COMPUTATIONAL RESEARCH PROGRESS IN APPLIED SCIENCE & ENGINEERING (CRPASE)

CRPASE: TRANSACTIONS OF CIVIL AND ENVIRONMENTAL ENGINEERING

Journal homepage: http://www.crpase.com

CRPASE: Transactions of Civil and Environmental Engineering 11 (3) Article ID: 2957, 1–16, September 2025 ISSN 2423-4591

Research Article

Exploring the Impact of Driver Sex, Driver Age, Area Type, and Lighting Conditions on Rear-End Collision Severity

Saber Naseralavi [1*], Mohammad Soltanirad [2], Mahdi Baghersad [3], Erfan Ranjbar [4], Keshav Jimee [2], Akram Mazaheri [5]

[1] *Department of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran 76169-14111*
[2] *Department of Civil, Environmental, and Construction Engineering, Texas Tech University, Lubbock, Texas, USA 79409*
[3] *Department of Civil, Construction, and Environmental Engineering, University of Alabama at Birmingham, Birmingham, Alabama USA 35294*
[4] *Department of Civil, Construction, and Environmental Engineering, University of Delaware, Newark, Delaware, USA 19716*
[5] *Department of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran 14115-111*

Keywords

Crash Severity, Traffic Safety, Rear-End Collision, Logistic Regression, HSIS Data.

Abstract

This study investigates the factors influencing injury severity in rear-end collisions in California using data from the Highway Safety Information System (HSIS). A total of 569,386 rear-end crashes recorded between 2015 and 2017 are analyzed. The dataset is divided into 24 subgroups based on driver sex (male, female), driver age (under 25, 25–65, over 65), area type (urban, rural), and lighting conditions (daylight, dark). For each subgroup, a binary logistic regression model is developed to examine the likelihood of injury (injury vs. no injury). Results indicate that crash severity is influenced by the number of vehicles, AADT, access control, surface type, terrain, lighting, time of crash, vehicle year, and season. Among all predictors, crashes involving three or more vehicles are consistently linked to lower injury odds, while high traffic volume and full freeway access control increase severity. Model performance is assessed using goodness-of-fit and discrimination measures, including deviance, AIC, pseudo R^2, and AUC. Of the 24 models, the best performance is observed for older female drivers in rural dark conditions, while the poorest is for young male drivers in rural dark conditions. These findings underscore the value of disaggregated modeling and suggest that traffic safety interventions can be tailored to specific demographic and environmental contexts.

1. Introduction

According to the World Health Organization's Global Report on Road Safety 2023, despite a 5% decrease in traffic deaths in 2021 compared to 2010, deaths and injuries caused by road traffic remain a major global health issue. Up to 2019, road crashes were considered the leading cause of death for children and young people aged 5 to 29 years, and the 12th leading cause of death among all age groups [1].

In the United States, motor vehicle crashes remain a significant public health concern. The National Highway Traffic Safety Administration (NHTSA) reported that an estimated 42,939 people lost their lives in traffic crashes in 2021, marking a 10% increase from the previous year and the highest number of fatalities since 2005. Additionally,

* Corresponding Author: Saber Naseralavi
E-mail address: saber_naseralavi@uk.ac.ir, ORCID: https://orcid.org/0000-0002-5392-910X

Received: 18 May 2025; Revised: 18 June 2025; Accepted: 01 July 2025
https://doi.org/10.82042/crpase.11.3.2957

Academic Editor: **Vahid Najafi Moghadam Gilani**

Please cite this article as: S. Naseralavi, M. Soltanirad, M. Baghersad, E. Ranjbar, K. Jimee, A. Mazaheri, Exploring the Impact of Driver Sex, Driver Age, Area Type, and Lighting Conditions on Rear-End Collision Severity, Computational Research Progress in Applied Science & Engineering, CRPASE: Transactions of Civil and Environmental Engineering 11 (2025) 1–16, Article ID: 2957.

around 2.5 million individuals sustained injuries in motor vehicle crashes during the same year, representing a 9.4% rise compared to 2020 [2]. According to a Centers for Disease Control and Prevention (CDC) report, the U.S. had the highest population-based motor vehicle crash death rate (11.1 per 100,000 population) among 29 high-income countries in 2019, and it was 2.3 times the average rate of the other countries examined [3].

California is the most populous state in the United States, with a population of over 39 million according to the 2023 census data [4]. It also has one of the highest volumes of traffic in the nation, with approximately 24 million licensed drivers and over 32 million registered vehicles annually [5]. It is home to major metropolitan areas like Los Angeles, San Francisco, and San Diego, which contribute to some of the busiest and most congested road networks in the nation [6]. This high vehicle density, along with extensive urban traffic, increases the likelihood and impact of road traffic crashes. California's vehicle density, urban traffic patterns, and extensive availability of high-quality crash data provide a valuable setting for targeted traffic safety research.

Among all the various types of crashes, rear-end crashes are the most frequent and predominant. This crash type accounts for nearly 29% of all automobile crashes in the United States and contributes to 7.2% of fatal crashes [7]. The severity of a rear-end collision can range from minor property damage, often resulting in inconvenient delays and insurance claims, to more severe outcomes such as serious injuries, permanent disabilities, or fatalities. This wide range of severity is influenced by a complex interplay of factors, including vehicle characteristics, road design, driver behavior, and environmental conditions. While prior research has explored crash frequency and overall crash severity, few studies have specifically examined the severity of rear-end collisions using disaggregated models that consider driver sex, age, area type, and lighting conditions. By identifying how these factors shape crash outcomes, we can better target interventions toward high-risk groups.

Although extensive research has addressed various aspects of traffic safety, a nuanced understanding of how specific driver demographics, namely sex and age, interact with environmental variables such as area type (urban vs. rural) and lighting conditions (daylight, dusk/dawn, and darkness) to influence rear-end collision severity remains limited.

To address this gap, this study analyzes 569,386 rear-end crashes in California between 2015 and 2017 using data from the Highway Safety Information System (HSIS). The dataset is divided into 24 distinct groups based on combinations of driver sex, age group (under 25, 25–65, and over 65), area type, and lighting conditions. A separate logistic regression model is developed for each group to evaluate the statistical significance and impact of variables influencing crash severity. Severity is categorized as either property damage only (PDO) or injury/fatality (NotPDO), where NotPDO refers to any crash in which at least one person is injured or killed. The findings aim to support traffic authorities and policymakers by identifying key factors associated with rear-end collision severity and suggesting targeted strategies to reduce injuries and fatalities.

To provide context, the next section reviews empirical findings on how demographic and environmental factors affect rear-end crash severity.

2. Literature Review

Rear-end collisions are among the most common crash types globally and significantly contribute to traffic injuries [8, 9]. For instance, one analysis revealed that rear-end collisions accounted for approximately 30% of all injury crashes and about 7% of crash fatalities [10]. Given that nearly 1.19 million lives are lost globally each year due to road traffic crashes [11, 12], understanding the factors that exacerbate or reduce injury severity in rear-end crashes remains a crucial goal of road safety research. Recent studies have increasingly focused on how driver characteristics and environmental conditions impact the outcomes of rear-end crashes. In particular, driver sex and driver age, as well as area type (e.g., urban vs. rural settings) and lighting conditions (e.g., daylight vs. darkness), have been examined as potential determinants of crash injury outcomes. This literature review synthesizes findings from recent research across different regions, highlighting the current state of knowledge for each factor.

2.1. Driver Sex and Collision Severity

Multiple studies in recent years have confirmed that driver sex is a significant factor in the injury severity of rear-end crashes. While traditional traffic statistics suggest that male drivers are more frequently involved in severe crashes due to greater exposure and riskier behaviors [13], contemporary studies have revealed that female drivers are more likely to sustain serious injuries when crashes occur [14]. For instance, Dabbour et al. (2020) analyzed over a decade of rear-end collision data and found that although female drivers were not associated with greater injury severity for the other driver involved, they were significantly more likely to suffer severe injuries, regardless of being the striking or struck driver [15].

Further support for this trend came from Ryan and Knodler (2022), who found that female drivers involved in rear-end crashes faced higher risks of neck, chest, and lower extremity injuries compared to males. The study highlighted that even when controlling for vehicle type and crash conditions, the probability of certain injuries remained elevated for women [16]. Sharafeldin et al. (2022) reinforced this point using data from signalized intersections in Wyoming. Their logistic regression model found that female drivers had a statistically significantly higher likelihood of experiencing injuries in rear-end crashes compared to male drivers, all else being equal [17].

2.2. Driver Age and Collision Severity

Driver age has long been recognized as a significant factor that influences both the likelihood and severity of motor vehicle crashes, including rear-end collisions. Numerous studies have shown that both the youngest and oldest drivers tend to be at higher risk, but for different reasons. Young drivers, particularly those under 25, are more prone to engage in risky driving behaviors such as speeding, distracted driving, and aggressive maneuvers, often due to inexperience and immature judgment [18]. Conversely, older drivers (generally over 65) are more likely to sustain serious or fatal injuries in crashes due to age-related physiological frailty, slower reaction times, and declines in visual and cognitive abilities [19, 20].

For rear-end collisions specifically, research has shown that older drivers are more vulnerable. Sharafeldin et al. (2022) found that age was a significant predictor of injury severity in rear-end crashes at signalized intersections, with older drivers experiencing more severe outcomes [17]. A study by Zou et al. (2023) using a latent class approach, similarly concluded that drivers aged 65 and older had significantly higher probabilities of sustaining incapacitating or fatal injuries in rear-end collisions [20]. This was echoed in the findings of Santolino et al. (2022), who observed that older drivers tended to face the highest severity risk in crashes involving rear-end impacts. The physiological frailty of older adults means that even moderate crashes can lead to disproportionately severe injuries. This finding reinforces the need for tailored interventions targeting this demographic [21].

