

MASTER'S THESIS

Impact of Pre-crash Scenarios on Cyclists' Crash Severities: Comparison of Crashes Involving Autonomous Vehicles and Conventional Vehicles

Author:

Asca Anne Amondi

Supervisor:

Dr. Ramandeep Singh

Prof. Dr. Constantinos Antoniou

Date of submission: June 10, 2025

Acknowledgement

I would like to thank my supervisor, Dr. Ramandeep Sign for the tremendous support she provided during this thesis. I would also like to thank my parents, Jane and Naphtali Ogal, my siblings Rachel, Judy and Angella and my friend Alexander for their endless support.

Abstract

This study examines how pre-crash scenarios differ between autonomous vehicles (AVs) and conventional vehicles (CVs) in their interactions with cyclists and how the associated pre-crash scenarios impact injury severity. Drawing on real world AV-to-cyclist crashes reported by the California DMV and NHTSA and CV-to-cyclist crashes from San Francisco's TIMS database, a comprehensive pre-crash scenario is constructed that includes vehicle and cyclist manoeuvres and trajectories, roadway type and vehicle model year. After collapsing low frequency variable categories, a multinomial logistic regression is fitted to estimate odds ratios.

Key findings reveal that straight-ahead cyclist movements increase the odds of more severe crash outcomes with this effect being over three times larger in AV-to-Cyclist crash scenarios compared to CV-to-Cyclist crash scenarios. A stationary AV reduces minor injury odds by about 94%, although one-third of AV collisions occur at zero speed, highlighting a substantial risk during start/stop phases of AVs. Conventional vehicles proceeding straight or turning significantly reduce minor injury odds, effects not mirrored for AV manoeuvres. Vehicle model year and roadway type exert no significant impact and AV involvement in a crash increases serious injury risk by about 164% relative to CVs, with no comparable impact on minor injuries.

These findings highlight the need for enhanced AV perception of straight-through cyclist movements and detection of cyclists in stationary states.

Table of Contents

List of Figures	vi
List of Abbreviations	viii
1. Introduction.....	1
1.1. Background of the Study.....	1
1.2. Objective of the Study.....	2
2. Literature Review	3
2.1. Cycling Trends in the USA.....	3
2.1.1. State and City-Level Cycling Trends.....	3
2.1.2. Demographic Patterns in Cycling	4
2.1.3. Cycling Volumes and Travel Patterns.....	4
2.1.4. Micro-Mobility	4
2.2. Cycling Safety.....	5
2.3. Autonomous Vehicles and Cycling Safety	11
2.4. Pre-crash Scenario.....	15
2.5. Gaps in Literature and Contribution of the Study.....	18
3. Methodology	20
3.1. Data Collection and Collation.....	20
3.1.1. California Department of Motor Vehicles (CA DMV)	20
3.1.2. National Highway Traffic Safety Administration (NHTSA)	21
3.1.3. California Transportation Injury Mapping System (TIMS).....	21
3.2. Crash Severity Classification	22
3.3. Descriptive Analysis of Crashes	24
3.3.1. Imputation of Missing Data.....	25
3.4. Pre-crash Scenario Construction.....	28
3.5. Statistical Methodology	33
3.5.1. Correlation Analysis of Dependent and Independent Variables	33
3.5.2. Model Selection	35
3.5.3. Correlation Analysis of Independent Variables	37
3.5.4. Parameter Estimation.....	38

3.5.5. Model Validation	40
4. Results.....	41
4.1. Imputation of Missing Data	41
4.2. Descriptive Analysis of Crashes	43
4.2.1. Autonomous Vehicle-to-Cyclist Crashes	43
4.2.2. Conventional Vehicle-to-Cyclist Crashes	52
4.3. Parameter Estimation	57
The parameter estimates are assessed at 90% confidence interval where p value = 0.01	57
4.3.1. Multinomial Logit Model for AV-to-cyclist Crashes	57
4.3.2. Multinomial Logit Model for CV-to-cyclist Crashes	59
4.3.3. Multinomial Logit Model for pooled AV and CV datasets.....	60
4.4. Model Validation Results	62
5. Discussion	64
5.1. Descriptive Analysis of Crashes	64
5.1.1. Autonomous Vehicle-to-Cyclist Crashes	64
5.1.2. Conventional Vehicle-to-Cyclist Crashes	66
5.1.3. Comparison of AV-to-Cyclist and CV-to-Cyclist Crashes	67
5.2. Parameter Estimation	68
5.2.1. Comparison of AV-to-Cyclist and CV-to-Cyclist Crashes	68
5.3. Model Validation	69
6. Conclusion.....	70
6.1. Main Findings	70
6.2. Study Limitations and Future Research.....	70
Appendix.....	72
Declaration concerning the Master’s Thesis.....	80
List of References.....	81

List of Figures

Figure 1 : Levels of Vehicle Automation; Source: Adapted from <i>Automated Vehicles for Safety</i> . NHTSA, n.d.....	11
Figure 2 :Box plots of continuous predictors versus crash severity	34
Figure 3 :Box plots of categorical predictors versus crash severity	35
Figure 4 : Density Plot of Observed and Imputed Values for AV Pre-crash Speed (mph)	41
Figure 5 : Density Plot of Observed and Imputed Values for Posted Speed Limit (mph).....	42
Figure 6 : Density Plot of Observed and Imputed Values for Mileage	43
Figure 7 : Frequency distribution of crashes among vulnerable road users (VRU).....	44
Figure 8 : Frequency distribution of crash severity in AV-to-cyclist crashes.....	45
Figure 9 : Frequency distribution of cyclist pre-crash manoeuvre in AV-to-cyclist crashes ...	46
Figure 10 : Frequency distribution of AV pre-crash manoeuvre in AV-to-cyclist crashes.....	47
Figure 11 : Frequency distribution of AV-to-cyclist pre-crash trajectory.....	48
Figure 12 : Weather condition in AV-to-cyclist crashes.....	48
Figure 13 : Lighting condition in AV-to-cyclist crashes.....	49
Figure 14 : Frequency distribution of type of roadway in AV-to-cyclist crashes.....	49
Figure 15 : Distribution of vehicle mileage in AV-to-cyclist crashes.....	50
Figure 16 : Distribution of posted speed limit in AV-to-cyclist crashes	51
Figure 17 : Distribution of AV pre-crash speed in AV-to-cyclist crashes	51
Figure 18 : Frequency distribution of crash severity in CV-to-cyclist crashes	53
Figure 19 : Frequency distribution of cyclist pre-crash manoeuvre in CV-to-cyclist crashes.	53
Figure 20 : Frequency distribution of driver pre-crash manoeuvre in CV-to-cyclist crashes..	54
Figure 21 : Frequency distribution of CV-to-cyclist pre-crash trajectory.....	55
Figure 22 : Weather condition in CV-to-cyclist crashes	55
Figure 23 : Lighting conditions in CV-to-cyclist crashes	56
Figure 24 : Type of roadway in CV-to-cyclist crashes	56

List of Tables

Table 1: Crash Risk Factors Influencing Cyclist-Conventional Vehicle Crashes	10
Table 2: Crash Risk Factors Influencing Severity of Crashes Involving Autonomous Vehicles	14
Table 3: A compilation of existing research on the most frequent vehicle-to-cyclist pre-crash scenarios	17
Table 4: Missing data for critical variables in AV-cyclist crash data set	25
Table 5: Correlation test results between predictor variables and variables with missing data	27
Table 6: Pre-crash Scenario Construction for Autonomous Vehicle Crashes	31
Table 7: Pre-crash Scenario Construction for Conventional Vehicle Crashes	33
Table 8: Correlation test results between predictor variables	38
Table 9: Parameter Estimates for Minor Injury AV-to-cyclist crashes.....	57
Table 10: Parameter Estimates for Serious Injury AV-to-cyclist crashes.....	58
Table 11: Parameter Estimates for Minor Injury CV-to-cyclist crashes	59
Table 12: Parameter Estimates for Serious Injury CV-to-cyclist crashes.....	60
Table 13: Parameter Estimates for Minor Injury AV+CV-to-cyclist crashes.....	61
Table 14: Parameter Estimates for Serious Injury AV+CV-to-cyclist crashes	62

List of Abbreviations

ACS	American Community Survey
ADAS	Advanced Driving Assistance System
ADS	Automated Driving System
AV	Autonomous Vehicle
CDMV	California Department of Motor Vehicles
CV	Conventional (Human-driven) vehicle
FARS	Fatality Analysis Reporting System
MAIS	Maximum Abbreviated Injury Scale
MICE	Multivariate Imputation by Chained Equations
NHTSA	National Highway and Transport Safety Administration
OECD	Organization for Economic Cooperation and Development
TIMS	Transportation Injury Mapping System
USDOT	United States Department of Transportation.
WHO	World Health Organization

1. Introduction

1.1. Background of the Study

Cycling as an active and sustainable mode of transportation has in the recent years garnered significant attention from policy makers and researchers. The health benefits attributed to cycling coupled with its contribution to reduced emissions, decreased congestion and enhanced social equity are among the key drivers for the promotion of cycling. Governments and organizations worldwide are making efforts to encourage cycling among the populace. However, despite these efforts, the uptake of cycling remains relatively low compared to other modes of transportation. In the United States of America, the share of cycling as a travel mode between 2007 and 2018 stood at a national average of 0.6% (League of American Bicyclists, 2018). This figure is considerably lower than in other high-income countries, such as Germany, where bicycles accounted for 10% to 11% of all journeys in 2017 (Nobis, 2019).

While there are many deterrents to cycling, fear of traffic-related injuries and death has been identified by research as a major factor discouraging the uptake of cycling (Aldred & Goodman, 2018; Cubbin et al., 2024). This perceived risk is reinforced by real-world statistics which show high rates of cyclist injuries and fatalities globally. According to WHO (2019) , every year approximately 41,000 cyclists are killed worldwide while commuting to school, work, a shopping centre or home. An analysis of trends in cycling among Organization for Economic Cooperation and Development (OECD) member countries revealed that the United States has significantly higher cyclist fatality and serious injury rates per kilometre travelled compared to other member countries (Buehler & Pucher, 2016).

