2001) *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. 1 edn. Springer-Verlag New York.

Hels, T. and Orozova-Bekkevold, I. (2007) 'The Effect of Roundabout Design Features on Cyclist Accident Rate', *Accident Analysis and Prevention*, 39(2), pp. 300-307.

Hosmer, D.W. and Lemeshow, S. (2000) *Applied Logistic Regression*. Second edn. John #Wiley & Sons, Inc.

Hotelling, H. (1933) 'Analysis of a Complex of Statistical Variables into Principal Components', *Journal of Education Psychology*, 24(6), p. 417.

Jensen, S.U. (2017) 'Safe Roundabouts for Cyclists', *Accident Analysis and Prevention*, 105, pp. 30-37.

Jolliffe, I. (2002) *Principal Component Analysis*. Wiley Online Library.

Jurisich, I., Asmus, D., Campbell, D. and Dunn, R. (2011) 'Reducing Speed: The C-Roundabout', *TRB International Roundabout Cconference* Carmel, Indiana, U.S.A.

Kline, P. (1994) *An easy guide to factor analysis*. Routledge.

Kornbrot, D. (2005) 'Point Biserial Correlation'.

Lamm, R., Psarianos, B. and Mailaender, T. (1999) 'Highway Design and Traffic Safety Engineering Handbook', *McGraw-Hill Professional Publishing*

Lawton, B.J., Webb, P.J., Wall, G.T. and Davies, D.G. (2003) 'Cyclists at Continental Style Roundabouts: Report on Four Trial Sites', *Transport Research Laboratory*, TRL report TRL584.

Lenters, M.S. (2004) 'Safety Auditing Roundabouts', *THE 2004 Annual Conference of the Transportation Association of Canada*. Quebec City, Quebec, Canada. SRM Associates.

Li, B., Lingsma, H., Steyerberg, E. and Lesaffre, E. (2011) 'Logistic Random Effects Regression Models: A Comparison of Statistical Packages for Binary and Ordinal Outcomes', *BMC Medical Research Methodology*, 11.

Lindenmann, H.P. (2006) 'Capacity of Small Roundabouts with Two-Lane Entries.', *Transportation Research Board*, (N'1988), pp. 119–126.

Litman, T. and Burwell, D. (2006) 'Issues in Sustainable Transportation', *International Journal of Glowbal Environmental Issues*, 6(4), pp. 331-347.

Manning, C. (2007) 'Logistic Regression (with R)'.

Mazari, M., Zayerzadeh, A. and Khandandel, H. (2008) 'Converting a Roundabout to an Intersection and Its Effects on Crash Rates: Case Study in Mashad City of Iran', *Transport Research Arena Europe 2008, Ljubljana*.

MEI (2007) 'Spearman's Rank Correlation'.

Møller, M. and Hels, T. (2008) 'Cyclists' Perception of Risk in Roundabouts', *Accident Analysis & Prevention*, 40(3), pp. 1055-1062.

Montella, A. (2011) 'Identifying Crash Contributory Factors at Urban Roundabouts and Using Association Rules to Explore Their Relationships to Different Crash Types', *Accident Analysis and Prevention*, 43(4), pp. 1451-1463.

Mukaka, M.M. (2012) 'Statistics Corner; A guide to appropriate use of correlation coefficient in medical research', *Malawi Medical Jornal*, 24(3), pp. 69-71.

Nambisan, S.S. and Parimi, V. (2007) 'A Comparative Evaluation of the Safety Performance of Roundabouts and Traditional Intersection Controls', *Institute of Transportation Engineers Journal*, 77(3), pp. 18-25.

NCSS (2011) 'Point-Biserial and Biserial Correlations'.

Ogundimu, E.O., Altman, D.G. and Collins, G.S. (2016) 'Adequate sample size for developing prediction models is not simply related to events per variable', *Journal of Clinical Epidemiology*, 76, pp. 175-182.

Parkin, J. (2018) *Designing for Cycle Traffic: International Priciples and Practice*. ICE Publishing.

Patterson, F. (2010) 'Cycling and Roundabouts: An Australian Perspective', *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice*, 19(2), pp. 4-19.

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. and Feinstein, A.R. (1996) 'A simulation study of the number of events per variable in logistic regression analysis', *Journal of clinical epidemiology*, 49(12), pp. 1373-1379.

Persaud, B.N., Retting, R.A., Garder, P.E. and Lord, D. (2001) 'Observational Before-After Study of the Safety Effect of U.S. Roundabout Conversions Using the Empirical Bayes Method', *Annual Meeting of the Transportation Research Board* (TRB ID: 01-0562).

Polders, E., Daniels, S., Casters, W. and Brijs, T. (2014) 'Identifying Crash Patterns on Roundabouts', *Traffic injury prevention*, 16(2), pp. 202-207.

Rasanen, M. and Summala, H. (2000) 'Car Drivers' Adjustments to Cyclists at Roundabouts', *Transportation Human Factors*, 2(1), pp. 1-17.

Razali, N.M. and Wah, Y.B. (2011) 'Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests', *Journal of Statistical Modelling and Analytics*, 2(1), pp. 21-33.