On the other hand, while young drivers are often involved in more crashes overall, the injury outcomes are typically less severe. This is attributed to a combination of better physical resilience and the fact that many youth-involved rear-end crashes occur in lower-speed, urban environments. However, when factors such as high speed or alcohol are involved, the severity escalates rapidly [15, 22-25]. Xi et al. (2019) [26] and Luo et al. (2020) [27] found that younger drivers were overrepresented in rear-end collisions, but these were often property-damage-only or minor-injury crashes unless compounded by additional risk factors.

Naseralavi et al. (2023) advanced this area of research by modeling crash severity across varying driver age groups using HSIS data and a multinomial logit framework. Their findings confirmed that rear-end crashes increased injury risk across all age categories, with older drivers showing greater sensitivity to environmental and roadway factors. This highlights the importance of age-specific modeling to reveal nuanced patterns of risk [28].

Therefore, while both young and older drivers contribute significantly to rear-end crash statistics, the nature and outcomes of these crashes differ, underscoring the need for age-targeted safety strategies and driver-assistance technologies that address these distinct risk profiles.

2.3. Area Type and Collision Severity

The geographical context of a collision, whether it occurs in an urban or rural area, significantly influences its characteristics and severity. Rural areas tend to experience more severe outcomes due to higher travel speeds, limited roadway lighting, and longer emergency response times. In contrast, urban crashes, although more frequent due to congestion, generally occur at lower speeds and result in lower injury severity [29, 30].

Champahom et al. (2020) applied hierarchical logistic models to analyze rear-end crash severity. They found that the rural model contained more significant predictors than the urban model, highlighting the need for geographically tailored crash models [31]. Their findings emphasized the importance of understanding geographical context when designing effective safety interventions.

Wei et al. (2021) used Bayesian binary logit models to highlight further disparities between rural and urban settings. The study found that drunk driving increased the likelihood of fatal crashes by 10.8% in urban areas and 16.4% in rural areas. Truck involvement raised fatality risk by 4.6% in urban areas and 9.8% in rural ones [32]. These differences demonstrate how identical crash factors can have varying impacts depending on area type. Furthermore, nighttime driving and limited visibility continue to be significant contributors to injury severity, particularly on rural roads where the infrastructure is less forgiving [30-33].

Urban expressways create distinct conditions that influence rear-end crash risks, particularly because of variations in traffic speed and volume. Zhang et al. (2020) showed that rear-end crash frequency on urban expressways increases under two conditions: low-speed, high-volume environments and high-speed, congested traffic. Their model highlighted how roadway function and flow characteristics intersect to affect crash risk [33]. Complementing this, Wang et al. (2022) used naturalistic driving data to show that car-following behaviors, such as reduced time gaps and speed variability, strongly predicted rear-end crashes and near-crashes in urban environments [34]. Together, these studies affirm the need for context-sensitive safety models and targeted countermeasures for both urban and rural roads.

2.4. Lighting Conditions and Collision Severity

Lighting conditions, particularly during nighttime or under low light, have consistently been identified as key factors affecting crash severity. Poor visibility impairs drivers' ability to detect hazards and respond promptly, increasing the likelihood of fatal or severe injuries in rear-end crashes. Liu et al. (2019) demonstrated that the probability of fatal nighttime crashes on unlit road segments was significantly higher than at intersections. Roads without street lamps accounted for the majority (93.5%) of nighttime crashes in their dataset [35]. Similarly, Chen et al. (2019) confirmed that inferior lighting conditions significantly increased injury severity in rear-end crashes, especially when compounded by factors such as high speed or poor weather [19].

In addition to reducing visibility, insufficient lighting can create visual illusions such as glare or tunnel vision, which impair a driver's ability to judge distance and relative speed. Liu et al. (2019) noted that the glare from oncoming headlights could cause momentary blindness, contributing to rear-end crashes on rural or poorly lit roads [35]. Despite these risks, many drivers fail to reduce their speed in low-light conditions, thereby increasing crash severity risks.

Swain and Larue (2024) provided more detailed and regional insights by analyzing rear-end crashes in Queensland, Australia. They found that lighting conditions were among the most significant environmental predictors for rear-end crash severity, particularly at intersections and in higher-speed zones [36]. Their machine learning model showed that nighttime crashes without adequate lighting were strongly associated with higher injury rates for not-at-fault (struck) drivers. This pattern of elevated nighttime crash severity, driven by increased vulnerability in dark, unlit environments, underscores the importance of improving roadway lighting as a preventive safety measure.

Similarly, Rahmaninezhad Asil et al. (2022) analyzed urban traffic crashes in Iran using a classification and regression tree model. They found that nighttime conditions, especially when combined with motorcycle involvement or driver inattention, substantially increased the risk of severe injuries. Their study reinforces the importance of disaggregating crash data by lighting condition to capture the complex ways in which the environment and behavior interact to influence injury outcomes [37].

Collectively, these studies highlight the urgent need for improved roadway lighting, especially on rural highways, at intersections, and in urban environments, to decrease injury severity in rear-end crashes under dark or low-light conditions**.**

2.5. Roadway and Environmental Factors

Beyond driver characteristics and contextual settings, numerous studies have emphasized the role of roadway and environmental variables in shaping crash severity. Variables such as average annual daily traffic (AADT), number of lanes, and access control influence both the occurrence and severity of crashes. Higher AADT is associated with increased crash frequency but not necessarily greater severity, while properly designed and positioned medians reduce severity by improving vehicle separation [38, 39]. Road surface conditions, particularly wet or icy roads, have also been linked to increased injury risk in rear-end crashes, especially when combined with poor lighting or high-speed environments [40]. Incorporating these variables into severity models enhances the ability to capture complex interactions between driver behavior and the built environment.

2.6. Modeling Crash Severity Using Binary Logit Models

Understanding and modeling crash severity are crucial in road safety research, as they allow analysts to identify factors that influence injury outcomes and to develop targeted safety interventions. Various statistical and machine learning models are used for this purpose, including binary and multinomial logit models, ordered probit/logit, latent class models, and, more recently, artificial neural networks and decision trees [17, 19, 31, 41-44]. Among these, the binary logit model is especially common in studies where the severity outcome is categorized, for instance, distinguishing between PDO and injury crashes.

The binary logit model is a type of generalized linear model that estimates the probability of a crash resulting in a severe outcome, based on a set of explanatory variables. It models the log-odds of a binary outcome using a logistic function, expressed as a linear combination of predictors. This approach is particularly useful in traffic safety research, where outcomes like injury or no injury are inherently binary. The coefficients of the model can be interpreted in terms of odds ratios, making them easier for policymakers and practitioners to understand [19, 43].

Binary logit models are particularly well-suited for crash severity studies for several reasons. First, crash datasets often support binary classification, particularly when severity is simplified into PDO versus injury/fatality. Second, the model accommodates both continuous and categorical predictors, such as driver age, driver sex, lighting condition, and road type, which are commonly recorded in crash databases. Third, it provides interpretable results and is computationally efficient, making it ideal for use with large-scale datasets such as those available from the HSIS [31].

Despite its strengths, the binary logit model has certain limitations. One major concern is its assumption of homogeneity; it does not account for unobserved heterogeneity among crash cases. If important variables are omitted or inaccurately measured, the model may yield biased estimates. Additionally, it assumes that the relationship between predictors and the log-odds of the outcome remains constant across all observations. To address these limitations, researchers use more advanced methods such as mixed logit, random parameters models, and latent class approaches, which allow variation in effects across individuals or subgroups [17, 19, 31, 41-44].

This study employs the binary logit model due to its suitability for binary outcomes (PDO vs. injury/fatality) and its ability to handle large crash datasets. This objective, investigating how demographic and environmental factors influence rear-end crash severity across disaggregated groups, aligns well with the strengths of this model. By applying it to 24 distinct subgroups, this approach offers interpretable and policy-relevant insights that can inform targeted safety interventions for specific populations and roadway contexts. This approach aligns with the work of Naseralavi et al. (2023), who applied binary logistic regression to HSIS motorcycle crash data, segmented by combinations of time of day and season. Their study demonstrated that the significance of variables varied across temporal subgroups, reinforcing the value of disaggregated modeling in revealing nuanced patterns in crash severity [45].

3. Methodology

3.1. Data Source

This study uses crash data from the HSIS, a nationally recognized database created and maintained by the U.S. Federal Highway Administration (FHWA). HSIS compiles comprehensive crash, roadway, and traffic volume data from selected states that maintain high-quality records, facilitating links between crash reports, roadway inventory, and traffic characteristics. Its consistent format and long-standing application in highway safety research make it a reliable foundation for empirical research on traffic injury [46, 47].

HSIS data are obtained from police-reported crash records and roadway inventories maintained by state transportation agencies. Researchers and transportation analysts frequently use the HSIS dataset to evaluate safety interventions and study factors associated with crash occurrence and severity. The database is publicly documented and available through the FHWA's official HSIS portal [46, 47].

This study uses rear-end crash data from California, one of the core HSIS states, covering the years 2015 to 2017. The extracted dataset comprises 569,386 rear-end crashes, with detailed information on crash severity, driver demographics, lighting conditions, vehicle characteristics, roadway features, and traffic volumes. The richness and volume of the dataset enable disaggregated analyses of crash severity across multiple dimensions, including driver sex, driver age, lighting condition, and area type.

