
Taylor & Francis
Taylor & Francis Group

Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ttra21

# Investigating older driver crashes on high-speed roadway segments: a hybrid approach with extreme gradient boosting and random parameter model

Ahmed Hossain, Xiaoduan Sun, Subasish Das, Monire Jafari & Julius Codjoe

**To cite this article:** Ahmed Hossain, Xiaoduan Sun, Subasish Das, Monire Jafari & Julius Codjoe (06 Jun 2024): Investigating older driver crashes on high-speed roadway segments: a hybrid approach with extreme gradient boosting and random parameter model, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2024.2362362

**To link to this article:** https://doi.org/10.1080/23249935.2024.2362362

Published online: 06 Jun 2024.

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2024.2362362

Taylor & Francis
Taylor & Francis Group



# Investigating older driver crashes on high-speed roadway segments: a hybrid approach with extreme gradient boosting and random parameter model

Ahmed Hossain [a], Xiaoduan Sun [a], Subasish Das [b], Monire Jafari [c] and Julius Codjoe [d]

[a]Department of Civil Engineering, University of Louisiana at Lafayette, Lafayette, LA, USA; [b]College of Science of Engineering, Texas State University, San Marcos, TX, USA; [c]Mathematics, Texas State University, San Marcos, TX, USA; [d]Special Studies Research Administrator, Louisiana Transportation Research Center, Baton Rouge, LA, USA

**ABSTRACT**
Older drivers are often more susceptible to crashes due to age-related physical and cognitive limitations, particularly in complex driving environments. Considering the limited research in this area, this study focuses on investigating crashes involving older drivers on high-speed roadways ( $\geq$ 45 mph). The analysis is based on data collected from Louisiana State, encompassing 18,300 older driver-involved crashes (2017-2021). For analysis, a two-step hybrid modelling approach is employed: a) Extreme Gradient Boosting (XGBoost) is used to classify top variable features and b) Correlated Random Parameter Ordered Probit with Heterogeneity in Means (CRPOP-HM) is used to predict the likelihood of crash injury severity. . Some of the critical factors increasing the likelihood of fatal-severe or injury crashes involving older drivers on high-speed segments include the manner of collision (rear-end, right-angle, single-vehicle), primary contributing factor (violation, pedestrian action), presence of passenger (s), location type (open country, residential, business with mixed residential), and weekend.

**ARTICLE HISTORY**
Received 7 January 2024
Accepted 27 May 2024

**KEYWORDS**
Older driver; unobserved heterogeneity; Correlated Random Parameter Ordered Probit with Heterogeneity in Means (CRPOP-HM); open country; weekend

## 1. Introduction

One of the most notable demographic trends in the US is the growth in the older population 65 years and above. According to the US Census Bureau, the older population is projected to nearly double from 46 million in 2016–95 million by 2060, with this specific age group accounting for 23% of the total population (Bureau n.d.). Along with the growth of the older population, the number of licensed older drivers is also on the rise. In 2021, there were approximately 49.6 million licensed older drivers in the US, which is a 38% increase since 2012 (NHTSA 2021). With the growing older population due to the aging baby boom, a significant safety concern is that a rise in fatal crashes involving older drivers would follow. Older drivers represent around 22% (6,134 out of 27,555) of all driver fatalities in the US

**CONTACT** Ahmed Hossain ahmed.hossain@txtstate.edu Department of Civil Engineering, University of Louisiana at Lafayette, Lafayette, LA 70503, USA

in 2021, a 13% increase compared to 2020 (NHTSA 2021). Due to age-related physical vulnerability, the driver fatalities per 1,000 licensed drivers involved in crashes follow a sharp rise with respect to driver age, particularly those aged 65 years and older (more details in Figure 1a on the left).

Speed is a major factor contributing to motor vehicle crashes and the resulting injury severity (Aarts and Van Schagen 2006; Bates et al. 2014; Elvik et al. 2009; Mohanty and Gupta 2015). When drivers drive at high speed, they are expected to have slower reaction time, longer stopping sight distance (SSD), less control over the vehicle, limited field of vision, difficulty in judging the distance and speed of other vehicles, and greater impact force (in the event of a crash). With relatively reduced cognitive skills, physiological limitations, and the decision-making ability of older drivers, high-speed settings can put them into extremely complex situations. In addition, complex environmental conditions including dark lighting conditions or rainy weather conditions can add to the complexity of the problem. The variation in older driver fatalities and posted speed limit [shown in Figure 1b on the right] clarifies the intriguing trend of older driver fatalities on roadways having higher posted speed limits. According to the FARS database, around 74% of the older driver fatalities in the U.S. occurred in the high-posted speed limit zone (45 mph or higher) in 2021. The over-representation of older driver fatalities in high-speed settings calls for immediate attention and this issue needs to be thoroughly addressed to improve the safety of older drivers.

To address the safety issue of older drivers, this study used a crash database (2017-2021) collected from the Louisiana Department of Transportation & Development (LADOTD) to investigate the factors contributing to older driver crashes in high-speed settings (posted speed limit of 45 mph or higher). For analysis purposes, this study used a two-step hybrid framework: Extreme Gradient Boosting (XGBoost) for selecting top variables in binary format (1, 0) and then using a random parameter model to investigate unobserved heterogeneity in crashes involving older drivers in high-speed settings. The random parameter modelling approach is widely applied in transportation safety research (Ali et al. 2020; Ali, Raadsen, and Bliemer 2023; Azimi et al. 2020; Fountas et al. 2021; Fountas and Anastasopoulos 2017, 2018; Jalayer et al. 2018; Se et al. 2021; Shao et al. 2020). The selection

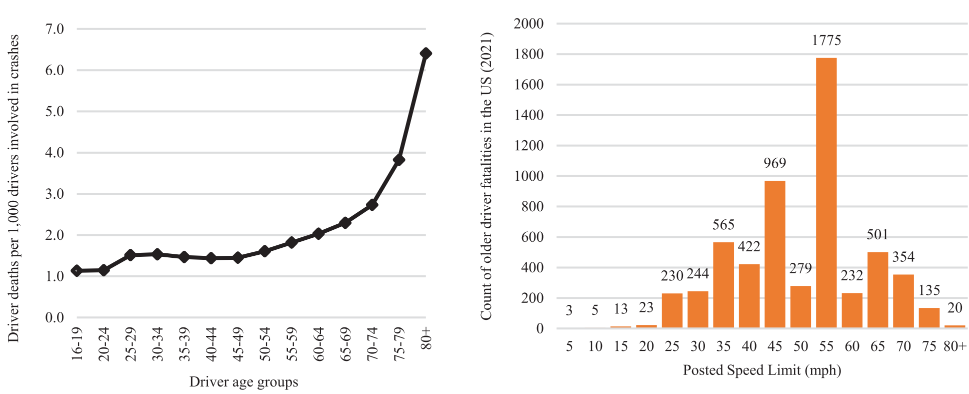


**Figure 1.** (a) Driver fatalities per 1,000 licensed drivers involved in crashes during the 2016–2021 period (Source: Insurance Institute of Highway Safety) (b) Count of older driver fatalities in the US according to the posted speed limit (year = 2021) [Source: FARS].

of this specific model (random parameter ordered probit) is significant from two different contexts: a) to account for the ordinal nature of injury severity data, and b) to address the 'unobserved heterogeneity' in crash data. In crash data analysis, heterogeneity can arise from various sources, such as differences in driver behaviour, road conditions, traffic volume, vehicle types, weather, and other environmental factors. These factors can vary across different regions, periods, or specific groups of road users (e.g. older drivers). Therefore, the complex interaction among the crash contributing factors and the resulting crash injury severity can be more accurately modelled by taking unobserved heterogeneity into account (Chen and Chen 2011). The findings of this research are expected to provide an improved understanding of older driver crash problems on high-speed settings and develop problem-specific countermeasures to reduce the target crashes.

## 2. Literature review

Investigation into older driver crashes has got extensive research attention over the last few decades. Some of the previous investigations addressed the physical and cognitive limitations of older drivers and their crash involvements. A previous study conducted by Suen and Mitchell reported several critical age-related impairments that contribute to crashes involving older drivers including increased reaction time, reduced vision (especially at night), difficulty in judging speed and distance at intersections, difficulty in perceiving and analyzing situations, difficulty in turning head, and more prone to fatigue (Ling Suen and Mitchell 1998). Another study focused on visual attention problems as a predictor of vehicle crashes involving older drivers (Ball et al. 1993). The study reported several factors significantly correlated with crashes including eye health status, visual sensory function, cognitive status, and chronological age.

Since older drivers are overrepresented in intersection crashes, some other studies addressed this issue in their investigation (Braitman et al. 2007; Broberg, Jakobsson, and Isaksson-Hellman 2008; Caird et al. 2005; Dotzauer et al. 2013; Lombardi, Horrey, and Courtney 2017; Oxley et al. 2006; Preusser et al. 1998; Samuel, Yamani, and Fisher 2016; Sifrit et al. 2010; Zafian et al. 2021). Most of these investigations focused on the likelihood of older driver crashes at intersections compared to other driver age groups (young or middle-aged). For example, a previous study explored older driver crashes at intersections by incorporating driver, vehicle, and roadway factors using crash data (2002-2006) collected from FARS and GES websites (Braitman et al. 2007). The study utilised the crash involvement ratio (CIR), a ratio of at-fault to not at-fault, as a measure of exposure. The major findings of this research were: (a) higher involvement of two-vehicle fatal crashes at the intersection (CIR = 2.4 for 70–79 and CIR = 5.4 for 80+) (b) difficulty in navigating intersections controlled by the stop sign and flashing lights, and (c) higher CIR for left-turn crashes at the signalised intersection.

Intersection control type and geometry play a significant role in crashes involving older drivers. Few of the previous studies focused on different traffic scenarios (e.g. approach roadway speed, signal timing, left turn phasing) at the intersection accounting for the driver age group. Using a driving simulator, a previous study investigated the performance of older and younger drivers at signalised intersections for a specific traffic scenario: traffic lights changed from green to yellow at the last second (Caird et al. 2007). The approach speed to the intersection was designed at 42 mph. The study found that older drivers who chose to go through the yellow light had a higher likelihood of being present at the intersection when the traffic light changed to red. Another study focused on three types of left-turn phasing (e.g. protected, permitted, and protected-permitted) at high-speed signalised intersections in Iowa considering three different age groups: 14-24, 25-64, and 65+ years (Mueller et al. 2007). Among all driver age groups, older drivers had the highest left-turn crash rates for all types of phasing with the highest crash rate at protected-permitted phasing, followed by permitted phasing.

Few other studies have addressed older driver crashes from other specific viewpoints (e.g. manner of collision, gender). A summary of these investigations is provided in the following Table 1. The notation of the arrow is interpreted as follows: a) the downward arrow suggests that older drivers are less likely to be involved (less represented) in crashes in the presence of such factors, b) the upward arrow suggests that older drivers are more likely to be involved (over-represented) in crashes in the presence of such factors, c) the horizontal arrow suggests that the factors do not significantly influence or are not related to the likelihood of older drivers being involved in crashes.

Nilsson (1981) suggested power relationships linking changes in traffic speeds to road crashes at varying injury severity levels (G 1981). According to the model, fatal crashes increase with the 4th power of mean speed, serious casualty crashes with the 3rd power, and casualty crashes with the 2nd power. The rise in crash victims at different injury severity levels is connected to crash increases and higher powers predicting the number of victims per crash. Higher speed is associated with more severe injuries. In recent years, many studies tried to examine the association between speed and the likelihood of crash occurrences (Das and Geedipally 2020; Park et al. 2021). Studies show that speed variation is associated with the likelihood of crash occurrences on different roadway facilities. Interstate and freeways show minimal relationship of speed-crash association due to the better road design. However, this finding has not been examined for different age groups of the driver population (Park et al. 2021). Older drivers are often more susceptible to crashes due to age-related physical and cognitive limitations, particularly in complex driving environments. The current study aims to examine this critical research gap by applying a hybrid approach (combining machine learning and statistical methods) to the crash data associated with older driver populations on high-speed roadways.

Focusing on injury severity modelling, a diverse range of statistical approaches has been implemented in the previous research. These models can be typically classified into unordered response model and ordered response model. While unordered response models are typically used for analysis (Al-Ghamdi 2002; Chen, Cao, and Logan 2012; Kononen, Flannagan, and Wang 2011; Sze and Wong 2007; Usman, Fu, and Miranda-Moreno 2016), these consider injury severity categories as unordered and independent, which might not fully capture the inherent ordinal structure of severity levels. Since injury severity levels have a natural order, such as the KABCO scale (K, A, B, C, O - which stands for fatal, incapacitating, non-incapacitating, possible injury, and no injury; most severe to least severe), ordered response models are more appropriate considering this fact (Eluru, Bhat, and Hensher 2008). These models explicitly account for the inherent ordinal nature of the injury severity, allowing for a more nuanced analysis of the factors influencing injury severity (Alrumaidhi and Rakha 2022; Asare and Mensah 2020; Pour-Rouholamin, Jalayer, and Zhou 2017). Few studies utilised the Ordered Probit (OP) approach for injury severity analysis (Ivan and Konduri 2018; Li et al. 2021; Wang et al. 2010; Yang et al. 2011). Addressing unobserved

**Table 1.** Key findings from selected studies.

| Study | Location; Data; Model | Key Focus | Factors contributing to older driver crashes |
|---|---|---|---|
| (McGwin and Brown 1999) | Alabama; 1996, Crash rate | Comparing three age groups (young, middle-aged, older) | Driver fatigue (↓), evening and early morning (↓), curved roads (↓), during adverse weather (↓), single-vehicle crash (↓), travelling at high speeds (↓), intersections (↑), failure to yield the right of way (↑), failure to pay attention at stop signs or signals (↑), unseen objects (↑), turning or changing lanes (↑) |
| (Dissanayake and Lu 2002) | Florida; Binary logit model | Fixed object passenger car crashes | Travel speed (↑), restraint usage (↓), point of impact as the front of the vehicle (↑), use of alcohol and drugs (↑), driver condition not normal (↑), female driver (↑), rural location (↑), presence of curves (↑) |
| (Viano et al. 1990) | US; 1975-1986; Summary statistics | Multi-vehicle side-impact crashes | Age (↑), intersection location (↑), traffic violation (↑), daylight (↑), dry roads (↑), no alcohol involvement (↑) |
| (Cicchino and McCartt 2015) | US; 2005-2007; Rate Ratio (RR) | Critical driving error | Driver error (↑), inadequate surveillance [looking but not seeing] (↑), gap misjudgment [gap between two vehicles, vehicle speed of oncoming vehicle] (↑), illegal maneuvers (↑), turning left at intersection (↑) |
| (Dissanayake 2004) | Florida; Logistic regression | Single-vehicle crashes (comparing young and older drivers) | Speeding (↑), no seat-belt use (↑), frontal impact point (↑), county of residence (→), weather condition (→) |
| (Alshehri and Dissanayake 2022) | US; logistic regression | Single-vehicle fatal intersection crashes | controlled intersection (↑), two-way undivided highways (↑), low posted speed limit (↑), urban locations (↑), nighttime (↑), roll-over (↑), hitting animals (↑), alcohol-impaired driving (→) |
| (Islam, Hossain, and Shaban 2023) | Alabama; 2012-2016; Random parameter logit model | Driver gender (unsignalized intersection) | Male older driver: alcohol-impaired driving (↑), horizontal curve (↑), stop sign (↑); Female older driver: intersection approaches on tangents with flat grades (↑), age older than 75 years (↑); Both gender: turning maneuver (↑), freeway-ramp junction (↑), high-speed approach (↑) |

heterogeneity in crash data, few recent research utilised more sophisticated Random Parameter Ordered Probit (RPOP) approach to allow for variability in the parameters of the OP model across individuals or observations (Babaei and Kunt 2024; Jomnonkwao et al. 2023; Rao et al. 2024; Se et al. 2021; Shangguan et al. 2022; Xing et al. 2020; Yu et al. 2020b).

