Exploring the Impact of Lighting Conditions on Injury Severity Utilizing Zero-Inflated Ordered Probit Model

University of
St Andrews

Ifeanyi Omeifejideofor 220024321

School of Mathematics and Statistics

August 2023

Supervised by Elham Mirfarah

Declaration

I hereby certify that this dissertation, which is approximately 12,000 words in length, has been composed by me, that it is the record of work carried out by me, and that it has not been submitted in any previous application for a degree. This project was conducted by me at the University of St Andrews from [May/2023] to [August/2023] towards the fulfillment of the requirements of the University of St Andrews for the degree of [Applied Statistics and Data Mining] under the supervision of Dr. Elham Mirfarah [ ].

Omeifejideofor

Abstract

This research aims to explore and identify various factors influencing collision injury severity for single vehicles across various lighting conditions. Utilizing the STATS19 datasets which records all traffic-related incidents in Britain from 2016 to 2021. The injury severity level are measured in hierarchical order, ranging from slight to severe and fatal, with the 'slight' injury being inflated. The study implements three distinct models to understand how the factors influencing accident severity vary across the different lighting conditions: Daylight, Darkness with streetlight lit, and complete Darkness. To model this, the Zero-Inflated Ordered Probit (ZIOP) was employed, and its performance was compared to that of the Standard Ordered Probit model using various evaluation metrics. The ZIOP surpassed the Standard Ordered Probit, indicating the possible presence of underlying factors and variability that influence the severity of collisions. The findings from this study indicate that a wide range of elements, including accident details, characteristics of vehicle drivers, weather conditions and trips, have varying degrees of impact on injury severity across the various lighting conditions.

Acknowledgement

I would like to thank my advisor, Dr. Elham Mirfarah for her invaluable support and insightful feedback that guided my research. I am equally grateful to the Transport Research Laboratory (TRL) for their support and guidance. A special acknowledgment goes to Caroline Wellbank for her unwavering commitment and support. My heartfelt thanks go out to my friends and family, who have supported me throughout my master's degree journey. I am especially grateful to John Hunja for his support and wise counsel that greatly assisted me during my studies.

"Until the lions have their own historians, the history of the hunt will always glorify the hunter."

- Chinua Achebe

"Age was respected among his people, but achievement was revered. As the elders said, if a child washed his hands he could eat with kings."

- Chinua Achebe

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Personal Interest

At the heart of this dissertation lies an insatiable thirst for knowledge. From a very young age, I have been captivated by various inventions and how they have evolved to become a critical need for human existence and societal functions. One invention that has piqued my interest is the automobile's engineering and the web of its numerous mechanical parts working in harmony to create an engineering masterpiece in motion. When in motion, the automobile interacts with the environment, infrastructure, and most importantly, these factors govern its direction, route, speed, and the safety of its occupants.

1.2 Background

Road safety remains a fundamental topic of interest commanding global attention. Cities have become more urbanized and have continued an upward growth trend as technology advances, leading to an ever-growing demand for infrastructure to support day-to-day activities and enhance the quality of life. Transportation has taken center stage in modern societies, becoming a need essential for daily commutes, economic growth, and global connectivity. Rapid urbanization is not without its challenges; it has led to an increase in traffic volume and also an increase in traffic incidents. Road traffic crashes now represent the eighth leading cause of death globally. They claim more than 1.35 million lives yearly and cause up to 50 million injuries; in 2021 alone, there were an estimated 1,558 reported road fatalities in the United Kingdom (Department for Transportation). In an effort to reduce the number of collisions and incidents affecting commuters, various agencies have invested resources into road safety research, creating policies and campaigns aimed at reducing collisions and making roadways safer for all users. The United Nations launched the Decade of Action campaign with the aim of reducing the number of road crashes by fifty percent in 2030 (United Nations, 2020). Similarly, The United Kingdom introduced the THINK campaign aimed at raising awareness and encouraging road users to act responsibly to reduce the risk for more vulnerable road users such as pedestrians, cyclists, and more (Department for

Transport, 2023).

These initiative has given rise to improved infrastructure, rigorous enforcement of traffic regulations, public awareness campaigns, and technological-driven solutions like Advanced Driver Assistance Systems (ADAS) to tackle the rise in traffic incidents. However, despite these efforts, a considerable gap exists in understanding the factors influencing road traffic collisions.

In recent years, there has been a rise in the number of literature and research on environmental factors and how they impact collision severity (Chan et al., 2022, Li et al., 2023, Yannis and Kondyli, 2013). These studies have demonstrated that environmental factors play a crucial role in influencing collision severity, acting as a significant determinant in the outcomes of various traffic accidents. One important environmental factor that influences collision severity is lighting conditions (Abbasi et al., 2022, Adanu et al., 2021); it affects visibility and also plays a crucial role in driver perception, reaction times, and overall vehicle control. Lighting conditions can vary from Daylight, Darkness with artificial illumination or complete Darkness. Each state impacts collision severity differently, influencing the potential severity outcome. lighting conditions hold a substantial promise as an area of research, its influence on visibility, a critical component of mobility, cannot be overemphasized. For instance, when visibility is reduced at night or in poorly illuminated areas can affect drivers' ability to recognize and react to potential hazards; on the other hand, excessive or inappropriate artificial lighting can cause glare and further impair a driver's vision. In this context, Anarkooli and Hosseinlou(2015) studied the varying effect of light conditions on injury severity; their findings indicated that the factors impacting collision severity varied under different lighting conditions. To illustrate, They found that during daylight and dark-lighted situations, collisions on longer road segments are associated with an increased probability of severe injuries since drivers are likely to travel faster. The study also indicated that in good lighting, driving straight could lead to a more severe collision, while in dark conditions, such collisions are more likely to result in minor injuries. Based on the study conducted, it is evident, beyond doubt, that the various lighting conditions significantly influence the severity of collisions.

The study of the impact of lighting conditions has been extensively explored. However, an area of research which have received limited focus is the underlying mechanism of collision generation in relation to the effect of various lighting conditions, especially when considering if they arise from two distinct factors, particularly in the case where a specific injury severity level is inflated.

1.3 Aim of the dissertation

This dissertation analyzes the effect of lighting conditions on injury severity in single-vehicle collisions. The zero-inflated ordered probit approach is used to model the data to gain insight into the complex interactions between the various lighting conditions and understand the variations of contributory factors found in the dataset.

- The core objective is to understand the impact of varying lighting conditions and their influence on injury severity, specifically under: Daylight, Darkness with streetlights lit, and Complete Darkness.

- Based on the aforementioned dimensions, the research seeks to identify the particular determinants influencing collision injury severity across various lighting conditions.

- Understand if the zero-inflated ordered probit model offers a more nuanced representation of the complexities of the data using its double hurdle structure compared to the standard ordered probit model.

1.4 Structure of the Dissertation

This dissertation is structured to allow for concise reading and understanding as follows: Chapter 2 presents an overview of existing literature on crash data and injury severity, drawing from the approach of existing literature to identify gaps and highlight the significance of the zero-inflated ordered probit models.

Chapter 3 explores the data utilized in this research, providing industry/statistical rationale and techniques for standardizing the data. Chapter 4 shed light on the logic behind the statistical estimation of the true and conditional state within the framework of the zero-inflated ordered probit model. It also explains in detail the components of the standard ordered probit model and the ZIOP model. Additionally, providing the resources and tools used for executing the research. Chapter 5 provides statistical estimates, model evaluation, and comparison, also providing detailed discussion of findings. Finally, it lays out pathways for future research.

Chapter 2

Literature Review

There has been a growing number of literature and various models used to evaluate factors that influence the severity of vehicle collision injuries. These models can be categorized into two main classes: Non-parametric models and Regression model

2.1 Non-Parametric Models

Non-parametric models refer to models that do not rely on assumptions regarding the functional relationship between variables; hence, they do not rely on data belonging to any particular parametric family of probability distributions but are rather flexible and can adapt to the complexities of the actual data. The non-parametric models applied to collision injury severity include decision tree-based models (Azhar et al., 2022, Aci and Özden, 2018, Sapri et al., 2017), Deep learning models (Silver et al., 2022), Neural networks (Siamidoudaran and İşçioğlu, 2019, Assi et al., 2020, Arhin and Gatiba, 2019). The findings from the nonparametric model have been great in providing insight into the factors that influence collision severity; however, these models do not provide interpretable parameters that quantify the impact of each variable of interest. To elaborate, the models might identify a variable that is important in making predictions, but they do not provide a precise measure of the extent the particular variable of interest will influence an outcome. This lack of quantification makes it challenging to understand the underlying structure of the data and the relationship between variables, which can act as a limitation in situations where interpretability is essential.

2.2 Regression models

Regression models are used as an alternative for researchers who are interested in modeling traffic-accident data on injury severity. This is because of the shortfalls of the non-parametric models. The two broad categories of regression models used in collision injury severity analysis are the unordered-response models, such as the multinomial logit model (Hausman and McFadden, 1984), and the ordered response model, such as the standard ordered probit model(Cameron and Trivedi, 2009). The multinomial logit model do not account for the natural ordering of the outcome, however, they are treated as separate categories with no implied order, while the standard ordered probit model accounts for the ordering of the outcomes.

2.2.1 Unordered-Response Models

The unordered response model has been widely adopted in analyzing collision injury severity data. Studies that have adopted the unordered response methods include multinomial logit models (Mahikul et al., 2022, Moomen et al., 2023, Adebisi et al., 2019), Mixed multinomial logit models (Liu et al., 2022), Nested multinomial logit model (patil et al., 2012). Using nominal models like mixed and nested multinomial logit models might provide insight into capturing the unobserved heterogeneity relating to collision injury severity data; however, they do not account for the ordering of the injury data which might lead to inefficient model estimates and loss of information.

2.2.2 Ordered-Response Models

The ordered response model has been widely adopted in analyzing the collision injury severity data. This is because these models can account for the ordinal nature of the inherent injury severity normally classified as slight, severe, and fatal injuries or using a similar classification scale that conveys the same meaning. Studies that have utilized the ordered response method include the standard ordered logit models (Rezapour et al., 2018, Mphekgwana 2022) and the standard ordered probit models (Abdel-Aty, 2003, Garrido et al., 2014). These conventional ordered response models have been widely criticized for having a biased estimate. A study by Ye and Lord (2011) investigated the effects of under-reporting of crash data utilizing models commonly used for analyzing collision severity. Their study revealed that these types of ordered response models are prone to issues relating to under-reporting. Furthermore, another source of bias in the estimates of the ordered response models may result from their inability to account for the underlying mechanism influencing collision severity and its heterogeneity, especially in cases where the outcome variable is inflated (Harris and Zhao, 2007).

In a bid to improve the shortfall of the ordered response models, researchers have constantly explored alternative variants of the standard ordered response models. For instance, the heteroscedastic ordered logit and probit model seeks to address the issue of constant variance (homoscedasticity); the ordered response model assumes a constant variance in the error term across all observations. However, this assumption may be violated, hence leading to unreliable and biased estimates. The heteroscedastic model relaxes this assumption, thus allowing for flexibility in how the data is modeled. One limitation of the model is the complexities surrounding how the heteroscedastic element is parameterized. In their study, Donnell and Connor (1996) considered factors such as occupant age, vehicle speed, vehicle year, and time of crash to determine the variability in their error terms. Another limitation, as described by Williams (2009), indicates that while these models can account for this variability, there is little evidence of how well these models accomplish this task.

Furthermore, The sequential logit and probit models are also extended versions of the standard ordered probit and logit models. According to Yamamoto et al (2011), the sequential models performed better than the standard ordered response model in analyzing accident data with less bias introduced from under-reporting, and the parameter estimates remained unbiased. One limitation associated with the model is that the error term associated with each injury severity is independent (Savolainen et al., 2011).

Finally, the generalized ordered logit model(GOL) extends the ordered response models. it relaxes the assumption that parameter estimates are constant across severity levels (Savolainen et al., 2011); this is also known as the proportional odds assumption (Quddus et al., 2009). The GOL model is particularly useful when the effect of some predictors differs across levels of collision severity. One major limitation of the model is being prone to generate negative probability estimates.

2.3 Zero-Inflated Models

Traffic incidents relating to collision injury severity are very complex to model and analyze due to many factors influencing the outcome; a particular challenge specific to modeling and analyzing these incidents involves accounting for inflation. Inflation is very common in records of traffic incidents, and this is caused by a myriad of events. An illustration of this could be caused by records of collisions resulting in no injuries which are often recorded as zero in the severity scale. Zero-inflated models are better suited in accounting for the abundance of zero since they introduce an additional component to model the excess zeros specifically. Additionally, these models offer more stability, consistent estimates, and a more nuanced approach to modeling probabilities relating to collision severity; hence, it is preferred over the ordered response models considering their limitations.