To ensure long-term uptake and sustainability of cycling as a travel mode, a holistic approach that not only highlights the benefits of cycling but also addresses the barriers that hinder its uptake is necessary. A critical component of this approach is an understanding of the factors that contribute to cycling-related injuries and fatalities and their impacts on crash outcomes. Cyclists are among the most vulnerable road users, as they frequently share infrastructure with motor vehicles while lacking any physical protection in the event of a crash. Additionally, the significant disparity in mass and speed between bicycles and motor vehicles amplifies the risk of injury for cyclists in collisions. Unlike motor vehicles, bicycles pose minimal danger to other road users, as impacts involving bicycles rarely result in injury or fatality for involved parties.

In recent years, research on autonomous vehicle (AV) technology has intensified, driven by its potential to significantly reduce crash frequency and severity, by eliminating driver error which accounts for approximately 94% of road crashes (Almaskati et al., 2024; Ren et al., 2022; Singh, 2018). While AV technology continues to evolve through ongoing testing and development, research specifically examining its impact on the safety of vulnerable road users (VRUs), including pedestrians and cyclists, remains scarce. Majority of existing research has focused on safety implications of AVs for motorized vehicles. A key challenge in assessing the safety impact of AVs for VRUs is the limited availability of empirical crash data, which has hindered comprehensive analysis. As a result, most of the existing literature has relied on simulations, virtual reality models or survey-based studies to predict the impacts of AV technology on collisions involving VRUs (Das, 2021; Tafidis et al., 2019). Moreover, existing research on AV-VRUs safety to date has focused on AV-pedestrian interactions and safety, while studies examining AV-cyclist interaction and safety remain particularly limited. The lack of real-world data makes it difficult to draw definitive conclusions on the extent to which AV technology could impact cyclist safety. Given the unique vulnerabilities cyclists face in mixed-traffic environments, further research is needed to bridge the gaps in existing research. In addition, the factors influencing crash type and severity in autonomous vehicles (AVs) may differ from those in human-driven vehicles, since autonomous driving is integrated and systematically based on big data and artificial intelligence, whereas conventional driving is personalized (Kockelman et al., 2016). Therefore, it is meaningful to explore the divergent effects of factors affecting cyclist safety for autonomous and conventional driving using crash data.

1.2. Objective of the Study

This study aims to address gaps in existing literature by answering the following research questions:

1. How do pre-crash scenarios involving cyclists differ between autonomous and conventional vehicles?
2. How do pre-crash scenarios involving cyclists impact crash outcomes?

2. Literature Review

2.1. Cycling Trends in the USA

Cycling in the United States has experienced significant developments in recent years with several cities implementing initiatives to promote and integrate cycling into urban mobility. However, cycling still has the least modal share compared to other transportation modes such as private vehicle and public transportation which have modal shares of 78.7% and 3.5% respectively (U.S. Census Bureau, 2023b). According to the League of American Bicyclists (2018), the share of commuters who bicycle to work in the United States averaged 0.6% nationally between 2007 and 2018. The lowest recorded average was 0.4% in 2007, while the highest was 0.6% from 2012 to 2018. Among the fifty largest U.S. cities, bicycle commuting rates were consistently higher than the national average, ranging from 0.7% in 2007 to 1.2% in 2018, indicating that urban areas generally exhibit higher cycling rates. In comparison, commuting by walking had a significantly higher national average of 2.7% during the same period. On the other hand, annual data from the American Community Survey (ACS) indicates a 50% increase in bicycle commuting between 2005 and 2016, rising from 4% to 6%. However, the League of American Bicyclists argues that the ACS data does not fully capture actual cycling rates due to limitations such as seasonal variations, non-commute trips, and multi-modal travel, which are not adequately accounted for in the surveys.

2.1.1. State and City-Level Cycling Trends

Bicycle commuting rates vary widely across U.S. states. In 2016, 1% of commuters in California bicycled to work, while Oregon had the highest rate among all states at 2.2%. Between 2007 and 2016, several states saw significant growth in cycling rates, with California experiencing a 27.1% increase and New York leading with a 70.6% increase (League of American Bicyclists, 2018). Among the nation's largest cities, the American Community Survey (ACS) reported that between 2008 and 2012, San Francisco and Seattle had the highest bicycle commuting rates at 3.4%, followed by Sacramento at 2.5% (U.S. Census Bureau, 2014). Notably, both San Francisco and Sacramento are in California thus highlighting the state's strong cycling culture.

2.1.2. Demographic Patterns in Cycling

Women are underrepresented among bicyclists in the United States. While they represent 50.9% of the population and 47.0% of all commuters, they account for only 30.3% of all bicycling trips and only 28.0% of bicycle commuters (League of American Bicyclists, 2018; Reish, 2021; U.S. Census Bureau, 2014, 2023a). Nevertheless, these figures reflect a gradual increase in women's participation in bicycling compared to previous years (U.S. Census Bureau, 2023a).

Age demographics also influence cycling trends. According to the 2017 National Household Travel Survey (NHTS), children and youth under the age of 16 accounted for 22.1% of bicycle trips, aligning closely with their 21.2% share of the U.S. population based on the 2020 Census. In contrast, senior citizens aged 65 and older accounted for 6% of all bicycle trips while representing 16.8 % of the population (McGuckin & Fucci A., 2018; U.S. Census Bureau, 2023a) (U.S. Census Bureau, 2023a)

2.1.3. Cycling Volumes and Travel Patterns

Nationwide, 3,789 million daily bicycling trips were recorded in 2017 with a total of 78 billion minutes spent cycling and 8.5 billion miles covered (McGuckin & Fucci A., 2018). Between 2008 and 2012, the average travel time by bicycle was 19.3 minutes with bicycle commuting being most prevalent for commutes between 10 and 14 minutes in length. The highest rate of bicycling commuting occurred between 9.00 a.m. and 11.59 a.m. (U.S. Census Bureau, 2014)

2.1.4. Micro-Mobility

While cycling and walking remain the primary means of non-motorized transportation, the use of micro-mobility devices such as skateboards, roller skates, and scooters is growing, particularly among younger age groups. In California users of these devices travel an estimated 48 million miles annually (Fang & Handy, 2019). The increasing popularity of micro-mobility highlights evolving trends in urban mobility and emphasizes the need for road safety interventions to integrate these emerging modes of transportation.

2.2. Cycling Safety

The growing shift from motorized travel to public transport, walking and cycling has made cycling safety a key focus of road safety initiatives (WHO, 2020). However, assessing bicycle-related injury trends remains challenging due to significant underreporting of non-fatal crashes. While police crash reports are the primary tool for evaluating traffic collisions and informing mitigation policies, they often fail to capture the full extent of cycling injuries (Lopez et al., 2012; Shinar et al., 2018).

Between 2013 and 2022, cyclist fatalities accounted for 2.2% to 2.6% of total traffic deaths in the USA (National Center for Statistics and Analysis, 2024). The League of American Bicyclists (2018) notes that while bicyclists represent over 2% of traffic fatalities, they only account for about 1% of trips. These deaths are increasingly concentrated in urban areas (National Center for Statistics and Analysis, 2024), while severe cyclist crashes occur more frequently in rural areas, where higher speed limits contribute to increased injury severity (Scarano et al., 2023).

Road classification, design and infrastructure significantly impact cycling safety. Major roads pose higher risks than minor roads, while designated cycling infrastructure such as on-road marked cycle lanes and off-road bike paths effectively reduce crash and injury risks (Ghoul & Sayed, 2025; Reynolds et al., 2009). Although streets with cycle tracks experience a higher concentration of injuries due to greater bicycle volumes, the individual risk per cyclist is lower, underscoring the benefits of separate cycling infrastructure (Strauss et al., 2015). Notably, most cyclist fatalities in the USA occur on principal or minor arterial roads, which constitute only 10% of the national roadway system (League of American Bicyclists, 2018). Additionally, damaged and uneven road surfaces, steep grades and sharp horizontal curves, elevate the likelihood of severe cyclist injuries (Robartes & Chen, 2017; Yan et al., 2011).

Intersections and non-intersections present different risk factors for cyclists. Most cyclist fatalities occur at non-intersections (Dash et al., 2022; National Center for Statistics and Analysis, 2023, 2024). Crash severity at intersections is influenced by factors such as helmet use and road geometry, with injuries worsening in the absence of helmets or on steep, curved roads. Conversely, crash severity decreases when bicycle-vehicle collisions involve front-to-rear impacts. At non-intersections, alcohol use and older age increase injury severity, while crashes involving the front of the bicycle and the vehicle's side tend to be less severe. Across all locations, factors such as high speeds, heavy trucks, driver intoxication, and side-impact

collisions heighten injury risks (Moore et al., 2011). Type of traffic control at intersections impact crash severities. Crashes at signalized intersections are generally less severe than those elsewhere (Eluru et al., 2008). However, some studies suggest that signalized intersections, which are often located along major arterials, experience higher injury rates and greater risk than non-signalized intersections (Chimba et al., 2012; Scarano et al., 2023; Strauss et al., 2015).

Lighting and environmental conditions play a significant role in cycling safety. Although most cyclist crashes occur during daylight and in clear weather conditions (Wisch et al., 2019), night-time and poor lighting conditions result in a higher probability of severe injuries (Chimba et al., 2012; Dash et al., 2022; Eluru et al., 2008; National Center for Statistics and Analysis, 2024). The absence of street lighting doubles the likelihood of fatal bicycle injuries (Dash et al., 2022). Additionally, cloudy or foggy weather and obstructed vision increase the severity of cyclist injuries (Chimba et al., 2012; Robartes & Chen, 2017; Wang et al., 2015). While most crashes occur on dry road surfaces, wet road conditions exacerbate injury severity (Kaplan et al., 2014; Rash-ha Wahi et al., 2018; Wang et al., 2015).