Recently, hybrid modelling approaches has been adopted which refers to the integration of data driven (i.e. machine learning) and statistical model to advance knowledge and predict safety outcomes on roadways (Mannering et al. 2020). This integration often involves combining the strengths of various modelling techniques to improve the overall interpretability of crash contributing factors. While machine learning models excel in precision and accuracy in classification tasks, statistical models offer interpretable coefficients or parameters (e.g. marginal effects) that offer insights into the influence of factors in crash injury severity. For example, some of the recent roadway crash investigations utilised hybrid modelling approaches incorporating random parameter modelling with other models including the Bayesian network (Sun et al. 2022), Decision Tree (Ali et al. 2021; Ali and Haque 2023; Subhan et al. 2023), Random Forest based SHAP algorithm (RF-SHAP) (Sun et al. 2023), and XGBoost (Goswamy, Abdel-Aty, and Islam 2023). In this research, we developed an innovative framework for crash data analysis by combining XGBoost as a classifier (i.e. feature selection tool) and then applying random parameter modelling for crash severity analysis. The proposed hybrid framework can offer advantages by leveraging the strengths of each component. XGBoost model is superior for the feature selection (Asselman, Khaldi, and Aammou 2023; Dhaliwal, Nahid, and Abbas 2018). While XGBoost is known for its predictive power, its black-box nature can sometimes limit interpretability. To address this, a random parameter modelling approach is used in the next step to report the marginal effects of key significant variables affecting crash injury severity while addressing unobserved heterogeneity in crash data. Overall, the proposed hybrid modelling approach provides flexibility in capturing complex multidimensional heterogenous relationships within the crash data.

### *2.1. Research gap and study objectives*

Based on the review of previous literature, it was identified that older driver-involved crashes on high-speed roadways have been limitedly explored, and therefore a separate investigation is required to address the problem. This research aims to identify the key contributing factors in crash injury severity involving older drivers on roadways having a posted speed limit of 45 mph or higher. For analysis purposes, a novel hybrid modelling approach combining XGBoost, and random parameter modelling is used. Another objective of this research is to recommend problem-specific safety countermeasures tailored to address the safety issues associated with older drivers. The findings of this research can assist transportation safety professionals and agencies by providing critical scenarios in which older drivers are more likely to be involved in severe crashes on high-speed roadways and develop problem-specific countermeasures.

### *2.2. Scope*

The scope of this research is limited to older drivers (age 65 years or higher). Roadways having a posted speed limit of 45 mph or higher were defined as 'high-speed settings'. The threshold of 45 mph was selected based on the recommendation provided by previous research conducted by the Louisiana Transportation Research Center (Codjoe et al. 2021).

## 3. Methods

In this section, details of the Extreme Gradient Boosting (XGBoost) and Random Parameter Ordered Probit (RPOP) model are discussed.

### 3.1. Extreme gradient boosting (XGBoost)

XGBoost, a supervised machine-learning technique, utilises an ensemble of decision trees and gradient boosting to make predictions (Chen and Guestrin 2016). The concept of ensemble learning involves integrating several low-precision decision tree models to produce strong, high-precision learners. Since its first release as an open-source software package in 2014, XGBoost has been extensively used in traffic safety research in recent times (Chen et al. 2020; Guo et al. 2021; Ma et al. 2019; Qu et al. 2019; Yang et al. 2022; Yang, Chen, and Yuan 2021). The choice of XGBoost lies in its ability to capture complex nonlinear relationships in crash data with high precision and less computational resources. As reported in previous literature, XGBoost is more accurate compared to other machine learning techniques (e.g. SVM, deep neural network) in predicting the likelihood of crashes (Mousa, Bakhit, and Ishak 2019; Schlögl et al. 2019). In addition, as an integrated algorithm, XGBoost is not affected by the multicollinearity of data (Chen et al. 2020).

The main purpose of the algorithm is to maximise the value of the objective function and apply the machine learning algorithm using the gradient boosting framework. The objective function consists of two parts: the training loss function and regularisation term and is provided by the following equation:

$$Obj(\theta) = Training\ loss\ function + Regularization\ term = L(\theta) + \Omega(\theta) \tag{1}$$

Model performance on training data is measured by the training loss, while model complexity (e.g. overfitting) is handled by the regularisation term. Overall, the core working principle of the XGBoost algorithm is based on a three-step process: a) apply a second-order Taylor expansion on the objective function, b) use a decision tree model using the second derivative information, and c) add the decision tree model's complexity as a regularisation term to the optimisation objective.

In this study, the classification function of the XGBoost algorithm is employed. In XGBoost, the multiclass classification function involves the use of a 'softmax transformation' to convert the raw predicted scores into class probabilities. Let's denote the raw predicted scores for observation I across K classes as $y_i = y_{i1}, y_{i2}, \ldots\ldots\ldots y_{ik}$. The softmax transformation is applied to each raw predicted score $y_{ik}$ to convert them into class probabilities:

$$P(y_{ik} = 1) = \frac{e^{y_{ik}}}{\sum_{j=1}^{K} e^{y_{ik}}} \tag{2}$$

Here, e is the base of the natural logarithm, and the denominator is the sum of the exponentials of all raw scores for observation 'i'. This ensures that the probabilities sum to 1 across all classes. The final predicted class for observation i is determined by selecting the class with the highest probability.

$$y_i' = argmax_k P(y_{ik} = 1) \tag{3}$$

This means that the class k with the highest softmax-transformed probability is chosen as the predicted class for observation i. For training, XGBoost minimises a multiclass extension

**Table 2.** Specification of XGBoost Model Hyperparameters.

| Hyperparameters | Range | Utilization |
|---|---|---|
| nrounds | User defined | Maximum number of boosting iterations |
| eta (learning rate) | 0 - 1 | It specifies the learning rate of the algorithm by controlling the step size shrinkage used in each boosting step. Lower values make the model more robust to overfitting but require more boosting rounds to achieve optimal performance. |
| max_depth | 0 - 6 | The maximum depth of a tree. Deeper trees can capture more complex patterns in the data but are prone to overfitting. |
| gamma | 0 - $\infty$ | Specifies the minimum loss reduction required to make a further partition on a leaf node. It acts as a regularisation parameter by controlling the complexity of the tree. |
| subsample | 0 - 1 | The fraction of training data to randomly sample during each boosting round. Setting it to lower values can prevent overfitting. For example, setting 'subsample' to 0.5 implies that XGBoost will randomly sample half of the training data before constructing trees. |
| colsample_by_tree | 0 - 1 | The fraction of features to randomly select for each tree. It helps introduce randomness and reduce correlation among trees. |
| min_child_weight | 0 - 1 | Specifies the minimum number of instances (or samples) required to create another node in the tree. If the number of instances in a leaf node is less than min_child_weight, the algorithm will stop further partitioning that node |

of the binary classification log-likelihood loss. The objective function can be redefined as:

$$Obj(\theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} softmax(y_{ik}) + \sum_{k=1}^{K} \Omega(f_k) \tag{4}$$

Here, $\theta$ represents the set of model parameters, and $\Omega(f_k)$ is the regularisation term for each tree in the ensemble. In summary, the multiclass classification function in XGBoost involves computing class probabilities using the softmax transformation and selecting the class with the highest probability as the final predicted class. This process is repeated for each observation during both training and prediction. The analysis was carried out using the open-source R package 'xgboost'.

The specification of hyperparameters for the XGBoost model involves tuning various parameters to optimise its performance focusing on an intricate balance between bias and variance. A grid search technique is adopted to systematically search through a predefined set of hyperparameter combinations, evaluating each combination to identify the optimal set of hyperparameters (Ali, Hussain, and Mazharul Haque 2024). A list of the model hyperparameters is provided in the following Table 2.

### 3.2. Random parameter ordered probit (RPOP)

In this sub-section, we provide a concise introduction to the theoretical framework of the Random Parameter Ordered Probit (RPOP) model, incorporating heterogeneity in the means. Among diverse regression models, we chose the 'ordered' version to handle the ordinal nature of crash severity levels. In injury severity modelling, severity levels are often ordinal specified by the KABCO scale (K = Fatal, A = Severe, B = Moderate, C = Complaint, O = Property Damage Only) ordering from most severe to least severe. This ordinality implies that there's a meaningful hierarchy among the severity levels. The ordered probit model acknowledges this ordinality structure by assuming that the underlying latent variable follows a normal distribution and that observed categories correspond to intervals along this latent continuum (Kockelman and Kweon 2002; Shao et al. 2020).

The RPOP model allows for testing unobserved heterogeneity by introducing random parameters and considering potential variations in the means of these parameters. To determine the injury severity value $Y*_i$ for the observation $i$ involved in the older driver crash (Eluru, Bhat, and Hensher 2008).

$$Y_*^i = \beta_i X_i + \varepsilon_i \tag{5}$$

Where $X_i$ is the vector of explanatory variables, $\beta_i$ represents the vector of estimated parameters, $\epsilon_i$ denotes a random distribution. The injury-severity $Y*_i$ for observation $i$ is defined as

$$Y_*^i = j, if u_{i,j-1} \leq Y_i^* \leq u_{i,j} \tag{6}$$

Where $j$ ($j = 0, 1, 2, \ldots$ ,J) denotes the injury-severity level, $u_{i,j}$ is the estimated threshold, and $u_{i,0} = -\infty$, $u_{i,j} = +\infty$. The values of thresholds distinguish the various injury severity categories (Fountas and Anastasopoulos 2018). The thresholds are assumed to be ascending in order (i.e. $u_{i,0} < u_{i,1} < \ldots < u_{i,j-1} < u_{i,j}$) (Xin et al. 2017). In this study, $j = 0, 1, 2$, respectively, for fatal and severe injury (FS), moderate and minor injury (IN), and no injury (PDO). The probability of an observation $i$ being $j^{\text{th}}$ injury-severity is defined as (Azimi et al. 2020; Eluru, Bhat, and Hensher 2008):

$$\begin{aligned} P(y = 0) &= \Phi(-\beta_i X_i) \\ P(y = 1) &= \Phi(u_1 - \beta_i X_i) - \Phi(-\beta_i X_i) \\ P(y = 2) &= \Phi(u_2 - \beta_i X_i) - \Phi(u_1 - \beta_i X_i) \end{aligned} \tag{7}$$

Where P ($y = j$) is the probability of the $j^{th}$ injury severity level; $\Phi$ denotes the cumulative function of the standard normal distribution. The likelihood function of individual $i$ is written as (Xin et al. 2017),

$$(Y_i|\beta_i) = \prod_{i=1}^{N} \prod_{j=0}^{J} P(Y_i - j)^{\delta_{ij}} \tag{8}$$

Where $\delta_{ij}$ is 1 if the injury severity for individual $i$ is $j$, and 0 otherwise. The log-likelihood function is expressed by the following equation:

$$LL(Y_i|\beta_i) = \sum_{i=1}^{N} \sum_{j=0}^{J} \delta_{ij} LN[P(Y_i = j)] \tag{9}$$

In the RPOP, the explanatory variables exhibit variability across the observations. This model considered three different types of distributions, namely normal, lognormal, and uniform, to test their suitability. The results indicated that the normal distribution yielded the best statistical fit, which aligns with findings from previous studies (Xin et al. 2017)**;** (Jalayer et al. 2018), and (Azimi et al. 2020).

In this study, we adopt the assumption that the random coefficients follow a normal distribution across individuals (Hsiao and Hashem Pesaran 2008). To estimate the coefficients of the random parameters, we tested the accuracy of Halton draws using 100, 200, and 500 draws. Among these, the results for 500 Halton draws demonstrated greater stability and statistical significance. While this data provides valuable insights, future studies could explore the option of testing with 1000 Halton draws to further enhance accuracy and comprehensiveness. The choice of 500 Halton draws in our study reflects a balance between computational efficiency and achieving robust and significant results.

Building upon previous research (Azimi et al. 2020; Behnood and Mannering 2017; Mannering, Shankar, and Bhat 2016; Xin et al. 2017; Yu, Zheng, and Ma 2020a), we account for heterogeneity in the means by treating the random parameters $\beta$i as estimable parameters, which are precisely defined by our modelling approach. $\beta$i can be defined by:

$$\beta_i = \beta + \Theta Z_i + \varphi_i \tag{10}$$

Here $\beta$ is the mean for all crashes, Zi is the vector of explanatory variables (e.g. driver characteristics, crash characteristics, roadway attributes, and environmental characteristics) for the individual i that affects the mean of $\beta$i, $\Theta$ is the vector of estimable parameters, and $\varphi$i is a randomly distributed term that captures unobserved heterogeneity across observations (Behnood and Mannering 2017). The marginal effect quantifies the change in probability resulting from an indicator switch from '0' to '1'. It is calculated using the following equation (Jalayer et al. 2018).

$$\frac{P(y=j)}{\partial X} = [\Phi(u_j - \beta_i X_i) - \Phi(u_{j+1} - \beta_i X_i)]\beta \tag{11}$$

The research team explored four different modelling techniques including:

- Random Parameter Ordered Probit (RPOP)
- Correlated Random Parameter Ordered Probit (CRPOP)
- Random Parameter Ordered Probit with Heterogeneity in Means (RPOP-HM)
- Correlated Random Parameter Ordered Probit with Heterogeneity in Means (CRPOP-HM)

The idea of using four variants of random parameter models is to check the suitability of the models and report the best model according to the AIC value.