3.2. Variable Definitions and Dataset Overview

To support disaggregated modeling of rear-end crash severity, this study employs a diverse range of demographic, environmental, vehicle, and roadway-related variables derived from the HSIS California dataset. The primary outcome variable is crash severity (dependent variable), classified as a binary outcome:

• Property Damage Only (PDO), recorded as 0

• Injury or Fatality (NotPDO), recorded as 1

All other variables are selected based on their theoretical relevance and availability in the HSIS. The predictor variables include:

• Driver Sex: Gender of the driver (Male or Female)

• Vehicle Year: Model year of the vehicle, organized into five-year intervals

• Hour Class: Time of crash occurrence, categorized into five classes:

Class 1: 12:01 AM – 6:00 AM

Class 2: 6:01 AM – 10:00 AM

Class 3: 10:01 AM – 4:00 PM

Class 4: 4:01 PM – 8:00 PM

Class 5: 8:01 PM – 12:00 AM

• Road Surface: Road surface condition (e.g., Dry, Snowy/Icy/Wet)

• Lighting condition: Daylight or Dark

• No. Vehicles: Number of vehicles involved in the crash (2, 3, or 3+)

• No. Lanes: Number of lanes on the roadway, categorized into three groups: fewer than five lanes, 6 to 7 lanes, and more than 7 lanes

• Surface Type: Surface type (Asphalt Concrete or Portland Cement Concrete)

• Median Type: Median type (Divided or Undivided)

• Access: Access control type (Full, Partial, or None)

• Terrain: Terrain type (Flat, Rolling, or Mountainous)

• AADT: Annual average daily traffic volume, categorized into four groups

• Rural/Urban: Area type where the crash occurred (Urban or Rural)

• Season: The season during which the crash took place

• Age Group: Driver's age group (under 25, 25–65, over 65)

Table 1 presents twenty randomly selected records from the cleaned dataset, showcasing the structure and value ranges of the chosen variables.

To better understand the distribution of each variable, Table 2 summarizes the frequency of categories observed in the dataset. Several notable patterns emerge. For example, male drivers are involved in approximately 61.6% of the crashes, and the 25–65 age group accounts for nearly 74% of cases, consistent with exposure trends. Most crashes occur in urban areas (91.7%) and under daylight conditions (75.3%), yet a significant portion of injury crashes (NotPDO) happen in darkness, highlighting lighting as an important risk factor.

Additionally, the vast majority of crashes occur on dry surfaces (95%) and on divided roads (96%), indicating that even in seemingly safer conditions, rear-end crashes remain common. AADT values are highest in the 175k-250k group, and eight-lane roadways are the most prevalent cross-section. Although Full Access Control roadways dominate the dataset, crashes on Conventional Roads (No Access Control) still represent a significant portion, highlighting the importance of access design in crash severity outcomes.

These patterns provide important context for the disaggregated models presented in the following sections, helping to explain variation in crash outcomes across different roadways and driver profiles.

3.3. Binary Logistic Regression

Binary logistic regression is a widely used statistical technique for modeling binary outcomes in public health, transportation safety, and social sciences. It is particularly effective when the dependent variable reflects a binary outcome, such as injury/fatality versus property damage only, as found in crash severity studies. This modeling technique provides a flexible and interpretable framework for quantifying the effects of explanatory variables on the likelihood of an event occurring [48, 49]. For instance, Saheli (2021) applied binary logistic regression to assess how driver route familiarity and contextual factors influenced injury severity in urban vehicle-to-vehicle crashes, demonstrating the method's applicability in disaggregated traffic safety analysis [50]

Unlike linear regression, which models a continuous outcome and relies on assumptions such as normality and homoscedasticity, logistic regression models the log odds of a binary response and imposes fewer distributional assumptions. It employs the logistic (sigmoid) function to convert any real-valued linear combination of inputs into a probability that is bounded between 0 and 1. The logistic function is expressed as shown in Eq. (1)

$$P(Y=1|X) = \frac{1}{1+\exp(-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k))} \tag{1}$$

Where, β_0 is the intercept and $\beta_1, \dots, \beta_k$ are the coefficients of the explanatory variables $X_1, \dots, X_k$, and $P(Y=1|X)$ represents the probability of the outcome occurring (e.g., an injury or fatal crash).

The logit transformation yields Eq. (2)

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \qquad (2)$$

Parameters in this model are estimated using Maximum Likelihood Estimation (MLE), which estimates the coefficient values that maximize the likelihood of the observed outcomes. MLE is preferred in logistic regression for its robustness and favorable asymptotic properties, including consistency and efficiency [51].

Another strength of logistic regression is its interpretability. The exponentiated coefficient e^{β_i} corresponds to the odds ratio (OR) for the variable X_i. An OR greater than 1 indicates that the variable increases the odds of the outcome, while an OR less than 1 implies a protective or reducing effect. This property makes logistic regression particularly useful in applied domains such as traffic safety, where policy implications often hinge on understanding relative risks [48, 49, 51, 52].

An additional advantage of the binary logistic regression model is its ability to accommodate both continuous and categorical variables without requiring them to follow a normal distribution. It also manages non-linear relationships by including interaction terms or transforming variables.

Table 1. Twenty randomly selected records from the HSIS dataset

Driver Sex	Vehicle Year	Hour Class	Severity	Road Surface	Lighting Condition	No. Vehicles	No. Lanes	Surface Type	Median Type	Access	Terrain	AADT	Rural/ Urban	Season	Age Group
F	-2000	2	PDO	Dry	Daylight	2	+8	PCC	Divided	Full Control	F	+250k	U	Spring	25-65
M	2010-2015	5	PDO	Dry	Daylight	2	+8	PCC	Divided	Full Control	F	175k-250k	U	Summer	25-65
F	2010-2015	5	PDO	Dry	Daylight	3	+8	PCC	Divided	Full Control	M	175k-250k	U	Summer	25-65
M	2000-2005	5	PDO	Dry	Dark	2	-5	PCC	Divided	Full Control	F	+250k	U	Winter	25-65
M	2000-2005	4	PDO	Snowy, Icy, Wet	Daylight	2	+8	AC	Divided	Full Control	F	+250k	U	Winter	-25
F	+2015	5	NotPDO	Dry	Daylight	3	+8	AC	Divided	Full Control	F	125k-175k	U	Summer	25-65
M	+2015	2	NotPDO	Dry	Daylight	+3	+8	PCC	Divided	Full Control	F	175k-250k	U	Fall	25-65
M	2010-2015	2	NotPDO	Dry	Daylight	3	+8	PCC	Divided	Full Control	F	-125K	U	Summer	25-65
M	+2015	2	NotPDO	Dry	Daylight	3	+8	PCC	Divided	Full Control	F	125k-175k	U	Fall	25-65
M	2010-2015	2	PDO	Dry	Daylight	2	6-7	PCC	Divided	Full Control	R	125k-175k	U	Winter	25-65
M	2005-2010	2	PDO	Dry	Daylight	+3	6-7	AC	Divided	Full Control	F	125k-175k	U	Fall	25-65
F	2010-2015	2	PDO	Dry	Daylight	3	6-7	PCC	Divided	Full Control	F	125k-175k	U	Summer	25-65
M	+2015	2	NotPDO	Dry	Daylight	3	+8	PCC	Divided	Full Control	F	+250k	U	Winter	25-65
M	2005-2010	5	NotPDO	Dry	Daylight	2	+8	PCC	Divided	Full Control	F	175k-250k	U	Summer	25-65
F	2005-2010	5	PDO	Dry	Dark	3	+8	PCC	Divided	Full Control	F	175k-250k	U	Winter	25-65
M	+2015	2	PDO	Dry	Daylight	2	+8	PCC	Divided	Full Control	F	+250k	U	Fall	25-65
F	+2015	5	PDO	Dry	Dark	2	+8	AC	Divided	Full Control	R	125k-175k	U	Winter	-25
F	2000-2005	2	NotPDO	Dry	Daylight	2	+8	PCC	Divided	Full Control	F	175k-250k	U	Fall	25-65
M	2010-2015	2	NotPDO	Dry	Daylight	2	+8	PCC	Divided	Full Control	F	+250k	U	Fall	25-65
F	2010-2015	4	NotPDO	Dry	Dark	2	+8	PCC	Divided	Full Control	F	175k-250k	U	Winter	25-65

Table 2. Frequency distribution of variables in the dataset

Variable	Category	Frequency (Percentage)
Driver Sex	Female	218739 (38.4%)
	Male	350647 (61.6%)
Vehicle Year	-2000	63591 (11.2%)
	2000-2005	104893 (18.4%)
	2005-2010	133005 (23.4%)
	2010-2015	165912 (29.1%)
	+2015	101985 (17.9%)
Hour Class	1	31396 (5.5%)
	2	193598 (34.0%)
	3	39856 (7.0%)
	4	126301 (22.2%)
	5	178235 (31.3%)
Severity	NotPDO	219378 (38.5%)
	PDO	350008 (61.5%)
Road Surface	Dry	541728 (95.1%)
	Snowy, Icy, Wet	27658 (4.9%)
Light	Dark	140648 (24.7%)
	Daylight	428738 (75.3%)
No. Vehicles	2	337758 (59.3%)
	3	158846 (27.9%)
	+3	72782 (12.8%)
No. Lanes	-5	97964 (17.2%)
	6-7	98618 (17.3%)
	+8	372804 (65.5%)
Surface Type	AC	179603 (31.5%)
	PCC	389783 (68.5%)
Median Type	Divided	548544 (96.3%)
	Undivided	20842 (3.7%)
Access	No Control	45258 (7.9%)
	Partial Control	11796 (2.1%)
	Full Control	512332 (90.0%)
Terrain	Flat	428564 (75.3%)
	Mountainous	20613 (3.6%)
	Rolling	120,209 (21.1%)
AADT	-125K	147,356 (25.9%)
	125k-175k	94,798 (16.6%)
	175k-250k	196,557 (34.5%)
	+250k	130,675 (23.0%)
Rural/Urban	Rural	47,236 (8.3%)
	Urban	522,150 (91.7%)
Season	Fall	155,624 (27.3%)
	Spring	137,047 (24.1%)
	Summer	144,304 (25.3%)
	Winter	132,411 (23.3%)
Age Group	-25	113,732 (20.0%)
	25-65	420,065 (73.8%)
	+65	35,589 (6.3%)

Common variables used in transportation safety literature include vehicle type, speed limit, lighting conditions, road geometry, driver demographics (e.g., age, sex), and environmental characteristics [49, 51-53].