Cyclist crash patterns vary based on time of day and day of the week. Crashes are more frequent on weekdays than weekends (Filipović et al., 2022; National Center for Statistics and Analysis, 2024). Additionally, most crashes occur during daylight hours, particularly in the late afternoon (Filipović et al., 2022; National Center for Statistics and Analysis, 2024).

Demographic factors including age and gender influence cyclist injury risks. Male cyclists have higher fatality and injury rates than female cyclists (Filipović et al., 2022; National Center for Statistics and Analysis, 2024; Wisch et al., 2019). The risk of fatal crashes is particularly elevated among younger and male drivers, who tend to engage in riskier driving behaviour (Scarano et al., 2023). Older cyclists have the highest fatality rates (Eluru et al., 2008; Filipović et al., 2022; National Center for Statistics and Analysis, 2024; Wisch et al., 2019), while younger cyclists have the highest injury rates (Kaplan et al., 2014; National Center for Statistics and Analysis, 2024; Wang et al., 2015).

Alcohol consumption by motor vehicle driver or the cyclist significantly increases cyclist injury severity (Kaplan et al., 2014; National Center for Statistics and Analysis, 2024). Intoxicated drivers elevate cyclist fatality risk sixfold and double the probability of severe injuries, while alcohol-impaired cyclists also face heightened fatality and severe injury risks (Robartes & Chen, 2017).

The type of vehicle involved in a crash influences injury severity. Cyclists struck by heavy vehicles, including trucks, buses, and SUVs, sustain more severe injuries compared to those hit by lighter vehicles (Eluru et al., 2008; Robartes & Chen, 2017; Yan et al., 2011). The direction of impact also affects injury outcomes, with frontal impacts resulting in more severe crashes than other impact types (Eluru et al., 2008).

Protective measures such as helmet use and high-visibility clothing significantly reduce injury risk (Moore et al., 2011; Zhuo Wang et al., 2021). Bicycle helmets provide significant protection against head and mid-face injuries, though they offer limited protection for the lower face (Thompson et al., 1996). Hard-shell helmets reduce head injury risks, while foam helmets have been found to increase face injury risks among children under nine years old (Hansen et al., 2003). Reflective jackets, vests and knee or ankle reflectors enhance cyclist conspicuity, especially at night, reducing the risk of multi-party crashes (Lahrmann et al., 2018; Wood et al., 2012).

Cyclist Behaviour plays a crucial role in crash occurrences. Failure by cyclists to yield right-of-way, disobeying traffic signs, poor visibility of cyclists to automobile drivers and riding against traffic significantly contribute to cyclist injuries and fatalities (Hamann et al., 2015; National Center for Statistics and Analysis, 2023; Yan et al., 2011; Zhuo Wang et al., 2021). Moreover, higher speed limits and excessive speeding by both cyclists and motorists significantly increase crash severity (Eluru et al., 2008; Isaksson-Hellman & Töreki, 2019; Kaplan et al., 2014; Robartes & Chen, 2017).

Like cyclists, micro-mobility devices such as scooters and skateboards also face safety challenges. Roads with speed limits exceeding 40 mph, rider violations, and intersection crashes significantly increase their risk of severe injuries (Qawasmeh et al., 2024). E-scooter riders are particularly vulnerable in mixed traffic, especially in work zones and around distracted or aggressive drivers. They also face a higher risk of severe injuries compared to bicyclists (Niemann et al., 2023).

In conclusion, factors influencing severity of cyclist-conventional vehicle crashes can be divided into six broad categories namely: roadway factors, environmental factors, cyclist factors, driver factors, motorized vehicle factors and crash factors.

The key findings from the literature on these risk factors are summarized in table 1 below.

Crash Risk Factor		Reference
Roadway factors	Major (arterial) road	Reynolds et al. (2009); League of American Bicyclists (2018)
	Higher speed limits	Isaksson-Hellman and Töreki (2019); Eluru et al. (2008) ; Yan et al. (2011); Kaplan et al. (2014)
	Damaged and uneven road surface	Yan et al. (2011)
	Curved and sloped road sections	Chimba et al. (2012); Robartes and Chen (2017)
	Non-intersection road sections	Dash et al. (2022); National Center for Statistics and Analysis (2023) ; National Center for Statistics and Analysis (2024)
	Signalized intersection	Chimba et al. (2012) ; Scarano et al. (2023); Strauss et al. (2015)
	Rural road	Scarano et al. (2023); Wisch et al. (2019)
Environmental factors	Night-time	Eluru et al. (2008)
	Inadequate lighting/Unlit roads	Eluru et al. (2008); Chimba et al. (2012); Dash et al. (2022)
	Weekday	Filipović et al. (2022)
	Foggy and rainy weather	Chimba et al. (2012); Wang et al. ; (2015) Filipović et al. (2022)
	Wet and slippery road surface	Kaplan et al. (2014); Wang et al. (2015)
Cyclist Factors	Older cyclist, age > 55	Wang et al. (2015); Filipović et al. (2022); Scarano et al. (2023); Yan et al. (2011); Kaplan et al. (2014); Wisch et al. (2019)
	Younger cyclist, age < 16	Wang et al. (2015); Kaplan et al. (2014); National Center for Statistics and Analysis (2024);

Crash Risk Factor**Reference**

Crash Risk Factor	Reference	
	Male cyclist	Wisch et al. (2019); Filipović et al. (2022); National Center for Statistics and Analysis (2024)
	Alcohol-impaired state	Robartes and Chen (2017); Kaplan et al. (2014) ; National Center for Statistics and Analysis (2024)
	Not wearing helmet	Hansen et al. (2003); Thompson et al. (1996); Moore et al. (2011); Zhuo Wang et al. (2021)
	Not wearing reflective clothing	Lahrmann et al. (2018); Wood et al. (2012)
	Failure to yield right-of-way	Yan et al. (2011) ; Hamann et al. (2015); National Center for Statistics and Analysis (2023)
	Red light violations	Zhuo Wang et al. (2021)
	Riding on wrong side of the road	National Center for Statistics and Analysis (2023)
	Speeding	Robartes and Chen (2017)
	Scooterist or Skateboarder	Niemann et al. (2023)
Driver Factors	Younger driver	Scarano et al. (2023)
	Male driver	Scarano et al. (2023)
	Alcohol-impaired state	Robartes and Chen (2017); National Center for Statistics and Analysis (2024)
	Speeding	Robartes and Chen (2017); Scarano et al. (2023)
Motorized Vehicle Factors	Large and heavy goods vehicle	Yan et al. (2011); Robartes and Chen (2017); Kaplan et al. (2014)

Crash Risk Factor**Reference**

Crash Risk Factor	Reference
Crash Factors	Frontal impact (Head-on collision) Eluru et al. (2008); National Center for Statistics and Analysis (2023) ;National Center for Statistics and Analysis (2024); Yan et al. (2011)
	Broad side impact (angular collision) Yan et al. (2011)

Table 1: Crash Risk Factors Influencing Cyclist-Conventional Vehicle Crashes

2.3. Autonomous Vehicles and Cycling Safety

The Society of Automotive Engineers (SAE) defines six levels of vehicle automation, ranging from 0 to 5 as illustrated in the figure below.







0	1	2	3	4	5
					
Momentary Driver Assistance	Driver Assistance	Additional Driver Assistance	Conditional Automation	High Automation	Full Automation
Driver performs all driving tasks while system provides momentary assistance like warnings or emergency safety interventions	Driver performs all driving tasks while system provides continuous assistance with either steering or acceleration/braking	Driver performs all driving tasks while system provides continuous assistance with both steering and acceleration/braking	System handles all aspects of driving while driver remains available to take over driving if system can no longer operate	When engaged, system is fully responsible for driving tasks under certain conditions. A human driver is not needed to operate the vehicle	When engaged, system is fully responsible for driving tasks under all conditions. A human driver is not needed to operate the vehicle

Figure 1 : Levels of Vehicle Automation; Source: Adapted from *Automated Vehicles for Safety*. NHTSA, n.d.

Advanced Driver Assistance Systems (ADAS) correspond to levels 1 and 2, while Automated Driving System (ADS) encompasses levels 3 to 5 of driving automation technology (*Automated Vehicles for Safety*. NHTSA, n.d.). A vehicle equipped with ADS is designed to perform the entire dynamic driving task on a sustained basis within a defined operational design domain without driver involvement. While these vehicles are in development and are being tested on public roads in limited capacities, they are not available for consumer purchase currently (*Automated Vehicles for Safety*. NHTSA, n.d.).

Autonomous Vehicles (AV) are expected to enhance road safety by reducing human error, a leading cause of road crashes (Fagnant & Kockelman, 2015; Singh, 2018). However, their

potential benefits remain largely unproven in real-world conditions as new risks continue to emerge alongside technological advancements. A major challenge in evaluating AV safety, particularly in relation to vulnerable road users (VRUs), is the scarcity of crash data (Kutela et al., 2022).

To address the lack of historical crash data, researchers have employed alternative methods, including modelling and simulation, to assess AVs' impact on cyclist safety. While simulations are commonly used to investigate AV interactions and support their development, they often lack accuracy due to reliance on assumptions (Tafidis et al., 2019). In a traffic simulation study modelling cyclist behaviour, Tafidis et al. (2019) found that AVs significantly reduce the total number and severity of conflicts between vehicles and cyclists. Additionally, conflict-based models, which extrapolate observed conflicts to estimate unobserved collisions, have been utilized for cyclist crash estimation (Ghoul & Sayed, 2025).