## 4. Data

The LADOTD crash database stored in MS Access format was utilised to pull all the police-reported crashes between 2017 and 2021. The primary database was created by merging four data tables - Crash table, Vehicle table, Highway table, and DOTD table and then merged with the help of using a 'crash number' as a matching criterion. A total of 790,673 police-reported crashes were collected during the 5-year study period (2017 = 165,941; 2018 = 163,592; 2019 = 160, 185; 2020 = 138,897; 2021 = 162, 058). At first, some of the crashes were removed based on: a) unknown injury severity level, and b) unknown driver age. In the next step, the final database was prepared using a three-step filtering process –

- Filter 1: Choose crashes involving older drivers (AGE = 65 or higher)
- Filter 2: Choose crashes that occurred on road segments (INTERSECTION = No)
- Filter 3: Choose crashes that occurred on roadways having posted speed limit of 45 mph or higher (POSTED_SPEED = 45 or higher)

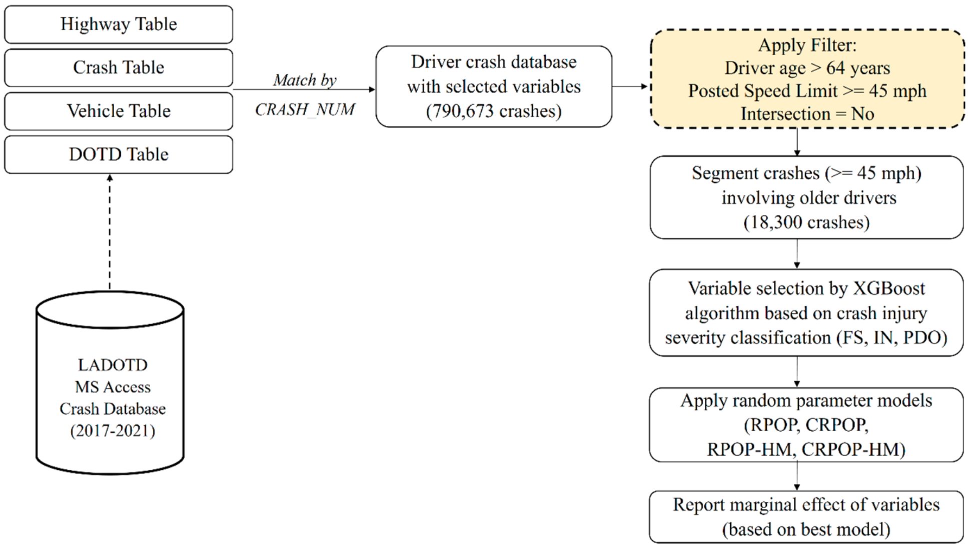


**Figure 2.** Data collection and analysis framework.

The study only considered the crashes that occurred on high-speed segments (posted speed limit = 45 mph or higher) and did not consider crashes that occurred at intersections with approaching roadways with a posted speed limit of 45 mph or higher. This is mainly because, the intersection location is expected to have multiple traffic controls (signal, stop, and yield) and maneuvers (left turn, right turn, and going straight), and the context of 'high speed' may play a different role while approaching an intersection. It is important to note that this study is designed to examine the driving behaviour of older individuals and its safety implications on high-speed roadways. By narrowing the focus to this demographic, this study can provide a more in-depth analysis and understanding of the unique challenges and patterns associated with older drivers. Including other age groups in the comparison would broaden the scope beyond the intended focus, potentially diluting the relevance of the study findings.

The final database contains 18,300 unique crashes involving older drivers (age 65 years or higher) that occurred on road segments having posted speed limits of 45 mph or higher. The data integration and analysis flowchart are shown in Figure 2.

This database utilises the 'KABCO' scale (K = Fatal, A = Severe, B = Moderate, C = Complaint, O = No injury) to categorise injury severity in a crash. In the sorted crash database, out of 18,300 unique crashes, the distribution crash injury severity was as follows: K = 218 crashes (1.19%), A = 135 crashes (0.74%), B = 1170 crashes (6.39%), C = 4307 crashes (23.54%), and O = 12,470 crashes (68.14%). Due to the low frequency of fatal (1.19%) and severe (0.74%) injury categories, these two injury severity levels are merged (Ashifur Rahman, Das, and Sun 2022; Meyer et al. 2021; Slikboer et al. 2020; Yu, Ma, and Shen 2021; Zeng et al. 2022). Since, 'no injury' (i.e. property damage only) is completely a distinct outcome of crashes, this category has been separately considered. Overall, 'crash severity' was regrouped into three categories: fatal and severe injury (K+A), moderate and complaint injury (B+C), and no injury (O). A similar 3-level injury severity categorisation technique has been intensively adopted in previous safety research allowing for more meaningful comparisons and insights into the factors contributing to different levels of crash severity (Alnawmasi, Ali, and Yasmin 2024; Alnawmasi and Mannering 2019; 2023). It should be mentioned that the following abbreviations are used in this research for ease of interpretation: FS stands for fatal and severe injury (K+A), IN for moderate and complaint injury (B+C), and PDO for Property Damage Only or no injury crashes. A total of 20 variables (including the target variable) were chosen by considering their availability in the Louisiana crash database and a comprehensive examination of relevant previous research on older drivers. The primary selected variables are:

- Driver characteristics (driver age, driver gender, driver distraction, driver condition)
- Roadway and traffic characteristics (highway type, road type, access control, number of lanes, ADT)
- Crash characteristics (primary contributing factor, manner of collision, passenger present, vehicle type)
- Temporal factors (day of the week, crash month, crash hour)
- Environmental characteristics (lighting condition, weather condition, location type)

Note that some of the variables were not considered due to their high skewness. Usually, older drivers are more likely to properly use their seat belts regularly while driving. According to seatbelt usage, the proper usage of seatbelt is found to be highly skewed (90.19%), and such variables are not considered in this research for meaningful model development and interpretation.

Table 3 presents the descriptive statistics of key variables. The data illustrates the distribution of 18,300 recorded crashes on high-speed segments throughout 2017-2021, categorised by crash injury severity (PDO, IN, FS). According to the driver age variable, the crashes are distributed as follows: 12,283 crashes involved drivers aged 65–74 years, 4,951 crashes involved drivers aged 75–84 years, and 1,099 crashes involved drivers aged 85 years and above. Regarding PDO crashes, the table provides insights into the involvement of older drivers based on their age groups. The analysis shows that 67.9% of property damage crashes were attributed to drivers aged 65–74 years, 26.8% to drivers aged 75–84 years, and 5.3% to drivers aged 85 years and above. Notably, the total percentage of these contributions sums up to 100% (67.9% + 26.8% + 5.3% = 100%).

## 5. Results and discussion

### 5.1. One-Hot encoding by XGBoost algorithm

In previous studies involving random parameter models, the conversion of variable categories into binary form (1 or 0) was often handled less systematically. For example, variables such as ‘lighting condition’ were frequently coded as ‘1 if dark, 0 otherwise’ solely based on the modeller’s engineering judgment, lacking a robust theoretical foundation to justify these coding choices. To address this limitation issue, this research utilised the eXtreme Gradient Boosting (XGBoost) algorithm to find the top variable features (i.e. categories) important for classifying crash injury severity (FS, IN, PDO).

Note that the crash injury severity (FS, IN, PDO) was set as the dependent variable in the XGBoost model. The ‘One Hot Encoding’ technique was used to convert primarily selected

**Table 3.** Descriptive Statistics of Key Variables.

| Type | Variable name | Variable Feature | Total crashes | PDO | | IN | | FS | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | # | % | # | % | # | % |
| Driver characteristics | driver_age | 65–74 | 12283 | 8462 | 67.9% | 3614 | 66.0% | 207 | 58.6% |
| | | 75–84 | 4951 | 3342 | 26.8% | 1502 | 27.4% | 107 | 30.3% |
| | | 85+ | 1066 | 666 | 5.3% | 361 | 6.6% | 39 | 11.0% |
| | driver_gender | male | 11192 | 7536 | 60.4% | 3404 | 62.2% | 252 | 71.4% |
| | | female | 7061 | 4897 | 39.3% | 2063 | 37.7% | 101 | 28.6% |
| | | unknown | 47 | 37 | 0.3% | 10 | 0.2% | 0 | 0.0% |
| | driver_distraction | yes | 12301 | 8675 | 69.6% | 3499 | 63.9% | 127 | 36.0% |
| | | no | 5999 | 3795 | 30.4% | 1978 | 36.1% | 226 | 64.0% |
| | driver_condition | inattentive_distracted | 12146 | 8579 | 68.8% | 3444 | 62.9% | 123 | 34.8% |
| | | normal | 4134 | 2999 | 24.0% | 1076 | 19.6% | 59 | 16.7% |
| | | illness_fatigued_asleep | 760 | 225 | 1.8% | 508 | 9.3% | 27 | 7.6% |
| | | alcohol_drug | 314 | 172 | 1.4% | 130 | 2.4% | 12 | 3.4% |
| | | physical_impairment | 54 | 30 | 0.2% | 23 | 0.4% | 1 | 0.3% |
| | | other_unknown | 892 | 465 | 3.7% | 296 | 5.4% | 131 | 37.1% |
| Roadway & Traffic characteristics | highway_type | city_street | 1459 | 1078 | 8.6% | 367 | 6.7% | 14 | 4.0% |
| | | state_highway | 8021 | 5212 | 41.8% | 2632 | 48.1% | 177 | 50.1% |
| | | parish_road | 773 | 525 | 4.2% | 227 | 4.1% | 21 | 5.9% |
| | | us_highway | 4136 | 2877 | 23.1% | 1194 | 21.8% | 65 | 18.4% |
| | | interstate | 3833 | 2736 | 21.9% | 1023 | 18.7% | 74 | 21.0% |
| | | others | 78 | 42 | 0.3% | 34 | 0.6% | 2 | 0.6% |
| | road_type | two_no_separation | 9430 | 6113 | 49.0% | 3097 | 56.5% | 220 | 62.3% |
| | | two_separation | 7534 | 5412 | 43.4% | 2001 | 36.5% | 121 | 34.3% |
| | | one_way | 1225 | 860 | 6.9% | 355 | 6.5% | 10 | 2.8% |
| | | other_unknown | 111 | 85 | 0.7% | 24 | 0.4% | 2 | 0.6% |
| | access_control | no_control | 13431 | 8997 | 72.1% | 4163 | 76.0% | 271 | 76.8% |
| | | full_control | 3612 | 2569 | 20.6% | 974 | 17.8% | 69 | 19.5% |
| | | partial_control | 1188 | 851 | 6.8% | 326 | 6.0% | 11 | 3.1% |
| | | other | 52 | 39 | 0.3% | 12 | 0.2% | 1 | 0.3% |
| | | unknown | 17 | 14 | 0.1% | 2 | 0.0% | 1 | 0.3% |

(*continued*).

**Table 3.** Continued.

| Type | Variable name | Variable Feature | Total crashes | PDO | | IN | | FS | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | # | % | # | % | # | % |
| | num_lanes | 2-lane | 5398 | 3244 | 26.0% | 1992 | 36.4% | 162 | 45.9% |
| | | 3-4-lane | 8640 | 6103 | 48.9% | 2408 | 44.0% | 129 | 36.5% |
| | | $> 4$ | 2523 | 1896 | 15.2% | 596 | 10.9% | 31 | 8.8% |
| | | unknown | 1739 | 1227 | 9.8% | 481 | 8.8% | 31 | 8.8% |
| | ADT | $< 5000$ | 2474 | 1391 | 11.2% | 985 | 18.0% | 98 | 27.8% |
| | | 5000–10000 | 2025 | 1224 | 9.8% | 741 | 13.5% | 60 | 17.0% |
| | | 10001–20000 | 2988 | 1996 | 16.0% | 934 | 17.1% | 58 | 16.4% |
| | | $> 20000$ | 9072 | 6630 | 53.2% | 2336 | 42.7% | 106 | 30.0% |
| | | unknown | 1741 | 1229 | 9.9% | 481 | 8.8% | 31 | 8.8% |
| Crash Characteristics | primary_contributing_factor | violation | 13746 | 9517 | 76.3% | 4012 | 73.3% | 217 | 61.5% |
| | | prior_movement | 2094 | 1501 | 12.0% | 545 | 10.0% | 48 | 13.6% |
| | | driver_condition | 1127 | 479 | 3.8% | 596 | 10.9% | 52 | 14.7% |
| | | vision_obscurement | 101 | 67 | 0.5% | 33 | 0.6% | 1 | 0.3% |
| | | pedestrian_actions | 63 | 8 | 0.1% | 34 | 0.6% | 21 | 5.9% |
| | | traffic_control | 14 | 9 | 0.1% | 5 | 0.1% | 0 | 0.0% |
| | | pedestrian_condition | 4 | 2 | 0.0% | 1 | 0.0% | 1 | 0.3% |
| | | other_factors | 1139 | 879 | 7.0% | 248 | 4.5% | 12 | 3.4% |
| | | unknown | 12 | 8 | 0.1% | 3 | 0.1% | 1 | 0.3% |
| | manner_of_collision | rear_end | 5837 | 3835 | 30.8% | 1940 | 35.4% | 62 | 17.6% |
| | | sideswipe | 4237 | 3607 | 28.9% | 606 | 11.1% | 24 | 6.8% |
| | | single_vehicle | 3751 | 2149 | 17.2% | 1433 | 26.2% | 169 | 47.9% |
| | | right_turn | 1590 | 993 | 8.0% | 572 | 10.4% | 25 | 7.1% |
| | | right_angle | 1296 | 755 | 6.1% | 509 | 9.3% | 32 | 9.1% |
| | | left_turn | 410 | 334 | 2.7% | 73 | 1.3% | 3 | 0.8% |
| | | head_on | 252 | 84 | 0.7% | 130 | 2.4% | 38 | 10.8% |
| | | other | 927 | 713 | 5.7% | 214 | 3.9% | 0 | 0.0% |
| | passenger_present | no | 14648 | 9998 | 80.2% | 4357 | 79.6% | 293 | 83.0% |
| | | yes | 3652 | 2472 | 19.8% | 1120 | 20.4% | 60 | 17.0% |
| | vehicle_type | light_truck | 5082 | 3444 | 27.6% | 1534 | 28.0% | 104 | 29.5% |
| | | passenger_car | 6419 | 4409 | 35.4% | 1905 | 34.8% | 105 | 29.7% |
| | | VAN_SUV | 5406 | 3742 | 30.0% | 1578 | 28.8% | 86 | 24.4% |
| | | other | 1333 | 840 | 6.7% | 437 | 8.0% | 56 | 15.9% |
| | | unknown | 60 | 35 | 0.3% | 23 | 0.4% | 2 | 0.6% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Temporal factors | crash_month | spring | 4374 | 2953 | 23.7% | 1329 | 24.3% | 92 | 26.1% |
| | | summer | 4349 | 2959 | 23.7% | 1301 | 23.8% | 89 | 25.2% |
| | | fall | 4916 | 3353 | 26.9% | 1468 | 26.8% | 95 | 26.9% |
| | | winter | 4661 | 3205 | 25.7% | 1379 | 25.2% | 77 | 21.8% |
| | day_of_week | weekday | 14810 | 10218 | 81.9% | 4333 | 79.1% | 259 | 73.4% |
| | | weekend | 3490 | 2252 | 18.1% | 1144 | 20.9% | 94 | 26.6% |
| | crash_hour | 12am-6am | 424 | 256 | 2.1% | 152 | 2.8% | 16 | 4.5% |
| | | 6am-12pm | 4481 | 3055 | 24.5% | 1349 | 24.6% | 77 | 21.8% |
| | | 12 pm-6 pm | 9231 | 6365 | 51.0% | 2705 | 49.4% | 161 | 45.6% |
| | | 6pm-12am | 4164 | 2794 | 22.4% | 1271 | 23.2% | 99 | 28.0% |
| Environmental factors | weather_condition | clear | 13512 | 9237 | 74.1% | 4007 | 73.2% | 268 | 75.9% |
| | | cloudy | 2940 | 1950 | 15.6% | 929 | 17.0% | 61 | 17.3% |
| | | rain | 1652 | 1145 | 9.2% | 487 | 8.9% | 20 | 5.7% |
| | | fog_sleet_snow | 171 | 121 | 1.0% | 48 | 0.9% | 2 | 0.6% |
| | | other_unknown | 25 | 17 | 0.1% | 6 | 0.1% | 2 | 0.6% |
| | lighting_condition | daylight | 14616 | 10056 | 80.6% | 4307 | 78.6% | 253 | 71.7% |
| | | dark_no_streetlight | 1859 | 1189 | 9.5% | 615 | 11.2% | 55 | 15.6% |
| | | dark_with_streetlight | 1373 | 925 | 7.4% | 412 | 7.5% | 36 | 10.2% |
| | | dusk_dawn | 415 | 268 | 2.1% | 139 | 2.5% | 8 | 2.3% |
| | | other_unknown | 37 | 32 | 0.3% | 4 | 0.1% | 1 | 0.3% |
| | location_type | business_mixed_residential | 4070 | 2770 | 22.2% | 1253 | 22.9% | 47 | 13.3% |
| | | residential | 2990 | 1743 | 14.0% | 1142 | 20.9% | 105 | 29.7% |
| | | business_industrial | 6853 | 5074 | 40.7% | 1707 | 31.2% | 72 | 20.4% |
| | | open_country | 3577 | 2324 | 18.6% | 1138 | 20.8% | 115 | 32.6% |
| | | other_locality | 810 | 559 | 4.5% | 237 | 4.3% | 14 | 4.0% |