In this study, binary logistic regression was applied to 569,386 rear-end crashes that occurred in California between 2015 and 2017. The binary outcome variable was defined as NotPDO (coded as 1) versus PDO (coded as 0). The dataset was divided into 24 disaggregated groups based on four key variables: driver age group (under 25, 25–65, and over 65), sex (male or female), area type (urban or rural), and lighting condition (daylight or dark). Each subgroup was modeled separately using a logistic regression model to evaluate how specific predictors influence crash severity in different contexts.

This stratified modeling framework not only offers a nuanced perspective on crash severity risks but also enables researchers and policymakers to tailor interventions for specific populations or environmental conditions. Analyzing crash severity separately for each of the 24 subgroups boosted interpretability and captured subgroup-specific effects that may be concealed in aggregated models. This approach aligned with the goal of identifying context-sensitive risk factors and supports the creation of more equitable and targeted traffic safety policies.

3.4. Model Evaluation and Goodness-of-Fit

To evaluate the performance and adequacy of the binary logistic regression models developed in this study, several goodness-of-fit and predictive accuracy metrics were considered. These included deviance, Akaike Information Criterion (AIC), McFadden's R^2, McKelvey and Zavoina's R^2, Cragg and Uhler's R^2, and the Area Under the Receiver Operating Characteristic Curve (AUC). Each metric provides a unique perspective on model performance, and together, they offer a robust assessment of validity.

Deviance is a general measure of model misfit and is based on the log-likelihood function. A lower deviance value indicates that the model fits the data more closely. It is analogous to the residual sum of squares in linear regression, but adapted for maximum likelihood estimation in logistic regression [48].

AIC further refines model comparison by penalizing complexity. AIC favors models with better fit while penalizing overfitting due to excessive parameters.

Among competing models, the one with the lowest AIC is preferred. It is calculated using Eq. (3)

$$\mathrm{AIC} = -2\log(L) + 2K \tag{3}$$

where L is the maximum likelihood and K is the number of estimated parameters [48].

McFadden's R^2 is one of the most widely used pseudo R^2 measures for logistic regression. It is defined as Eq. (4)

$$\mathrm{R}^2_{McFadden} = 1 - \frac{\log L_{full}}{\log L_{null}} \tag{4}$$

where $\log L_{full}$ is the log-likelihood of the fitted model and $\log L_{null}$ is the log-likelihood of the model with only an intercept. While its values are often lower than the R^2 from linear regression (with values of 0.2 to 0.4 considered excellent), it still offers a valuable index of relative improvement over a baseline model [51].

Cox and Snell's R^2 (R^2ML) attempts to replicate the interpretation of variance explained from linear regression but is constrained by a theoretical maximum below 1. To overcome this limitation, Cragg and Uhler's R^2 (R^2CU) adjusts Cox and Snell's measure to range from 0 to 1, improving interpretability across datasets and models of varying sizes [51]. These two measures are especially useful for understanding the proportion of outcome variability explained by the predictors in a likelihood-based framework. They provide broader context when McFadden's R^2 alone may seem too conservative.

AUC is a widely accepted measure of model discrimination, meaning the model's ability to distinguish between injury and property-damage-only outcomes. An

AUC of 0.5 indicates random classification, while 1.0 represents perfect separation [54]. In applied crash modeling, AUC values between 0.6 and 0.7 are considered acceptable, particularly when using observational data that includes natural noise and imbalances in class frequencies.

Together, these indicators suggest that the logistic regression models developed in this study provide a reasonable balance of fit, interpretability, and predictive capacity. Although the pseudo R^2 values were modest in magnitude, this was consistent with other transportation safety research relying on binary outcomes and large observational datasets. The AUC values further supported the model's utility in distinguishing crash severity outcomes.

4. Results

This section presents the results of binary logistic regression models developed to examine injury severity in rear-end collisions across stratified demographic and environmental subgroups. The dataset is divided into 24 distinct subgroup-specific models. These models are based on combinations of four key characteristics: driver sex (male or female), age group (<25, 25–65, 65+), lighting condition (daylight or dark), and the road environment (urban or rural). This disaggregated modeling approach is selected to reveal subtle patterns in crash severity that might be obscured in aggregated analyses.

Table 3 presents the performance metrics for all 24 stratified logistic regression models. These models are evaluated using several measures of fit and predictive ability, including model deviance, AIC, McFadden's pseudo R^2, McKelvey and Zavoina's R^2 (R^2ML), Cragg and Uhler's R^2 (R^2CU), and the area under the receiver operating characteristic curve (AUC).

4.1. Overview of Model Performance

The evaluation metrics presented in Table 3 reveal significant differences in both data distribution and model quality across the 24 stratified subgroups. The number of observations in each model refers to the total number of rear-end crashes used to estimate the injury severity model for that specific subgroup. These counts range from 178,069 crashes for male drivers aged 25-65 in urban daylight conditions (Model 1) to 165 crashes for female drivers aged 65 and over in rural areas during dark conditions (Model 24). While larger sample sizes provide greater statistical stability, they do not always guarantee improved model performance. For instance, despite the substantial volume of data in Model 1, it has an AUC of 0.622 and a McFadden's R^2 of only 0.033, moderate values compared to the smaller-sample Model 24, which attains the highest AUC (0.671) and McFadden R^2 (0.077) among all models. This contrast highlights how homogeneity and behavioral consistency within a subgroup may influence model performance more strongly than data volume alone.

To visually highlight model discrimination performance across all stratified subgroups, Figure 1 presents a heatmap of AUC values. The rows represent different lighting and area conditions, while the columns represent combinations of driver sex and age group. Darker shades indicate higher AUC scores, corresponding to better model performance.

To further evaluate model capability, we examine the discrimination power and explanatory strength of each model.

4.2 Model Discrimination and Explanatory Power

In terms of discrimination power, as measured by AUC, most models fell within the range of 0.60 to 0.62, which is generally accepted in the context of transportation safety modeling. These AUC values are common in transportation safety research, particularly due to the presence of unobserved behavioral heterogeneity in crash data. This heterogeneity, which arises from factors such as driver distraction, emotional state, or in-vehicle behavior, none of which are captured in typical datasets, often limits the predictive power of injury severity models. Nevertheless, certain subgroups demonstrated stronger model performance. Female drivers aged 65 and older consistently had higher AUC values across all lighting and environmental conditions (Models 11, 21, and 24), suggesting that crash injury outcomes in these groups are more systematically associated with measurable predictors. In contrast, younger drivers under 25, particularly males in rural dark environments (Model 20), exhibited notably lower model performance, with an AUC of just 0.552 and a McFadden R^2 of 0.007. This stark contrast underscores the difficulty in modeling injury severity among high-risk groups, where behavioral factors are likely more variable and less observable.

4.3. Evaluation of Goodness-of-Fit Metrics

The model deviance and AIC exhibited expected patterns that corresponded with the size of each subgroup. Larger models, such as Model 1 and Model 2, showed higher deviance and AIC values, primarily due to the greater volume of data. However, interpreting these metrics alongside R^2 and AUC clarifies that a lower deviance does not necessarily indicate stronger predictive power, particularly in behaviorally heterogeneous subgroups. The R^2ML and R^2CU values added further context. For example, the R^2CU exceeded 0.06 in only a few instances, primarily for elderly female drivers in rural areas, again emphasizing the predictability within this group. Notably, these subgroups had R^2ML values nearing 0.10, the highest in the dataset. This suggests that approximately 10% of the variation in

Table 3. Model Performance Metrics for Stratified Logistic Regression Models

Model	Driver Sex	Age Group	Observations	Deviance	AIC	McFadden R^2	R^2ML	R^2CU	AUC
Daylight and Urban:									
1	M	-25	42,860	53,455	53,491	0.028	0.035	0.048	0.609
2	F	-25	33,948	43,166	43,182	0.027	0.035	0.048	0.608
3	M	25-65	178,069	226,084	226,132	0.033	0.043	0.059	0.622

Model	Driver Sex	Age Group	Observations	Deviance	AIC	McFadden R^2	R^2ML	R^2CU	AUC
4	F	25-65	114,092	148,602	148,644	0.029	0.039	0.052	0.613
5	M	+65	16,853	22,148	22,168	0.032	0.043	0.058	0.618
6	F	+65	8,811	11,644	11,676	0.034	0.045	0.060	0.622
Daylight and Urban:									
7	M	-25	4,005	5,184	5,212	0.026	0.034	0.047	0.604
8	F	-25	2,828	3,698	3,706	0.030	0.039	0.053	0.604
9	M	25-65	16,034	21,015	21,037	0.029	0.039	0.053	0.606
10	F	25-65	8,207	10,940	10,978	0.028	0.038	0.051	0.609
11	M	+65	2,080	2,760	2,776	0.033	0.044	0.059	0.621
12	F	+65	951	1,251	1,259	0.041	0.054	0.073	0.625
Dark and Urban:									
13	M	-25	16,388	21,155	21,183	0.026	0.034	0.046	0.608
14	F	-25	11,151	14,475	14,505	0.031	0.040	0.054	0.618
15	M	25-65	60,548	79,494	79,542	0.025	0.034	0.046	0.605
16	F	25-65	33,254	44,049	44,091	0.026	0.035	0.048	0.608
17	M	+65	4,264	5,732	5,750	0.024	0.032	0.043	0.594
18	F	+65	1,912	2,552	2,580	0.037	0.049	0.066	0.624
Dark and Rural:									
19	M	-25	1,696	2,298	2,308	0.007	0.010	0.013	0.552
20	F	-25	856	1,111	1,130	0.035	0.046	0.062	0.615
21	M	25-65	7,297	9,778	9,804	0.022	0.029	0.039	0.599
22	F	25-65	2,564	3,435	3,455	0.025	0.034	0.045	0.598
23	M	+65	553	746	756	0.025	0.035	0.046	0.602
24	F	+65	165	201	213	0.077	0.097	0.132	0.671