Although a few studies have used real-world AV crash data to examine AV-cyclist interactions (Boggs et al., 2020; Das, 2021; Kutela et al., 2022), the existing literature remains limited, with most research focusing on AV collisions involving motorized vehicles. Prior studies have primarily analyzed crashes between AVs and automobiles (Abdel-Aty & Ding, 2024; Almaskati et al., 2024; Boggs et al., 2020; Channamallu et al., 2024; Fu et al., 2025; Kohanpour et al., 2024; Kuo et al., 2024), and to a lesser extent, crashes between AVs and pedestrians (Fu et al., 2025; Kutela et al., 2022).

Factors influencing AV crash types and severity differ from those in human-driven vehicles (Abdel-Aty & Ding, 2024; Kockelman et al., 2016; Ren et al., 2022). While AVs are generally not at fault in most crashes, they exhibit a higher propensity for rear-end collisions compared to conventional vehicles (Almaskati et al., 2024). Research has identified various predictors of AV crash severity. Channamallu et al. (2024) and Kuo et al. (2024) highlighted factors such as AV manufacturer, vehicle type, collision type and AV movement prior to impact as key determinants of crash severity. Conversely, the number of vehicles involved, the condition of the other party's vehicle and weather and road surface conditions were found to have minimal impact on crash severities (Channamallu et al., 2024). Contrary to the later finding on the influence of environmental conditions on AV crash severity, Fu et al. (2025) noted that adverse weather (e.g., cloudy, foggy, rainy, or snowy conditions) and non-dry road surfaces increase the likelihood of severe injury crashes.

Kohanpour et al. (2024) found that most AV-involved crashes are rear-end collisions, while angular crashes are more common at night. Moreover, angular and head-on collisions pose the highest injury risk, particularly for VRUs such as motorcyclists, cyclists, and scooterists. While ADS disengagement helps reduce rear-end and object collisions, it also leads to an increase in other crash types. Poor AV sensor performance in low-visibility conditions, including darkness, fog, and glare, further exacerbates crash risks.

Among vulnerable road users (VRUs), bicyclists and scooterists are more likely to be directly involved in AV crashes, with bicyclists often found at fault, while pedestrians tend to be involved in AV collisions indirectly (Kutela et al., 2022). Intersection-related factors, such as crosswalks, traffic signals, and AV manoeuvres (e.g., turning, slowing down, stopping), are key predictors of AV-VRU crashes (Kutela et al., 2022). Additionally, lighting conditions, road type, weather and road surface conditions significantly impact the severity of AV-involved crashes, with the likelihood of severe injuries increasing when AVs collide with Vulnerable Road Users (VRUs) on urban streets (Fu et al., 2025).

In conclusion, factors influencing severity of crashes involving autonomous vehicles can be divided into five broad categories namely: roadway factors, environmental factors, autonomous vehicle factors, crash factors and Vulnerable Road Users (VRUs).

The key findings from the literature on these risk factors are summarized in the table below.

Crash Risk Factor		Reference
Roadway factors	Road type ¹	Fu et al. (2025)
	Road surface condition ¹	Fu et al. (2025); Channamallu et al. (2024)
	Urban street ¹	Fu et al. (2025)
	Intersection ²	Kutela et al. (2022)
	Signalized intersection ²	Kutela et al. (2022)
	Crosswalk ²	Kutela et al. (2022)
Environmental factors	Cloudy or foggy weather ¹	Kohanpour et al. (2024); Fu et al. (2025)
	Rainy or snowy weather ¹	Fu et al. (2025)
	Wet and slippery road surface ¹	Fu et al. (2025)
	Nighttime / Darkness ¹	Kohanpour et al. (2024)
Autonomous Vehicle factors	AV manufacturer ¹	Channamallu et al. (2024); Kuo et al. (2024)
	Vehicle type ¹	Channamallu et al. (2024); Kuo et al. (2024)
	Collision type ¹	Channamallu et al. (2024); Kuo et al. (2024)
Crash factors	AV pre-crash manoeuvre ²	Kohanpour et al.; Kutela et al. (2022); Channamallu et al. (2024); Kuo et al. (2024)
	Head-on collision ²	Kohanpour et al. (2024)
	Broad side (angular) collision ²	Kohanpour et al. (2024)
Vulnerable Road Users	Cyclist /Scooterist/Motor cyclist/ Pedestrian ²	Kohanpour et al. (2024); Fu et al. (2025)

Table 2: Crash Risk Factors Influencing Severity of Crashes Involving Autonomous Vehicles

¹ Represent crash risk factors that are related to AV crashes in general

² Represent crash risk factors that are specific to Cyclist-AV crashes

2.4. Pre-crash Scenario

Pre-crash scenarios are critical for understanding factors contributing to crashes and developing effective countermeasures. Klug et al. (2020) define pre-crash scenario as specification of a conflict situation and all crash parameters, including a detailed description of the environment (roadway, lighting and weather conditions), as well as velocity profile and trajectory of road users. Najm G. et al. (2007) further elaborates that pre-crash scenarios not only describe crash types but also the vehicle movements and critical events leading up to a collision.

Crash causation analysis based on pre-crash scenarios can provide more specific identification of the causes of crashes, which can inform targeted safety interventions. Traditionally, pre-crash scenario typologies, which categorize crashes into scenario types, have been developed for conventional vehicles. However, significant differences exist between pre-crash scenarios of autonomous and conventional vehicles, and inadequate AV data has put limits on research into these differences (Liu et al., 2024). As driving practices and technology evolve, pre-crash scenario typologies are continuously revised to reflect these changes (Najm G. et al., 2007), making them especially useful for analyzing crashes involving autonomous vehicles.

A key aspect of pre-crash scenarios considered in most research, is the manoeuvring actions made by the parties involved before impact. Various studies have categorized pre-crash manoeuvres based on the trajectories, viewpoints, or intentions of the vehicle and cyclist. While MacAlister and Zuby (2015) and Lindman et al. (2015) focus on vehicle trajectory, Kuehn et al. (2015) and Char and Serre (2020) emphasize cyclist trajectory as the primary classification criterion.

Char and Serre (2020) identified the four main pre-crash movements in car-to-cyclist crashes to be 'crossing nearside', 'crossing far side', 'longitudinal' and 'turning' (both right and left). Kuehn et al. (2015) reported that most common pre-crash movements are "cyclist coming from the right" and "cyclist coming from the left," often occurring at property entrances, parking lots, and intersections.

Certain crash configurations result in higher fatality rates. Car-to-cyclist crashes in which the vehicle and cyclist were travelling longitudinally in traffic were found to result in the greatest number of fatalities. (Wisch et al., 2019). MacAlister and Zuby (2015) identified the most

frequent fatal vehicle-cyclist pre-crash movement combinations as 'straight-in-line', 'straight-crossing', and 'straight-against' collisions, with fatalities more likely in non-daylight conditions and on roads with speed limits of 40 mph or higher.

Lindman et al. (2015) identified 'straight-crossing,' 'car turning with an oncoming cyclist,' and 'dooring' (when a cyclist collides with an opening car door) as the leading pre-crash movements resulting in MAIS2+ cyclist injuries. The most prevalent impact configurations were, in order of prevalence, car front to cyclist side, cyclist front to car side, car front to cyclist front, and dooring incidents.

Fernández et al. (2022) identified three main crash scenarios: same-direction travel without turns, a bicycle crossing a vehicle's path by turning and a car crossing a bicycle's path by turning. Among these, same-direction car-to-bicycle crashes posed the highest risk of severe or fatal injuries.

Schindler and Jeppsson (2024) reported that crashes between cyclists and large vehicles such as buses and heavy goods vehicles typically occur in urban areas, during daylight conditions, on dry surfaces and with clear weather conditions. The most common crash scenarios are crossing scenarios and turning-off-a-road crashes, with young male cyclists under 18 years old being the primary victims. Additionally, crashes between turning trucks and bicycles significantly contribute to high-severity injuries (Johannsen et al., 2015).

Puller et al. (2023) found that most cyclist-to-vehicle crashes occur when cyclists are on designated cycling infrastructure or riding against the expected direction of traffic. These crashes are most frequent at intersections where drivers are required to yield, during right-turn manoeuvres, and occur predominantly at low traffic densities and speeds below 13 km/h.

For autonomous vehicles, Liu et al. (2024) reported that the most common crash scenarios at intersections are rear-end collisions and lane-change crashes.

The findings on common pre-crash scenarios from existing literature are summarized in the table and list below.

Vehicle-to-cyclist pre-crash scenarios	Classification criteria	Reference
<ul style="list-style-type: none"> ▪ Crossing nearside: the vehicle drove straight, and a cyclist crossed from the closest side of the road ▪ Crossing far side: the vehicle drove straight, and a cyclist crossed at least one lane of road before being hit by the vehicle ▪ Longitudinal: the vehicle and the cyclist both travelled in the same direction on the road, and the car hit the cyclist in the rear during the travel or laterally during an overtaking manoeuvre. ▪ Turning left/Turning right: the vehicle was turning left or right at an intersection and hits a cyclist whatever the trajectory of the cyclist 	Pre-crash movement	Char and Serre (2020)
<ul style="list-style-type: none"> ▪ Cyclist coming from the right ▪ Cyclist coming from the left <p>Both scenarios occurring at property entrances, parking lots and intersections</p>	Pre-crash movement	Kuehn et al. (2015)
<ul style="list-style-type: none"> ▪ Straight-in-line ▪ Straight-crossing ▪ Straight-against 	Pre-crash movement	MacAlister and Zuby (2015)
<ul style="list-style-type: none"> ▪ Same-direction travel without turns ▪ A bicycle crossing a vehicle's path by turning ▪ A car crossing a bicycle's path by turning 	Pre-crash movement	Fernández et al. (2022)
<ul style="list-style-type: none"> ▪ Crossing scenarios ▪ Turning-off-a-road ▪ Young male cyclists under 18 years 	Pre-crash movement	Schindler and Jeppsson (2024)
<ul style="list-style-type: none"> ▪ Cyclists riding on designated cycling infrastructure ▪ Cyclist riding against the expected direction of traffic ▪ Intersections where drivers are required to yield and during right-turn manoeuvres ▪ Low traffic densities and speeds below 13 km/h 	Roadway infrastructure and traffic environment	Puller et al. (2023)
<ul style="list-style-type: none"> ▪ Lane change manoeuvres 	Pre-crash movement	Liu et al. (2024)

Table 3: A compilation of existing research on the most frequent vehicle-to-cyclist pre-crash scenarios

The literature review also identified the following vehicle-to-cyclist pre-crash scenarios as leading to high severity injuries:

- Vehicle and cyclist travelling longitudinally in traffic (Wisch et al., 2019)
- Straight-in-line, straight-crossing, and straight-against (MacAlister & Zubby, 2015)
- Straight-crossing, car turning with an oncoming cyclist and dooring (Lindman et al., 2015)
- Same-direction car-to-bicycle crashes (Fernández et al., 2022)
- Crashes between turning trucks and bicycles (Johannsen et al., 2015)
- Non-daylight conditions and roads with speed limits of 40 mph or higher (MacAlister & Zubby, 2015)

2.5. Gaps in Literature and Contribution of the Study

A review of existing literature has revealed gaps that require further research to enhance the understanding of cycling safety. These gaps are elaborated below.