∗PDO = Property Damage Only, IN = Injury, FS = Fatal or severe

| Id | lighting_condition |
|---|---|
| 1 | dark_with_streetlight |
| 2 | daylight |
| 3 | daylight |
| 4 | dark_no_streetlight |
| 5 | other/unknown |
| 6 | dusk/dawn |
| 7 | daylight |

One Hot Encoding

| Id | dark_with_streetlight | daylight | dusk/dawn | dark_no_streetlight | other/unknown |
|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 | 0 |
| 3 | 0 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 0 |
| 5 | 0 | 0 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 | 0 | 0 |
| 7 | 0 | 1 | 0 | 0 | 0 |

**Figure 3.** Creation of one hot encoded matrix from the original crash dataset.

19 variables into 87 variable categories (87 columns) which are used as independent variables in the XGBoost model. In this technique, two major steps are performed: 1) create a new binary feature for each possible category and 2) assign a value of '1' to the feature of each sample that corresponds to its original category. The following Figure 3 illustrates the creation of one hot encoded matrix from the original crash data.

### 5.2. *XGBoost model performance*

A critical stage in the XGBoost model development involves evaluating its performance through the visualisation of learning curves (Figure 4). This step aims to prevent overfitting by pinpointing the 'inflection point' where the model's performance on the test dataset begins to decline (the curve remains flat or goes upward), while its performance on the training dataset continues to improve (the curve follows a downward direction). Initially, the model was run for a total of 1,500 iterations. The learning curves, depicting the logarithmic loss of the XGBoost model for each iteration on both the training and test datasets, are illustrated in Figure 5. Note that, logarithmic loss is the price paid for the inaccuracy of predictions in classification problems. At the 1271st iteration (marked by a black dotted line), the test dataset exhibited the lowest logarithmic loss value of 0.624771. Consequently, the final XGBoost model was run up to the 1271st iteration to achieve optimal results. The classification accuracy of this model was determined to be 70.96% and the corresponding confusion matrix is provided (Figure 5).

### 5.3. *Variable feature selection*

The obtained XGBoost model results in terms of 'gain value' for the top 25 variable categories are provided in the following Figure 6. The gain value denotes the relative contribution of the associated feature to the model, which is computed by dividing the contribution of each feature by the number of trees in the model. A higher gain value suggests that it is more significant for making predictions. For example, 'manner of collision = sideswipe' (gain value = 0.1095) was identified as the topmost important predictor for classifying crash injury severity involving older drivers on roadways having a posted speed limit of 45 mph or higher.

Based on the obtained results by XGBoost algorithms, the following variables were selected for the random parameter models (listed in Table 4). Note that, a few of the 'other'

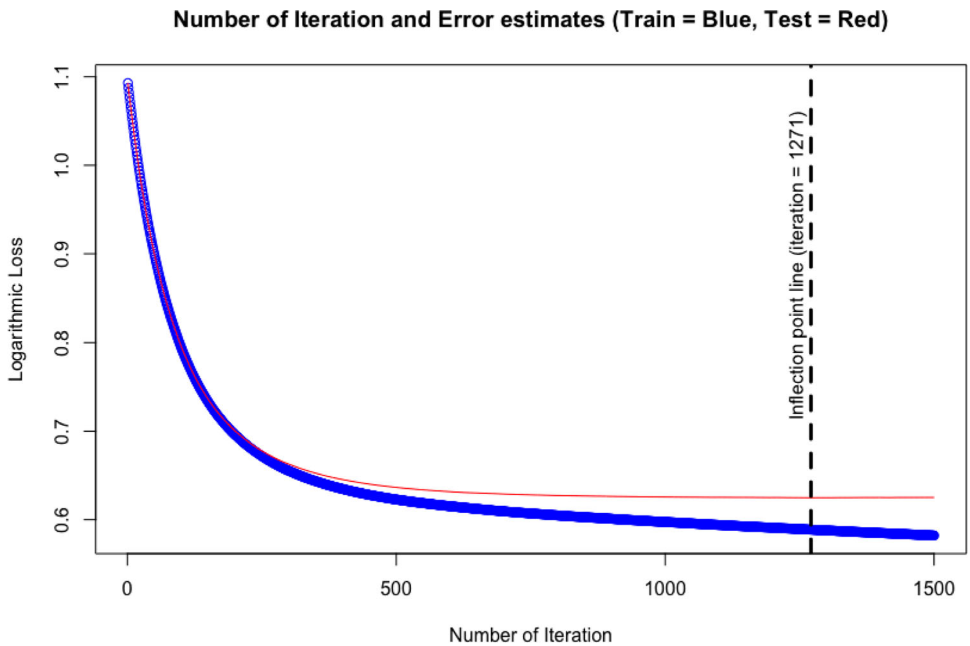


**Figure 4.** XGBoost Model iteration details and error estimates.


**Figure 5.** Confusion Matrix.

and 'unknown' variable categories were not selected to improve the interpretability of the random parameter models. The variables were coded in binary format: 1 for the presence of that specific factor and 0 for the absence of that specific factor.

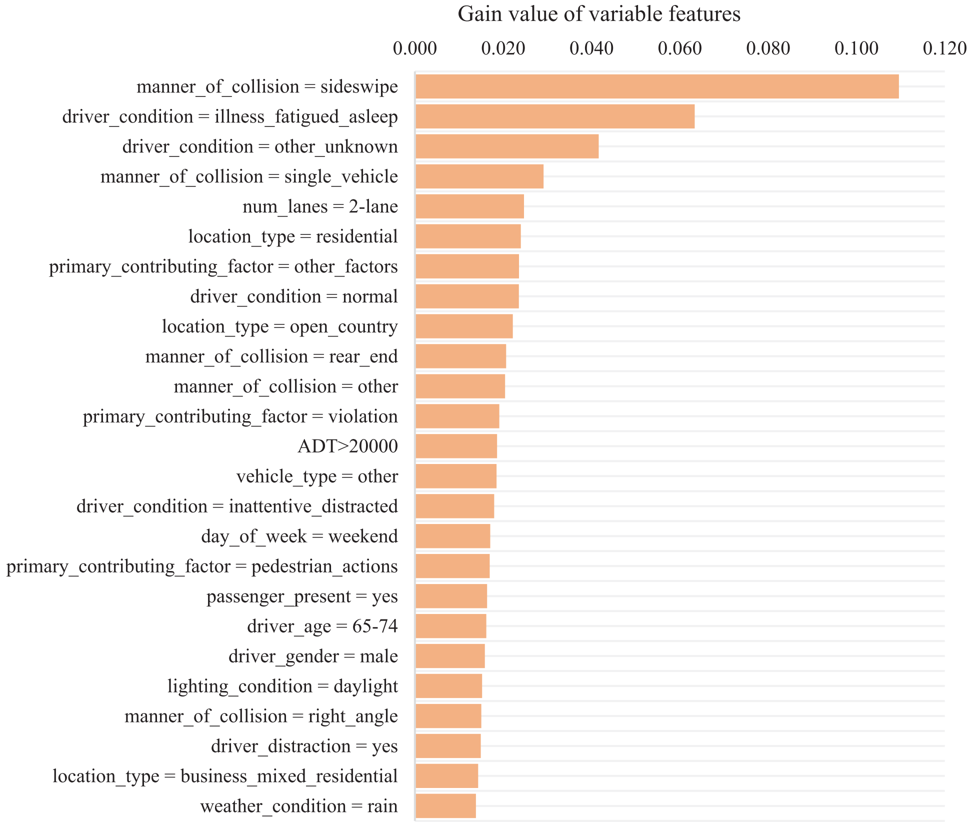


**Figure 6.** Variable feature selection by XGBoost algorithm.

**Table 4.** Details of Variables Selected for Random Parameter Model.

| # | Coding of variables | Interpretation | Mean | SD |
|---|---|---|---|---|
| 1 | crash injury severity | Dependent variable in the model | 0.338 | 0.512 |
| 2 | m_side_swipe | 1 if manner of collision = sideswipe, 0 otherwise | 0.232 | 0.422 |
| 3 | d_condition_ifa | 1 if driver condition = Illness/Fatigued/Asleep, 0 otherwise | 0.042 | 0.200 |
| 4 | m_single_vehicle | 1 if manner of collision = single vehicle, 0 otherwise | 0.205 | 0.404 |
| 5 | lanes_two | 1 if number of lanes = 2, 0 otherwise | 0.295 | 0.456 |
| 6 | location_r | 1 if location = residential area, 0 otherwise | 0.163 | 0.370 |
| 7 | d_condition_normal | 1 if driver condition = normal, 0 otherwise | 0.226 | 0.418 |
| 8 | location_oc | 1 if location = open country, 0 otherwise | 0.195 | 0.397 |
| 9 | m_rear_end | 1 if manner of collision = rear-end, 0 otherwise | 0.319 | 0.466 |
| 10 | pcf_violation | 1 primary contributing factor = violation, 0 otherwise | 0.751 | 0.432 |
| 11 | ADT_20000 | 1 if ADT > 20,000, 0 otherwise | 0.496 | 0.500 |
| 12 | d_condition_id | 1 if driver condition is inattentive/distracted, 0 otherwise | 0.664 | 0.472 |
| 13 | dow_weekend | 1 if day of the week = weekend, 0 otherwise | 0.191 | 0.393 |
| 14 | pcf_ped_actions | 1 if primary contributing factor = pedestrian action, 0 otherwise | 0.003 | 0.059 |
| 15 | passenger_present | 1 if a passenger is present, 0 otherwise | 0.200 | 0.400 |
| 16 | d_age_65_to_74 | 1 if driver age = 65-74, 0 otherwise | 0.671 | 0.470 |
| 17 | d_gender_male | 1 driver gender = male, 0 otherwise | 0.612 | 0.487 |
| 18 | light_daylight | 1 if lighting condition = daylight, 0 otherwise | 0.799 | 0.401 |
| 19 | m_right_angle | 1 if manner of collision = right angle, 0 otherwise | 0.071 | 0.257 |
| 20 | d_distraction | 1 if the driver is distracted, 0 otherwise | 0.672 | 0.469 |
| 21 | location_br | 1 if location = business/residential area, 0 otherwise | 0.222 | 0.416 |
| 22 | weather_rain | 1 weather condition = rain, 0 otherwise | 0.090 | 0.287 |

In the next step, crash injury severity (FS, IN, PDO) was set as the dependent variable, and 21 variable categories (details in Table 3) were set as the independent variable in the model. In the process of model building, we employed a stepwise approach for model development. Firstly, we build an ordered probit (OP) model, incorporating significant variables determined through a backward elimination process. This initial model serves as the foundation, with all terms considered in the OP model being statistically significant. It is essential to highlight that throughout the entire process of model building, selection, and comparison, a significance level of 90% was consistently employed. Subsequently, we construct a random parameter model on top of the ordered probit model. This involves testing all parameters in the model for randomness to identify the best-fitting model with random parameters. Once the optimal random parameter ordered probit model is established, we proceed to test all parameters in the dataset to identify instances of heterogeneity in means within the random parameter framework. For models featuring multiple random parameters, correlated models were also explored. The goodness of fit for these models was assessed using AIC, AIC/N, McFadden Pseudo R-Squared, and the likelihood ratio test throughout the process (see Appendix Table A1, A2 for more details).

Among the four models, the CRPOP-HM model exhibited superior performance, leading to its selection for further analysis. The measure AIC/N (AIC value normalised by sample size) was found to be the same for RPOP-HM and CRPOP-HM; thus, the best model was selected based on the lowest AIC value of 23562.1 (CRPOP-HM model). In the final model, setting the variables 'manner of collision = sideswipe', driver condition = normal', and 'location = open country', as random parameters yield significant model results. In the next step, the other variables were checked (trial-and-error) if they affected the 'means' of the random parameter in a significant way. The research team found three cases, where a certain variable affects the means of random parameters:

- 'Manner of collision' affecting the mean of 'driver condition = normal'
- 'ADT' affecting the mean of the variable 'location = open country'
- Variable 'lanes = two' affecting the mean of the variable 'manner of collision = side swipe'

Following the identification of the CRPOP-HM model, a comprehensive examination of all parameters in the dataset was conducted to assess their impact on the variances of the random parameter. However, no models were found to have a significant impact on the variances of the random parameters. Consequently, the 'heterogeneity in variance' context was not considered in this study. The obtained model results are provided in Table 5.

In the next step, the marginal effects of key variables in terms of the CRPOP-HM model were calculated and reported in Table 6. Note that marginal effects show the likelihood of crash injury severity (FS, IN, PDO separately) involving older drivers on roadways having a posted speed limit of 45 mph or higher.