In recent years, there has been a growing number of literature that criticizes the zero-inflated approach for modeling traffic incidents. Lord et al.(2005) suggested that a zero-inflated model assumes that a traffic location is fundamentally in a safe or an unsafe state. This is in reference to the additional zero component in the model that has a long-term average of zero; the authors supported this argument with a large body of literature indicating that a large percentage of collisions precisely around 70 - 90% occurs due to human error, and this invalidates the assumptions of safe sites. In their follow-up paper (Lords et al., 2007); they reiterated their argument and further stated that the assumption of a safe site contradicts established theoretical expectations. However, some researchers oppose the critique of the zero-inflated models, suggesting that models should be selected on the conditions regarding their effectiveness in achieving the objectives of the research(Pew et al., 2020); this also aligns with Buckland et al.(1997) stating that model selection should be based on identifying the best approximating model and also accepting that the data can never support, and we can never identify, the true model. Additionally, the authors (Pew et al., 2020) strengthened their argument by asserting that their analysis of the zero-inflated model shows that the model does not group traffic sites into safe or unsafe states; this is because from a detailed examination of the probability mass function of a zero-inflated distribution reveals that no site is fundamentally safe as there is a non-zero probability assigned to each integer observed.

Many studies have employed the zero-inflated models to address the problem of inflation in the outcome variable in crash data. For instance, Lord et al.(2005) employed the zero-inflated approach in modeling vehicle crash data. The models include the zero-inflated negative binomial (ZINB) and the zero-inflated Poisson (ZIP). The findings indicated that the ZINB model had a superior fit when compared to the standard Poisson and the negative binomial model; one reason for this might be that the ZINB models account for overdispersion even after accounting for the excess zeroes (Lee et al., 2002).

Furthermore, the Markov Switching Negative Binomial (MSNB) model created by Malyshkina et al.(2010) also addresses the heterogeneity in crash data. the model introduces a latent state variable that determines the state of all traffic sites at a specific time period. At each time period, the state can only assume two values $st = 0$ and $st = 1$, referred to as more frequent and less frequent states. The model allows for switching between states, hence enabling the model to capture the dynamics of the traffic crash data, and a Markov chain governs the transitions. In their finding, the MSNB model was compared to the ZINB model. Results show that the MSNB model is a viable alternative and produces a superior fit compared to the ZINB model.

Finally, the zero-inflated model developed by Harris and Zhao (2007) is an extension of the standard ordered probit model. The model is particularly useful when the outcome variable has an ordinal structure. The model incorporates a two-level framework to account for the excess zero value in a dependent variable while recognizing that the inflated zero value could originate from two distinct sources. The authors demonstrated this in their research on tobacco consumption, however, suggesting that non-tobacco consumers may be formed by two distinct groups. the first being non-consumers with perfectly inelastic demand to prices and income, and the second being non-consumers who may consume once the price is right. Building on this rationale, the authors further elaborated that the former may relate to personal demographics and socioeconomic status, whilst the latter group may exhibit behavior similar to other non-consumers and may be more responsive to economic factors such as prices and income; additionally, their findings demonstrated a few advantages of the ZIOP model over the standard ordered probit model.

Chapter 3

Data Exploration

3.1 Data Description

Over the years documenting and reviewing traffic-related incidents has become crucial for public safety policies, city planning, and infrastructure investments in countries. Britain is no exception, as it strives to improve road safety by implementing the STATS19 data collection system. The STATS19 database is a standardized police crash report of all road traffic accidents that resulted in personal injury (Department for Transport); it is essential for conducting studies and gaining insights into the complex dynamics of road safety. Researchers use statistical and machine learning methods to analyze this dataset thereby uncovering connections between different variables, identifying patterns, and understanding the underlying relationships that influence accident outcomes. However, the dataset records injury-only accidents and this limits the dataset scope, as accidents without injuries are not captured in the records(Imprialou and Quddus, 2019).

The STATS19 dataset measures injury severity using three distinct classification methods, which are slight, severe, and fatal. However, these classifications are similar to the KABCO scale, commonly used by professionals in road safety and accident analysis, which categorizes injury severity as non-capacitating, incapacitating, and fatal. The STATS19 dataset contains layers of information related to various accident scenarios. These can be categorized into details concerning casualties, vehicle-related information, pedestrian-related information, environmental factors, contributory factors, and all road information.

To explore in-depth, the accident data consist of vehicle-related information, such as maneuverability, skidding and overturning, point of impact, and vehicle age. Additionally, the dataset records the vehicle's location post-collision, indicating whether the vehicle was on the carriageway or off the carriageway.

Moving to casualty-related information, the dataset includes information such as casualty severity, age of casualty, and details relating to pedestrian involvement. This includes pedestrian location, pedestrian movement, whether pedestrian were road maintenance workers.

Environmental factors are essential in determining the dynamics of an accident (Grigorios et al., 2020); The dataset records essential environmental variables at the time of the accident, such as weather conditions and lighting conditions.

The dataset includes elements known as contributory factors, which are conditions or actions the police officer attending the scene of the accident believed contributed to the accident. Finally, the dataset captures all other essential variables that are pivotal in determining the dynamics of an accident, such as speed limit, junction details, road types, road class, junction control, carriageway hazards, and urban or rural areas.

3.2 Data Preparation

The STATS19 dataset containing accident information from 2016 - 2021 was used for this analysis, and it includes accident data, vehicle, and casualty information within the five-year period. The accident data includes 562,439 observations, casualties 728,541 observations, while the vehicle data contains 1,034,534 observations. However, the researcher noticed that the higher count of observations in the casualties and vehicle data resulted from the comprehensive recording of each individual casualty severity.

In an incident involving multiple vehicles or people, the dataset captures details for every vehicle and person involved in the accident; therefore, to address these instances in the vehicle data, only accidents or collision involving a single vehicle was chosen for this analysis. Additionally, to account for the higher count in the casualty data, the approach taken was to select the highest casualty severity value. This decision was motivated by cases where duplicates occurred due to accidents involving multiple individuals. The files were linked together using the accident index variable, which is common across all three distinct files. The combined dataset was cleaned to ensure accuracy and reliability utilizing statistical and industry knowledge.

One aspect of the cleaning process involves the removal of negative values. These negative values do not exist in the current encoding of the STATS19 accident values and represent likely missing values, which was done for all variables in the combined dataset. To enhance the dataset consistency and mitigate the introduction of bias in the model, all encoding representing unknown values was removed from variables like drivers' sex, weather conditions, and road types. For example, encoding such as 'unknown driver sex' was used to depict hit-and-run accidents where police could not trace the driver. Studies have shown that hit-and-run accidents involved pedestrians, and factors like drinking and driving were some of the key reasons why hit-and-run drivers left the scene as their decision-making became impaired (Hopkins et al., 2017). Therefore, using variables like this in the model can introduce bias.

Additionally, variables that do not contribute to driving the research were removed, including accident reference, casualty reference, if a police officer attended the scene of the accidents, home area type, etc. Accidents involving pedestrians and pedal cyclists were also excluded from the model because pedestrians have different injury mechanisms (Jiang et al., 2013). To elaborate further, pedestrians are vulnerable and exposed, thus, lacking the protection that a vehicle provides for its occupants; hence accidents involving pedestrians might lead to a higher and different injury severity compared to vehicle passengers.

Furthermore, collisions that occurred at an intersection and intersection-related variables were excluded because intersections involve complex interactions as vehicles, pedestrians, and cyclists cross paths in different directions. This leads to different types and patterns of accidents compared to those on continuous road segments where the movement is largely in one direction. Large trucks, coach buses, and non-traditional vehicles were also excluded from the dataset because their size and weight subject them to varying accident types, and the injury mechanism differs in these vehicles.

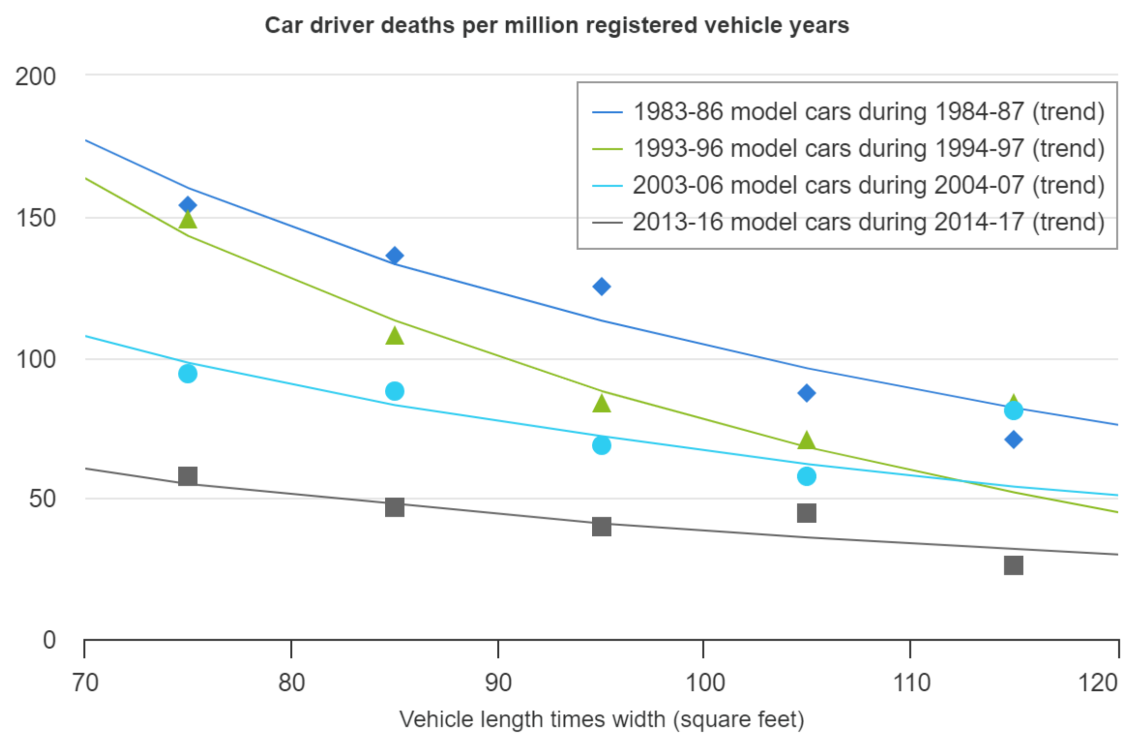

Figure 3.1: Decline in Crash Deaths Relative to Vehicle Size. Source: (28)

An illustration of this difference in injury mechanism can be observed in Figure 3.1, where crash deaths decline as vehicle size increases. According to the Insurance Institute for Highway Safety (IIHS), larger vehicles offer better protection for their occupants because the part of the vehicle between the front bumper and the occupant compartment absorbs energy from crashes by crumpling. As a result, longer front ends offer better protection in frontal crashes. Heavier vehicles also tend to continue moving forward in crashes with lighter vehicles and other obstacles, so the people inside them are subject to less force. Additionally, motorcycles and other non-conventional vehicles were excluded due to their limited representation within the dataset, hence, leaving only accidents involving cars.

The vehicle age was also restricted to a maximum of 24 years due to the evolution of safety regulations and advancements in car safety standards. This is because cars within this range share similar safety features such as airbags, seat belts, and whiplash systems which have become standard for auto manufacturers.

The Pearson correlation coefficient was calculated to avoid the potential issue of multicollinearity between the variables, as strong correlations could lead to instability in the parameter estimates and inflated standard errors, hence causing difficulties in discerning the effect of each individual predictor. A threshold of 70 percent was set to determine the correlation. Correlation coefficients falling between 70 and 90 percent indicates a high correlation.

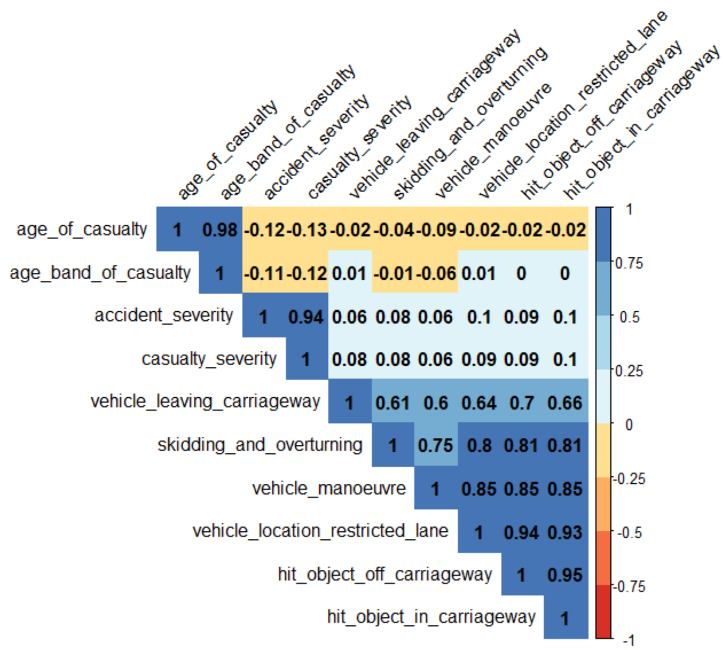

Figure 3.2: Correlated Variables in the Dataset

In Figure 3.2, a group of highly correlated variables can be observed. However, certain variables were omitted from the model to address multicollinearity and reliability of the model. Variables such as hit objects off the carriageway, vehicle location restricted lane, and vehicles leaving the carriageway were excluded; the exclusion of these variables was not limited to their significant correlation but also because their impact on injury severity may be less related to lighting conditions and more influenced by other factors such as driver behavior, distractions and more. This is consistent with studies (McLaughlin et al., 2009) that highlight how accidents involving vehicles leaving the carriageway are often caused by distracted drivers.