- i. There is limited research analysing empirical data on crashes involving autonomous vehicles and cyclists, mainly due to data availability limitations
- ii. There is limited literature on comparative analyses of pre-crash risk factors for conventional vehicles and autonomous vehicles in cyclist-involved crashes, and their impact on cyclist crash severities
- iii. Existing research on AV safety has primarily focused on the safety benefits of AVs for motorized vehicles and, to a lesser extent, pedestrians. However, other vulnerable road users, particularly cyclists, have not been given sufficient priority.
- iv. Existing pre-crash scenario typologies for cyclist-to-vehicle crashes primarily focus on the pre-crash movements, with lesser emphasis on other contributing factors.

This study contributes to addressing the above-identified gaps in literature by:

1. Integrating multiple databases and extracting crash details from narratives to create a multisource dataset of real-world cyclist-to-autonomous vehicle crashes, thereby addressing data availability challenges
2. Comparing pre-crash risk factors in cyclist-involved crashes between conventional and autonomous vehicles and evaluating their comparative impact on crash severity
3. Evaluating the safety implications of autonomous vehicles (AVs) for cyclists by analyzing their risk factors, crash outcomes and potential safety benefits and challenges

4. Integrating vehicle, roadway and environmental factors to develop comprehensive pre-crash scenario typology for cyclist-to-vehicle crashes.

Overall, the study aims to contribute new literature on implications of autonomous vehicle technology on cycling safety.

3. Methodology

The study methodology is described in detail in the subsequent sub-sections.

3.1. Data Collection and Collation

Data on motor vehicle-to-cyclist collisions were drawn from three federal and state-maintained crash databases, to identify crash scenarios involving both autonomous and conventional vehicles. The specific sources, the composition of each dataset, and the data extraction process are described in detail in the subsequent sections.

3.1.1. California Department of Motor Vehicles (CA DMV)

The California Department of Motor Vehicles started documenting the testing of autonomous vehicles (AVs) in 2014 under the Autonomous Vehicle Tester Program. As of February 2025, CA DMV had received a total of 797 Autonomous Vehicle Collision Reports. Manufacturers testing autonomous vehicles are required to report any collision that resulted in property damage, bodily injury, or fatality within 10 days of the incident. Every report comprises a narrative account of the crash alongside other details about the crash.

CA DMV collision reports involving cyclists and occurring between 2017 and June 2021 were isolated, and the details of the crashes extracted. The data extracted from these crash reports were consistent with the data provided in the crash reports available in the National Highway Traffic Safety Administration (NHTSA) database. The extracted variables included:

- i. Reporting details: reporting entity, report type and operating entity.
- ii. Vehicle information: vehicle make, model, model year and mileage.
- iii. Operator status: driver/operator classification and an indication of ADS (Automated Driving System) engagement at time of crash.
- iv. Temporal and spatial data: incident date and time, city and state.
- v. Roadway characteristics: roadway type (street or intersection) and road surface type
- vi. Environmental conditions: lighting and weather
- vii. Crash dynamics: parties involved, highest injury severity, property-damage assessment, pre-crash trajectories and speeds and any traffic violations.

A representative crash report is presented in *Appendix A: Sample crash report obtained from California DMV Database*.

3.1.2. National Highway Traffic Safety Administration (NHTSA)

In June 2021, the National Highway Traffic Safety Administration (NHTSA) issued a Standing General Order requiring designated manufacturers and operators to report, within ten days, any crash involving a vehicle equipped with an Automated Driving System (ADS) or a SAE Level 2 Advanced Driver Assistance System (ADAS). This Order was designed to ensure timely and transparent notification of real-world collisions associated with ADS and Level 2 ADAS vehicles (NHTSA).

The NHTSA database comprises of detailed summaries of autonomous vehicle crashes reported from July 2021 to 2024 in the USA. Crashes involving cyclists were extracted for the study.

Prior to July 2021, collision data for autonomous vehicles under testing were only recorded by the California Department of Motor Vehicles (CA DMV). Beginning in June 2021, NHTSA began its own parallel recording, and CA DMV continues to collect autonomous vehicle collision reports. To maximize the number of AV-to-cyclist crashes while avoiding duplication, the study combined:

- i. CA DMV reports from 2014 through June 2021, and
- ii. NHTSA reports from July 2021 through 2024.

Only those collisions in which the ADS was engaged at the time of impact were included in the final dataset.

3.1.3. California Transportation Injury Mapping System (TIMS)

The California Transportation Injury Mapping System (TIMS) is a statewide crash-reporting platform, covering all collisions involving human-driven vehicles from 2017 onward. TIMS integrates raw collision records, complete with geospatial coordinates, roadway attributes, environmental conditions, participant details and injury outcomes, with higher-level summaries and mapping tools.

For this study, we queried TIMS for every conventional vehicle-to-cyclist crash occurring within the City of San Francisco between January 2017 and December 2024. San Francisco was selected because 75% percent of the AV-to-cyclist collisions identified in Sections 3.1.1

and 3.1.2 took place there, thus enabling a consistent within-jurisdiction comparison under the same regulatory, infrastructural and environmental context.

The extracted TIMS dataset includes the following fields, harmonized with the CA DMV and NHTSA datasets for cross-comparison:

- i. Vehicle information: vehicle make, model and model year
- ii. Temporal and spatial data: crash date and time
- iii. Roadway characteristics: roadway type (street or intersection) and road surface type
- iv. Environmental conditions: lighting and weather
- v. Participant information: driver and cyclist demographics, manoeuvres and contributing factors
- vi. Injury severity levels

Although, driver and cyclist factors were available in this dataset, this information was not utilized in the study as there was no similar information available in the AV dataset for comparison.

The final datasets comprise of 52 AV-to-cyclist crashes and 2238 conventional vehicle-to-cyclist crashes. Each dataset includes a set of independent variables and crash severity as the dependent variable.

3.2. Crash Severity Classification

Crash severity is coded according to the NHTSA Standing General Order on crash reporting, which defines five categories namely:

- i. Fatality: Confirmed or alleged incident where an involved party was fatally injured.
- ii. Serious injury: Confirmed or alleged incident where an involved party sustained serious injuries that required hospitalization or emergency treatment.
- iii. Moderate injury: Confirmed or alleged incident where an involved party sustained injuries and sought medical treatment, excluding hospitalization and emergency treatment.
- iv. Minor injury: Confirmed or alleged incident where an involved party sustained possible injuries that did not require medical treatment.
- v. No injury: No injuries were reported or alleged from the incident. Some of these cases involve property damage while some do not.

The above categorization aligns with the coding in the CA DMV reports and closely mirrors the KABCO injury classification scale. The KABCO injury scale categorizes injuries into five groups namely: K–Fatal injury, A–Incapacitating injury(suspected serious injury), B–Non-incapacitating injury (suspected minor injury), C–Possible injury and O–No apparent injury.

The California TIMS dataset uses four crash severity categories namely: fatal injury, (suspected) serious injury, (suspected) minor injury and possible injury (complaint of pain). Non-injury crashes are not captured in the database. Therefore, in later stages of the study, while no injury will be used as the baseline for AV-to-cyclist crashes, possible injury will be used as the baseline for CV-to-cyclist crashes.

For CA DMV reports, crash narratives were reviewed to assign severity. In instances where the narrative did not clearly distinguish between minor and moderate injuries, particularly when subsequent reports contradicted initial accounts and cyclists later reported injuries , cases were conservatively coded as moderate injury.

Imbalanced data across injury severity categories can introduce a bias toward the majority thus causing less accurate predictions of severe crashes which are often the minority. To mitigate the effect of imbalanced data on subsequent analysis, adjacent injury categories were merged on a new scale to gain more balanced data. Severity recorded in the California TIMS as fatal injury were combined with serious injury, while those recorded as serious injury in the NHTSA database were combined with moderate injury.

Consequently, the final injury severity coding was as follows. Conventional vehicle-to-cyclist crash severities comprised of three classes:

- Fatal/Serious (K+A)
- Minor (B)
- Possible (C)

AV-to-cyclist crash severities comprised of three classes:

- Serious/Moderate (A)
- Minor (B)
- No injury(C)

3.3. Descriptive Analysis of Crashes

An initial analysis of crash datasets for both autonomous vehicle-to-cyclist and conventional vehicle-to-cyclist crashes was undertaken to understand the characteristics of the data and inform the subsequent causal analysis. This preliminary analysis revealed key patterns in the crashes that informed subsequent decisions about variable inclusion, category grouping and model selection and specification. It also enabled assessment of data completeness through identification of missing values in key variables.