### 5.4. Explanation of random parameters

Interpretation of random parameters is the most critical step, and it helps to identify the complex characteristics of variables and how they affect the crash involvement of older drivers on high-speed roadways. Given the outcomes of previous studies, which

**Table 5.** CRPOP-HM Model Results.

| Variable | Coding | Model Statistics $\beta$ | t-value |
|---|---|---|---|
| *Manner of collision = sideswipe* | *m_side_swipe* | *−0.82* | *−21.28* |
| | *standard deviation of m_side_swipe* | *0.06* | *2.24* |
| Driver condition = inattentive/fatigued/asleep | d_condition_ifa | 0.18 | 3.36 |
| Manner of collision = single vehicle | m_single_vehicle | 0.21 | 6.02 |
| Number of lanes = 2 | lanes_two | - | - |
| Location type = residential | location_r | 0.26 | 8.11 |
| *Driver condition = normal* | *d_condition_normal* | *−0.68* | *−18.11* |
| | *standard deviation of d_condition_normal* | *0.53* | *21.97* |
| *Location type = Open country* | *location_oc* | *0.33* | *8.04* |
| | *standard deviation of location_oc* | *0.33* | *14.07* |
| Manner of collision = Rear end | m_rear_end | 0.07 | 2.71 |
| Primary contributing factor = violation | pcf_violation | 0.08 | 3.11 |
| ADT > 20,000 | ADT_20000 | −0.06 | −2.29 |
| Driver condition = inattentive/distracted | d_condition_id | −0.61 | −21.13 |
| Day of the week = weekend | dow_weekend | 0.08 | 3.23 |
| Primary contributing factor = pedestrian actions | pcf_ped_actions | 2.11 | 14.52 |
| Passenger present = Yes | passenger_present | 0.11 | 4.45 |
| Driver age = 65–74 years | d_age_65_to_74 | −0.08 | −4.09 |
| Manner of collision = right angle | m_right_angle | 0.27 | 6.71 |
| Location type = business/residential | location_br | 0.10 | 3.67 |
| Weather condition = rain | weather_rain | −0.07 | −2.00 |
| Variables affecting the means of random parameter | Effect of 'two-lane roadway' on the mean of 'manner of collision = sideswipe' | 0.427 | 6.16 |
| | Effect of 'manner of collision = single vehicle on means of 'driver condition = normal' | −0.411 | −7.54 |
| | Effect of ADT > 20,000 on the mean of 'location = open country' | −0.276 | −5.38 |
| Model statistics | LL (0) | −12784.038 | |
| | Log-likelihood at convergence, LL($\beta$) | −11754.0535 | |
| | Akaike Information Criterion (AIC) | 23562.1 | |
| | AIC/N | 1.288 | |
| | McFadden Pseudo R-Squared | 0.0806 | |

*Note: Random parameters are marked in gray colour in the table

demonstrated that aligning with a normal distribution statistically provides a better fit, we opted for a normal distribution for the random parameters in our model. This choice is rooted in the observed statistical performance favouring the normal distribution in similar contexts (Al-Bdairi, Behnood, and Hernandez 2020; Behnood and Mannering 2015; 2016; Shaheed et al. 2013). For the normal distribution of the random parameters, the probability that the distribution is less than zero can be calculated using the common standard normal

**Table 6.** The marginal effect of key significant variables (Model: CRPOP-HM).

| Type | Variable | PDO | IN | FS |
|---|---|---|---|---|
| Crash Factors | ***Manner of collision = sideswipe | 0.2463 | −0.2333 | −0.0130 |
| | Manner of collision = rear end | −0.0263 | 0.0245 | 0.0018 |
| | Manner of collision = right angle | −0.0991 | 0.0906 | 0.0084 |
| | Manner of collision = single vehicle | −0.0750 | 0.0692 | 0.0057 |
| | Primary contributing factor = violation | −0.0264 | 0.0247 | 0.0017 |
| | Primary contributing factor = pedestrian action | −0.6441 | 0.2597 | 0.3843 |
| Driver & Occupant | ***Driver condition = normal | 0.2088 | −0.1976 | −0.0112 |
| | Driver condition = illness/Fatigued/Asleep | 0.0601 | 0.0052 | – |
| | Driver condition = inattentive/distracted | 0.2214 | −0.2019 | −0.0195 |
| | Driver age = 65–74 years | 0.0293 | −0.0273 | −0.0020 |
| | Passenger presence | −0.0399 | 0.0370 | 0.0029 |
| Spatial, Temporal & Environmental Factors | ***Location type = open country | −0.1202 | 0.1102 | 0.0100 |
| | Location type = residential | −0.0926 | 0.0852 | 0.0074 |
| | Location = business with mixed residential | – | 0.0318 | 0.0024 |
| | Day of the week = weekend | −0.0283 | 0.0263 | 0.0020 |
| | Weather condition = rain | 0.0241 | −0.0225 | −0.0015 |
| Traffic | ADT > 20,000 | 0.0202 | −0.0188 | −0.0014 |

Orange colour means increasing the likelihood of a specific injury severity class; no colour means decreasing the likelihood of a specific injury severity class; *** specifies random parameters

**Table 7.** Explanation of Random Parameters.

| Variable Description | μ & σ Significant? | μ | σ | % of distribution below zero | % of distribution above zero |
|---|---|---|---|---|---|
| 1 if manner of collision = sideswipe, 0 otherwise | Yes | −0.82 | 0.06 | 100.00% | 0% |
| 1 if driver condition = normal, 0 otherwise | Yes | −0.68 | 0.53 | 90.03% | 9.97% |
| 1 if crash location = open country, 0 otherwise | Yes | 0.33 | 0.33 | 15.87% | 84.13% |
| Variables affecting means of random parameter | | | | | |
| Effect of 'two-lane roadway' on the mean of 'manner of collision = sideswipe' | | | | $\beta = 0.427$ | ↑ |
| Effect of 'manner of collision = single vehicle on means of 'driver condition = normal' | | | | $\beta = -0.411$ | ↓ |
| Effect of ADT > 20,000 on the mean of 'location = open country' | | | | $\beta = -0.276$ | ↓ |

distribution (Z-formula):

$$Zscore = \frac{Zero - \mu}{\sigma} \tag{12}$$

Here $\mu$ is the mean parameter estimate, and $\sigma$ is the estimate's standard deviation from the mean. The mean parameter estimate in the equation represents the estimated value of the random parameter in the CRPOP-HM model. After calculating the Z-value, the percentage of the distribution being less than zero or greater than zero can be determined from the Z-tables (more details in Table 7).

Overall, the model results are summarised under the following sub-headings.

### 5.5. Crash-related factors

The variable 'manner of collision = sideswipe' was identified as a random parameter with an estimated mean of −0.82 and a standard deviation of 0.06, suggesting that 100% of this distribution is below zero. This implies that almost all older drivers are less likely to be involved in sideswipe collisions resulting in FS injury on high-speed roadways. This is also supported by the marginal effects results: sideswipe crashes involving older drivers on high-speed settings were found to increase the likelihood of PDO crashes (0.2463), but less likely to result in IN (−0.2333) or FS crashes (−0.013). This is consistent with a prior investigation (Adanu et al. 2021). Sideswipe crashes often occur when drivers fail to properly judge the distance between their vehicle and nearby objects or vehicles. As people age, their vision tends to decline, leading to reduced depth perception and peripheral vision increasing the chances of being involved in such crashes and making it a common crash experience for older drivers. In contrast, a previous study suggested that crossing centreline or median crashes due to sideswipe collisions were less likely to occur in all young driver age groups (15-16, 17-19, and 20–24 years) (Rahman et al. 2021).

Interestingly, the mean of the indicator variable 'manner of collision = sideswipe' was found to be shifted upward (0.427) by the presence of a 'two-lane roadway'. This suggests a higher likelihood of FS injury when older drivers are involved in side-swipe collisions on two-lane high-speed roadways. Two-lane roadways often have narrower lanes and limited space, which leaves little room for drivers to maneuver and avoid a collision. In high-speed situations, older drivers have less time to react to sudden movements or obstacles, increasing the likelihood of sideswipe collisions and FS injury outcomes.

A different pattern was observed if the manner of collision is rear-ended, right angle, or single-vehicle all of which increased the likelihood of IN crashes (rear-end = 0.0245, right-angle = 0.0906, single vehicle = 0.0692) or FS crashes (rear-end = 0.0018, right-angle = 0.0084, single vehicle = 0.0057) involving older drivers on high-speed segments. This is intuitive that, involving a rear-end/right-angle/single-vehicle crash on high-speed segments is expected to increase the risk of serious injuries. High-speed roadways often require handling complex driving tasks, such as navigating interchanges, merging, and changing lanes safely. Older drivers struggle with multitasking or managing these intricate maneuvers, increasing the likelihood of single-vehicle crashes (Kim et al. 2013). The study also reported that older drivers have a higher risk of suffering fatal injuries in single-vehicle crashes compared to young (16-24 years) or working age group (25-64 years) drivers.

Older drivers may struggle to maintain situational awareness, leading to traffic rule violations (Scott-Parker et al. 2020). When crashes occur due to violations of older drivers (not complying with traffic rules and regulations), these are more likely to result in IN (0.0247) or FS (0.0017) crashes on high-speed roadways. It was interesting to notice that, if the primary contributing factor is 'pedestrian action', those crashes were more likely to result in FS crashes (0.3843, highest among all other marginal effects), or IN crashes (0.2597, highest among all other marginal effects). The involvement of older drivers in pedestrian crashes on high-speed roadways is identified by a previous study (Hossain et al. 2023). Older drivers are less likely to perform evasive maneuvers to abrupt pedestrian movements on high-speed roadways due to less reaction time and more stopping sight distance.

### 5.6. Driver and occupant-related factors

If the older driver's condition is normal (alert to the task of driving), their involvement in crashes on high-speed segments is more likely to result in PDO crashes (0.2088) and less likely to result in IN (−0.1976) or FS crashes (−0.0112). The variable 'driver condition = normal' was identified as a random parameter with an estimated mean of −0.68 and a standard deviation of 0.53, suggesting that 90.03% of this distribution is below zero and only 9.97% of this distribution is above zero. This implies that most of the older drivers in normal physical conditions are less likely (90.03%) to be involved in FS crashes on high-speed roadways, and only 9.97% of the older drivers in normal physical conditions have higher odds of being involved in FS injury crashes on high-speed roadways. Interestingly, the mean of the indicator variable 'driver condition = normal' was found to be shifted downward (−0.411) by the presence of 'manner of collision = single vehicle'. This suggests a lower odds of FS injury when older drivers are involved in single-vehicle collision on high-speed roadways in 'normal' physical conditions. This is consistent with a previous investigation (Zhang et al. 2000).

However, if the older driver's physical condition is categorised as 'illness/fatigued/asleep', in such cases, their involvement in crashes on high-speed roadways is more likely to result in IN (0.0052) or PDO crashes (0.0601). According to a prior study by Smolensky et al., 'nodding-off' related crashes are most frequent in older drivers (during afternoon hours) highlighting the significance of excessive daytime fatigue and sleepiness in these driving incidents (Smolensky et al. 2011). It was interesting to notice that, if an older driver's physical condition is 'inattentive/distracted', their involvement in crashes on high-speed roadways is more likely to result in PDO crashes (0.2214) but less likely to result in IN (−0.2019) or FS crashes (−0.0195). This is consistent with a previous investigation that suggests that older drivers try to compensate for their driving performance while being distracted as they feel more vulnerable compared to young or middle-aged drivers (Papantoniou et al. 2016).

In this study, the older drivers were divided into three age groups: 65–74 years (12,283 crashes), 75–84 years (4,951 crashes), and 85+ years (1,099 crashes). Since the age group 65–74 years drivers are expected to be more physically stable (compared to the 74–84 or 85+ years age group), their involvement in crashes on high-speed roadways is more likely to result in PDO crashes (0.0293) and less likely to result in IN (−0.0273) or FS (−0.002) crashes. The presence of passengers with older drivers is more likely to result in IN (0.037) or FS (0.0029) crashes but less likely to result in PDO crashes (−0.0399). This is contrary to the findings of a few of the previous studies which generally imply that the presence of passengers was found to be beneficial to older drivers (Braitman, Chaudhary, and McCartt 2014; Hing, Stamatiadis, and Aultman-Hall 2003). However, the scenario could be different taking high-speed roadways and associated risky driving situations into account. In similar high-speed roadways (e.g. freeways), young drivers (16-24 years) tend to be careless and exhibit reckless driving behaviour when they are accompanied by their peers (Lee and Abdel-Aty 2008).

### 5.7. Spatial factors

The research revealed several intriguing insights into several spatial and temporal factors and how they affect older driver-involved crashes on high-speed roadways. For instance, older drivers have a higher risk of IN (0.1102) or FS collisions (0.01) but a lower risk of PDO crashes (−0.1202) when involved in crashes in open rural high-speed settings. The variable 'location = open country' was identified as a random parameter with an estimated mean of 0.33 and standard deviation of 0.33, suggesting that only 15.87% of this distribution is below zero and 84.13% of this distribution is above zero. This implies that most of the older drivers (84.13%) in high-speed open country locations have higher odds of being involved in FS crashes, and only a small proportion (15.87%) of the older drivers in high-speed open country locations have lower odds of being involved in FS injury crashes. This is consistent with a previous investigation (Rahman, Das, and Sun 2023). Open country locations typically involve longer travel distances between destinations. Longer trips can lead to hypovigilance (highly predictable and uneventful driving tasks) in older drivers, which can impair driving performance and increase the risk of severe crashes. In addition to the physical vulnerability of older drivers, the longer EMS response times in a rural crash are more likely to increase the severity of crashes in such settings (Gonzalez et al. 2007). In contrast, the 20–45 years age group drivers have higher probabilities of 'no injury' when they were involved in crashes in ROR (run-off-road) crashes in rural areas which can be attributed to their physiological capabilities (Al-Bdairi and Hernandez 2020).

Interestingly, the mean of the indicator variable 'location = open country' was found to be shifted downward (−0.276) by the presence of ADT greater than 20,000. This suggests lower odds of FS injury when older drivers are involved in crashes on high-speed open country locations when average daily traffic is more than 20,000. In high-traffic conditions, older drivers tend to be more alert and aware of their surroundings. The presence of more vehicles and potential hazards on the road requires heightened attention, which can lead to reducing the likelihood of being involved in a crash in high-traffic conditions. In residential areas, crashes involving older drivers are more likely to result in IN (0.0852) or FS (0.0074) crashes but less likely to result in PDO crashes (−0.0926). Similar results are obtained for businesses with mixed residential areas as well (IN = 0.0318, FS = 0.0024).

## *5.8. Temporal and environmental factors*

Older driver-involved crashes on high-speed roadways occurring on weekends were found to increase the likelihood of FS (0.0020), or IN (0.0263) crashes but decreased the likelihood of PDO crashes (−0.0283). This is consistent with the findings reported in a previous study (Islam, Hossain, and Shaban 2023). Older driver involvement in crashes on weekends potentially suggests their driving routine and patterns. Older drivers may use weekends to visit friends, and family, or take longer trips, leading to increased exposure to potential crash scenarios. Older drivers are expected to avoid driving in adverse weather conditions (e.g. rainy) and tend to be more cautious while driving in such situations (Abdel-Aty, Chen, and Essam Radwan 1999). That is why, older driver's involvement in crashes in rainy weather conditions is more likely to result in PDO crashes (0.0241) but less likely to result in IN (−0.0225) or FS crashes (−0.0015). Similar findings are reported in another research focusing on young drivers ( $<$ 25 years), who were associated with a lower risk of serious injuries and fatalities in crashes under rainy weather conditions possibly due to lower average speed in such environments (Li et al. 2019).