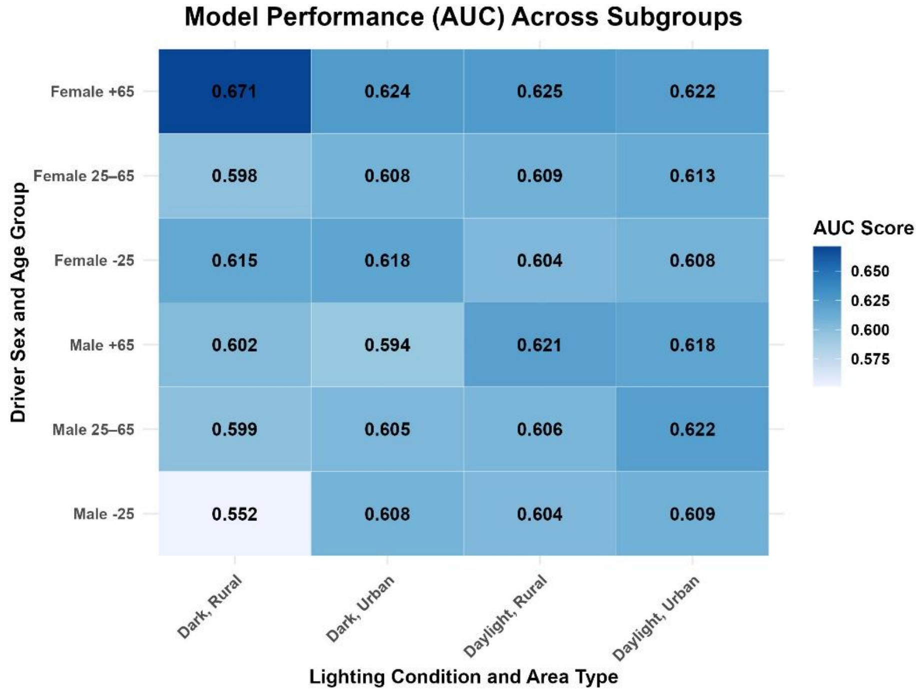

Figure 1. Heatmap of AUC Scores Across 24 Logistic Regression Models

crash severity within those groups was explained by the model, which is substantial for crash-level data.

4.4. Environmental and Demographic Patterns

An intriguing pattern emerged when comparing the urban and rural subgroups in terms of model quality. While the urban models had more observations and often reflected typical roadway characteristics (e.g., divided highways, signalized intersections), the rural models exhibited greater variation in terrain, lighting conditions, and access control. This environmental diversity likely contributed to improved model performance for elderly drivers in rural areas, where specific roadway features, such as slope, lighting, and access, interact directly with age-related driver vulnerabilities. For example, Model 21 (female, rural, daylight, +65) and Model 18 (male, rural, daylight, +65) achieved relatively high R^2 and AUC values, despite having sample sizes below 2,100. These results suggest that injuries in these groups are not only more likely but also more explainable based on measurable factors.

Another compelling observation is the effect of lighting conditions on different age groups. While dark lighting conditions generally reduced model performance for younger and middle-aged drivers, they seemed to enhance the predictability of crashes involving older adults. For instance, Model 24 (female, rural, dark, +65), Model 19 (female, urban, dark, +65), and Model 14 (male, urban, dark, +65) all exhibited relatively strong AUC and R^2 metrics compared to their respective counterparts. This may be attributed to increased caution, limited exposure, or the tendency to avoid night driving among older adults, leading to fewer but more structured crash events that are easier to model. In contrast, dark conditions likely introduce randomness for young drivers due to behavioral factors not captured in the dataset, such as fatigue, alcohol use, and risk-taking behaviors. To enhance the interpretation, we now analyze performance trends across specific demographic and environmental subgroups.

4.5. Subgroup-Specific Interpretations

Performance trends varied significantly across subgroups. Models for older female drivers, particularly in rural dark conditions, achieved the highest AUC and R^2 values, indicating more consistent and predictable injury outcomes. Conversely, younger male drivers, especially in rural dark settings, showed the weakest model performance, likely due to greater behavioral variability. Lighting and area type also influenced predictability, with rural and dark conditions generally associated with either clearer or more variable risk patterns, depending on the subgroup.

4.6. Predictor Effects Across Subgroup Models

To identify which variables most consistently influence injury severity in rear-end crashes, we analyze the results across various driver and environmental subgroups. Table 4 presents the descriptive statistics for all predictors that were statistically significant in at least one of the 24 estimated models. When interpreting the regression coefficients, a positive sign indicates that the variable is associated with a higher likelihood of injury, while a negative sign suggests a lower probability of injury in rear-end crashes. This directional insight helps identify factors that either increase or decrease crash severity across different subgroups. The variables are organized by thematic categories (e.g., crash characteristics, traffic volume and access control) to enhance interpretability. Within each group, the frequency of significance, average coefficient, and variability across models are provided to illustrate the strength and consistency of each predictor. For example, “No. Vehicles: +3” is statistically significant in 23 models, indicating a robust and

Table 4. Coefficient statistics for significant variables in crash severity regression modeling

Variable	Frequency	Mean Coefficient	Standard Deviation	Minimum Coefficient	Maximum Coefficient
Crash Characteristics:					
No. Vehicles: +3	23	-1.001	0.131	-1.244	-0.625
No. Vehicles: 3	22	-0.567	0.092	-0.874	-0.445
(Intercept)	16	0.275	0.445	-0.692	0.900
Traffic Volume and Access Control:					
AADT: 175k-250k	13	0.319	0.185	0.143	0.699
AADT: +250k	11	0.330	0.090	0.234	0.492
AADT: 125k-175k	8	0.196	0.289	-0.322	0.687
Access: Freeway Full Control	12	0.470	0.117	0.292	0.697
Access: Expressway Partial Control	3	-0.264	0.387	-0.556	0.175
Temporal Factors:					
Hour Class: 4	10	0.494	0.307	-0.208	0.897
Hour Class: 5	9	0.284	0.412	-0.282	1.054
Hour Class: 2	6	0.156	0.938	-1.063	1.515
Hour Class: 3	6	0.172	0.253	-0.305	0.389
Season: Summer	5	-0.064	0.221	-0.372	0.176
Season: Winter	3	0.360	0.529	0.055	0.971
Roadway Features:					
Terrain: R	11	0.015	0.297	-0.156	0.816
Terrain: M	9	0.350	0.392	0.117	1.365
No. Lanes: +8	7	-0.282	0.225	-0.716	-0.054
No. Lanes: 6-7	2	-0.403	0.314	-0.626	-0.181
Median Type: Undivided	6	-0.206	0.211	-0.459	0.170
Surface Type: PCC	1	0.031	N/A	0.031	0.031
Road Surface: Snowy, Icy, Wet	3	0.194	0.103	0.079	0.277

Variable	Frequency	Mean Coefficient	Standard Deviation	Minimum Coefficient	Maximum Coefficient
Vehicle Characteristics:					
Vehicle Year: +2015	9	0.184	0.163	-0.136	0.371
Vehicle Year: 2010-2015	7	0.138	0.106	-0.070	0.219
Vehicle Year: 2005-2010	5	0.055	0.113	-0.087	0.164
Vehicle Year: 2000-2005	1	0.119	N/A	0.119	0.119

recurring association with lower injury severity. This consistency suggests that it is among the most influential factors in determining injury severity across various combinations of driver sex, age group, and lighting conditions. The minimum and maximum coefficients for each variable illustrate the range of influence this variable has across contexts.

Some variables in Table 4 exhibit a change in coefficient sign across different models, showing both positive and negative effects depending on the subgroup. These cases are highlighted in red to indicate directional inconsistency. Such variations may stem from subgroup-specific differences in behavior, driving environments, or interactions with unobserved factors. For example, a variable may be associated with increased injury severity in one context but reduced severity in another, depending on the characteristics of the drivers or roadway conditions. This underscores the importance of stratified modeling and reflects the complex, context-dependent nature of crash severity outcomes.

Notably, for “No. Vehicles: +3”, the coefficients range from -1.244 to -0.625, always negative, indicating a strong and consistently significant association with reduced injury severity. This finding aligns with theoretical expectations that multi-vehicle rear-end crashes, while common, often distribute collision forces in a way that lessens individual injury risk.

In contrast, variables such as “Hour Class: 2” and “Season: Winter” exhibit wide ranges of effects, with minimum and maximum coefficients even switching signs. This suggests that their impact on injury severity is more context-dependent, sometimes increasing severity (e.g., winter night crashes with low visibility), and at other times showing no consistent pattern. Such inconsistencies may reflect interactions with unmeasured factors such as driver behavior, roadway treatment, or vehicle maneuverability in specific subgroups, and they warrant further targeted modeling.

Variables representing AADT levels, particularly “AADT: 175k-250k” and “AADT: +250k”, show positive mean coefficients of 0.319 and 0.330, respectively, across 13 and 11 models. These values indicate that high traffic volumes are associated with an increased risk of injury severity in rear-end crashes. The consistency and magnitude of these effects suggest a possible link to higher speeds and more complex driving environments. These factors may, in turn, amplify the impact of crashes.

Different access control types also reveal notable patterns. Full access control on freeways (Access: Freeway Full Control) is associated with a positive mean coefficient of 0.470 across 12 models. This finding supports the hypothesis that while limited access is generally intended to reduce crash frequency, it may also correspond to higher-speed environments, leading to more severe outcomes when crashes occur. Conversely, partial access control (Access: Expressway Partial Control) appeared in only three models, with a negative mean coefficient, suggesting it may function as a transitional design with mixed safety impacts.