For AV-to-cyclist crashes, the following crash attributes were identified and frequency bar charts plotted to visualize their distribution:

- i. Highest injury severity (crash severity)
- ii. Cyclist and AV Pre-crash Manoeuvre (PCM)
- iii. Pre-crash Trajectory (PCT) of the cyclist relative to the vehicle
- iv. Weather and lighting condition
- v. Roadway type (intersection, street or parking lot)
- vi. Property damage
- vii. City and state
- viii. Vehicle manufacturer (operator)
- ix. Vehicle make, model and manufacture/model year
- x. Cyclist traffic violation
- xi. Time of crash (time of day and year of crash)
- xii. Collision types
- xiii. Mileage of vehicle at the time of crash
- xiv. Posted Speed Limit
- xv. Pre-crash speed of vehicle

Similarly, for CV-to-cyclist crashes, frequency bar charts for the following attributes were plotted:

- i. Highest injury severity (crash severity)
- ii. Cyclist and CV Pre-crash Manoeuvre (PCM)
- iii. Pre-crash Trajectory (PCT) of the cyclist relative to the vehicle
- iv. Weather and lighting condition
- v. Roadway type (intersection or non-intersection)

- vi. Vehicle manufacture/model year
- vii. Time of crash (time of day and year of crash)

3.3.1. Imputation of Missing Data

Following the initial analysis of AV-to-cyclist crash data, values were found to be missing for the following three key variables, as shown in the table below.

Variable	Number of missing values	Percentage of the total values in data set
Mileage	14	27%
AV Pre-crash Speed	9	17%
Posted Speed Limit	12	23%

Table 4: Missing data for critical variables in AV-cyclist crash data set

The missing values in the three variables were classified as “Missing at Random” (MAR) as they appear to be unrelated to the missing data itself but are instead associated with other observed variables (Zhou et al., 2024). A significant portion of the missing data stems from crash narratives reported by the California DMV during a period when a standardized directive outlining required AV crash attributes to be captured in crash reports, had not yet been established.

All the variables having missing values are numerical. The Multivariate Imputation by Chained Equations (MICE) method was implemented in R using the mice package, to predict missing values. The mice function constructs separate imputation models for each variable with missing data and creates multiple imputed datasets using chained equations. Each variable with missing data is imputed conditionally using a regression-like model based on other variables. Categorical predictor variables like manoeuvres, lighting etc. are imputed using polytomous regression (polyreg), while both discrete and continuous numerical predictor variables are imputed using normal linear regression prediction (norm.predict). For numeric response variables, mice utilizes the predictive mean matching (pmm) method. The pmm method is grounded in linear regression but enhances realism by selecting observed values with predicted means closest to those of the missing data (Azur et al., 2011). This approach provides greater robustness than methods relying solely on linear regression (e.g., "norm").

All variables that will be used in the final model as well as auxiliary variables that are predictive of the missing values (even if they may not be used in final model) are included in the imputation process. The later helps to improve imputation and reduces bias (Azur et al., 2011).

Predictor variables for the missing data were identified based on statistical correlation and logical relationships between the variables. The Pearson correlation test, which assumes linear relationship and normality was used to assess correlation between numerical versus numerical variables. For numerical versus numerical variables which appear non-linear and skewed, the spearman test was used. The Kruskal-Wallis test, used when the data is not normally distributed and skewed was used to assess correlation between numerical versus categorical variables, having more than two levels. The summary of the correlation test results for predictor variables found to be statistically significant are shown in the table below.

Response variable (Variable with missing data)	Predictor Variable	Test Statistic Values	Inference
Mileage	Crash Year	Test: Pearson Correlation p-value: 0.04919 Correlation Coefficient (r): 0.3213 95% confidence interval: [0.0018, 0.5813]	There is a weak to moderate, but statistically significant, positive correlation between crash year and mileage, suggesting that vehicles in more recent crashes tend to have slightly higher mileage, possibly due to faster accumulation or more complete recent data.
	Reporting Entity	Test: Kruskal-Wallis p-value: 0.04184	Since $p = 0.04184 < 0.05$, the result is statistically significant. That implies that there is evidence that at least one entity reports vehicles with significantly different mileage compared to others.
Posted Speed Limit	City	Test: Kruskal-Wallis p-value: 0.02507	There is statistically significant difference in posted speed limits across cities ($p < 0.05$).
	State	Test: Kruskal-Wallis p-value: 0.02242	There is statistically significant difference in posted speed limits across states ($p < 0.05$).

Response variable (Variable with missing data)	Predictor Variable	Test Statistic Values	Inference
AV Pre-crash Speed	AV Precrash Manoeuvre	Test: Kruskal-Wallis p-value: 0.0001044	There is statistically significant difference in pre-crash speeds depending on the manoeuvre being performed ($p < 0.001$).
	Collision Type	Test: Kruskal-Wallis p-value: 0.01192	There is statistically significant difference in AV speeds depending on the type of collision ($p < 0.05$)
	AV-Cyclist Precrash Trajectory	Test: Kruskal-Wallis p-value = 0.002642	There is statistically significant difference in AV Precrash Speed across different AV-cyclist trajectories before the crash.

Table 5: Correlation test results between predictor variables and variables with missing data

Logical relationships between variables with complete data and variables with missing data were also taken into consideration even when statistical correlations were found to be weak or missing.

The variables considered to predict vehicle mileage, included:

- Crash Year: More recent crashes may be associated with lower mileage for newer vehicles.
- Reporting Entity / Operating Entity: Different agencies or companies may report data with varying completeness or vehicle usage patterns.
- Vehicle Model (Manufacture) Year: Older vehicles may have higher mileage compared to newer vehicles. The technological sophistication associated with newer vehicles may also affect speed and consequently mileage.

The variables considered to predict AV pre-crash speed included:

- AV Pre-Crash Manoeuvre: Types of manoeuvres naturally affect vehicle speed prior to crash e.g. vehicles moving straight are likely to travel at higher speeds compared to turning vehicles.
- AV-Cyclist Pre-crash Trajectory: Certain trajectories may be associated with certain speed ranges
- Collision Type: Different crash types imply different speeds (e.g., rear-end versus side-impact).

- Incident (Crash) Time: Driving speed patterns vary by time of day e.g. lower pre-crash speeds may be associated with nighttime and vice versa.
- Lighting: Night driving likely correlates with lower speed and maybe cautious driving thus impacting pre-crash speed.

The variables considered to predict Posted Speed Limit included:

- Roadway Type: The type of roadway strongly influences both posted speed limit and AV pre-crash speed. Streets compared to intersections and parking lots impose different speeds.
- City/State: Different jurisdictions have different default speed limits.
- AV Pre-crash Speed: Posted Speed Limits often correlates with driving speed

The final variables included in the imputation process included, Highest Injury Severity, AV Pre-crash Speed, Posted Speed Limit, Mileage, Cyclist Precrash Manoeuvre, AV Pre-crash Manoeuvre, AV-Cyclist Precrash Trajectory, Collision Type, Lighting, Incident Time, Reporting Entity, Crash Year, Model Year, City and Roadway Type.

Five multiple imputed datasets were generated ($m=5$) with 50 iterations per imputation ($maxit=50$) to stabilize estimates, using mice.

3.4. Pre-crash Scenario Construction

In this study, a pre-crash scenario is defined as the set of circumstances leading to a motor-vehicle-to-cyclist crash. This includes the trajectories of both the cyclist and the vehicle, as well as contributing external factors such as driver and cyclist characteristics, vehicle attributes, roadway features and environmental conditions.

The study integrates risk factors identified in existing literature, such as cyclist and driver behavior and additional elements specific to autonomous vehicles (AVs) and combines them with pre-crash movements observed in our dataset to develop a comprehensive framework for defining pre-crash scenarios. These scenarios are subsequently used to assess crash severity outcomes.

Key factors considered in constructing these pre-crash scenarios are categorized as follows:

i. Pre-crash movement

This encompasses the relative positions and trajectories of the vehicle and cyclist prior to impact (e.g. Stopped – Straight, Straight – Straight, Straight – Crossing or Parked – Straight). It also includes driver and cyclist manoeuvres before the crash, such as braking, stopping, starting, parking, leaving a parked position, passing, turning, reversing, merging or changing lanes.

ii. Roadway factors

These refer to road infrastructure attributes including roadway segments and presence of intersections.

iii. Environmental factors

These include prevailing environmental conditions at the time of the crash, such as weather, lighting and road surface conditions.

iv. Motorized vehicle factors

This entails vehicle manufacture/model year.

Crash factors, particularly impact locations (types of collision) were not included in pre-crash scenarios as they describe circumstances occurring during impact rather than before impact.

Categorical variables exhibiting high class/level dominance were not included in the construction to avoid unstable coefficient estimates in the subsequent steps. Such variables included weather conditions where 94% of crashes occurred in clear weather and only 6% of crashes occurred in cloudy weather, lighting conditions where 79% of crashes occurred during daylight and only 21% during dark lighted conditions or during dawn/dusk, and road surface conditions where all (100%) of crashes occurred on dry road surfaces. The independent variable: roadway type had three categories namely street (50%), intersection (48%) and parking lot (2%). To avoid unstable coefficient estimates due to sparse data for the parking lot category; while maintaining completeness of the overall dataset, the one (1) parking lot observation was merged with street observations. Additionally, categories with fewer observations were merged into "Other" to prevent unstable coefficient estimates for the respective categorical variables.

The final AV-to-cyclist dataset comprised of fifteen (15) variables as described in section 3.3. While ensuring alignment with the types of variables available in the CV-to-cyclist dataset to enable comparison of estimates and leaving out variables exhibiting high class/level dominance, seven out of the fifteen variables including crash severity were selected for pre-crash scenario construction as listed below:

Dependent variable: Highest injury severity/crash severity (nominal categorical)

Categorical independent variables:

- Cyclist pre-crash manoeuvre
- AV pre-crash manoeuvre
- AV-to-cyclist pre-crash trajectory
- Roadway type

Numerical independent variables:

- Incident time
- Model year

The tables below summarize the final pre-crash scenario constructions for the AV and CV datasets, detailing the dependent variable, six independent variables, categorical variable encodings and each scenario's share of total crashes.