### 5.9. Traffic characteristics

Average Daily Traffic (ADT) was found as a significant predictor of older driver-involved crashes on high-speed segments. According to the model, if ADT is greater than 20,000, older drivers' involvement in crashes is more likely to result in PDO crashes (0.0202) and less likely to result in IN (−0.0188) or FS crashes (−0.0014). This is consistent with a previous study (Chang and Xiang 2003). The investigation reported that crashes that occur during more congested traffic conditions tend to be at a less severe level; as such traffic conditions increase driver awareness and/or encourage more cautious driving.

## 6. Conclusions

To keep older drivers mobile without compromising safety, it is critical to promote safe driving practices. Using the most recent 5-year crash data (2017-2021) collected from Louisiana state, the current study investigated the driving challenges of older drivers on roadways having posted speed limits of 45 mph or higher. The research team explored a two-step hybrid approach with the combination of machine learning (XGBoost) and statistical models (random parameter approach). The XGBoost model was used to classify top variable features in binary format (1,0) and later used in the random parameter models. The choice of random parameter modelling technique helped to address the 'unobserved heterogeneity' (i.e. the presence of individual-specific characteristics or factors that are not directly measured) in the crash data. Among several candidate random parameter models, the Correlated Random Parameter Ordered Probit with Heterogeneity in Means (CRPOP-HM) performed better, and later marginal effects of significant key variables were reported. Some of the major findings are summarised below:

- Manner of collision rear end, right angle, or single vehicle found to increase the likelihood of fatal-severe or injury crashes involving older drivers on high-speed segments.
- Due to reduced cognitive skills, older drivers are prone to violation of traffic rules in complex driving environments like high-speed roadways. Such crashes (primary contributing factor = violation) are expected to increase the likelihood of fatal-severe or injury crashes.
- If the primary contributing factor leading to the crash is 'pedestrian action', such crashes are expected to increase the likelihood of 0.3843 in fatal-severe crashes, or a 0.2597 increase in injury crashes involving older drivers. This potentially suggests unique safety challenges faced by older drivers in high-speed settings involving pedestrians.
- The presence of passengers with older drivers on high-speed roadways was found as contributing factor to fatal-severe or injury crashes.
- Older drivers involved in crashes in high-speed open country locations are more likely to result in injury or fatal-severe crashes.
- Temporal factors like the weekend increased the likelihood of fatal-severe or injury crashes involving older drivers on high-speed segments.

### 6.1. Recommended countermeasures

The enhancement of safety on high-speed roads necessitates a holistic strategy that includes infrastructure and technology improvements, educational initiatives, and policies while keeping in mind the physical limitations of older drivers. This research identified the safety issues associated with older drivers on two-lane high-speed highways leading to serious injury outcomes. Adding positive barriers in transition zones (e.g. divided highway to undivided highway, rural to urban areas) and positive separation (channelization) between opposing two-lane traffic is recommended (Staplin 2004). Technology-based solutions, such as the Advanced Driver Assistance System (ADAS), are helpful for older drivers in identifying their travel lane by directing their attention more toward the middle of the road than either left or right (Dotzauer et al. 2013). Specifically, older drivers using Forward Collision Warning (FCW) showed significantly longer time-to-collision (TTC) when approaching the critical event than those who did not. With the rapid development of ADAS, tailoring ADAS technologies specifically for older drivers is also critical (Eby 2023). Focus is required considering the assistance required to overcome potential aging-related declines, and ease of understanding and use of ADAS.

This study identified the safety challenges of pedestrians on high-speed speed roadways and the involvement of older drivers in such settings. The recognition of dynamic targets, including walking or jogging pedestrians with fluorescent-activated retroreflective delineators is significantly effective for all age group drivers including older drivers (Turner, Nitzburg, and Knoblauch 1998). Educational campaigns aimed at addressing the safety of older drivers can be effective in raising awareness, providing information, and promoting safe driving habits. Tailoring these campaigns to the specific needs and challenges faced by older drivers is also essential. For example, it has been discovered that the presence of passengers with older drivers contributes to serious injury crashes in high-speed settings. Education campaigns can be used to raise awareness of these important safety issues. Previous research revealed that older drivers who had participated in Driver Safety Programmes (DSPs) performed better in head movements and braking (Bao and Boyle 2009). For example, the older drivers stopped significantly earlier while approaching intersections and they had significantly more head movements checking for traffic conflicts. A similar investigation is required to be conducted for high-speed settings for further clarification.

### *6.2. Limitations and recommendations*

The study is not without limitations. XGBoost model was used as a classifier to select top variable categories, which can be further explored by comparing its performance with other machine learning classifiers, including Random Forest, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Neural Networks, and so on. This comparative analysis would provide a more comprehensive understanding of the effectiveness of different classifiers in selecting relevant variable categories. The study considered crashes involving older drivers aged 65 years or higher. Within this specific driver age group, three separate age groups were considered including 65-74, 75-84, and 84+ years. Separate models can be developed to see how crash pattern varies within different age groups of older drivers. In addition, the exploration of all driver age groups (young, middle, and older) is expected to provide additional insights into crash involvement in high-speed segments, and this broader-scope study is thus recommended as a future study. The study only considered crashes on high-speed segments, and it did not consider crashes occurring on high-speed intersections (approaching roadway to the intersection have a high posted speed limit like 45 mph or higher). This is another unique driving challenge for older drivers and is recommended for exploration in future studies.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

*Ahmed Hossain* http://orcid.org/0000-0003-1566-3993
*Xiaoduan Sun* http://orcid.org/0000-0001-7282-1340
*Subasish Das* http://orcid.org/0000-0002-1671-2753
*Monire Jafari* http://orcid.org/0009-0002-5298-493X
*Julius Codjoe* http://orcid.org/0000-0003-1958-8695

## References

Aarts, Letty, and Ingrid Van Schagen. 2006. "Driving Speed and the Risk of Road Crashes: A Review." *Accident Analysis & Prevention* 38 (2): 215–224.

Abdel-Aty, Mohamed A., Chien L. Chen, and A. Essam Radwan. 1999. "Using Conditional Probability to Find Driver Age Effect in Crashes." *Journal of Transportation Engineering* 125 (6): 502–507.

Adanu, Emmanuel Kofi, Abhay Lidbe, Elsa Tedla, and Steven Jones. 2021. "Factors Associated with Driver Injury Severity of Lane Changing Crashes Involving Younger and Older Drivers." *Accident Analysis & Prevention* 149: 105867. https://doi.org/10.1016/j.aap.2020.105867.

Al-Bdairi, Nabeel Saleem Saad, Ali Behnood, and Salvador Hernandez. 2020. "Temporal Stability of Driver Injury Severities in Animal-Vehicle Collisions: A Random Parameters with Heterogeneity in Means (and Variances) Approach." *Analytic Methods in Accident Research* 26: 100120.

Al-Bdairi, Nabeel Saleem Saad, and Salvador Hernandez. 2020. "Comparison of Contributing Factors for Injury Severity of Large Truck Drivers in Run-off-Road Crashes on Rural and Urban Roadways: Accounting for Unobserved Heterogeneity." *International Journal of Transportation Science and Technology* 9 (2): 116–127. https://doi.org/10.1016/j.ijtst.2020.01.004.

Al-Ghamdi, Ali S. 2002. "Using Logistic Regression to Estimate the Influence of Accident Factors on Accident Severity." *Accident Analysis & Prevention* 34 (6): 729–741.

Ali, Yasir, Michiel C. J. Bliemer, Zuduo Zheng, and Md. Mazharul Haque. 2020. "Cooperate or Not? Exploring Drivers' Interactions and Response Times to a Lane-Changing Request in a Connected Environment." *Transportation Research Part C: Emerging Technologies* 120: 102816. https://doi.org/10.1016/j.trc.2020.102816.

Ali, Yasir, and Md Mazharul Haque. 2023. "Modelling the Response Times of Mobile Phone Distracted Young Drivers: A Hybrid Approach of Decision Tree and Random Parameters Duration Model." *Analytic Methods in Accident Research* 39: 100279. https://doi.org/10.1016/j.amar.2023.100279.

Ali, Yasir, Fizza Hussain, and Md Mazharul Haque. 2024. "Advances, Challenges, and Future Research Needs in Machine Learning-Based Crash Prediction Models: A Systematic Review." *Accident Analysis & Prevention* 194: 107378. https://doi.org/10.1016/j.aap.2023.107378.

Ali, Yasir, Md. Mazharul Haque, Zuduo Zheng, and Michiel C. J. Bliemer. 2021. "Stop or Go Decisions at the Onset of Yellow Light in a Connected Environment: A Hybrid Approach of Decision Tree and Panel Mixed Logit Model." *Analytic Methods in Accident Research* 31: 100165. https://doi.org/10.1016/j.amar.2021.100165.

Ali, Yasir, Mark P. H. Raadsen, and Michiel C. J. Bliemer. 2023. "Modelling Speed Reduction Behaviour on Variable Speed Limit-Controlled Highways Considering Surrounding Traffic Pressure: A Random Parameters Duration Modelling Approach." *Analytic Methods in Accident Research* 40: 100290. https://doi.org/10.1016/j.amar.2023.100290.

Alnawmasi, Nawaf, Yasir Ali, and Shamsunnahar Yasmin. 2024. "Exploring Temporal Instability Effects on Bicyclist Injury Severities Determinants for Intersection and Non-Intersection-Related Crashes." *Accident Analysis & Prevention* 194: 107339. https://doi.org/10.1016/j.aap.2023.107339.

Alnawmasi, Nawaf, and Fred Mannering. 2019. "A Statistical Assessment of Temporal Instability in the Factors Determining Motorcyclist Injury Severities." *Analytic Methods in Accident Research* 22: 100090.

Alnawmasi, Nawaf, and Fred Mannering. 2023. "An Analysis of Day and Night Bicyclist Injury Severities in Vehicle/Bicycle Crashes: A Comparison of Unconstrained and Partially Constrained Temporal Modeling Approaches." *Analytic Methods in Accident Research* 40: 100301. https://doi.org/10.1016/j.amar.2023.100301.

Alrumaidhi, Mubarak, and Hesham A. Rakha. 2022. "Factors Affecting Crash Severity among Elderly Drivers: A Multilevel Ordinal Logistic Regression Approach." *Sustainability* 14 (18): 11543. https://doi.org/10.3390/su141811543.

Alshehri, Abdulaziz, and Sunanda Dissanayake. 2022. "Comparison of Risk Factors Associated with Older and Other Drivers in Single-Vehicle Fatal Intersection Crashes." 184–96.

Asare, Isaac Ofori, and Alice Constance Mensah. 2020. "Crash Severity Modelling Using Ordinal Logistic Regression Approach." *International Journal of Injury Control and Safety Promotion* 27 (4): 412–419. https://doi.org/10.1080/17457300.2020.1790615.

Ashifur Rahman, M., Subasish Das, and Xiaoduan Sun. 2022. "Using Cluster Correspondence Analysis to Explore Rainy Weather Crashes in Louisiana." *Transportation Research Record: Journal of the Transportation Research Board* 2676 (8): 159–173. https://doi.org/10.1177/03611981221082582.

Asselman, Amal, Mohamed Khaldi, and Souhaib Aammou. 2023. "Enhancing the Prediction of Student Performance Based on the Machine Learning XGBoost Algorithm." *Interactive Learning Environments* 31 (6): 3360–3379.

Azimi, Ghazaleh, Alireza Rahimi, Hamidreza Asgari, and Xia Jin. 2020. "Severity Analysis for Large Truck Rollover Crashes Using a Random Parameter Ordered Logit Model." *Accident Analysis & Prevention* 135: 105355.

Babaei, Zaniar, and Mehmet Metin Kunt. 2024. "A Correlated Random Parameters Ordered Probit Approach to Analyze the Injury Severity of Bicycle-Motor Vehicle Collisions at Intersections." *Accident Analysis & Prevention* 196: 107447.

Ball, Karlene, Cynthia Owsley, Michael E. Sloane, Daniel L. Roenker, and John R. Bruni. 1993. "Visual Attention Problems as a Predictor of Vehicle Crashes in Older Drivers." *Investigative Ophthalmology & Visual Science* 34 (11): 3110–3123.

Bao, Shan, and Linda Ng Boyle. 2009. "Driver Safety Programs: The Influence on the Road Performance of Older Drivers." *Transportation Research Record* 2096 (1): 76–80. https://doi.org/10.3141/2096-10.

Bates, Lyndel J., Jeremy Davey, Barry Watson, Mark J. King, and Armstrong. Kerry. 2014. "Factors Contributing to Crashes among Young Drivers." *Sultan Qaboos University Medical Journal* 14 (3): e297.

Behnood, Ali, and Fred L. Mannering. 2015. "The Temporal Stability of Factors Affecting Driver-Injury Severities in Single-Vehicle Crashes: Some Empirical Evidence." *Analytic Methods in Accident Research* 8: 7–32. https://doi.org/10.1016/j.amar.2015.08.001.

Behnood, Ali, and Fred L. Mannering. 2016. "An Empirical Assessment of the Effects of Economic Recessions on Pedestrian-Injury Crashes Using Mixed and Latent-Class Models." *Analytic Methods in Accident Research* 12: 1–17.

Behnood, Ali, and Fred Mannering. 2017. "The Effect of Passengers on Driver-Injury Severities in Single-Vehicle Crashes: A Random Parameters Heterogeneity-in-Means Approach." *Analytic Methods in Accident Research* 14: 41–53.

Braitman, Keli A., Neil K. Chaudhary, and Anne T. McCartt. 2014. "Effect of Passenger Presence on Older Drivers' Risk of Fatal Crash Involvement." *Traffic Injury Prevention* 15 (5): 451–456.

Braitman, Keli A., Bevan B. Kirley, Susan Ferguson, and Neil K. Chaudhary. 2007. "Factors Leading to Older Drivers' Intersection Crashes." *Traffic Injury Prevention* 8 (3), https://doi.org/10.1080/15389580701272346.

Broberg, T., Lotta Jakobsson, and I. Isaksson-Hellman. 2008. "Pre-Crash Factors Influencing Drivers of Older Ages in Intersection Collisions." 149–62 in Vol. 36. International Research Council on Biomechanics of Injury.

Bureau, US Census. n.d. "Census.Gov." *Census.Gov*. Retrieved April 9, 2023 (https://www.census.gov/en.html).

Caird, Jeff K., S. L. Chisholm, Christopher J. Edwards, and Janet I. Creaser. 2007. "The Effect of Yellow Light Onset Time on Older and Younger Drivers' Perception Response Time (PRT) and Intersection Behavior." *Transportation Research Part F: Traffic Psychology and Behaviour* 10 (5): 383–396.