Terrain classification (Terrain: M and Terrain: R) shows less consistency but still notable trends. “Terrain: M” has a relatively high mean coefficient of 0.350, while “Terrain: R” remains close to zero (0.015), indicating mixed or limited effects on injury severity. These effects may be influenced by region-specific factors such as road curvature, slope, or visibility, which are not directly captured in the model.

The time of the crash also plays a significant role in injury severity. “Hour Class: 4” (evening rush hours) has the highest positive mean coefficient among all time-based variables (0.494), suggesting that this period presents a heightened risk for severe injuries, likely due to congestion, reduced visibility, and driver fatigue. However, the large standard deviation (0.307) also suggests variability across subgroups, pointing to potential behavioral or environmental moderators.

The vehicle model year shows an interesting and somewhat counterintuitive pattern. While newer vehicles (Vehicle Year: +2015) are generally expected to reduce injury severity due to advanced safety features, they exhibit slightly positive mean coefficients in this dataset. These small but consistent coefficients (e.g., 0.184 for Vehicle Year: +2015) may suggest that newer vehicles are more frequently involved in higher-speed environments, or that certain safety features cannot fully mitigate the dynamics of severe rear-end collisions in such contexts.

Cross-section geometry, as indicated by lane count, has mixed results. Roads with eight or more lanes (No. Lanes: +8) show a negative coefficient (-0.282), while those with 6–7 lanes (No. Lanes: 6-7) have an even more negative mean (-0.403), although based on fewer observations. These findings may reflect the relationship between roadway width, speed expectations, and conflict points, although the specific mechanisms likely vary by roadway design and driver familiarity.

Environmental variables, such as season (Winter, Summer) and road surface (Snowy, Icy, Wet), show relatively inconsistent effects, as expected from their broad and regionalized nature. For instance, “Season: Winter” has a wide coefficient range, indicating that winter conditions can significantly elevate injury severity, but only in particular subgroups or contexts (e.g., rural locations, nighttime). Meanwhile, “Season: Summer” shows a slight negative mean, consistent with the idea that minor crashes occur more frequently during summer travel periods, possibly due to congestion and lower average speeds.

Finally, variables such as "Median Type: Undivided" and "Surface Type: PCC" appeared less frequently but demonstrated directionally consistent effects. "Median Type: Undivided" has a negative mean coefficient (–0.206), implying reduced injury severity in simpler road configurations. In contrast, "Surface Type: PCC" was included in only one model and is not generalizable.

In Table 4, some variables have "N/A" in the standard deviation (SD) column. These entries refer to variables that were statistically significant in only one model. Since standard deviation reflects variability across multiple estimates, it is undefined for variables that appear in only one model. Thus, the absence of variation results in a non-applicable (N/A) SD value.

Overall, the disaggregated modeling approach reveals significant differences in the predictability of injury severity in rear-end collisions across demographic and environmental subgroups. While most models demonstrate modest discrimination and explanatory power, partly limited by unobserved behavioral factors, certain populations, such as elderly female drivers, consistently show higher predictability. The observed trends underscore the value of stratified modeling in uncovering nuanced relationships and highlight the need for tailoring safety interventions to subgroup-specific risk profiles. These results provide a solid foundation for the following discussion and its associated policy implications.

5. Discussion

This study presents a comprehensive analysis of the factors influencing injury severity in rear-end collisions, using 24 logistic regression models adjusted for combinations of driver sex, driver age, lighting conditions, and area type. The results confirm that the diverse dynamics of rear-end crashes, particularly in relation to driver demographics and environmental contexts, require subgroup-specific modeling rather than a one-size-fits-all approach. The discussion below summarizes the key findings and places them in context with previous studies and real-world implications.

5.1. Variation in Model Explanatory Power

As reported in the results, model explanatory power varied widely across demographic and environmental subgroups. Notably, older female drivers in rural daylight conditions exhibited the highest predictability, while young male drivers in rural dark settings had the lowest. This disparity is not only statistical but also conceptually significant. The high predictability observed for older female drivers may reflect consistent, cautious driving behavior, limited exposure to complex conditions (e.g., avoiding nighttime or urban driving), and a closer alignment between recorded variables and real-world risk. In contrast, the low explanatory power for young male drivers likely reflects unobserved behavioral variability, including speeding, distraction, substance use, and aggression, factors not directly captured in crash records but known to influence injury outcomes. This finding highlights a critical methodological challenge: traditional data sources may fail to capture behavior-based risks among high-variability groups, emphasizing the need to integrate behavioral or telematics data in future research.

5.2. Key Predictors: Consistent and Counterintuitive Effects

Across the 24 stratified models, several predictors consistently influence injury severity in rear-end crashes. These variables appear frequently with stable coefficient signs, indicating their broad relevance. A few predictors also demonstrate counterintuitive trends that warrant further discussion.

Number of Involved Vehicles: Multi-vehicle crashes are consistently associated with lower injury severity. This aligns with previous findings that in chain-reaction crashes, impact forces are distributed across multiple vehicles, reducing injury risk for individual occupants [19, 55]. These crashes often occur in urban or congested conditions with lower speeds, which may also contribute to less severe outcomes.

Traffic Volume: Higher traffic volumes, especially on roads with AADT above 175,000, are linked to increased injury severity. Contributing factors include dense traffic, short headways, and increased speed variance, which reduce driver reaction time and raise crash impact levels [17, 56]. These findings emphasize the importance of traffic calming and speed management strategies on high-volume corridors.

Access-Controlled Freeways: Notably, crashes on fully access-controlled freeways are associated with higher injury severity in several models. While such roads generally reduce crash frequency through controlled access and steady flow, the crashes that occur may involve higher speeds and reduced attentiveness [57]. This highlights a paradox in freeway safety: fewer but more severe crashes.

Vehicle Model Year: Newer vehicles (+2015) show a slight but unexpected increase in injury odds. Despite advanced safety features, this pattern may reflect behavioral adaptation, known as risk compensation, where drivers take more risks due to a perceived increase in vehicle protection [58, 59]. It is also possible that newer vehicles are overrepresented on high-speed roads, which can diminish the effectiveness of safety technologies.

Roadway Geometry: Wider roads (six or more lanes) occasionally exhibit a protective effect, with lower injury severity. These roads offer greater maneuverability, multiple lanes to distribute traffic, and often better-designed infrastructure, which collectively enhance safety [17, 60]. Additionally, recurring congestion may reduce crash impact by limiting vehicle speeds.

Environmental Factors: Mountainous terrain increases crash severity, likely due to difficult geometry, reduced visibility, and limited safety infrastructure [61]. Other factors, such as rolling terrain or seasonal conditions, show inconsistent patterns across subgroups. This variability suggests complex interactions between environment, driver behavior, and road design.

5.3. Subgroup-Specific Patterns: Age, Sex, Lighting Conditions, and Area Type

Age emerges as one of the most influential factors in crash severity outcomes. Elderly drivers (+65) consistently exhibit higher injury severity across models, even in moderate crashes. This is attributed to physiological frailty and slower reaction times, which make recovery more difficult and injuries more likely [62].

In contrast, young drivers (<25) are involved in more frequent but less predictable crashes. The variability in model performance reflects inconsistent driving patterns, often influenced by overconfidence, distraction, or impairment [61]. Their involvement in nighttime and high-speed rural crashes further increases injury risk.

Middle-aged drivers (25–64) show moderate injury outcomes, with relatively balanced behavior and predictability. While not as high-risk as younger drivers or as physically vulnerable as older ones, they still benefit from subgroup-based analysis to account for varying roadway and temporal conditions.

Sex-based differences in crash severity are apparent in both the statistical models and behavioral patterns. Female drivers aged 65 and above produce models with the highest predictive accuracy, suggesting more consistent driving behaviors. This demographic tends to avoid high-risk settings, such as rural nighttime driving, and adheres more closely to traffic rules [62].

Male drivers under 25, on the other hand, have the most unpredictable crash outcomes. Their models consistently show weaker performance, indicating high variability in behavior and risk exposure. Risk-taking tendencies, such as mobile phone use and peer-influenced driving, are key contributors to this inconsistency [63].

Additionally, female drivers across all age groups are more likely to be injured, possibly due to physical vulnerability rather than behavioral factors. This aligns with broader crash literature and highlights the need to consider sex-based physiological differences when evaluating injury risk.

Lighting conditions significantly influence injury severity across all age and sex groups. Crashes occurring in darkness, both with and without street lighting, are associated with higher injury severity compared to crashes during daylight. This finding is consistent with prior studies and reflects the inherent risks posed by reduced visibility, such as limited reaction time and lower situational awareness [64].

Notably, the effect of lighting is more pronounced for young male drivers, whose nighttime crashes often occur under recreational or low-supervision circumstances. Behavioral factors like speeding, distracted driving, and alcohol use are especially prevalent in this group, compounding the risks of low-light environments [63].

Older drivers, in contrast, are less frequently involved in nighttime crashes, possibly due to self-regulation. However, when such crashes occur, their injury severity is elevated, likely due to age-related frailty and decreased nighttime vision [62].

The urban-rural divide has a significant impact on injury severity. Rural crashes are associated with higher severity due to longer emergency response times, higher travel speeds, limited lighting, and fewer roadway safety features [65]. These contextual disadvantages make rural areas a key target for infrastructure improvement and emergency response enhancements.

Urban crashes, although more frequent, typically result in lower injury severity due to traffic calming, shorter travel distances, and faster access to medical care. The frequent presence of traffic control devices and denser road networks also contributes to safer outcomes.

Importantly, the interplay between area type and demographic characteristics, such as the presence of young drivers in rural areas or older drivers in urban settings, proves critical in understanding injury severity patterns. These findings support the need for spatially targeted safety strategies that address both environmental and behavioral risk factors.