Pre-crash Scenarios for AV-to-cyclist Crashes		
Variable/Variable Categories	Variable encoding	% of total crashes
Crash Severity (Highest Injury Severity)		
No injury	1 if highest injury severity is no injury, 0 otherwise	54%
Minor injury	1 if highest injury severity is minor, 0 otherwise	29%
Serious injury (including moderate injury)	1 if highest injury severity is serious or moderate, 0 otherwise	17%
AV Pre-crash Manoeuvre		
Proceeding Straight	Proceeding straight indicator (1 if AV pre-crash manoeuvre is proceeding straight, 0 otherwise)	36%
Stopped	Stopped indicator (1 if AV pre-crash manoeuvre is stopped, 0 otherwise)	32%
Other	Other indicator (1 if AV pre-crash manoeuvre is turning left or turning right, making evasive manoeuvre, backing, braking, parked or merging, 0 otherwise)	32%
Cyclist Pre-crash Manoeuvre		
Proceeding Straight	Proceeding straight indicator (1 if cyclist pre-crash manoeuvre is proceeding straight, 0 otherwise)	46%

Pre-crash Scenarios for AV-to-cyclist Crashes		
Variable/Variable Categories	Variable encoding	% of total crashes
Other	Other indicator (1 if cyclist pre-crash manoeuvre is crossing intersection, changing lanes, crossing lanes, turning left, turning right, passing, travelling on sidewalk or travelling on bike lane, 0 otherwise)	54%
AV-to-Cyclist Pre-crash Trajectory		
Stopped – Straight	Stopped – Straight indicator (1 if AV-Cyclist pre-crash trajectory is Stopped – Proceeding Straight, 0 otherwise)	29%
Straight – Straight	Straight – Straight indicator (1 if AV-Cyclist pre-crash trajectory is Proceeding Straight – Proceeding Straight, 0 otherwise)	27%
Other	Other indicator (1 if AV-Cyclist pre-crash trajectory is Straight – Crossing, Turning Right – Turning Right, Turning Left – Turning Left, Turning Right – Turning Left, Turning Left – Straight, Backing – Straight, Parked – Straight, Merging – Straight, Turning Right – Straight or Straight – Turning Left, 0 otherwise)	44%
Roadway Type		
Street	Street indicator (1 if a street or parking lot, 0 otherwise)	52%
Intersection	Intersection indicator (1 if an intersection, 0 otherwise)	48%
Motorized vehicle factors		
AV Model year	Numerical data (not categorized)	
Incident time		
Incident time	Numerical data in decimal hours (not categorized)	

Table 6: Pre-crash Scenario Construction for Autonomous Vehicle Crashes

Pre-crash Scenarios for CV-to-cyclist Crashes		
Variable/Variable Categories	Variable encoding	% of total crashes
Crash Severity (Highest Injury Severity)		
Possible injury	1 if highest injury severity is possible injury, 0 otherwise	48.5%
Minor injury	1 if highest injury severity is minor, 0 otherwise	43.5%
Serious injury (including fatal injury)	1 if highest injury severity is serious or fatal, 0 otherwise	8.0%
CV Pre-crash Manoeuvre		
Proceeding Straight	Proceeding straight indicator (1 if CV pre-crash manoeuvre is proceeding straight, 0 otherwise)	35%
Turning	Turning indicator (1 if CV pre-crash manoeuvre is turning right or turning left, 0 otherwise)	35%
Other	Other indicator (1 if CV pre-crash manoeuvre is neither proceeding straight nor turning, 0 otherwise)	30%
Cyclist Pre-crash Manoeuvre		
Proceeding Straight	Proceeding straight indicator (1 if cyclist pre-crash manoeuvre is proceeding straight, 0 otherwise)	81%
Other	Other indicator (1 if cyclist pre-crash manoeuvre is crossing intersection, changing lanes, crossing lanes, turning left, turning right, passing, travelling on sidewalk or travelling on bike lane, 0 otherwise)	19%
CV-to-Cyclist Pre-crash Trajectory		
Straight - Straight	Straight – Straight indicator (1 if CV-Cyclist pre-crash trajectory is Proceeding Straight – Proceeding Straight, 0 otherwise)	24%
Turning (Left or Right) - Straight	Turning – Straight indicator (1 if CV-Cyclist pre-crash trajectory is Turning (left or right) – Proceeding Straight, 0 otherwise)	31%
Other	Other indicator (1 if CV-Cyclist pre-crash trajectory is Parked – Straight, Stopped –	45%

Pre-crash Scenarios for CV-to-cyclist Crashes		
Variable/Variable Categories	Variable encoding	% of total crashes
	Straight, Making U-turn – Straight, Straight – Turning, Entering Traffic – Straight or other movements, 0 otherwise)	
Roadway Type		
Intersection	Intersection indicator (1 if an intersection, 0 otherwise)	63%
Non-Intersection	Non-intersection indicator (1 if a non-intersection, 0 otherwise)	37%
Motorized vehicle factors		
CV Model Year	Numerical data (not categorized)	
Incident time		
Crash/Incident time	Numerical data in decimal hours (not categorized)	

Table 7: Pre-crash Scenario Construction for Conventional Vehicle Crashes

3.5. Statistical Methodology

3.5.1. Correlation Analysis of Dependent and Independent Variables

In addition to the frequency bar charts presented for crashes in Section 3.3, box plots and stacked bar charts were used to visualize each predictor’s distribution and to compare these distributions across the three injury severity levels. For continuous predictors, a monotonic relationship, a key prerequisite for ordinal models, was inferred when medians or interquartile ranges of box plots consistently increased or decreased from “No Injury” to “Minor Injury ” to “Serious Injury”. For categorical predictors, stacked bar charts highlighted proportional shifts in each category across the severity levels. A clear, unidirectional trend supported treating severity as ordinal, whereas the absence of such a trend or the presence of non-monotonic reversals indicated that a nominal approach would be more appropriate.

The plotted box plots and stacked bar charts for AV-to-cyclist crashes are shown in figures 2 and 3 below.

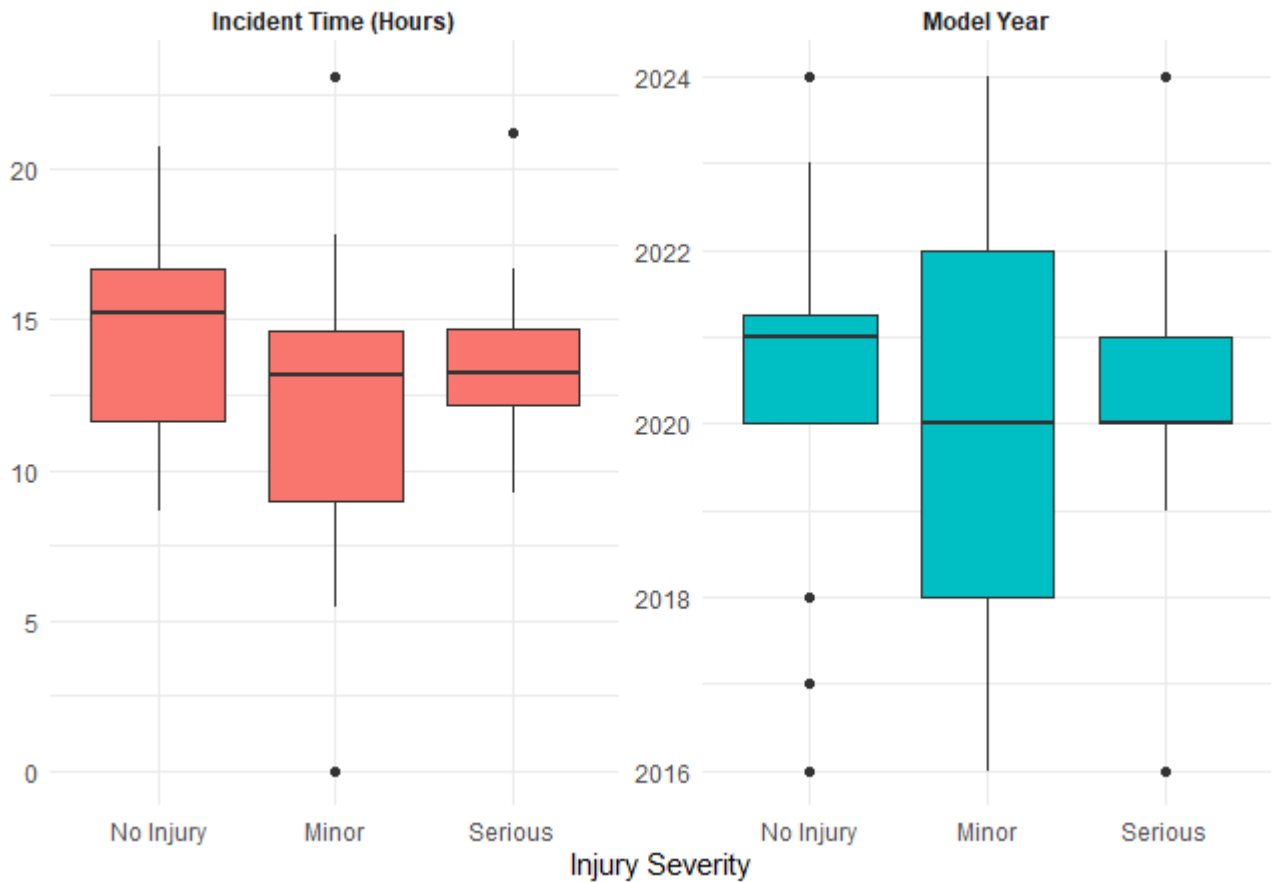


Figure 2 :Box plots of continuous predictors versus crash severity

An assessment of boxplots for each of the continuous predictor variables across the three injury severity levels in figure 2 above shows no consistent increase or decrease in their medians and interquartile ranges. For instance, median incident time decreased then slightly increased while median model year decreased then remained constant. The interquartile ranges overlapped substantially at every level with no clear “stretching” or “shrinking” pattern that would imply a clear monotonic trend.