Additionally, variables sharing similarities, such as accident severity and age band of casualty, which are closely linked to severity of casualty and age, were excluded; hit objects in the carriageway and vehicle maneuvers were removed to reduce the effect of collinearity in the model. The final dataset used for model estimation includes 16,821 observations of single-vehicle collisions. In terms of injury severity, slight injury accounted for 74.55% of collisions, severe injury accounted for 23.95%, and fatal injury constituted 1.49% of collision injury severity.

Figures 3.3 and 3.4 depict the variations in injury severity and its distribution across different lighting conditions. The collision data consistently exhibit a higher frequency of slight injuries. This distribution pattern might suggest the possibility of underlying injury severity patterns influencing the mechanism of collision occurrence.

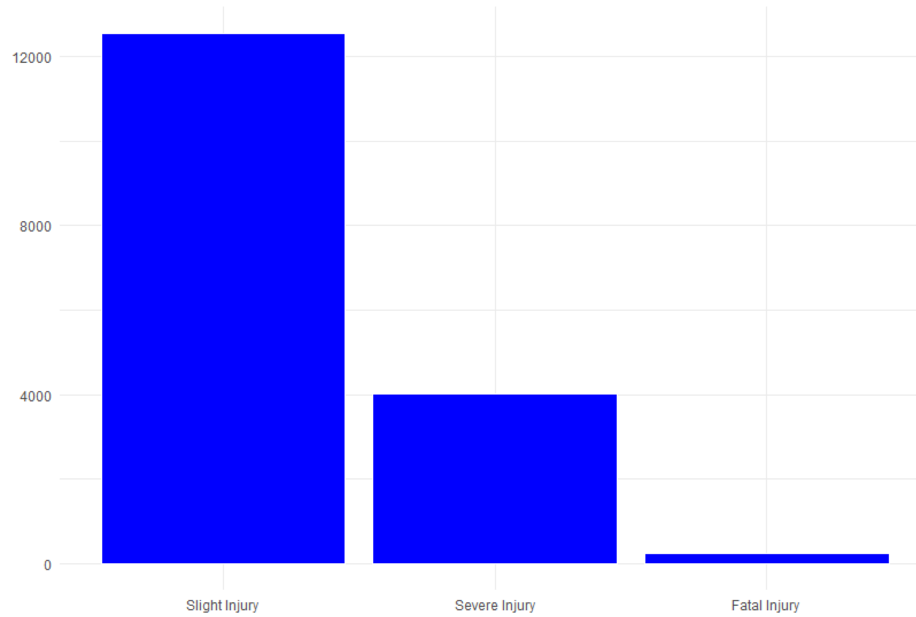

Figure 3.3: Histogram of Injury Severity

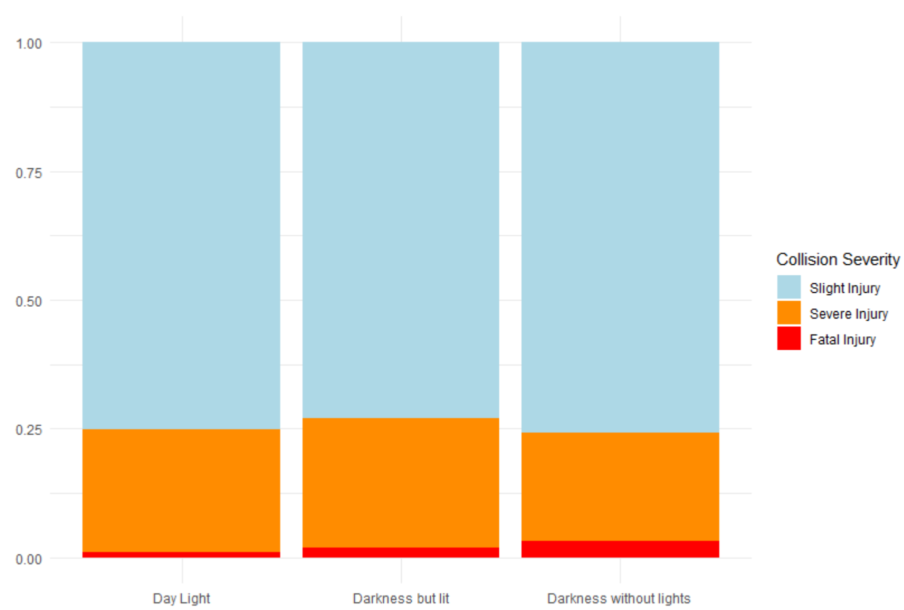

Figure 3.4: Histogram of Injury Severity Distribution across various Lighting Conditions

The collision data used for this research contains an exhaustive list of information. They are summarized and presented in Table 3.1 below:

Table 3.1: Variable Descriptions and Summary Statistics

Variable	Description	Mean or % of 1	Min	Max
Road type indicator	1 if collision occurred on a dual/single motorway, 0 otherwise	93.13%	0	1
Speed limit indicator	1 if the speed limit is above 40 mph, 0 otherwise	14.22%	0	1
Weather condition indicator	1 if collision occurred in fine weather, 0 otherwise	81.72%	0	1
Road surface conditions indicator	1 if collision occurred in wet roads, 0 otherwise	29.98%	0	1
Special conditions at site indicator	1 if road hazards present (Debris present, Defective Road), 0 otherwise	1.68%	0	1
Carriageway hazard indicator	1 if collision involves carriageway hazards, 0 otherwise	2.68%	0	1
Urban or Rural area indicator	1 if the collision occurred in a rural area, 0 otherwise	16.74%	0	1
Vehicle type indicator	1 if collision involved mini buses and vans under 3.5 tonnes, 0 otherwise	7%	0	1
Skidding and Overturning indicator	1 if vehicle skidded, overturned, and jackknifed during the collision, 0 otherwise	11.59%	0	1
Point of impact indicator	1 if the vehicle was impacted (front, back, and side), 0 otherwise	96.96%	0	1
Journey purpose of driver indicator	1 if collision is work related, 0 otherwise	26.58%	0	1
Sex of driver indicator	1 if the driver is female, 0 otherwise	32.25%	0	1
Age of driver indicator	1 if the driver was younger than 23 years, 0 otherwise	18.55%	0	1
Age of vehicle indicator	1 if the age of the vehicle is above 10 years, 0 otherwise	35.24%	0	1
Sex of casualty indicator	1 if the casualty is female, 0 otherwise	44.4%	0	1
Age of casualty indicator	1 if casualty age is younger than 30 years, 0 otherwise	46.33%	0	1
Car passenger indicator	1 if collision involved a passenger car, 0 otherwise	4.51%	0	1

The use of an indicator variable allows for the comparison of groups within a categorical variable. The coefficient associated with an indicator variable provides an estimate of the average difference in the dependent variable between the groups identified by the indicator variable and reference group after taking into account other variables in the model. The reference group consists of observations where the indicator variable is always zero. In the zero-inflated ordered probit model, they represent categorical or binary variables however this method also transforms qualitative data into a format suitable for statistical analysis. By representing these variables as indicators the model can include their impact on the outcome variable while preserving the structure required for the modeling process.

Chapter 4

Methodology

4.1 Brief Description

This section provides an overview of the standard ordered probit model and the zero-inflated ordered probit model(ZIOP) in addressing the critical issue of modeling a zero-inflated collision severity. It explains the underlying framework of the model and covers aspects of model evaluation and test of model fitness.

4.2 Standard Ordered Probit Model

The ordered probit model is a statistical model that handles ordinal categorical variables, it has been widely used in the analysis of traffic-related incidents and collision injury severity.

The structure of the model can be described as follows: (Washington et al., 2003)

$$r_i = \beta x_i + \epsilon_i \, ,$$

where r_i is a latent variable determining the injury severity of a collision event involving an individual i, x_i represents a vector of the observed explanatory variables, β represents the estimated parameters, and ϵ_i is the random error term that follows a standard normal distribution.

The latent variable r_i is unobserved; to account for this, an observable variable y_i is introduced depicting the injury severity which can be expressed as follows: (O'Donnell and Connor, 1999)

$$yi = \begin{cases} 1 & \text{if } r_i \leq \mu_1 \\ z & \text{if } \mu_{z-1} < r_i \leq \mu_z \\ Z & \text{if } r_i > \mu_{z-1} \, , \end{cases}$$

here, $\mu = \{\mu_1, \ldots, \mu_z, \ldots, \mu_{z-1}\}$ represents the threshold for each severity level of injury; essentially, it denotes the boundaries that define y_i, where μ_{z-1} is the maximum ordered threshold value which corresponds to the highest injury severity level Z.

The probability that a collision results in an injury severity z is equal to the probability that the latent variable r_i falls within a range defined by two thresholds. To illustrate further, provided the value of x_i the probability associated with each severity level can be expressed as follows:

$$Pr(y = z) = \Phi(u_z - \beta x_i) - \Phi(u_{z+1} - \beta x_i).$$

here Φ represents the cumulative distribution function, u_z and u_{z+1} represents the lower and upper threshold for the injury severity level z.

4.3 Limitation of the Ordered Probit Model

The standard ordered probit model is quite effective, in handling the nature of the data as it takes into account the varying levels of injury severity. However, one major criticism of this model is that it has a limited capacity in explaining the abundance of an inflated injury severity variable especially when this variable is generated from two this distinct sources. However, this is because it assumes a uniform severity function for all types of injuries. This oversimplified assumption can lead to an inaccurate representation of the complexities within the data, potentially limiting the model's ability to accurately predict and explain injury severity across the various types of injuries.

4.4 Zero-Inflated Ordered Probit Model

This research utilizes the ZIOP model to address the limitations of the standard ordered probit model when analyzing accident-related data; however, this is because accident data contains an overwhelming case of zero inflation; such cases could be likened to near-misses or when safety measures prevent serious outcomes hence creating an excessive amount of zeros which presents a substantial challenge for researchers as conventional models struggle with this imbalance.

The concept of zero inflation originates from the Poisson model of count data being overdispersed by zeros (Lambert, 1992). The ZIOP model was developed by Harries and Zhao(2007) as an extension to the standard ordered probit model, the model accounts for the possibility of zeros generated from two distinct sources. The generation of zeros from the two separate sources causes an abundance of zeros in the dataset, leading to the term zero-inflated. The ZIOP model utilizes a double hurdle structure which consists of the binary split and ordered probit, the binary split, often referred to as the first level of the model, distinguishes between observations whose outcome is always fixed and those that follow the ordered probit process, while the ordered probit or the second level of the model estimates the probabilities of the different ordinal outcomes.

In this research, the researcher relied on the STATS19 dataset, derived from the standardized police crash reports (Department for Transport, 2022), which records injury-only accidents. A large portion of the injury data comprises of slight injuries; this accounts for roughly over 74 percent of injuries reported. This imbalance suggests the underlying process generating injuries may not be uniform and could originate from two sources. The standard ordered probit model can account for the ordinal nature of the data, but it assumes a consistent severity function across all types of injuries. However, this assumption becomes problematic and may result in biased estimates. To address the excessive presence of slight injuries and explore sub-groups within the data, the ZIOP model is recommended. This approach allows for an understanding of the hidden heterogeneity in the generation process, thus providing a more accurate analysis of the injury severity levels.

The police crash report is often prone to inaccurate classification and under-reporting of accident instances and injury severity; many slight or non-fatal injuries are not reported to the police, as this is evident in hospital records, surveys, and compensation claims data, which show relatively higher injury numbers than those documented in the police accident report (Department for Transport, 2022). Previous studies have demonstrated that accident severity can often be over- and under-classified. These results were obtained by comparing the data from hospital records with that of police reports. It was found that 19.3 percent of cases involving victims who needed to be hospitalized for longer than 12 hours (a situation usually associated with a severe injury) had incorrectly been classified in the police records as slight injuries. Alternatively, police report misclassified 9.7 percent of victims as having serious injuries despite spending less than 12 hours in the hospital, which is often suggestive of slight injuries (Tsui et al., 2008).

The two distinct populations of accidents could explain the excessive representation of slight injuries in the data. The ZIOP model excels in handling these types of data by utilizing its double hurdle structure, which considers two underlying states, The first state - A minor injury state that can be formed by minor collisions with less severe consequences(minor or possible injuries), This could be likened to minor accidents or collisions where damage upon impact is very low, resulting in minor injuries. As an illustration, consider the following scenario when a vehicle bumps into a small stationary object at a slow speed, The energy released upon impact is low. Hence, occupants may not sustain injuries, and even if there is a chance of injury, it would likely be extremely minor, however, recorded under the category of slight injuries (Grigorios et al., 2020; Fountas and Rye, 2019). The second state - The ordered injury state, is concerned with slight injury accidents that could lead to more severe consequences(more severe injuries) under certain unfavorable conditions, Building upon the aforementioned illustration, consider a similar scenario where the vehicle bumps into a larger stationary object at slow speed, making the impact much more significant hence could lead to more severe injuries. The ordered probit state also accounts for serious and fatal injuries (Grigorios et al., 2020; Fountas and Rye, 2019).