5.4. Infrastructure-Related Factors

Roadway infrastructure plays a crucial role in determining injury outcomes in rear-end collisions. Two key elements, pavement type and geographic terrain, consistently emerge as significant across models and warrant deeper discussion from a design perspective.

Pavement Type: Crashes on PCC surfaces are associated with higher injury severity. This may stem from PCC’s rigidity and reduced shock absorption compared to asphalt. In adverse weather conditions, PCC's lower skid resistance can further increase crash risk, particularly in high-speed corridors [17]. These findings suggest that materials chosen for roadway surfaces should be evaluated not only for durability but also for their role in impact energy absorption and crash recovery.

Terrain: Mountainous and hilly terrain substantially increases crash severity due to sharp curves, steep grades, and limited sight distances. These conditions challenge driver control and often lack compensatory infrastructure like shoulders, signage, or lighting. Previous studies recommend measures such as high-friction surface treatments, improved delineation, and slope stabilization to mitigate risks in these areas [65]. Integrating such solutions into roadway planning may be especially important in rural or isolated regions where emergency response times are longer.

These results support the integration of context-sensitive design strategies in transportation safety planning. Engineering improvements tailored to environmental and surface conditions can significantly reduce injury risk in high-severity zones.

5.5. Policy Implications

The findings of this study offer valuable insights for transportation policymakers, planners, and safety engineers seeking to mitigate injury severity in rear-end collisions. First, the consistent influence of variables such as traffic volume, number of vehicles involved, and access-controlled roadways indicates that high-risk roadway types can be proactively targeted for injury mitigation. Interventions such as dynamic speed management, improved road surface treatments, and enhanced pavement markings or lane delineation may help reduce crash impact severity, particularly in high-volume corridors and areas with challenging terrain.

Second, the subgroup-specific modeling approach highlights the need for tailored strategies that address the distinct risk profiles of demographic groups. For example, behavioral programs or in-vehicle technologies targeting young male drivers, especially in rural nighttime settings, could improve safety in ways that infrastructure changes alone may not achieve. Likewise, enhancing lighting and signage in rural and mountainous regions could substantially benefit older drivers, who remain physically vulnerable even in moderate-speed crashes.

For instance, the Washington State Department of Transportation implements Active Traffic Management strategies, including variable speed limits and lane control signage, on multiple freeways in Seattle and Vancouver, such as I-5, I-90, and SR-520. A case study on southbound I-5 in Vancouver reports a 24% reduction in rear-end crashes following the deployment of ATM systems [66]. Similarly, the Texas Department of Transportation's "Talk. Text. Crash." public education campaign, aimed at youth and young adults, addresses over 91,000 distracted-driving crashes in 2024, including 373 fatalities and 2,587 serious injuries [67]. These examples demonstrate how targeted infrastructure and behavioral interventions can effectively reduce subgroup-specific risks highlighted in this study.

Although the dataset used in this study originates from California, the implications extend beyond state boundaries. The behavioral and environmental patterns identified, such as the influence of lighting, area type, and driver age, likely apply to comparable urban, suburban, and rural contexts across the United States and potentially in other countries. Therefore, the evidence-based strategies derived from this research can be adapted and implemented across diverse settings to improve traffic safety outcomes.

Finally, the insights gained underscore the importance of integrating infrastructure design, enforcement, education, and vehicle technology into a comprehensive approach for reducing rear-end crash injury severity. Tailoring interventions to match context and demographic characteristics enhances effectiveness and supports more equitable, data-driven transportation safety policies.

6. Conclusion

This study applied 24 disaggregated binary logistic regression models to analyze rear-end crash injury severity using 569,386 crash records from California (2015–2017), stratified by driver sex, age, lighting condition, and area type. The results demonstrate that injury severity in rear-end collisions was influenced by a complex interaction of environmental, roadway, temporal, and driver-related factors, with distinct effects across subgroups.

Multi-vehicle crashes involving three or more vehicles were consistently linked to lower injury odds, likely due to distributed impact forces in chain-reaction scenarios. In contrast, high AADT levels and fully access-controlled freeway segments significantly increased injury severity, underscoring the role of speed and exposure in crash outcomes. Roadway conditions, especially those involving snowy, icy, or wet surfaces, and PCC pavement types were also associated with an increased risk of injury, highlighting the importance of surface friction and energy absorption. Terrain also played a major role, with rolling and mountainous areas increasing severity due to visibility limitations and challenging roadway geometry.

Lighting conditions and the timing of crashes emerge as key contributors, with nighttime and late-afternoon crashes, particularly in rural areas, consistently linked to higher injury severity. Seasonal effects, especially during winter, further amplified risks. The role of vehicle model year was complex; newer vehicles displayed an increased risk of injury in some subgroups, suggesting potential behavioral risk compensation despite safety advancements. Median type and access control also influenced outcomes, with undivided medians and partial-access roads associated with higher severity. In contrast, wider roads (eight lanes or more) were generally associated with lower severity, likely due to better traffic dispersion and roadway infrastructure.

Model performance, measured by AIC, deviance, pseudo R^2, and AUC, varied significantly across subgroups. The most robust model was associated with older female drivers in rural dark conditions, indicating more consistent behavior and predictable outcomes. In contrast, the weakest model was linked to young male drivers in the same environment, suggesting the presence of unmeasured behavioral factors such as speeding, distraction, or substance use. These findings highlight the limitations of aggregated models and emphasize the importance of disaggregated approaches in revealing context-specific risk dynamics.

By identifying both consistent and context-dependent predictors of injury severity, this study provides actionable insights for traffic safety policy, road design, and behavioral interventions. Tailored countermeasures, such as improved rural lighting, season-specific infrastructure treatments, and targeted behavioral programs for high-risk drivers, could be more effectively implemented using disaggregated model outputs. Overall, this research underscores the importance of subgroup-specific modeling in developing equitable, data-driven strategies to reduce the severity of rear-end collisions.

Limitation: This study provides valuable insights into rear-end collision severity using disaggregated logistic regression models; however, several limitations should be acknowledged. First, the analysis was based solely on crash data from California, which may limit generalizability to other regions with different traffic laws or driving behaviors. Second, key behavioral factors, such as distraction, seatbelt use, and impairment, were missing or only indirectly represented. This limitation likely decreased model accuracy, especially for high-risk groups like young drivers. For instance, a crash involving a young male driver late at night might have been caused by texting or alcohol use, but without this information, the model attributed the injury outcome solely to observed factors like lighting or road type. As a result, the actual impact of risky behaviors on driving injury severity might be underestimated or misrepresented. Third, the exclusive focus on rear-end collisions may not capture dynamics relevant to other crash types. Lastly, while the categorical binning of continuous variables was useful for interpretation, it might reduce granularity and obscure subtle nuanced trends.

Future works: To address these gaps, future research could incorporate behavioral data from telematics or naturalistic driving studies to better capture risk-related behaviors. Expanding the model to include other crash types (e.g., angle or sideswipe) would enable a broader application of the findings. Methodologically, using advanced techniques such as machine learning could uncover nonlinear patterns and complex interactions among predictors. Additionally, adopting continuous modeling approaches or refining binning strategies could enhance detail and precision. Longitudinal studies could also help assess how crash severity evolves over time in response to changing policies, technologies, or driver demographics.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