Caird, Jeff K., Christopher J. Edwards, Janet I. Creaser, and William J. Horrey. 2005. "Older Driver Failures of Attention at Intersections: Using Change Blindness Methods to Assess Turn Decision Accuracy." *Human Factors* 47 (2): 235–249.

Chang, Gang-Len, and Hua Xiang. 2003. *The Relationship between Congestion Levels and Accidents*.

Chen, Huiqin, Libo Cao, and David B. Logan. 2012. "Analysis of Risk Factors Affecting the Severity of Intersection Crashes by Logistic Regression." *Traffic Injury Prevention* 13 (3): 300–307. https://doi.org/10.1080/15389588.2011.653841.

Chen, Feng, and Suren Chen. 2011. "Injury Severities of Truck Drivers in Single- and Multi-Vehicle Accidents on Rural Highways." *Accident Analysis & Prevention* 43 (5): 1677–1688. https://doi.org/10.1016/j.aap.2011.03.026.

Chen, Hengrui, Hong Chen, Zhizhen Liu, Xiaoke Sun, and Ruiyu Zhou. 2020. "Analysis of Factors Affecting the Severity of Automated Vehicle Crashes Using XGBoost Model Combining POI Data." *Journal of Advanced Transportation* 2020: 1–12.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785–794.

Cicchino, Jessica B., and Anne T. McCartt. 2015. "Critical Older Driver Errors in a National Sample of Serious US Crashes." *Accident Analysis & Prevention* 80: 211–219.

Codjoe, Julius, Elisabeta Mitran, Paul Eyram Kornyoh, and Kwabena Abedi. 2021. "Louisiana Transportation Research Center, Louisiana Department of Transportation and Development, and Federal Highway Administration". *Evaluating Pedestrian Crossings on High-Speed Urban Arterials*. Final Report. *Digital/other*.

Das, Subasish, and Srinivas Geedipally. 2020. "Rural Speed Safety Project for USDOT Safety Data Initiative: Findings and Outcomes." *ITE Journal* 90 (9): 38–44.

Dhaliwal, Sukhpreet Singh, Abdullah-Al Nahid, and Robert Abbas. 2018. "Effective Intrusion Detection System Using XGBoost." *Information* 9 (7): 149.

Dissanayake, Sunanda. 2004. "Comparison of Severity Affecting Factors between Young and Older Drivers Involved in Single Vehicle Crashes." *IATSS Research* 28 (2): 48–54. https://doi.org/10.1016/S0386-1112(14)60108-4.

Dissanayake, Sunanda, and Jian John Lu. 2002. "Factors Influential in Making an Injury Severity Difference to Older Drivers Involved in Fixed Object–Passenger Car Crashes." *Accident Analysis & Prevention* 34 (5): 609–618. https://doi.org/10.1016/S0001-4575(01)00060-4.

Dotzauer, Mandy, Simone R. Caljouw, Dick De Waard, and Wiebo H. Brouwer. 2013. "Intersection Assistance: A Safe Solution for Older Drivers?" *Accident Analysis & Prevention* 59: 522–528.

Eby, David W. 2023. *Recommended Advanced Driver Assistance System (ADAS) Technologies for Older Drivers*. *Technical Report*. UMTRI.

Eluru, Naveen, Chandra R. Bhat, and David A. Hensher. 2008. "A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes." *Accident Analysis & Prevention* 40 (3): 1033–1054.

Elvik, Rune, Alena Høye, Truls Vaa, and Michael Sørensen. 2009. "Factors Contributing to Road Accidents." In *The handbook of road safety measures*, 35–80. Leeds: Emerald Group Publishing Limited.

Fountas, Grigorios, and Panagiotis Ch Anastasopoulos. 2017. "A Random Thresholds Random Parameters Hierarchical Ordered Probit Analysis of Highway Accident Injury-Severities." *Analytic Methods in Accident Research* 15: 1–16.

Fountas, Grigorios, and Panagiotis Ch Anastasopoulos. 2018. "Analysis of Accident Injury-Severity Outcomes: The Zero-Inflated Hierarchical Ordered Probit Model with Correlated Disturbances." *Analytic Methods in Accident Research* 20: 30–45.

Fountas, Grigorios, Achille Fonzone, Adebola Olowosegun, and Clare McTigue. 2021. "Addressing Unobserved Heterogeneity in the Analysis of Bicycle Crash Injuries in Scotland: A Correlated Random Parameters Ordered Probit Approach with Heterogeneity in Means." *Analytic Methods in Accident Research* 32: 100181.

Gonzalez, Richard P., Glenn R. Cummings, Herbert A. Phelan, Shanna Harlin, Madhuri Mulekar, and Charles B. Rodning. 2007. "Increased Rural Vehicular Mortality Rates: Roadways With Higher Speed Limits or Excessive Vehicular Speed?" *Journal of Trauma and Acute Care Surgery* 63 (6): 1360. https://doi.org/10.1097/TA.0b013e31815b83b3.

Goswamy, Amrita, Mohamed Abdel-Aty, and Zubayer Islam. 2023. "Factors Affecting Injury Severity at Pedestrian Crossing Locations with Rectangular RAPID Flashing Beacons (RRFB) Using XGBoost and Random Parameters Discrete Outcome Models." *Accident Analysis & Prevention* 181: 106937. https://doi.org/10.1016/j.aap.2022.106937.

Guo, Manze, Zhenzhou Yuan, Bruce Janson, Yongxin Peng, Yang Yang, and Wencheng Wang. 2021. "Older Pedestrian Traffic Crashes Severity Analysis Based on an Emerging Machine Learning XGBoost." *Sustainability* 13 (2): 926. https://doi.org/10.3390/su13020926.

Hing, Jason Yaw Cheuk, Nikiforos Stamatiadis, and Lisa Aultman-Hall. 2003. "Evaluating the Impact of Passengers on the Safety of Older Drivers." *Journal of Safety Research* 34 (4): 343–351.

Hossain, Ahmed, Xiaoduan Sun, Niaz Mahmud Zafri, and Julius Codjoe. 2023. "Investigating Pedestrian Crash Patterns at High-Speed Intersection and Road Segments: Findings from the Unsupervised Learning Algorithm." *International Journal of Transportation Science and Technology* S2046043023000369), https://doi.org/10.1016/j.ijtst.2023.04.007.

Hsiao, Cheng, and M. Hashem Pesaran. 2008. "Random Coefficient Models." In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice, Advanced Studies in Theoretical and Applied Econometrics*, edited by L. Mátyás, and P. Sevestre, 185–213. Berlin, Heidelberg: Springer.

Islam, Samantha, Akhter B. Hossain, and Mohamed Shaban. 2023. "Older Driver At-Fault Crashes at Unsignalized Intersections in Alabama: Injury Severity Analysis with Supporting Evidence from a Deep Learning Based Approach." *Journal of Safety Research* 85: 419–428. https://doi.org/10.1016/j.jsr.2023.04.009.

Ivan, J. N., and K. C. Konduri. 2018. "Crash Severity Methods." Safe Mobility: Challenges, Methodology and Solutions.

Jalayer, Mohammad, Ramin Shabanpour, Mahdi Pour-Rouholamin, Nima Golshani, and Huaguo Zhou. 2018. "Wrong-Way Driving Crashes: A Random-Parameters Ordered Probit Analysis of Injury Severity." *Accident Analysis & Prevention* 117: 128–135.

Jomnonkwao, Sajjakaj, Thanapong Champahom, Chamroeun Se, Natthaporn Hantanong, and Vatanavongs Ratanavaraha. 2023. "Contributing Factors to Near-Miss Experiences of Motorcyclists in Thailand: A Random Parameter Probit Model Approach." *Heliyon* 9 (12).

Kim, Joon-Ki, Gudmundur F. Ulfarsson, Sungyop Kim, and Venkataraman N. Shankar. 2013. "Driver-Injury Severity in Single-Vehicle Crashes in California: A Mixed Logit Analysis of Heterogeneity Due to Age and Gender." *Accident Analysis & Prevention* 50: 1073–1081. https://doi.org/10.1016/j.aap.2012.08.011.

Kockelman, Kara Maria, and Young-Jun Kweon. 2002. "Driver Injury Severity: An Application of Ordered Probit Models." *Accident Analysis & Prevention* 34 (3): 313–321.

Kononen, Douglas W., Carol AC Flannagan, and Stewart C. Wang. 2011. "Identification and Validation of a Logistic Regression Model for Predicting Serious Injuries Associated with Motor Vehicle Crashes." *Accident Analysis & Prevention* 43 (1): 112–122.

Lee, Chris, and Mohamed Abdel-Aty. 2008. "Presence of Passengers: Does It Increase or Reduce Driver's Crash Potential?" *Accident Analysis & Prevention* 40 (5): 1703–1712. https://doi.org/10.1016/j.aap.2008.06.006.

Li, Zhenning, Yusheng Ci, Cong Chen, Guohui Zhang, Qiong Wu, Zhen (Sean) Qian, Panos D. Prevedouros, and David T. Ma. 2019. "Investigation of Driver Injury Severities in Rural Single-Vehicle Crashes under Rain Conditions Using Mixed Logit and Latent Class Models." *Accident Analysis & Prevention* 124: 219–229. https://doi.org/10.1016/j.aap.2018.12.020.

Li, Jing, Shouen Fang, Jingqiu Guo, Ting Fu, and Min Qiu. 2021. "A Motorcyclist-Injury Severity Analysis: A Comparison of Single-, Two-, and Multi-Vehicle Crashes Using Latent Class Ordered Probit Model." *Accident Analysis & Prevention* 151: 105953. https://doi.org/10.1016/j.aap.2020.105953.

Ling Suen, S., and C. G. B. Mitchell. 1998. "The Value of Intelligent Transport Systems to Elderly and Disabled Travellers.".

Lombardi, David A., William J. Horrey, and Theodore K. Courtney. 2017. "Age-Related Differences in Fatal Intersection Crashes in the United States." *Accident Analysis and Prevention* 99: 20–29. https://doi.org/10.1016/j.aap.2016.10.030.

Ma, Jun, Yuexiong Ding, Jack CP Cheng, Yi Tan, Vincent JL Gan, and Jingcheng Zhang. 2019. "Analyzing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective." *IEEE Access* 7: 148059–72.

Mannering, Fred, Chandra R. Bhat, Venky Shankar, and Mohamed Abdel-Aty. 2020. "Big Data, Traditional Data and the Tradeoffs between Prediction and Causality in Highway-Safety Analysis." *Analytic Methods in Accident Research* 25: 100113.

Mannering, Fred L., Venky Shankar, and Chandra R. Bhat. 2016. "Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data." *Analytic Methods in Accident Research* 11: 1–16.

McGwin, Gerald, and David B. Brown. 1999. "Characteristics of Traffic Crashes among Young, Middle-Aged, and Older Drivers." *Accident Analysis & Prevention* 31 (3): 181–198. https://doi.org/10.1016/S0001-4575(98)00061-X.

Meyer, Denny, Samuel Muir, Sampathawaduge Sandun Malpriya Silva, Reneta Slikboer, Allison McIntyre, Kelly Imberger, and Victoria Pyta. 2021. "Modelling the Relationship of Driver License and Offense History with Fatal and Serious Injury (FSI) Crash Involvement." *Journal of Safety Research* 79: 83–93. https://doi.org/10.1016/j.jsr.2021.08.008.

Mohanty, Malaya, and Ankit Gupta. 2015. "Factors Affecting Road Crash Modeling." *Journal of Transport Literature* 9: 15–19.

Mousa, Saleh R., Peter R. Bakhit, and Sherif Ishak. 2019. "An Extreme Gradient Boosting Method for Identifying the Factors Contributing to Crash/near-Crash Events: A Naturalistic Driving Study." *Canadian Journal of Civil Engineering* 46 (8): 712–721.

Mueller, Kim, Shauna L. Hallmark, Huaiqing Wu, and Michael Pawlovich. 2007. "Impact of Left-Turn Phasing on Older and Younger Drivers at High-Speed Signalized Intersections." *Journal of Transportation Engineering* 133 (10): 556–563.

NHTSA. 2021. "Older Driver Traffic Safety Facts 2021 Data | National Highway Traffic Safety Administration.".

Nilsson, G. 1981. "The Effects of Speed Limits on Traffic Accidents." OECD Proceedings of Symposium on the Effects of Speed Limits on Traffic Accidents & Transport Energy Use.

Oxley, Jennifer, Brian Fildes, Bruce Corben, and Jim Langford. 2006. "Intersection Design for Older Drivers." *Transportation Research Part F: Traffic Psychology and Behaviour* 9 (5): 335–346.

Papantoniou, Panagiotis, George Yannis, Constantinos Antoniou, Eleonora Papadimitriou, Dimosthenis Pavlou, and John Golias. 2016. "Investigating the Effect of Area Type and Traffic Conditions on Distracted Driving Performance." *Transportation Research Procedia* 14: 3839–3848.

Park, Eun Sug, Kay Fitzpatrick, Subasish Das, and Raul Avelar. 2021. "Exploration of the Relationship among Roadway Characteristics, Operating Speed, and Crashes for City Streets Using Path Analysis." *Accident Analysis & Prevention* 150: 105896. https://doi.org/10.1016/j.aap.2020.105896.

Pour-Rouholamin, Mahdi, Mohammad Jalayer, and Huaguo Zhou. 2017. "Modelling Single-Vehicle, Single-Rider Motorcycle Crash Injury Severity: An Ordinal Logistic Regression Approach." *International Journal of Urban Sciences* 21 (3): 344–363. https://doi.org/10.1080/12265934.2017.1311801.

Preusser, David F., Allan F. Williams, Susan A. Ferguson, Robert G. Ulmer, and Helen B. Weinstein. 1998. "Fatal Crash Risk for Older Drivers at Intersections." *Accident Analysis and Prevention* 30 (2): 151–159. https://doi.org/10.1016/S0001-4575(97)00090-0.

Qu, Yi, Zhengkui Lin, Honglei Li, and Xiaonan Zhang. 2019. "Feature Recognition of Urban Road Traffic Accidents Based on GA-XGBoost in the Context of Big Data." *IEEE Access* 7: 170106–15.

Rahman, M. Ashifur, Subasish Das, and Xiaoduan Sun. 2023. "Understanding the Drowsy Driving Crash Patterns from Correspondence Regression Analysis." *Journal of Safety Research* 84: 167–181. https://doi.org/10.1016/j.jsr.2022.10.017.

Rahman, M. Ashifur, Md. Mahmud Hossain, Elisabeta Mitran, and Xiaoduan Sun. 2021. "Understanding the Contributing Factors to Young Driver Crashes: A Comparison of Crash Profiles of Three Age Groups." *Transportation Engineering* 5: 100076. https://doi.org/10.1016/j.treng.2021.100076.