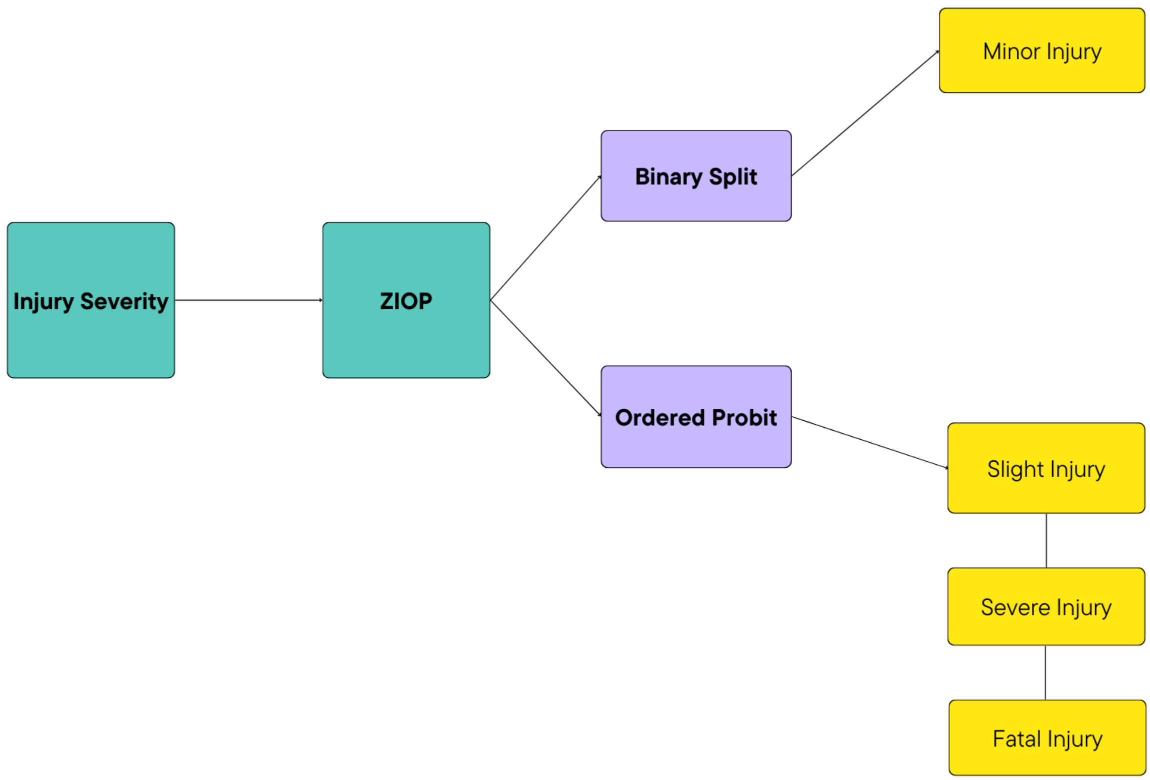

Figure 4.1: A sketch of the zero-inflated ordered probit model

4.5 Components of the Zero-inflated ordered Probit(ZIOP)

The following describes the components of the zero-inflated ordered probit model (Harris and Zhao, 2007). Let's denote a binary variable indicating a split in the first level of the ZIOP model where $g_i = 1$ (accident associated with minor injuries) and $g_i = 0$ (accidents that do not result in minor injuries)

$$g_i^* = a_i x_i + \varepsilon_i,$$

where g_i^* is a latent variable that indicates the propensity of an accident i to be associated with the minor injury state and g_i is derived from the latent variable g_i^*, x_i represents a vector of determinants for the choice between both states in the binary split, a_i represents the vector of parameter estimates, and ε_i represents an error term that follows a standard normal distribution.

The probability of being in an accident that does not result in a minor injury is

$$\Pr(g_i = 0|x) = \Pr(g_i^* \leq 0|x) = 1 - \Phi(a_i x_i),$$

here Φ represents the CDF of the standard normal distribution. To determine factors influencing the severity of injuries resulting from accidents in the ordered probit category which is conditional on $g_i = 0$, the injury level $\tilde{y}$ ($\tilde{y} = 0,1,\ldots,J$) is connected to a latent variable y^* through the ordered probit regression function:

$$y_i = w_i \gamma_i + u_i,$$

where w_i is a set of unknown parameters representing the variables influencing the frequency of non-minor injury levels, γ_i represents explanatory variables in the non-minor injury state, and u_i refers to a standard normally distributed error term independent of ε_i.

The mapping between y_i and $\tilde{y}$ is obtained by

$$\tilde{y} = \begin{cases} 0 & \text{if } y_i \leq \mu_0 \\ 1 & \text{if } \mu_0 < y_i \leq \mu_1 \\ 2 & \text{if } y_i > \mu_1 \,, \end{cases}$$

where μ represents the threshold for each severity level of injury; essentially, it denotes the calculated boundaries partitioning the underlying latent response variable y_i. The threshold values adhere to constraints such as $\mu_0 \leq \mu_1 \leq \mu_2 \leq \ldots \leq \mu_j$, where μ_j is the maximum ordered threshold value.

The ordered probit state allows for a minor injury level(i.e., accident associated with minor injury or even lower severity e.g., possible injuries), and there is no requirement that $x_i = \gamma_i$ hence allowing for a separate explanatory variable to be used in both part of the model. The probability of each injury level in the ordered probit states is given as:

$$\begin{aligned} P(\tilde{y} = 0) &= \Phi(-w_i \gamma_i), \\ P(\tilde{y} = 1) &= \Phi(\mu_2 - w_i \gamma_i) - \Phi(-w_i \gamma_i) \\ P(\tilde{y} = J) &= 1 - \Phi(\mu_J - w_i \gamma_i). \end{aligned}$$

The modeling for the binary state(minor injury) and ordered probit states(injury level) can be synthesized using the equation:

$$y = \tilde{y} \cdot g_i.$$

To observe a $y = 0$ outcome (minor injury with a less severe consequence, i.e., a possible injury), we need either that $g_i = 1$ (an accident resulted in a minor injury) or jointly that $g_i = 0$ (a non-minor injury accident) and $\tilde{y}_i = 0$ (under certain conditions the outcome of the accident was, however, a minor injury). However, to observe a $\tilde{y}_i > 0$ outcomes (injury level), we require jointly that the accident did not result in a minor injury ($g_i = 0$) and that the accident led to a more severe injury. Furthermore, the ZIOP assumes ε_i and u_i are identical and independently follow the standard Gaussian distributions.

$$\Pr(y) = \begin{cases} \Pr(y=0|\gamma,x) & = \Pr(g_i=1|x) + \Pr(g_i=0|x)\Pr(\tilde{y}=0|\gamma,g_i=0) \\ \Pr(y=j|\gamma,x) & = \Pr(g_i=0|x)\Pr(\tilde{y}=j|\gamma,g_i=0), (j=1,...,J) \\ \Pr(y=0|\gamma,x) & = [1-\Phi(a_ix_i)] + \Phi(a_ix_i)\Phi(-w_i\gamma_i) \\ \Pr(y=j|\gamma,x) & = \Phi(a_ix_i)[\Phi(u_j-w_i\gamma_i) - \Phi(u_{j-1}-w_i\gamma_i)], (j=1,...,J-1) \\ \Pr(y=J|\gamma,x) & = \Phi(a_ix_i)[1-\Phi(u_{J-1}-w_i\gamma_i)]\,. \end{cases}$$

The maximum log-likelihood estimation (MLE) method estimates the model's parameters. The log-likelihood function expresses this estimation as follows:

$$l(\theta) = \sum_{i=1}^{n}\sum_{j=0}^{J} h_{ij}\ln[Pr(y_i=j|x,\gamma,\theta)]\,,$$

where θ represents a vector of parameters to be estimated and h_{ij} is the indicator function expressed as:

$$h_{ij} = \begin{cases} 1 & \text{if an individual } i \text{ chooses ordered outcome } j, \\ 0 & \text{otherwise .} \end{cases} \qquad (i=1,\ldots,n;\; j=0,1,\ldots,J)$$

4.6 Marginal effect

The marginal effect is a fundamental part of the ZIOP model; this is because the zero-inflated model measures the impact of each independent variable on the severity of injuries in a single-vehicle collision. However, the model is limited in measuring the probability of a specific injury severity changes alongside changes in the independent variable hence the marginal effect explores the influence of a specific independent variable while holding other variables constant; it provides a comprehensive understanding of how changes in an independent variable affect the expected outcome.

In a binary model, the marginal effect for a continuous variable x is computed as follows:

$$ME_{\Pr(gi=0)} = \frac{\partial \Pr(g_i=0)}{\partial x} = \phi\Big(ax\Big)a\,,$$

where ϕ represents the probability density function (pdf) of the standard normal distribution. In the Binary marginal effects the coefficients merely indicate probabilities based on specific covariates. In an ordered probit model, the marginal effect shows the effect of a change in a specific variable on the probability of each ordered category occurrence, given the specific values of covariates denoted by their parameter estimates. Nonetheless, when conditions are met, like when covariates don't appear in non-linear terms, a positive (or negative) coefficient correlates with a decreased (or increased) likelihood of the initial category and an elevated (or decreased) chance of the topmost category. Given this 'single crossing attribute'—where some probabilities decrease while others increase—the direction of the probability shift can be deduced from the coefficient's sign. The marginal effects in the ordered model can be calculated as follows:

$$ME_{Pr(y=j)} = \frac{\partial Pr(y=j)}{\partial w} = \left[\phi(u_{j-1} - w\gamma) - \phi(u_j - w\gamma)\right] w \, .$$

where ϕ represents the pdf for a standard normal distribution, it is used to calculate the probability of a random variable is less than or equal to a given value, u_{j-1} represents the threshold or cutoff point for the j-1 level of injury severity; it determines the boundary between different levels or categories of the dependent variable and w are coefficients that correspond to specific factors or variables in the model and γ represents the explanatory variables in the ordered model. The marginal effect of the ordered model calculates the difference in the probability density function between thresholds when multiplied by the parameter estimates w it reveals the change in the predicted probability of being in each category for a one-unit increase in the respective predictor, assuming other predictors are held constant.

The marginal effect of the ZIOP model can be decomposed into the binary probit marginal effect and ordered probit marginal effect(Harris and Zhao, 2007). The Binary probit marginal effect measures the effect of a change in an independent variable on the predicted probability of identifying an injury as minor, given all other variables are held constant. The binary probit marginal effect is calculated as follows:

$$ME_{Pr(y=j)} = \left[\Phi\left(\frac{u_j - w\gamma + \rho a x}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{u_{j-1} - w\gamma + \rho a x}{\sqrt{1-\rho^2}}\right)\right]\phi(ax)a^* \, ,$$

The ordered probit marginal effect measures the effect of a change in an independent variable on the predicted probability of categorizing an injury into various severity levels, such as slight, severe, and fatal. The marginal effect for the ordered probit process of the ZIOP model is calculated as:

$$ME_{Pr(y=j)} = \left[\Phi\left(\frac{ax + \rho(u_{j-1} - w\gamma)}{\sqrt{1-\rho^2}}\right)\phi(u_{j-1} - w\gamma) - \Phi\left(\frac{ax + \rho(u_j - w\gamma)}{\sqrt{1-\rho^2}}\right)\phi(u_j - w\gamma)\right]\phi(w\gamma)w^* \, ,$$

The Overall marginal effect of the ZIOP model combines both the effect at the binary split level and the effect at the ordered probit level. The overall marginal effect can be calculated as follows:

$$ME_{Pr(y=j)} = \left[\Phi\left(\frac{u_j - w\gamma + \rho a x}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{u_{j-1} - w\gamma + \rho a x}{\sqrt{1-\rho^2}}\right)\right]\phi(ax)a^* +$$

$$\left[\Phi\left(\frac{ax + \rho(u_{j-1} - w\gamma)}{\sqrt{1-\rho^2}}\right)\phi(u_{j-1} - w\gamma) - \Phi\left(\frac{ax + \rho(u_j - w\gamma)}{\sqrt{1-\rho^2}}\right)\phi(u_j - w\gamma)\right]\phi(w\gamma)w^* ,$$

where Φ represents the CDF of the standard normal distribution and ϕ represents the pdf of the standard univariate normal distribution. Additionally, w^* and a^* include coefficients linked to the variables present in either the minor injury state or ordered probit state, as well as those common variables shared by both equations. ρ measures the correlation coefficient between the binary probit and ordered probit components; however, the ZIOP model assumes independence between binary probit and ordered probit components; hence $\rho = 0$.

To understand the impact of factors determining injury severity $Pr(\tilde{y} = j|g_i = 0)$, the marginal effect for the ordered probit component of the ZIOP model is used in this research to illustrate the change in the probability of an accident leading to a particular severity due to a unit change in the value of the independent variable. The Marginal effect can be described as follows:

$$ME_{Pr(y=j)} = \left[\Phi\left(\frac{ax + \rho(u_{j-1} - w\gamma)}{\sqrt{1-\rho^2}}\right)\phi(u_{j-1} - w\gamma) - \Phi\left(\frac{ax + \rho(u_j - w\gamma)}{\sqrt{1-\rho^2}}\right)\phi(u_j - w\gamma)\right]\phi(w\gamma)w^* .$$

4.7 Model Comparison and Evaluation

The AIC(Akaike Information Criterion) and BIC(Bayesian information Criterion) are used in statistical modeling and analysis for model selection; however, they provide insight of the balance between model fit and complexity hence aiding the selection of the most appropriate model. The study also aimed to evaluate and compare the goodness of fit of two different models: the standard Ordered Probit model and the ZIOP model. The objective was to identify the model that exhibits the highest level of fit to the data.