[1] World Health Organization, Global status report on road safety 2023, 2023.
[2] T. Stewart, Overview of motor vehicle traffic crashes in 2021, National Highway Traffic Safety Administration, 2023.
[3] M.A. Yellman, E.K. Sauber-Schatz, Motor vehicle crash deaths - United States and 28 other high-income countries, 2015 and 2019, Centers for Disease Control and Prevention (CDC), 2022.
[4] U.S. Census Bureau, QuickFacts: California, 2024.
[5] California Department of Transportation, California transportation plan 2040, 2016.
[6] California Office of Traffic Safety, California annual report 2023: fiscal year 2023, 2023.
[7] National Center for Statistics and Analysis, Traffic safety facts 2022: a compilation of motor vehicle traffic crash data, 2025.
[8] O.A. Abaza, M. Arafat, T.D. Chowdhury, Study on pedestrian road crossing compliance at high pedestrian crash locations of Anchorage, Alaska, International Conference on Transportation and Development, American Society of Civil Engineers, 2018, pp. 19–29.
[9] M. Arafat, S.R. Nafis, E. Sadeghvaziri, F. Tousif, A data-driven approach to calibrate microsimulation models based on the degree of saturation at signalized intersections, Transportation Research Interdisciplinary Perspectives 8 (2020) 100231.
[10] Insurance Information Institute, Facts + statistics: highway safety.
[11] World Health Organization, Road traffic injuries, 2023.
[12] M. Soltanirad, S. Naseralavi, K. Jimee, Q. Luo, T. Bataineh, M. Igene, H. Liu, Development of a new safety indicator for predictive safety analysis based on different arbitrary surrogate safety measures using LiDAR sensor data, Transportation Research Board 104th Annual Meeting, Transportation Research Board, Washington, D.C., United States, 2025.
[13] C.S. Parenteau, D.C. Viano, Size and age of fatal drivers by crash type, vehicle type and gender, Traffic Injury Prevention 24(3) (2023) 203–207.
[14] A. Ryan, T. Francis, F. Cole, G. Jennifer, R. Robin, M. Knodler, The impact of sex on motor vehicle crash injury outcomes, Journal of Transportation Safety & Security 14(5) (2022) 818–842.
[15] E. Dabbour, O. Dabbour, A.A. Martinez, Temporal stability of the factors related to the severity of drivers' injuries in rear-end collisions, Accident Analysis & Prevention 142 (2020).
[16] A. Ryan, M. Knodler, Influential crash conditions leading to injury differences experienced by female and male drivers, Journal of Transport & Health 24 (2022) 101293.
[17] M. Sharafeldin, A. Farid, K. Ksaibati, Injury severity analysis of rear-end crashes at signalized intersections, Sustainability 14(21) (2022) 13858.
[18] N.S. Kusumastutie, B. Patria, S. Kusrohmaniah, T.D. Hastjarjo, Drivers' decision-making when experiencing a traffic conflict: A scoping review, Transportation Research Record 2678(7) (2023) 950–964.
[19] F. Chen, M. Song, X. Ma, Investigation on the injury severity of drivers in rear-end collisions between cars using a random parameters bivariate ordered probit model, International Journal of Environmental Research and Public Health 16 (2019) 2632.
[20] R. Zou, H. Yang, W. Yu, H. Yu, C. Chen, G. Zhang, D.T. Ma, Analyzing driver injury severity in two-vehicle rear-end crashes considering leading-following configurations based on passenger car and light truck involvement, Accident Analysis & Prevention 193 (2023) 107298.
[21] M. Santolino, L. Céspedes, M. Ayuso, The impact of aging drivers and vehicles on the injury severity of crash victims, International Journal of Environmental Research and Public Health 19(24) (2022) 17097.
[22] A. Adebisi, J. Ma, J. Masaki, J. Sobanjo, Age-related differences in motor-vehicle crash severity in California, Safety 5(3) (2019) 48.
[23] A. Taheri, K. Switala, G. Fountas, A. Sheykhfard, N. Dadashzadeh, S. Müller, An empirical analysis of crash injury severity among young drivers in England: accounting for data imbalance, Applied Sciences 15(9) (2025).
[24] M. Yang, Q. Bao, Y. Shen, Q. Qu, R. Zhang, T. Han, H. Zhang, Determinants influencing alcohol-related two-vehicle crash severity: A multivariate Bayesian hierarchical random parameters correlated outcomes logit model, Analytic Methods in Accident Research 44 (2024) 100361.
[25] C. Wang, M. Abdel-Aty, L. Han, Effects of speed difference on injury severity of freeway rear-end crashes: Insights from correlated joint random parameters bivariate probit models and temporal instability, Analytic Methods in Accident Research 42 (2024) 100320.
[26] J. Xi, H. Guo, J. Tian, L. Liu, W. Sun, Analysis of influencing factors for rear-end collision on the freeway, Advances in Mechanical Engineering 11(7) (2019).
[27] Q. Luo, X. Chen, J. Yuan, X. Zang, J. Yang, J. Chen, Study and simulation analysis of vehicle rear-end collision model considering driver types, Journal of Advanced Transportation 2020 (2020) 1–11.
[28] S.S. Naseralavi, M. Soltanirad, A. Mazaheri, Modeling the severity of motor-vehicle accidents by different age groups of the driver using multinomial logit regression, Quarterly Journal of Transportation Engineering 14 (2023) 3061–3083.
[29] National Highway Traffic Safety Administration, Traffic safety facts annual report tables, 2022.
[30] M. Li, H. Xie, P. Shu, Study on the impact of traffic accidents in key areas of rural roads, Sustainability 13(14) (2021) 7802.
[31] T. Champahom, S. Jomnonkwao, D. Watthanaklang, A. Karoonsoontawong, V. Chatpattananan, V. Ratanavaraha, Applying hierarchical logistic models to compare urban and rural roadway modeling of severity of rear-end vehicular crashes, Accident Analysis & Prevention 141 (2020) 105537.
[32] F. Wei, Z. Cai, Y. Guo, P. Liu, Z. Wang, Z. Li, Analysis of roadside accident severity on rural and urban roadways, Intelligent Automation & Soft Computing 28 (2021) 753–767.
[33] H. Zhang, S. Li, C. Wu, Q. Zhang, Y. Wang, Predicting crash frequency for urban expressway considering collision types using real-time traffic data, Journal of Advanced Transportation 2020 (2020) 1–8.
[34] X. Wang, X. Zhang, F. Guo, Y. Gu, X. Zhu, Effect of daily car-following behaviors on urban roadway rear-end crashes and near-crashes: A naturalistic driving study, Accident Analysis & Prevention 164 (2022) 106502.
[35] J. Liu, J. Li, K. Wang, J. Zhao, H. Cong, P. He, Exploring factors affecting the severity of night-time vehicle accidents under low illumination conditions, Advances in Mechanical Engineering 11 (2019) 168781401984094.

[36] R. Swain, G.S. Larue, Looking back in the rearview: Insights into Queensland's rear-end crashes, Traffic Injury Prevention 25(2) (2024) 138–146.
[37] M. Rahmaninezhad Asil, H. Toroghi, I. Bargegol, Analysis of factors associated with traffic injury severity on urban roads in different lighting conditions, CRPASE: Transactions of Civil and Environmental Engineering 08 (2022).
[38] M. Hosseinpour, K. Haleem, Examining crash injury severity and barrier-hit outcomes from cable barriers and strong-post guardrails on Alabama's interstate highways, Journal of Safety Research 78 (2021) 155–169.
[39] A.K. Hoye, I.S. Hesjevoll, Traffic volume and crashes and how crash and road characteristics affect their relationship - A meta-analysis, Accident Analysis & Prevention 145 (2020) 105668.
[40] Y. Li, C. Liu, L. Ding, Impact of pavement conditions on crash severity, Accident Analysis & Prevention 59 (2013) 399–406.
[41] G. Shiran, R. Imaninasab, R. Khayamim, Crash severity analysis of highways based on multinomial logistic regression model, decision tree techniques, and artificial neural network: A modeling comparison, Sustainability 13(10) (2021) 5670.
[42] M.-H. Wang, Investigating the difference in factors contributing to the likelihood of motorcyclist fatalities in single motorcycle and multiple vehicle crashes, International Journal of Environmental Research and Public Health 19(14) (2022) 8411.
[43] L. Gong, W. Fan, Modeling single-vehicle run-off-road crash severity in rural areas: Accounting for unobserved heterogeneity and age difference, Accident Analysis & Prevention 101 (2017) 124–134.
[44] F.L. Mannering, C.R. Bhat, Analytic methods in accident research: Methodological frontier and future directions, Analytic Methods in Accident Research 1 (2014) 1–22.
[45] S. Naseralavi, M. Soltanirad, M. Igene, K. Jimee, E. Ranjbar, Determining factors affecting motorcycle crash severity based on time of day and season combinations, ENG Transactions 6 (2025) 1–11.
[46] Federal Highway Administration, Highway safety information system 2023, 2023.
[47] R.J. Porter, I. Hamilton, C. Chestnutt, K. Kersavage, T. Cohen, Highway safety information system guidebook for the California data files, Federal Highway Administration, Washington, D.C., 2024.
[48] D.W. Hosmer Jr., S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, 3rd ed., Wiley, 2013.
[49] A. Agresti, Categorical Data Analysis, 3rd ed., Wiley, 2013.
[50] M. Vahedi Saheli, Understanding the factors affecting urban vehicle-to-vehicle crash severity with focus on drivers' route familiarity, CRPASE: Transactions of Civil and Environmental Engineering 08 (2022).
[51] S. Menard, Logistic Regression: From Introductory to Advanced Concepts and Applications, SAGE Publications, Inc., Thousand Oaks, California, 2010.
[52] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer, New York, 2013.
[53] C.-Y.J. Peng, K.L. Lee, G.M. Ingersoll, An introduction to logistic regression analysis and reporting, The Journal of Educational Research 96(1) (2002) 3–14.
[54] K. Train, Discrete Choice Methods with Simulation, 2nd ed., Cambridge University Press, 2009.
[55] C. Wang, F. Chen, Y. Zhang, J. Cheng, Analysis of injury severity in rear-end crashes on an expressway involving different types of vehicles using random-parameters logit models with heterogeneity in means and variances, Transportation Letters 15(7) (2023) 742–753.
[56] S. Bogue, R. Paleti, L. Balan, A modified rank ordered logit model to analyze injury severity of occupants in multivehicle crashes, Analytic Methods in Accident Research 14 (2017) 22–40.
[57] M. Yu, C. Zheng, C. Ma, 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 (2020) 100126.
[58] National Highway Traffic Safety Administration, How vehicle age and model year relate to driver injury severity in fatal crashes, 2013.
[59] J.T. Shope, C.R. Bingham, Teen driving: Motor-vehicle crashes and factors that contribute, American Journal of Preventive Medicine 35(3, Supplement) (2008) S261–S271.
[60] O. Almutairi, Analyzing rear-end crash counts on Ohio interstate freeways using advanced multilevel modeling, Systems 12(10) (2024) 438.
[61] K. Sun, Y. Li, Q. Zeng, X. Wang, Exploring the factors affecting the injury severity of crashes on mountainous freeways with high proportions of heavy vehicles considering their heterogeneous and interactive effects, Sustainability 17(4) (2025) 1624.
[62] G. Li, E.R. Braver, L.-H. Chen, Fragility versus excessive crash involvement as determinants of high death rates per vehicle-mile of travel among older drivers, Accident Analysis & Prevention 35(2) (2003) 227–235.
[63] B.G. Simons-Morton, M.C. Ouimet, Z. Zhang, S.E. Klauer, S.E. Lee, J. Wang, R. Chen, P. Albert, T.A. Dingus, The effect of passengers and risk-taking friends on risky driving and crashes/near crashes among novice teenagers, Journal of Adolescent Health 49(6) (2011) 587–593.
[64] G. Zhang, K.K.W. Yau, G. Chen, Risk factors associated with traffic violations and accident severity in China, Accident Analysis & Prevention 59 (2013) 18–25.
[65] S. Islam, S.L. Jones, D. Dye, Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama, Accident Analysis & Prevention 67 (2014) 148–158.
[66] Washington State Department of Transportation, Active traffic management: Case study: I-5 ramp metering & ATM, Vancouver, WA., 2024.
[67] Texas Department of Transportation, Talk. Text. Crash. Campaign fact sheet, 2024.