Rao, Xiaohua, Changwei Yuan, Shengxuan Ding, Xinjie Fang, and Renteng Yuan. 2024. “Analyzing Injury Severity of School Bus Crashes Considering Temporal and Spatial Heterogeneity Using Random Parameters Ordered Probit Model.”.

Samuel, Siby, Yusuke Yamani, and Donald L. Fisher. 2016. “Large Reductions Are Possible in Older Driver Crashes at Intersections.” *Clinical and Experimental Optometry* 99 (5), https://doi.org/10.1111/cxo.12443.

Schlögl, Matthias, Rainer Stütz, Gregor Laaha, and Michael Melcher. 2019. “A Comparison of Statistical Learning Methods for Deriving Determining Factors of Accident Occurrence from an Imbalanced High Resolution Dataset.” *Accident Analysis & Prevention* 127: 134–149.

Scott-Parker, Bridie, Tamara De Regt, Christian Jones, and Jamie Caldwell. 2020. “The Situation Awareness of Young Drivers, Middle-Aged Drivers, and Older Drivers: Same but Different?” *Case Studies on Transport Policy* 8 (1): 206–214. https://doi.org/10.1016/j.cstp.2018.07.004.

Se, Chamroeun, Thanapong Champahom, Sajjakaj Jomnonkwao, Palaphorn Chaimuang, and Vatanavongs Ratanavaraha. 2021. “Empirical Comparison of the Effects of Urban and Rural Crashes on Motorcyclist Injury Severities: A Correlated Random Parameters Ordered Probit Approach with Heterogeneity in Means.” *Accident Analysis & Prevention* 161: 106352.

Shaheed, Mohammad Saad B., Konstantina Gkritza, Wei Zhang, and Zachary Hans. 2013. “A Mixed Logit Analysis of Two-Vehicle Crash Severities Involving a Motorcycle.” *Accident Analysis & Prevention* 61: 119–128. https://doi.org/10.1016/j.aap.2013.05.028.

Shangguan, Qiangqiang, Junhua Wang, Ting Fu, and Shou’en Fang. 2022. “Quantification of Cut-in Risk and Analysis of Its Influencing Factors: A Study Using Random Parameters Ordered Probit Model.” *Journal of Transportation Safety & Security* 14 (12): 2029–2054.

Shao, Xiaojun, Xiaoxiang Ma, Feng Chen, Mingtao Song, Xiaodong Pan, and Kesi You. 2020. “A Random Parameters Ordered Probit Analysis of Injury Severity in Truck Involved Rear-End Collisions.” *International Journal of Environmental Research and Public Health* 17 (2): 395.

Sifrit, Kathy J., Jane Stutts, Loren Staplin, and Carol Martell. 2010. “Intersection Crashes among Drivers in Their 60s, 70s and 80s.” In *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 54, 2057–2061. Los Angeles, CA: SAGE Publications Sage CA.

Slikboer, Reneta, Samuel D. Muir, Sandun M. Silva, and Denny Meyer. 2020. “A Systematic Review of Statistical Models and Outcomes of Predicting Fatal and Serious Injury Crashes from Driver Crash and Offense History Data.” *Systematic Reviews* 9: 1–15.

Smolensky, Michael H., Lee Di Milia, Maurice M. Ohayon, and Pierre Philip. 2011. “Sleep Disorders, Medical Conditions, and Road Accident Risk.” *Accident Analysis & Prevention* 43 (2): 533–548. https://doi.org/10.1016/j.aap.2009.12.004.

Staplin, Loren. 2004. “Highway Enhancements to Improve Safety and Mobility of Older Road Users.” *Transportation in an Aging Society* 155.

Subhan, Fazle, Yasir Ali, Shengchuan Zhao, and Oscar Oviedo-Trespalacios. 2023. “Understanding and Modeling Willingness-to-Pay for Public Policies to Enhance Road Safety: A Perspective from Pakistan.” *Transport Policy* 141: 182–196. https://doi.org/10.1016/j.tranpol.2023.07.016.

Sun, Zhiyuan, Duo Wang, Xin Gu, Mohamed Abdel-Aty, Yuxuan Xing, Jianyu Wang, Huapu Lu, and Yanyan Chen. 2023. “A Hybrid Approach of Random Forest and Random Parameters Logit Model of Injury Severity Modeling of Vulnerable Road Users Involved Crashes.” *Accident Analysis & Prevention* 192: 107235. https://doi.org/10.1016/j.aap.2023.107235.

Sun, Zhiyuan, Yuxuan Xing, Jianyu Wang, Xin Gu, Huapu Lu, and Yanyan Chen. 2022. “Exploring Injury Severity of Vulnerable Road User Involved Crashes across Seasons: A Hybrid Method Integrating Random Parameter Logit Model and Bayesian Network.” *Safety Science* 150: 105682.

Sze, N. N., and S. C. Wong. 2007. “Diagnostic Analysis of the Logistic Model for Pedestrian Injury Severity in Traffic Crashes.” *Accident Analysis & Prevention* 39 (6): 1267–1278. https://doi.org/10.1016/j.aap.2007.03.017.

Turner, Dan, Marsha Nitzburg, and Richard Knoblauch. 1998. “Ultraviolet Headlamp Technology for Nighttime Enhancement of Roadway Markings and Pedestrians.” *Transportation Research Record: Journal of the Transportation Research Board* 1636 (1): 124–131. https://doi.org/10.3141/1636-20.

Usman, Taimur, Liping Fu, and Luis F. Miranda-Moreno. 2016. "Injury Severity Analysis: Comparison of Multilevel Logistic Regression Models and Effects of Collision Data Aggregation." *Journal of Modern Transportation* 24: 73–87.

Viano, David C., Clyde C. Culver, Leonard Evans, Michael Frick, and Robert Scott. 1990. "Involvement of Older Drivers in Multivehicle Side-Impact Crashes." *Accident Analysis & Prevention* 22 (2): 177–188. https://doi.org/10.1016/0001-4575(90)90068-V.

Wang, Zhenyu, Jian J. Lu, Qing Wang, Linjun Lu, and Zhao Zhang. 2010. "Modeling Injury Severity in Work Zones Using Ordered Probit Regression." ICCTP 2010: Integrated Transportation Systems: Green, Intelligent, Reliable, 1058–1067.

Xin, Chunfu, Rui Guo, Zhenyu Wang, Qing Lu, and Pei-Sung Lin. 2017. "The Effects of Neighborhood Characteristics and the Built Environment on Pedestrian Injury Severity: A Random Parameters Generalized Ordered Probability Model with Heterogeneity in Means and Variances." *Analytic Methods in Accident Research* 16: 117–132.

Xing, Yingying, Shengdi Chen, Shengxue Zhu, Yi Zhang, and Jian Lu. 2020. "Exploring Risk Factors Contributing to the Severity of Hazardous Material Transportation Accidents in China." *International Journal of Environmental Research and Public Health* 17 (4): 1344.

Yang, Chao, Mingyang Chen, and Quan Yuan. 2021. "The Application of XGBoost and SHAP to Examining the Factors in Freight Truck-Related Crashes: An Exploratory Analysis." *Accident Analysis & Prevention* 158: 106153. https://doi.org/10.1016/j.aap.2021.106153.

Yang, Yang, Kun Wang, Zhenzhou Yuan, and Dan Liu. 2022. "Predicting Freeway Traffic Crash Severity Using XGBoost-Bayesian Network Model with Consideration of Features Interaction." In *Journal of Advanced Transportation*, edited by Gen Li, 1–16. Hindawi. https://doi.org/10.1155/2022/4257865.

Yang, Zhang, L. I. Zhibin, Liu Pan, and Z. H. A. Liteng. 2011. "Exploring Contributing Factors to Crash Injury Severity at Freeway Diverge Areas Using Ordered Probit Model." *Procedia Engineering* 21: 178–185.

Yu, Miao, Changxi Ma, and Jinxing Shen. 2021. "Temporal Stability of Driver Injury Severity in Single-Vehicle Roadway Departure Crashes: A Random Thresholds Random Parameters Hierarchical Ordered Probit Approach." *Analytic Methods in Accident Research* 29: 100144. https://doi.org/10.1016/j.amar.2020.100144.

Yu, Miao, Changjiang Zheng, and Changxi Ma. 2020a. "Analysis of Injury Severity of Rear-End Crashes in Work Zones: A Random Parameters Approach with Heterogeneity in Means and Variances." *Analytic Methods in Accident Research* 27: 100126.

Yu, Miao, Changjiang Zheng, Changxi Ma, and Jinxing Shen. 2020b. "The Temporal Stability of Factors Affecting Driver Injury Severity in Run-off-Road Crashes: A Random Parameters Ordered Probit Model with Heterogeneity in the Means Approach." *Accident Analysis & Prevention* 144: 105677.

Zafian, Tracy, Alyssa Ryan, Ravi Agrawal, Siby Samuel, and Michael Knodler. 2021. "Using SHRP2 NDS Data to Examine Infrastructure and Other Factors Contributing to Older Driver Crashes during Left Turns at Signalized Intersections." *Accident Analysis & Prevention* 156: 106141–106141.

Zeng, Qiang, Qianfang Wang, Fangzhou Wang, and N. N. Sze. 2022. "Revisiting Spatial Correlation in Crash Injury Severity: A Bayesian Generalized Ordered Probit Model with Leroux Conditional Autoregressive Prior." *Transportmetrica A: Transport Science* 18 (3): 1084–1102.

Zhang, Jun, Joan Lindsay, Kathy Clarke, Glenn Robbins, and Yang Mao. 2000. "Factors Affecting the Severity of Motor Vehicle Traffic Crashes Involving Elderly Drivers in Ontario." *Accident Analysis & Prevention* 32 (1): 117–125.

## Appendix

**Table A1.** Comparison among random parameter modelling techniques

| | RPOP | | CRPOP | | RPOP-HM | | CRPOP-HM | |
|---|---|---|---|---|---|---|---|---|
| Variables | $\beta$ | t-value | $\beta$ | t-value | $\beta$ | t-value | $\beta$ | t-value |
| *m_side_swipe* | *−0.70* | *−20.25* | *−0.72* | *−20.48* | *−0.81* | *−21.12* | *−0.82* | *−21.28* |
| *standard deviation of m_side_swipe* | *0.47* | *17.46* | *0.41* | *15.02* | *0.52* | *18.72* | *0.06* | *2.24* |
| d_condition_ifa | 0.24 | 4.46 | 0.24 | 4.52 | 0.18 | 3.34 | 0.18 | 3.36 |
| m_single_vehicle | 0.09 | 3.02 | 0.10 | 3.18 | 0.21 | 5.99 | 0.21 | 6.02 |
| lanes_two | 0.09 | 3.30 | 0.09 | 3.31 | - | - | - | - |
| location_r | 0.21 | 6.61 | 0.22 | 6.68 | 0.26 | 8.16 | 0.26 | 8.11 |
| *d_condition_normal* | *−0.81* | *−24.83* | *−0.82* | *−25.13* | *−0.66* | *−17.85* | *−0.68* | *−18.11* |
| *standard deviation of d_condition_normal* | *0.50* | *21.46* | *0.58* | *23.81* | *0.46* | *19.98* | *0.53* | *21.97* |
| *location_oc* | *0.18* | *6.18* | *0.18* | *6.19* | *0.33* | *7.98* | *0.33* | *8.04* |
| *standard deviation of location_oc* | *0.39* | *16.93* | *0.02* | *0.70* | *0.43* | *18.58* | *0.33* | *14.07* |
| m_rear_end | 0.07 | 2.72 | 0.08 | 2.80 | 0.07 | 2.62 | 0.07 | 2.71 |
| pcf_violation | 0.11 | 4.64 | 0.11 | 4.38 | 0.08 | 3.16 | 0.08 | 3.11 |
| ADT_20000 | −0.09 | −3.66 | −0.09 | −3.67 | −0.06 | −2.36 | −0.06 | −2.29 |
| d_condition_id | −0.64 | −22.06 | −0.63 | −21.90 | −0.61 | −21.16 | −0.61 | −21.13 |
| dow_weekend | 0.08 | 3.20 | 0.08 | 3.28 | 0.08 | 3.18 | 0.08 | 3.23 |
| pcf_ped_actions | 1.95 | 12.95 | 1.99 | 13.62 | 2.09 | 14.11 | 2.11 | 14.52 |
| passenger_present | 0.11 | 4.19 | 0.11 | 4.19 | 0.11 | 4.46 | 0.11 | 4.45 |
| d_age_65_to_74 | −0.08 | −3.88 | −0.08 | −3.96 | −0.08 | −4.02 | −0.08 | −4.09 |
| m_right_angle | 0.27 | 6.81 | 0.28 | 6.88 | 0.27 | 6.66 | 0.27 | 6.71 |
| location_br | 0.08 | 2.98 | 0.08 | 2.98 | 0.10 | 3.68 | 0.10 | 3.67 |
| weather_rain | −0.07 | −2.00 | −0.07 | −1.96 | −0.07 | −2.04 | −0.07 | −2.00 |
| *Heterogeneity in means* | | | | | | | | |
| Effect of 'two-lane roadway' on the mean of 'manner of collision = sideswipe' | | | | | 0.410 | 5.93 | 0.427 | 6.16 |
| Effect of 'manner of collision = single vehicle on means of 'driver condition = normal' | | | | | −0.438 | −8.03 | −0.411 | −7.54 |
| Effect of ADT > 20,000 on the mean of 'location = open country' | | | | | −0.264 | −5.15 | −0.276 | −5.38 |
| Model statistics | | | | | | | | |
| LL (0) | −12784.038 | | −12784.038 | | −12784.038 | | −12784.038 | |
| Log-likelihood at convergence, LL($\beta$) | −11809.582 | | −11799.876 | | −11760.689 | | −11754.0535 | |
| Akaike Information Criterion (AIC) | 23663.2 | | 23649.8 | | 23569.4 | | 23562.1 | |
| AIC/N | 1.293 | | 1.292 | | 1.288 | | 1.288 | |
| McFadden Pseudo R-Squared | 0.0762 | | 0.0770 | | 0.0800 | | 0.0806 | |

Table A2, summarises model comparison results using the likelihood ratio test with test statistics as the first value, degrees of freedom in brackets, and *p*-value in parentheses. The results show that CRPOP, RPOPHM, and CRPOPHM are a better fit than the RPOP model (first row), RPOPHM and CRPOPHM are a better fit than CRPOP (second line), and the CRPOPHM model is a better fit than RPOPHM. As a result, CRPOPHM is the best fit overall.

**Table A2.** Likelihood Ratio Test Results for Model Comparison.

| | RPOP | CRPOP | RPOPHM | CRPOP-HM |
|---|---|---|---|---|
| RPOP | - | 19.42 [3] (0.0002) | 97.79 [2] (0.0000) | 111.05 [5] (0.0000) |
| CRPOP | - | - | 78.37 [1] (0.0000) | 91.65 [3] (0.0000) |
| RPOPHM | - | - | - | 13.27 [3] (0.0046) |
| CRPOPHM | - | - | - | - |