The Model is evaluated using AIC(Akaike, 1973) as follows:

$$\text{AIC}(M_k) = -2\log L(M_k) + 2k .$$

where $L(M_k)$ represents the likelihood of corresponding to the model M_k and k represents the number of parameters in the model (Kisslinger 1996).

The model is also evaluated using BIC(Schwarz, 1978):

$$\text{BIC} = -2 \log L(M_k) + \log(n) \cdot k \, .$$

where $L(M_k)$ represents the likelihood of corresponding to the model M_k, and k represents the number of parameters in the model and n is the sample size. The BIC incorporates a larger penalty compared to the AIC hence the BIC selects a model with fewer parameters, resulting in sparse model selection (Kisslinger 1996).

The Vuong test is important when comparing non-nested models; It is designed to address the challenge of comparing models with different underlying assumptions or functional forms. This is particularly useful for comparing the standard ordered and the ZIOP model. The model is evaluated using the Voung Test(Voung 1989):

$$V = \frac{\sqrt{n}\,((1/n)\sum_{i=1}^{n} m_i)}{\sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(m_i - \bar{m})^2}} \, .$$

where m_i is calculated as follows:

$$m_i = \log\left(\frac{s_1(y_i)}{s_2(y_i)}\right) \, .$$

$s_1(y_i)$ and $s_2(y_i)$ are the estimated probabilities of the observed injury severity level using model 1 and model 2, respectively, n represents the number of observations or sample size. The Vuong test is interpreted as $v < -1.96$ favors the second model, $-1.96 < v < 1.96$ lends no support to either model, and $v > 1.96$ supports the first model. The standard ordered probit and the ZIOP model must have the same number of observations.

4.8 Resource and Tools

This project has been executed using R version 4.2.2. The standard probit model was implemented utilizing the Ordinal package using the probit link. The zero-inflated ordered probit (ZIOP) was implemented using the ZMIOP package. The researcher developed codes for the components of the ZIOP model, which includes the Vuong test, Marginal effects, AIC, and BIC, to answer the research question and steer the narrative of this study. The detailed description of the code can be accessed under the respective **GitHub** repository. To execute the analysis, ensure the custom codes for the Vuong test, Marginal effects, AIC, and BIC are initiated, and then run the ZIOP file corresponding to each lighting condition.

Chapter 5

Result

This section provides a detailed description and interpretation of the parameter estimates resulting from the application of the ZIOP model. To understand the factors that contribute to the injury severity of the single-vehicle collision and the effect of the various lighting conditions, three combinations of lighting conditions were analyzed. These combinations are daylight, darkness with street lights lit, and darkness (without street lights); by studying these combinations, the analysis aims to uncover patterns and connections that could provide insight into how various lighting conditions and injury severity are interconnected. The STATS19 dataset defines darkness as the timeframe between half an hour after sunset to half an hour before sunrise. Alternatively, daylight includes all time intervals outside this period while darkness with street lights lit covers all instances where street lights and lamps are illuminated.

The first step involves a backward selection technique by identifying variables significant at the 90% confidence level in all three combinations of the standard ordered probit model and the ZIOP model. The statistically significant variables are selected as ideal variables for the ZIOP model. The ZIOP model is fitted with these variables and the ones that are not significant in both processes of the first fitting of the ZIOP model are removed and the ZIOP model is fitted again. Finally, the variables in the final ZIOP model are significant in either or both processes of the ZIOP model.

In the analysis of each of the combinations of lighting effects for the ZIOP model, the variables are organized into indicator variables where "0" represents the reference attribute and the parameter estimates indicate changes in injury severity compared to the reference attribute. Additionally, due to the inflated cases of slight injuries, it was coded as zero on the severity scale in the collision dataset hence allowing the use of the zero-inflated approach in modeling collision injury severity, and the ZIOP model can effectively account for the potential presence of an underlying state. In order to compare the performance of the standard ordered probit model and the zero-inflated ordered probit model, the same variables were used in the final version of both models. The estimates of the ZIOP model on the effect of the various lighting conditions (Daylight, Darkness with street light lit, and Darkness without street light) are presented in the tables below.

The table below presents parameter estimates of the effect of daylight conditions on injury severity for single-vehicle collisions. However, numerous combinations of explanatory variables have been extensively explored as potential candidates for the presented model. Additionally, interacting terms between the sex of the driver and speed limit, weather and road surface conditions were explored, and they were not statistically significant. The presented variables are statistically significant at either the binary or the ordered probit or both sections of the model.

Table 5.1: Parameter estimates on the effect of Daylight on injury severity.

	Binary probit process			Ordered probit process		
	Coef.	Std. Err.	P	Coef.	Std. Err.	P
Road type (1 if collision occurred on a dual/single motorway, 0 otherwise)	0.189	0.058	**0.001**	0.418	0.239	**0.081**
Speed limit (1 if the speed limit is above 40 mph, 0 otherwise)	0.070	0.047	0.132	0.256	0.130	**0.049**
Weather conditions (1 if collision occurred in fine weather, 0 otherwise)	0.090	0.047	**0.058**	0.427	0.181	**0.019**
Carriageway hazards (1 if the collision involves carriageway hazards, 0 otherwise)	0.141	0.325	0.664	-2.191	1.240	**0.077**
Vehicle type (1 if collision involved buses and vans under 3.5 tonnes, 0 otherwise)	0.134	0.051	**0.009**	0.225	0.155	0.147
Skidding and Overturning (1 if the vehicle skidded, overturned, and jackknifed during the collision, 0 otherwise)	-0.286	0.059	< **0.001**	0.482	0.158	**0.002**
Point of impact (1 if the vehicle was impacted (front, back, and side), 0 otherwise)	0.001	0.101	0.991	0.723	0.430	**0.093**
Journey Purpose of driver (1 if collision is work-related, 0 otherwise)	0.032	0.036	0.371	-0.312	0.119	**0.009**
Sex of driver (1 if the driver is female, 0 otherwise)	-0.086	0.029	**0.003**	-0.058	0.100	0.564
Age of Casualty (1 if casualty age is younger than 30 years, 0 otherwise)	-0.224	0.038	< **0.001**	-0.716	0.122	< **0.001**
Car Passenger (1 if collision involved a passenger car (front or rear seat), 0 otherwise)	-0.544	0.101	< **0.001**	-0.207	0.442	0.640
/cut1				-1.491	1.277	0.243
/cut2				3.027	0.244	< 0.001
Number of observations						10,712

The numbers in bold are significant at the 10% level or better (including 5% and 1% levels).

cut1 and cut2 values are the threshold parameters.

In the binary component of the ZIOP model, $g_i = 1$ indicates a collision associated with a minor injury; however, this transcends into the interpretation of the estimates from the model. In the binary probit part of the model, a positive and statistically significant estimate indicates an increased likelihood of a collision resulting in a minor injury, while a negative estimate suggests a decreased likelihood of a collision resulting in a minor injury. Additionally, In the ordered probit part of the ZIOP model, a positive statistically significant estimate means that there is an increased likelihood of collisions leading to more severe injury in a progressive order, alternatively a negative statistically significant estimate results in a decreased tendency of a collision leading to a less severe injury in a progressive order.

Under daylight conditions, collisions involving instances where a vehicle skidded and overturned is associated with a decreased probability of resulting in minor injury; hence having a likelihood of belonging to an ordered injury state and it is highly statistically significant ($p < 0.001$); however, in the ordered probit process which indicates that conditional on being a non-minor injury, skidding and overturning has a positive coefficient 0.482 ($p = 0.002$) indicating that injuries associated with skidding and overturning are more likely to be severe. This is consistent with studies by Bener et al.(2009) which indicates that most road traffic incident occurs during sunny days, and drivers were more injured from overturning skid crashes and hitting fixed objects.

Collisions that occurred on the motorway during daylight conditions are associated with an increased likelihood of resulting in a minor injury having a positive coefficient of 0.189($p = 0.001$); however, its associated increase in the ordered probit state is marginally significant.

Additionally, the age of casualty depicts collisions involving individuals younger than 30 years have a decreased likelihood of leading to a minor injury; instead, it is likely to fall into an ordered injury category, and this is highly statistically significant with an estimate of -0.224 ($p < 0.001$), Remarkably, at a 1% confidence, the ordered probit section indicates that collision with individuals younger than 30 years results in a less severe injury as compared to older casualties. One reason for this might be that older individuals, especially people over the age of 65, are generally more frail, increasing their likelihood of having more severe injuries. This is also consistent with the finding from The National Highway Traffic Safety Administration(NHTSA), which states that fatal crash rates increase noticeably starting at age 70-74 and are highest among drivers 85 and older.

Furthermore, collisions involving female drivers are associated with a decreased probability of resulting in minor injuries with a coefficient of -0.224($p = 0.003$); instead, it is likely to result in a more severe injury; however, its associated decreased in probability in the ordered probit section is not significant. The choice of vehicle driven by females might contribute to the increased likelihood of being in a collision that results in severe injury. Studies have illustrated that female drivers are prone to injuries, as evidenced by their increased chances of experiencing moderate injuries, They are twice more likely to sustain injuries affecting arms and legs compared to male drivers; however, these differences between gender collision injuries arise from vehicles typically driven by women and the nature of accident rather than physiological differences (Matthew and Jessica, 2021).

Finally, mini buses and vans under 3.5 tonnes are associated with an increased likelihood of resulting in a minor injury; this can be attributed to size and weight differences as collisions involving larger vehicles often lead to less impact hence offering better protection for their occupants (Insurance Institute for highway safety).

Moving to determinants of collision injury-severities associated with the ordered injury state, collision at a speed above 40 mph increases the likelihood of resulting in a more severe injury having a positive estimate of $0.256(p = 0.049)$. Additionally, fine weather increases the odds of severe injury in the event of a collision; however, it's associated increase in minor injury is very minimal. this could be because drivers tend to adopt behaviors to compensate for the perceived benefit associated with fine weather, similarly in brighter lighting conditions, the average value of speed distribution increases (Bassani et al., 2016). Work-related collisions decrease the likelihood of a collision leading to more severe injuries, remarkably significant at 1% confidence level. However, this variable might constitute a significant source of unobserved variations, and its impact on collision injury severity would require a more in-depth study.

To compare and evaluate the performance of the zero-inflated ordered probit and the standard ordered probit, a set of performance criteria was employed to assess each model's performance. The results are presented in Table 5.2

	Standard Ordered Probit	**ZIOP**
Number of observation	10,712	10,712
Number of Parameters	14	25
AIC	12775.11	12728.80
BIC	12869.74	12910.78
Log-Likelihood	-6374.56	-6339.40
Vuong Test(OP/ZIOP)		-4.19

Table 5.2: Performance metrics comparison of OP and ZIOP model for Daylight

Based on the model evaluation, the zero-inflated ordered probit model is the preferred model when compared to the standard ordered probit model. A higher log-likelihood indicates a better fit of the model to the data; the ZIOP model has a higher (less negative) log-likelihood value when compared to the standard ordered probit, also a smaller AIC value suggests that the ZIOP model is the better model. The Vuong test provides a more rigorous statistical insight since it is specifically designed for comparing two non-nested models. However, the test score lends support to the more general ZIOP model. Based on the model performance, it appears that the model in which slight injuries are attributed to two sources performs better on these data.

5.1 Marginal Effects of Daylight on Injury Severity

To understand the impact of certain variables on the effect of daylight on collision injury severity, marginal effects are calculated. Marginal effects are simply derivatives evaluated at a particular point of the covariate space. There are two ways of calculating marginal effects; the first option is to evaluate the marginal at the mean of the variable, while the other process involves evaluating them separately and taking the average across all observations. In this study, the marginal effect is evaluated at the mean of the covariates. Marginal effects show how much the probability of an accident resulting in a specific injury severity outcome will be affected by a unit change in the value of an independent variable while holding other variables constant.

Table 5.3: Marginal effect of daylight on injury severity for single vehicle collision

	Slight Injury	**Severe Injury**	**Fatal Injury**
road_type(1 if collision occurred on a dual/single motorway, 0 otherwise)	-0.00246	-0.01178	0.01424
speed_limit (1 if the speed limit is above 40 mph, 0 otherwise)	-0.00151	-0.00722	0.00873
weather_conditions (1 if collision occurred in fine weather, 0 otherwise)	-0.00252	-0.01202	0.01454
carriageway_hazards (1 if the collision involves carriageway hazards, 0 otherwise)	0.01293	0.06176	-0.07468
vehicle_type (1 if collision involved buses and vans under 3.5 tonnes, 0 otherwise)	-0.00133	-0.00634	0.00766
skidding_and_overturning (1 if the vehicle skidded, overturned, and jack-knifed during the collision, 0 otherwise)	-0.00284	-0.01357	0.01641
point_of_impact (1 if the vehicle was impacted (front, back, and side), 0 otherwise)	-0.00426	-0.02036	0.02463
journey_purpose_of_driver (1 if the collision is work-related, 0 otherwise)	0.00184	0.00879	-0.01063
sex_of_driver (1 if the driver is female, 0 otherwise)	0.00034	0.00163	-0.00197
age_of_casualty (1 if casualty age is younger than 30 years, 0 otherwise)	0.00422	0.02018	-0.02441
car_passenger (1 if collision involved a passenger car (front or rear seat), 0 otherwise)	0.00122	0.00583	-0.00705

The marginal effect reveals some interesting findings on covariates and how they influence the ordered injury severity state under daylight conditions. To illustrate this, collisions on dual/single motorways are found to be associated with a 0.246%, 1.178% decrease in the probability of slight and serious injuries while increasing the risk of fatal injury by 1.424%. Additionally, speed limits above 40 mph are associated with a 0.151%, 0.722% decrease in the probability of resulting in a slight and severe injury while having an increased likelihood of resulting in a fatal injury.