Retting, R.A., Persaud, B.N., Garder, P.E. and Lord, D. (2001) 'Crash and Injury Reduction following 17 Installation of Roundabouts in the United States', *American Journal of Public Health*, 91.4, pp. 628-31.

Robinson, B.W., Rodegerdts, L., Scarborough, W., Kittelson, W., Troutbeck, R., Brilon, W., Bondzio, L., Courage, K., Kyte, M., Mason, J., Flannery, A., Myers, E., BUNKER, J. and Jacquemart, G. (2000) 'Roundabouts: An Informational Guide', *FHWA-RD-00-067,, Project 2425,, Informational Guide Book.*

Rolison, J.J., Regeva, S., Moutari, S. and Feeney, A. (2018) 'What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records', *Accident Analysis & Prevention*, 115, pp. 11-24.

Rummel, R.J. (1988) *Applied factor analysis*. Northwestern University Press.

Sakshaug, L., Laureshyn, A., Svensson, A. and Hyden, C. (2010) 'Cyclists in Roundabouts: Different Design Solutions', *Accid Anal Prev*, 42(4), pp. 1338-51.

Schneider, A., Hommel, G. and Blettner, M. (2010) 'Linear regression analysis: part 14 of a series on evaluation of scientific publications', *Deutsches Arzteblatt International*, 107(44).

Silva, A.B., Santos, S., Vasconcelos, L., Seco, Á. and Silva, J.P. (2014) 'Driver Behavior Characterization in Roundabout Crossings', *Transportation Research Procedia*, 3, pp. 80-89.

Silvano, A.P. and Linder, A. (2017) *Traffic Safety for Cyclists in Roundabouts: Geometry, Traffic and Priority Rules* (VTI notat 31A-2017).

Silvano, A.P., Ma, X. and Koutsopoulos, H.N. (2015) 'When do drivers yield to cyclists at unsignalized roundabouts', *Transportation Research Record: Journal of the Transportation Research Board*, Volume 2520 (DOI: 10.3141/2520-04).

Sommet, N. and Morselli, D. (2017) 'Keep Calm and Learn Multilevel Logistic Modeling: A Simplified Three-Step Procedure Using Stata, R, Mplus, and SPSS.', *International Review of Social Psychology*, 30(1).

Sperandei, S. (2013) 'Understanding Logistic Regression', *Croatian Society of Medical Biochemistry and Laboratory Medicine*.

St-Aubin, P., Saunier, N., Miranda-Moreno, L.F. and Ismail, K. (2013) 'Detailed Driver Behaviour Analysis and Trajectory Interpretation at Roundabouts Using Computer Vision Data', *Transportation Research Board - 92nd Annual Meeting*. Washigton D. C., USA.

Steele, F. (2009) *Multilevel Models for Binary Responses*. Centre for Multilevel Modelling, Bristol.

Streiner, D.L. (2003) 'Being inconsistent about consistency: When coefficient alpha does and doesn’t matter', *Journal of Personality Assessment*, 80(3), pp. 217-222.

Suhr, D.D. (2005) 'Principal component analysis vs. exploratory factor analysis', *SUGI 30 proceedings*, 203, p. 230.

Tabachnick, B.G., Fidell, L.S. and Osterlind, S.J. (2014) *Using Multivariate Statistics*. 6th edn.

Tavakol, M. and Dennick, R. (2011) 'Making Sense of Cronbach's Alpha', *International Journal of Medical Education*.

Taylor, G. (2011) 'Roundabout Geometric Design'. SunCam Online Education Course.

Tollazzi, T. (2015) *Alternative Types of Roundabouts: An Informational Guide*. Springer International Publishing.

TS (2013) *Procedures for Submitting Road Accident Data to Transport Scotland from non-CRASH Sources* Transport Scotland.

Turner, S.A. and Roozenburg, A.P. (2009) *Roundabout Crash Prediction Models* (Agency Research Report 386). New Zealand: NZ Transport.

UCLA (2019) Logistic Regression Analysis: STATA Annotated Output, Institue for Digital Research and Education, URL: https://stats.idre.ucla.edu/stata/output/logistic-regression-analysis/

Wang, H., Peng, J., Wang, B., Lu, X., Zheng, J.Z., Wang, K., Tu, X.M. and Feng, C. (2017) 'Inconsistency Between Univariate and Multiple Logistic Regressions', *Shanghai Archives of Psychiatry*, 29(2), pp. 124-128.

Yap, B.W. and Sim, C.H. (2011) 'Comparisons of Various Types of Normality Tests', *Journal of Statistical Computation and Simulation*, 81(12), pp. 2141-2155.

Yor, I., Helman, S. and Vermaat, P. (2015) *Dutch Style Roundabout Safety-DRAFT PROJECT REPORT RPN751*. Transport Research Laboratory. [Online]. Available at: http://www.trl.co.uk/media/839260/ppr751_dutch_roundabout_safety_v1.pdf.

Yurdugül, H. (2008) 'Minimum Sample size for Cronbach's Alpha: A Monte-Carlo Study', *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, H. U. Journal of Education*, 35, pp. 397–405.