Additionally, collisions that occurred in fine weather conditions, where the vehicle was impacted either in the front, rear, or side, and collisions where the vehicle skidded and overturned are associated with a decreased probability of resulting in a slight and severe injury while having an increased likelihood of resulting in a fatal injury under daylight conditions. Carriageway hazards such as involvement of previous accidents and dislodged vehicle load, among others, are associated with an increased probability of resulting in a slight and severe injury while having a decreased likelihood of leading to a fatal injury. A reason for this could be that driver might reduce their speed as a precautionary measure upon witnessing a carriageway hazard.

Furthermore, work-related trips are associated with 0.184%, 0.879% probability of resulting in a slight and severe injury while having a 1.063% decrease in the probability of resulting in a fatal injury. Female drivers and young drivers are also associated with an increased probability of resulting in a slight or severe injury and a reduced likelihood of resulting in a fatal injury. Finally, passenger cars involved in a collision are associated with 0.122%, 0.583% increased probability of resulting in a slight and severe injury, and decreased probability of 0.705% being a fatal collision.

5.2 Parameter estimates for the effect of darkness with street light lit

The following are parameter estimates for the effects of darkness with street light lit on collision injury severity; numerous combinations of variables were extensively explored. The presented variables are statistically significant at either the binary, the ordered probit, or both sections of the model.

Table 5.4: Parameter estimates on the effect of Darkness with street light lit on injury severity.

	Binary probit process			Ordered probit process		
	Coef.	Std. Err.	P	Coef.	Std. Err.	P
Road type (1 if collision occurred on a dual/single motorway, 0 otherwise)	0.434	0.260	**0.096**	0.099	0.115	0.391
Speed limit (1 if the speed limit is above 40 mph, 0 otherwise)	-0.551	0.192	**0.004**	0.333	0.112	**0.003**
Carriageway hazards (1 if the collision involves carriageway hazards, 0 otherwise)	-0.756	0.328	**0.021**	-0.011	0.189	0.954
Skidding and Overturning (1 if the vehicle skidded, overturned, and jackknifed during the collision, 0 otherwise)	0.309	0.281	0.272	-0.395	0.097	< **0.001**
Sex of driver (1 if the driver is female, 0 otherwise)	4.482	34.423	0.896	-0.395	0.085	< **0.001**
Sex of casualty (1 if casualty is female, 0 otherwise)	1.593	1.330	0.231	-0.347	0.072	< **0.001**
/cut1				0.262	0.152	0.085
/cut2				1.857	0.038	< 0.001
Number of observations						5,015

The numbers in bold are significant at the 10% level or better (including 5% and 1% levels).

cut1 and cut2 values are the threshold parameters.

At night time with the illumination of streetlights, collisions at speeds above 40mph are associated with a lower probability (-0.551) of resulting in a minor injury hence suggesting that higher speed might lead to more severe injury; also, in the ordered probit process, the positive coefficient of 0.333 ($p = 0.003$) suggests that collisions at a high-speed result in more severe injuries. Additionally, collisions on dual/single motorways are associated with an increased likelihood of resulting in a minor injury accident; however, this should be taken cautiously; Given its statistical significance is marginal ($p = 0.096$), however, it increase in the probability of severe injuries in the ordered probit process is not significant. Furthermore, the decrease associated with carriageway hazard on injury severity level prediction is not significant, but it is significantly associated with a reduced likelihood of resulting in a minor injury, hence suggesting collisions involving carriageway hazards in dark conditions with illuminated street lights are more likely to belong in the ordered state hence leading to more severe injuy. This finding could be connected to insufficient illumination or poor lighting conditions because carriageway hazards such as accident sites have been identified as impacting night-time collisions, leading to more severe injury (Li et al., 2019). Collisions where the vehicle skidded and overturned at the time of the accident are found to decrease the probability of a serious or fatal injury. Still, its increase in the prediction of minor injury is not statistically significant. Additionally, collisions involving females are associated with a decreased probability of resulting in severe or fatal injury; however, it's increase in the minor injury state is not significant. Similarly, female driver-related collisions are associated with a decreased probability of resulting in a severe injury; likewise, its increase in minor injury is not statistically significant.

The standard ordered probit model and the zero-inflated ordered probit model are compared using a range of evaluation metrics to compare model performance and determine the effectiveness of both models in addressing the inflation of slight injuries in relation to the effect of darkness with illuminating street lights on collision injury severity. The results are presented in Table 5.5

	Standard Ordered Probit	**ZIOP**
Number of observation	5,015	5,015
Number of Parameters	9	15
AIC	6512.52	6495.50
BIC	6564.68	6593.30
Log-Likelihood	-3248.26	-3232.75
Vuong Test (OP/ZIOP)		-2.72

Table 5.5: Performance metrics comparison of OP and ZIOP for Darkness with streetlights lit

The ZIOP model surpasses the standard ordered probit model in every evaluation criterion except for BIC. The ZIOP model has a higher log-likelihood indicating a better fit of the model to the data; a smaller AIC value suggests that the ZIOP model is the better model. The Vuong test score also suggests that the ZIOP model outperforms the standard ordered probit model. Considering the model's performance, it is clear that the ZIOP model outperforms the ordered probit model in comprehending the inflated data and capturing its complexities.

5.3 Marginal Effects of Darkness with Streetlight on Injury Severity

Table 5.6: Marginal effect of darkness with streetlight lit on injury severity for single vehicle collision

	Slight Injury	**Severe Injury**	**Fatal Injury**
road_type (1 if collision occurred on a dual/single motorway, 0 otherwise)	-0.02901	0.02375	0.00526
speed_limit (1 if the speed limit is above 40 mph, 0 otherwise)	-0.09742	0.07975	0.01766
carriageway_hazards (1 if the collision involves carriageway hazards, 0 otherwise)	0.00317	-0.00260	-0.00058
skidding_and_overturning (1 if the vehicle skidded, overturned, and jack-knifed during the collision, 0 otherwise)	0.11560	-0.09464	-0.02096
sex_of_driver (1 if the driver is female, 0 otherwise)	0.11564	-0.09467	-0.02097
sex_of_casualty (1 if casualty is female, 0 otherwise)	0.10169	-0.08325	-0.01844

The marginal effect describes the effects of darkness with functional streetlights on the impact of injury severity. Under darkness with illuminated streetlights, when a collision occurs on a dual/single motorway, it is associated with a 2.9% decrease in the probability of resulting in a slight injury and a 2.375% percent increase in the probability of resulting in a severe injury. The effect on fatal injuries is minimal, with a 0.526% increase in probability. High-speed collisions are associated with a 9.742% decrease in the probability of slight injuries, a 7.975% increase in the probability of severe injuries, and a 1.766% increase in the probability of fatal injuries.

Additionally, carriageway hazards lead to a 0.317% increase in the probability of slight injuries, a 0.260% decrease in the probability of severe injuries, and a minimal 0.058% decrease in the probability of fatal injuries. In a collision where the vehicle skidded and overturned, slight injuries have an increased probability of 11.560%, a 9.464% decrease in the probability of severe injuries, and a 2.096% decrease in the probability of fatal injuries. a collision involving a female occupant and female driver has an increased probability of resulting in a slight injury and a decreased probability of resulting in severe and fatal injury.

5.4 Parameter estimates for the effect of darkness on injury severity

Numerous studies have focused on the influence of environmental conditions and their impact on collision injury severity. Darkness often stands out as a significant factor influencing collision injuries' severity. This section explores parameter estimates of variables that quantify collision injury severity under dark lighting conditions. These variables are presented in Table 5.7

Table 5.7: Parameter estimates on the effect of darkness on collision injury severity

	Binary probit process			Ordered probit process		
	Coef.	Std. Err.	P	Coef.	Std. Err.	P
Weather conditions (1 if collision occurred in fine weather, 0 otherwise)	-1.764	1.082	0.103	0.832	0.239	**0.001**
Urban or rural area (1 if the collision occurred in a rural area, 0 otherwise)	-0.347	0.212	0.102	0.327	0.195	**0.094**
Skidding and overturning (1 if the vehicle skidded, overturned, and jackknifed during the collision, 0 otherwise)	0.471	0.257	**0.068**	-0.701	0.210	**0.001**
Age of driver (1 if the driver was younger than 23 years, 0 otherwise)	-0.779	0.233	**0.001**	0.389	0.223	**0.081**
Road type (1 if collision occurred on a dual/single motorway, 0 otherwise)	-1.225	1.191	0.304	0.960	0.374	**0.010**
Vehicle type (1 if collision involved mini buses and vans under 3.5 tonnes, 0 otherwise)	0.101	0.383	0.792	0.513	0.265	**0.053**
Journey purpose of driver(1 if the collision is work-related, 0 otherwise)	0.667	0.444	0.134	-0.516	0.187	**0.006**
Sex of driver (1 if the driver is female, 0 otherwise)	0.336	0.295	0.255	-0.496	0.163	**0.002**
Age of vehicle (1 if the age of the vehicle is above 10 years, 0 otherwise)	-0.257	0.218	0.240	0.249	0.145	**0.087**
Car passenger (1 if collision involved a passenger car (front or rear seat), 0 otherwise)	-0.114	0.443	0.798	-0.916	0.303	**0.003**
/cut1				1.479	0.365	<0.001
/cut2				3.071	0.131	< 0.001
Number of observations						1,094

The numbers in bold are significant at the 10% level or better (including 5% and 1% levels).

cut1 and cut2 values are the threshold parameters.

Under darkness or dark lighting conditions, the variables influencing collision injury severity vary. To illustrate this, collisions involving drivers younger than 23 years are associated with a decreased probability of resulting in a minor injury, indicating that such injuries belong in the ordered state, and this is highly statistically significant with a coefficient of -0.779 (p =0.001), also in the ordered state, the increased prediction in injury severity indicates that such injury has the likelihood of being severe. This is also consistent with studies by Rice et al.(2003), stating that the injury crash rate for young drivers aged 16 or 17 increases during nighttime hours and in the absence of adult supervision, with or without other passengers, also driving between 10 pm and midnight is particularly dangerous for young drivers. furthermore, these results align with the findings and perspectives of the Royal Society for the Prevention of Accidents (RoSPA), stating that young drivers are more likely to be involved in a collision that results in fatal or serious injury at night. Collisions where vehicles skidded and overturned are associated with an increased probability of being a minor injury, evidenced by its positive statistically significant coefficient of 0.471 (p=0.068), also in the ordered probit state, its negative highly significant coefficient of -0.701 (p=0.001) is associated with a decreased likelihood of resulting in a severe or fatal injury. One possible reason for skidding and overturning being associated with minor injuries during nighttime in complete darkness could be due to reduced visibility. To elaborate further, studies have shown that the mean speed of drivers is reduced at night time due to poor visibility conditions (Zolali et al., 2021). Since drivers are generally aware that the visibility at night is reduced, this expectation might make them more cautious, alert, and prepared for potential hazards; hence this heightened caution might influence how they react, thus leading to minor injury outcomes.

Moving to the ordered probit state, the decreased probability prediction of minor injury for fine weather conditions is not statistically significant, but it is significantly associated with an increase in the probability of resulting in severe injury outcomes in the ordered probit state. Similarly, the decreased probability of older vehicles resulting in minor injuries is not statistically significant, but in the ordered probit state, collisions involving vehicles older than 10 years are associated with an increased injury severity with a borderline significant value (p = 0.087). Additionally, older vehicles have an increased collision injury risk; for every year older a vehicle is, there's a slight increase in the risk of it being involved in a severe car crash. This increase in risk is about 5% for each year of an increase in the vehicle's age (Blows et al., 2003). Furthermore, collisions involving mini busses and vans under 3.5 tonnes are associated with an increased probability of resulting in a severe injury in the ordered probit state, but its increase in minor injuries is not statistically significant; Similarly, female drivers involved in a collision during nighttime in complete darkness have a reduced likelihood of having severe or fatal injuries, however it's increase in the minor injury state is not significant.

The decrease in minor injuries associated with collisions in rural areas is not statistically significant; however, the positive coefficient 0.327(p =0.094) in the ordered probit state is marginally significant hence indicating collision which occurs in rural areas during times of darkness or nighttime are associated with severe injuries. According to Wu et al. (2021), some of the highest numbers of fatal collisions in the rural-urban fringe occurred between 19:00 pm - 19:59 pm; the time frame between 4:00 am and 5:00 am have the highest fatality rate with 0.78 death per crash; interestingly, more than half of the fatal collision occurred between 17:00 pm and 05:00 am. These timeframes align with the STATS19 definition of darkness; finally, Collisions with passenger cars reduce severe injury likelihood in the ordered section but don't significantly affect minor injuries.

The standard ordered probit model and the zero-inflated ordered probit model are compared using a range of evaluation metrics to compare model performance and determine the effectiveness of both models in addressing the inflation of slight injuries in relation to the effect of darkness on collision injury severity. The results are presented in Table 5.8

	Standard Ordered Probit	**ZIOP**
Number of observation	1,094	1,094
Number of Parameters	13	23
AIC	1383.99	1371.97
BIC	1443.96	1486.92
Log-Likelihood	-679.99	-662.99
Vuong Test (OP/ZIOP)		-4.06

Table 5.8: Performance metrics comparison of OP and ZIOP for Complete Darkness

The summarized results from comparing the ZIOP and OP model indicates that the ZIOP model has a superior fit to the data. The ZIOP surpasses the OP model in every evaluation metric except for BIC. The Vuong test results also favor the ZIOP model over the standard ordered probit model. Given these results, it's evident that the ZIOP model is more adept at understanding the data's nuances and addressing its inflated nature. This reinforces the notion that slight injuries from nighttime collisions may be generated from two distinct sources.

5.5 Marginal the effect of darkness on injury severity

The Table provides a detailed overview of the marginal effect of darkness on injury severity resulting from single-vehicle collisions.

Table 5.9: Marginal effect of complete darkness on injury severity for single vehicle collision

	Slight Injury	Serious Injury	Fatal Injury
Weather conditions (1 if collision occurred in fine weather, 0 otherwise)	-0.12556185	0.12556151	0.000000342
Urban or rural area (1 if the collision occurred in a rural area, 0 otherwise)	-0.04930277	0.04930264	0.000000134
Skidding and overturning (1 if the vehicle skidded, overturned, and jackknifed during the collision, 0 otherwise)	0.10579055	-0.10579026	-0.000000288
Age of driver (1 if the driver was younger than 23 years, 0 otherwise)	-0.05874351	0.05874335	0.000000160
Road type (1 if collision occurred on a carriageway, 0 otherwise)	-0.14497394	0.14497355	0.000000395
Vehicle type (1 if collision involved buses and vans under 3.5 tonnes, 0 otherwise)	-0.07745900	0.07745878	0.000000211
Journey purpose of driver (1 if the collision is work-related, 0 otherwise)	0.07792144	-0.07792123	-0.000000212
Sex of driver (1 if the driver is female, 0 otherwise)	0.07480967	-0.07480946	-0.000000204
Age of vehicle (1 if the age of the vehicle is above 10 years, 0 otherwise)	-0.03765247	0.03765236	0.000000103
Car passenger (1 if collision involved a passenger car (front or rear seat), 0 otherwise)	0.13830845	-0.13830807	-0.000000377

The marginal effects in Table 5.9 highlight the influence of the covariates and how they might affect the outcome of a collision in darkness. To elaborate, fine weather conditions are associated with a decrease in the probability of slight injuries while increasing the probability of serious and fatal injuries. Additionally, collisions in rural areas decrease the likelihood of slight injuries and increase the likelihood of serious and fatal injuries. Collisions where vehicles skidded, overturned, and jackknifed, are associated with decreased probability of leading to slight injuries; alternatively, serious and fatal injuries have an increased probability. Drivers younger than 23 years involved in a collision have an increased probability of serious or fatal injuries compared to slight injuries associated with a decreased probability.

Collisions on a carriageway and collisions involving buses and vans are associated with a decrease in the probability of leading to a slight injury and an increase in the probability of resulting in a severe or serious injury. Additionally, collisions involving female drivers, passenger cars, and work-related collisions are associated with an increased probability of resulting in slight injuries while having a reduced probability of resulting in a severe or fatal injury. collision with a vehicle older than 10 years increases the probability of being severe or fatal.

5.6 Discussion

The need to improve road safety and fostering a safer environment has driven research into understanding collision severity and the mechanism influencing collision. over the years, studies have considered the effect originating from the no-injury mechanism; but there has been little consideration in exploring the possibility of the latent injury severity state underpinning the mechanism of injury and also the impact of environment variables such as lighting conditions on injury severity. This study aims to identify the various lighting conditions and how they influence injury severity outcomes for single-vehicle collisions. It focuses on three key scenarios: Daylight, Darkness with street lights lit, and complete Darkness. On the basis of collision data collected from the STATS19 crash reports, a ZIOP model has been developed to estimate the effects of the various lighting conditions.

The results suggest that the factors influencing injury mechanisms are statistically unique across the various lighting conditions. To elaborate, during daylight conditions, collisions where a vehicle skidded, overturned, or jackknifed are associated with a decreased probability of minor injury, and in the ordered probit process, its positive coefficient signifies an increase in the probability of severe or fatal injuries; however, this is different during darkness where the streetlight is lit and in total darkness where the results suggest that such collision are associated with an increased probability of minor injury and in the ordered state, its prediction in the decreased injury severity is statistically significant, hence indicating skidding, overturning and jackknife collisions in conditions of darkness with streets light and total darkness are likely to result in minor injuries compared to daylight.

Collisions involving speeds above 40 mph during conditions of daylight are associated with an increased probability of causing severe or fatal injury, though its impact on minor injuries is not significant. similarly, in conditions where street lights are lit during darkness, such speed is associated with the decreased probability of leading to minor injuries and an increase in the ordered injury severity leading to severe or fatal injuries. However, this does not influence the injury severity in conditions of complete darkness. This is consistent with studies revealing that vehicle speeds are impacted by different variables one of which is lighting conditions; Findings from studies suggest that vehicle speed decrease was higher on roads without lights compared to roads with lights (Jägerbrand and Sjöbergh, 2016).

Gender is a prominent factor influencing collision severity under varying conditions. To elaborate, under daylight conditions, female drivers involved in a collision are associated with a decreased probability of resulting in minor injury, hence suggesting that such injuries are likely to be severe; however, in the ordered probit state, its decrease is not statistically significant. Interestingly, under different lighting conditions, such as darkness with street lights lit and complete darkness, the same collision is associated with a decrease in the probability of resulting in severe or fatal injuries. Additionally, a collision involving a female occupant during dark conditions with street lights is likely to result in less severe injury. Remarkably, this variable does not impact collision severity for Daylight conditions and complete Darkness.

Age is a prominent factor influencing injury severity in both daylight and total darkness. Individuals younger than 30 years involved in daylight collisions tend to be associated with a decreased probability of sustaining minor injuries. Yet, there is a significant decrease in the injury severity levels compared to individuals older than 30 years in the ordered probit process. In total darkness, younger drivers involved in collisions are associated with a decreased probability of minor injuries; however, the ordered probit process indicates a high probability of sustaining severe or fatal injuries.

Weather conditions also show notable variations in their influence. In daylight, collisions occurring in fine weather, while marginally significant in the binary probit process, show a strong positive association with injury severity in the ordered probit process. However, during complete darkness, fine weather collisions show a non-significant trend towards decreased injury severity in the binary probit process but are significantly associated with increased severity in the ordered process. This indicates collision occurring in fine weather conditions is associated with the probability of severe or fatal injury.

Furthermore, injury severity outcome varies under different lighting conditions for collisions that occur on a motorway; for instance, during complete darkness, collisions on motorways are more likely to result in more severe or fatal injuries. This is not the case in daylight and darkness with street lights lit, where similar collisions are associated with an increased likelihood of minor injuries. An explanation for this might be the benefits associated with well-lit and daylight conditions, such as improved visibility, which allows drivers to assess road conditions and potential hazards better, thereby reducing the severity of collisions.

The complexities of how collisions occur and the severity of resulting injuries are greatly affected by the varying lighting conditions present. However, some variables that impact collision severity are specific to a particular lighting condition. To elaborate, in darkness with streetlights lit, the presence of a carriageway hazard indicates a decreased probability of resulting in a minor injury, signifying that collision involving hazards on the carriageway belongs to the ordered injury state. However, it did not bring about a significant change in injury severity levels.

Similarly, the effect of location on injury severity is prominent only in complete darkness conditions. For example, collisions in rural areas are more likely to result in severe or fatal injuries compared to urban areas. However, this location-based effect does not seem to influence injury outcomes in daylight and well-lit darkness conditions. A possible reason for this could be that rural areas often lack adequate street lighting and emergency medical services, making it more challenging to prevent or respond to severe injuries in complete darkness. In contrast, the presence of better lighting and quicker emergency response in urban areas and during daylight could mitigate the severity of injuries, rendering the location effect less significant.

On the other hand, in conditions of complete darkness, the age of the vehicle is a statistically significant factor influencing injury severity. Collisions involving vehicles over 10 years old show a non-significant trend towards a decrease in minor injury in the binary probit process, alternatively having a significant increase in the probability of resulting in severe or fatal injury in the ordered probit process. This is also consistent with findings suggesting that drivers of older vehicles are more likely to be fatally injured than those of a driver less than 3 years old (National Highway Traffic Safety Administration, 2013). One possible reason for this might be that older vehicles lack modern safety features, which are found in newer vehicles, making them more vulnerable in situations where visibility is compromised. Furthermore, the structural integrity of these older vehicles may be compromised over time from experiencing wear and tear, making them less effective in safeguarding passengers in the event of a collision.

5.7 Conclusion

The analysis focuses on understanding the impact of lighting conditions on collision injury severity under different conditions, thus distinguishing how various elements impact lighting conditions. The findings have shown that a vast array of elements, such as accident details, vehicle details, and driver characteristics, influence collision severity differently, and the presence of these distinct variables under different lighting conditions highlights the significance of a multi-faceted approach to understanding collision severity and road safety measures. Additionally, the model evaluation supports the ZIOP model in uncovering the complexities underpinning collision severity compared to the standard ordered probit model.

Further Study

The empirical analysis and modeling results may be subject to possible data-related biases. The bias may arise from the limitation of the collision dataset, which stems from the omission of no-injury collision. it is worth noting that the impact of injury severity factors can vary across lighting conditions. The findings from this study can be particularly useful for the automobile industry to design better lighting, and improve safety features to reduce the potential risk of collision severity; also, urban planners can utilize the findings to design better street lighting systems, ensuring safer roads and reducing the risk of severe collision in various lighting conditions. Moving forward, future research could focus towards on delving into the relationship, between different lighting technologies and how they affect driver behavior. This kind of investigation would help uncover factors and subtle nuances that might not be apparent in a broader analysis. By taking this approach, the study can fully explore the intricacies that may not have been fully captured within the limitations of the study methodology.

Bibliography

[1] 16th Annual road safety performance index (PIN) report (2022). *ETSC*. Available at: https://etsc.eu/16th-annual-road-safety-performance-index-pin-report/.

[2] Abbasi, M. et al. (2022). "Analysis of Crash Severity of Texas Two-Lane Rural Roads Using Solar Altitude Angle Based Lighting Condition." *MDPI*. Available at: https://www.mdpi.com/2071-1050/14/3/1692.

[3] Abdel-Aty, M. (2003) Analysis of driver injury severity levels at multiple locations using ordered Probit models, Journal of Safety Research. Available at: https://www.sciencedirect.com/science/article/pii/S0022437503000811.

[4] Aci, C. and Özden, C. (2018) Predicting the severity of motor vehicle accident injuries in Adana, Turkey Using Machine Learning Methods and Detailed Meteorological Data, Research Gate. Available at: https://www.researchgate.net/publication/324101053_Predicting_the_Severity_of_Motor_Vehicle_Accident_Injuries_in_Adana-Turkey_Using_Machine_Learning_Methods_and_Detailed_Meteorological_Data.

[5] Adanu, E.K. et al. (2021) Full article: A comprehensive analysis of factors that influence. *Journal of Transportation Safety & Security*. Available at: https://www.tandfonline.com/doi/full/10.1080/19439962.2021.1949414.

[6] Adebisi, A. et al. (2019) Age-related differences in motor-vehicle crash severity in California, MDPI. Available at: https://doi.org/10.3390/safety5030048.

[7] Akaike, H. (1998) Information theory and an extension of the maximum likelihood principle. *SpringerLink*. Available at: https://link.springer.com/chapter/10.1007/978-1-4612-1694-0_15.

[8] Arhin, S.A. and Gatiba, A. (2019) Predicting injury severity of angle crashes involving two vehicles at unsignalized intersections using artificial neural networks, Engineering, Technology & Applied Science Research. Available at: https://etasr.com/index.php/ETASR/article/view/2551.

[9] Assi, K. et al. (2020) Predicting crash injury severity with machine learning algorithm synergized with Clustering Technique: A Promising Protocol, MDPI. Available at: https://www.mdpi.com/1660-4601/17/15/5497.

[10] Azhar, A. et al. (2022) Classification of driver injury severity for accidents involving heavy vehicles with decision tree and Random Forest, MDPI. Available at: https://www.mdpi.com/2071-1050/14/7/4101.

[11] Bassani, M. et al. (2016) Night-time and daytime operating speed distribution in urban arterials, Transportation Research Part F: Traffic Psychology and Behaviour. Available at: https://www.sciencedirect.com/science/article/pii/S1369847816301206#s0060.

[12] Bener, A. et al. (2009) Road traffic injuries and risk factors, Californian Journal of Health Promotion. Available at: https://www.academia.edu/26289152/Road_traffic_injuries_and_risk_factors.

[13] Blows, S. et al. (2003) Vehicle year and the risk of Car Crash Injury, Injury prevention: journal of the International Society for Child and Adolescent Injury Prevention. Available at: https://pubmed.ncbi.nlm.nih.gov/14693899/.

[14] Buckland, S.T., Burnham, K.P. and Augustine, N.H. (1997) Model Selection: An integral part of inference, JSTOR. Available at: https://sci-hub.se/10.2307/2533961.

[15] Chan, T.-C. et al. (2022). Association of air pollution and weather factors with traffic injury severity: A study in Taiwan. *International journal of environmental research and public health*. Available at: https://pubmed.ncbi.nlm.nih.gov/35742691/.

[16] Charbotel, B., Martin, J.L., and Chiron, M. (2010) Work-related versus non-work-related road accidents, developments in the last decade in France, Accident; analysis and prevention. Available at: https://pubmed.ncbi.nlm.nih.gov/20159085/.

[17] Departments for Transport - Road Safety Data. Available at: https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data.

[18] Donnell, C.J.O. and Connor, D.H. (1999) Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice, Accident Analysis & Prevention. Available at: https://www.sciencedirect.com/science/article/pii/S0001457596000504.

[19] Edwards, J. (1998) The relationship between road accident severity and recorded weather, *Journal of Safety Research*. Available at: https://www.sciencedirect.com/science/article/pii

[20] Fountas, G., Tom, R. (2019) A note on accounting for underlying injury-severity states in statistical modeling of Injury Accident Data, Procedia Computer Science. Available at: https://www.sciencedirect.com/science/article/pii/S1877050919304922.

[21] Garrido, R. et al. (2014) Prediction of road accident severity using the ordered probit model, Transportation Research Procedia. Available at: https://www.sciencedirect.com/science/article/pii/S2352146514002701.

[22] Grigorios Founta, G. et al. (2020) The joint effect of weather and lighting conditions on injury severities of single-vehicle accidents, Analytic Methods in Accident Research. Available at: https://www.sciencedirect.com/science/article/pii/S2213665720300142.

[23] Harris, M. and Zhao, X. (2007) A zero-inflated ordered probit model, with an application to modelling tobacco consumption, Journal of Econometrics. Available at: https://www.sciencedirect.com/science/article/pii/S0304407607000048.

[24] Hausman, J. and McFadden, D. (1984) Specification tests for the multinomial logit model - JSTOR, Econometrica. Available at: https://www.jstor.org/stable/1910997.

[25] Highway Safety Improvement Program (HSIP) Highway Safety Improvement Program (HSIP) — FHWA. Available at: https://highways.dot.gov/safety/hsip.

[26] Hopkins, M., Chivers, S. and Steveson-Freer, G.(2017) Hit-and-run: Why do drivers fail to stop after an accident? - MIB, University of Leicester. Available at: https://www.mib.org.uk/media/350114/hit-and-run-why-do-drivers-fail-to-stop-after-an-accident.pdf.

[27] Imprialou, M. and Quddus, M. (2019) Crash Data Quality for road safety research: Current State and future directions, *Accident Analysis & Prevention*. Available at: https://www.sciencedirect.com/science/article/pii/S0001457517300842.

[28] IIHS (n.d.). *Vehicle Size and Weight*. Available at: https://www.iihs.org/topics/vehicle-size-and-weight.

[29] Jafari Anarkooli, A. and Hadji Hosseinlou, M. (2015) Analysis of the injury severity of crashes by considering different lighting conditions on two-lane rural roads. *Journal of Safety Research*. Available at: https://www.sciencedirect.com/science/article/pii/S0022437515001073.

[30] Jägerbrand, A.K. and Sjöbergh, J. (2016) Effects of weather conditions, light conditions, and road lighting on vehicle speed - springerplus, SpringerOpen. Available at: https://springerplus.springeropen.com/articles/10.1186/s40064-016-2124-6.

[31] Jiang, X., et al. (2013) Investigating the influence of curbs on single-vehicle crash injury severity utilizing zero-inflated ordered probit models. *Accident Analysis & Prevention*. Available at: https://www.sciencedirect.com/science/article/pii/S0001457513001164#bib0005.

[32] Keane, M. (1992) A Note on Identification in the Multinomial Probit Model, A note on identification in the multinomial probit model on JSTOR. Available at: https://doi.org/10.2307/1391677.

[33] Kisslinger, C. Advances in Geophysics, Advances in Geophysics - an overview — ScienceDirect Topics. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0065268708600199.

[34] Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in Manufacturing. *JSTOR*. Available at: https://www.jstor.org/stable/1269547.

[35] Lee, A.H. et al. (2002) Modeling young driver motor vehicle crashes: Data with extra zeros, Accident Analysis & Prevention. Available at: https://www.sciencedirect.com/science/article/pii/S0001457501000495.

[36] Li, J. et al. (2019) Exploring factors affecting the severity of night-time vehicle accidents. Available at: https://journals.sagepub.com/doi/full/10.1177/1687814019840940.

[37] Li, J. et al. (2023). Predicting the severity of traffic accidents on mountain freeways with dynamic traffic and weather data. *Academic.oup.com*. Available at: https://academic.oup.com/tse/advance-article/doi/10.1093/tse/tdad001/6998550.

[38] Liu, S., Li, Y., and Fan, W. (David) (2021) Mixed logit model based diagnostic analysis of bicycle-vehicle crashes at daytime and nighttime, International Journal of Transportation Science and Technology. Available at: https://www.sciencedirect.com/science/article/pii/S2046043021000769.

[39] Lord, D., Washington, S. and Ivan, J.N. (2007) Further notes on the application of zero-inflated models in Highway Safety, Accident Analysis & Prevention. Available at: https://www.sciencedirect.com/science/article/pii/S0001457506001072.

[40] Lord, D., Washington, S.P. and Ivan, J.N. (2005) Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory, Accident Analysis & Prevention. Available at: https://www.sciencedirect.com/science/article/pii/S0001457504000521.

[41] Mahikul, W. et al. (2022) Factors affecting bus accident severity in Thailand: A multinomial logit model, PLOS ONE. Available at: https://doi.org/10.1371/journal.pone.0277318.

[42] Mahikul, W. et al. (2022) Factors affecting bus accident severity in Thailand: A multinomial logit model, PLOS ONE. Available at: https://journals.plos.org/plosone/articleid=10.1371%2Fjournal.pone.0277318.

[43] Mannering, F., Shankar, V. and Bhat, C. (2016a) UNOBSERVED heterogeneity and the statistical analysis of Highway Accident Data, Analytic Methods in Accident Research. Available at: https://www.sciencedirect.com/science/article/pii/S2213665716300100.

[44] Matthew, B. and Jessica, J. (2021) Injury risks and crashworthiness benefits for females and males: Which differences are physiological?, Traffic injury prevention. Available at: https://pubmed.ncbi.nlm.nih.gov/34874809/.

[45] McLaughlin, S.B. et al. (2009) Contributing factors to run-off-road crashes and near crashes, National Highway Traffic Safety Administration. Available at: https://tinyurl.com/National-highway.

[46] Moomen, M., Molan, A.M. and Ksaibati, K. (2023) A random parameters multinomial logit model analysis of median barrier crash injury severity on Wyoming Interstates, MDPI. Available at: https://www.mdpi.com/2071-1050/15/14/10856.

[47] Mphekgwana, P.M. (2022) Influence of environmental factors on injury severity using ordered logit regression model in Limpopo Province, South Africa, Journal of Environmental and Public Health. Available at: https://doi.org/10.1155/2022/5040435.

[48] National Highway Traffic Safety Administration. (2013) How vehicle age and model year relate to driver injury severity in fatal crashes DOT HS 811 825, Traffic Safety Facts - Research Note. Available at: https://trid.trb.org/view/1262716.

[49] O'Donnell, C.J. and Connor, D.H. (1999) Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. *Accident Analysis & Prevention.* Available at: https://www.sciencedirect.com/science/article/pii/S0001457596000504

[50] Patil, S., Geedipally, S.R. and Lord, D. (2012) Analysis of crash severities using nested Logit Model–accounting for the underreporting of crashes, Accident; analysis and prevention. Available at: https://pubmed.ncbi.nlm.nih.gov/22269553/.

[51] Pew, T. et al. (2020) Justification for considering zero-inflated models in crash frequency analysis, Transportation Research Interdisciplinary Perspectives. Available at: https://www.sciencedirect.com/science/article/pii/S2590198220301603#b0155.

[52] Quddus, M., Wang, C., and Ison, S. (2009) Road traffic congestion and crash severity: Econometric ...- asce library, ASCE library. Available at: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29TE.1943-5436.0000044.

[53] Reported road casualties in Great Britain, Provisional Estimates: Year ending June 2022 GOV.UK. Available at: https://www.gov.uk/government/statistics/reported-road-casualties-in-great-britain-provisional-estimates-year-ending-june-2022/reported-road-casualties-in-great-britain-provisional-estimates-year-ending-june-2022.

[54] Reported road casualties in Great Britain, Provisional Estimates: Year ending June 2021. *GOV.UK.* Available at: https://www.gov.uk/government/statistics/reported-road-casualties-in-great-britain-provisional-estimates-year-ending-june-2021/reported-road-casualties-in-great-britain-provisional-estimates-year-ending-june-2021.

[55] Rezapour, M., Moomen, M., and Ksaibati, K. (2018) Ordered logistic models of influencing factors on crash injury severity of single and multiple-vehicle downgrade crashes: A case study in Wyoming, Journal of Safety Research. Available at: https://www.sciencedirect.com/science/article/pii/S0022437518300707.

[56] Rice, T.M., Peek-Asa, C. and Kraus, J.F. (2003) Nighttime driving, passenger transport, and injury crash rates of young drivers, Injury prevention: journal of the International Society for Child and Adolescent Injury Prevention. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1730980/.

[57] RoSPA When and where are young drivers at risk?, RoSPA. Available at: https://www.rospa.com/resources/hubs/young-drivers/after-the-test/young-drivers-at-risk.

[58] Sapri, F.E. et al. (2017) Decision tree model for non-fatal road accident injury - researchgate, Research Gate. Available at: https://www.researchgate.net/publication/314085824_Decision_Tree_Model_for_Non-Fatal_Road_Accident_Injury.

[59] Savolainen, P.T. et al. (2011) The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives, Accident Analysis & Prevention. Available at: https://www.sciencedirect.com/science/article/pii/S0001457511000765.

[60] Schwarz, G. Estimating the dimension of a model. *Project Euclid*. Available at: https://projecteuclid.org/journals/annals-of-statistics/volume-6/issue-2/Estimating-the-Dimension-of-a-Model/10.1214/aos/1176344136.full?tab=ArticleFirstPage.

[61] Siamidoudaran, M. and İşçioğlu, E. (2019) Injury severity prediction of traffic collision by applying a series of Neural Networks: The City of London Case Study, Research Gate. Available at: https://www.researchgate.net/publication/338249030_Injury_Severity_Prediction_of_Traffic_Collision_by_Applying_a_Series_of_Neural_Networks_The_City_of_London_Case_Study.

[62] Silver, D. et al. (2022) Estimating automobile crash characteristics from images using Deep Learning, The International FLAIRS Conference Proceedings. Available at: https://journals.flvc.org/FLAIRS/article/view/130629.

[63] Tsui, K. et al. (2008) Misclassification of injury severity among road casualties in police reports, Accident Analysis & Prevention. Available at: https://www.sciencedirect.com/science/article/pii/S0001457508001863.

[64] United Nation (2020) Decade of Action 2020 - Road safety. Available at: https://www.un.org/en/safety-and-security/road-safety.

[65] Vuong, Q. Likelihood ratio tests for model selection and non-nested. *JSTOR*. Available at: https://www.jstor.org/stable/1912557?read-now=1.

[66] Wang, H., et al. (2022) Analysis of the injury-severity outcomes of maritime accidents using a zero-inflated ordered probit model. *Ocean Engineering*. Available at: https://www.sciencedirect.com/science/article/pii/S0029801822011428#sec4.

[67] Washington, S.P., Karlaftis, M.G. and Mannering, F. (2003) Statistical and Econometric Methods for Transportation Data Analysis, Taylor & Francis. Available at: https://www.taylorfrancis.com/books/mono/10.1201/9780203497111/statistical-econometric-methods-transportation-data-analysis-simon-washington-matthew-karlaftis-fred-manner

[68] Williams, R. (2009) Using heterogeneous choice models to compare logit and probit, Sage Journals Home. Available at: https://journals.sagepub.com/doi/10.1177/0049124109335735.

[69] Wu, B. et al. (2021) Exploring factors contributing to crash injury severity in the rural-urban fringe of the Central City, Journal of Advanced Transportation. Available at: https://www.hindawi.com/journals/jat/2021/8453465/.

[70] Yamamoto, T., Shankar, V., and Hashiji, J. (2008) Underreporting in traffic accident data, bias in parameters and the structure of injury severity models, Accident Analysis & Prevention. Available at: https://www.sciencedirect.com/science/article/pii/S0001457508000237.

[71] Yannis, G. and Kondyli, A. (2013) Effect of road lighting conditions on the frequency and severity of road accidents. *Research Gate*. Available at: https://www.researchgate.net/publication/260187176_Effect_of_Road_Lighting_Conditions_on_the_Frequency_and_Severity_of_Road_Accidents.

[72] Ye, F. and Lord, D. (2011) Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models, sage journals. Available at: https://journals.sagepub.com/doi/10.3141/2241-06.

[73] Zolali, M. et al. (2021) A behavioral model of drivers' mean speed influenced by weather conditions, road geometry, and driver characteristics using a driving simulator study, Advances in Civil Engineering. Available at: https://www.hindawi.com/journals/ace/2021/5542905/.