Of the four categorical predictors visualized in the stacked bar charts in figure 3 below, only cyclist pre-crash manoeuvre shows a consistent monotonic trend across “No Injury”, “Minor” and “Serious”. The other three predictors exhibit non-monotonic reversals thus violating the prerequisite for an ordinal model.

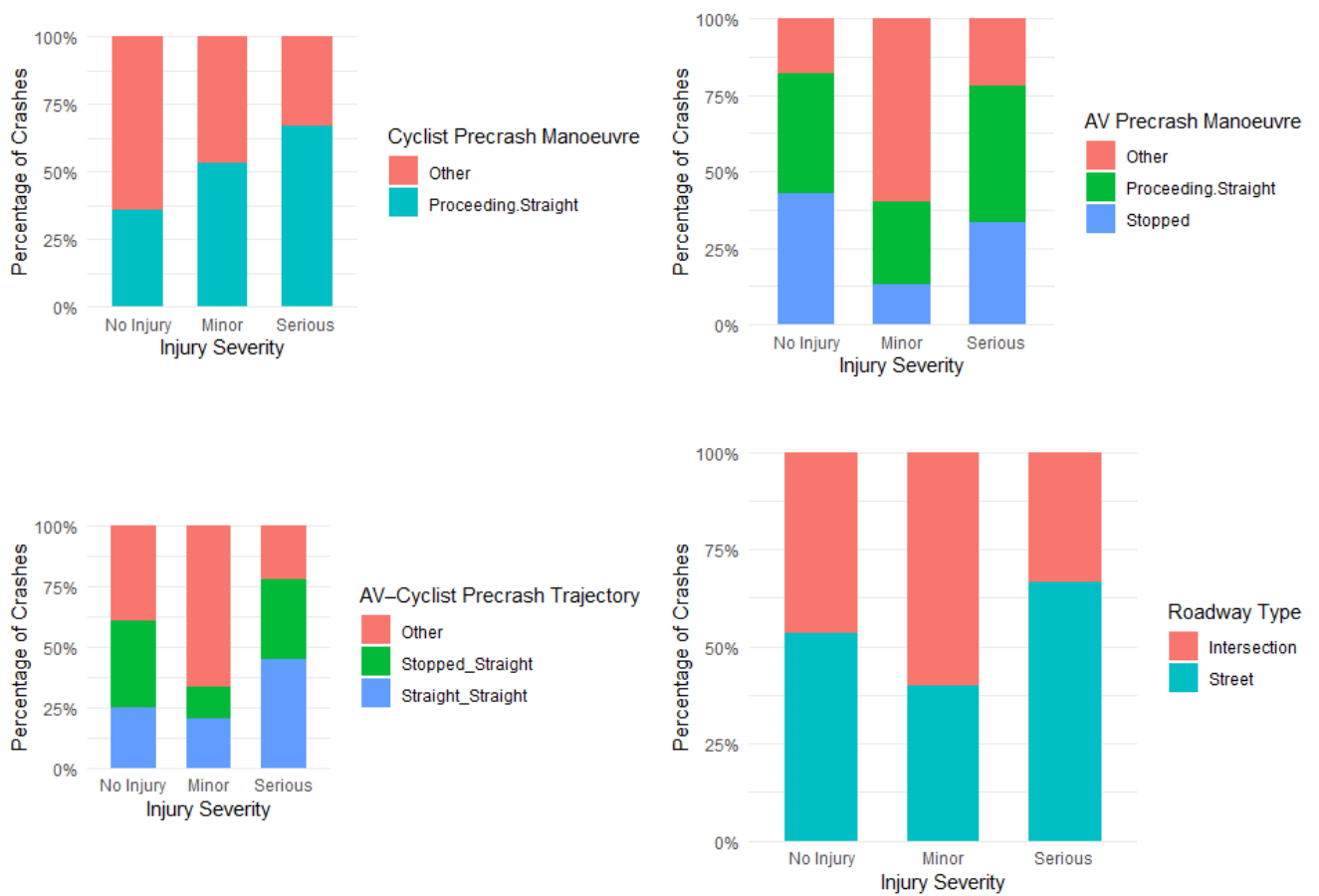


Figure 3 :Box plots of categorical predictors versus crash severity

Similar plots were made for conventional vehicle crashes which also revealed no consistent monotonic trend across injury levels. Accordingly, these findings justify modeling highest injury severity as an unordered (nominal) outcome.

3.5.2. Model Selection

In crash severity research, various statistical methods have been employed to identify the relationships between dependent and independent variables, with logistic and probit models and their variations being among the most common approaches. Each of these techniques has its own methodological strengths and weaknesses which are often driven by data limitations or the specific nature of the variables involved. Because injury severity is inherently categorical, predictive modeling has traditionally relied on logistic-regression approaches including binary logistic regression, ordered logistic regression and multinomial logistic regression (MNL). More recent studies have also utilized machine-learning techniques.

Initial data exploratory analyses of AV crashes, including box plots for continuous predictor variables and stacked bar charts for categorical factors, revealed no consistent, monotonic relationship between most predictors and the three severity levels (No Injury, Minor, Serious). Only “Cyclist Pre-crash Manoeuvre” showed a clear ordering. All other variables failed to meet the proportional-odds assumption required for ordered logistic models. Consequently, treating injury severity as a nominal (unordered) outcome is justified.

Furthermore, the multinomial logit is the most extensively used model in crash severity research (Chen & Fan, 2019; Scarano et al., 2023; Shiran et al., 2021; Vilaça et al., 2020). It balances ease of interpretation by providing straight-forward odds-ratio estimates for each severity level and predictive performance. For these reasons, MNL was found to be the most appropriate method for the study.

The multinomial logistic regression model is a discrete outcome model for predicting a categorical dependent variable with three or more unordered levels which in this study is the injury severity level. The model estimates the log odds of each crash severity category relative to a reference category which is often the lowest injury severity level. The general framework for modelling the level of injury severity sustained in a crash begins by defining functions that determine the injury severity outcome n for crash m as follows (McFadden, 1974; Train, 2009):

Given there are $N+1$ injury severity levels, indexed as $n=0, 1, \dots, N$, where:

- $n = 0$ corresponds to the reference category “No Injury”
- $n = 1, \dots, M$ correspond to the other injury severity levels (“Minor Injury” or “Serious Injury”)

and

$X_m = (1, X_{m1}, X_{m2}, \dots, X_{mk})$ is the $(K+1)$ vector of explanatory variables associated with crash m

$\beta_j = (\beta_{n0}, \beta_{n1}, \dots, \beta_{nk})$ is the vector of coefficients associated with injury severity level n , relative to the reference $n = 0$

The probability that crash m falls into severity level n is given by:

$$P(Y_m = n|X_m) = \frac{\exp(X_m\beta_n)}{1 + \sum \exp(X_m\beta_m)} \quad (n = 1, \dots, N) \quad (1)$$

Equivalently, taking logs against the reference injury level:

$$\text{Log} \frac{P(Y_m = n | X_m)}{P(Y_m = 0 | X_m)} = X_m\beta_n, \quad (n = 1, \dots, N) \quad (2)$$

3.5.3. Correlation Analysis of Independent Variables

The relationship between the variables selected for parameter estimation was evaluated to assess which independent variables would be appropriate for fitting the model. Multicollinearity is a statistical phenomenon that occurs when two or more independent variables in a regression model are highly correlated, indicating a strong linear relationship among the predictor variables. This issue complicates regression analysis by making it difficult to accurately determine the individual effects of each independent variable on the dependent variable. The presence of multicollinearity can lead to unstable and unreliable coefficient estimates, making it challenging to interpret the results and draw meaningful conclusions from the model.

Multicollinearity can be detected using correlation matrices, scatter plots or the Variance Inflation Factor (VIF) (Shrestha, 2020). While correlation matrices and scatter plots reveal the pairwise relationships between variables, they only reflect bivariate correlations as multicollinearity can hide in higher dimensions. VIF, on the other hand, is more robust, as it assesses how much the variance of a regression coefficient is increased due to multicollinearity with a group of other variables. VIF is computed as per the following formula:

$$\text{Variance Inflation Factor, VIF} = \frac{1}{1 - R_p^2} \quad (3)$$

Where R_p^2 is the regression predictor p on all other predictors

The closer the R^2 value to 1, the higher the value of VIF and the higher the multicollinearity among the independent variables. Dropping one of the correlated variables helps in bringing down the multicollinearity between correlated variables.

Results of the collinearity test between predictor variables using VIF technique is shown in the table below. There was no evidence of problematic multicollinearity in the set of predictors as all VIF values were below the commonly used threshold value of 5.

Independent Variable	Generalized VIF	Collinearity
Cyclist Pre-crash Manoeuvre	1.59	Very low
AV Pre-crash Manoeuvre	3.85	Low
Roadway Type	1.64	Very low
AV-to-Cyclist Pre-crash Trajectory	2.27	Low
Incident Time	1.22	Negligible
Model Year	1.45	Very low

Table 8: Correlation test results between predictor variables

3.5.4. Parameter Estimation

The model fitting and parameter estimation process comprised of several sub-steps as described below.

i. Encoding of categorical variables into dummy variables

Categorical variables were encoded into dummy variables prior to model fitting. The first column which contained the original variable header and values was dropped to avoid the dummy variable trap.

ii. Lumping of low frequency variables

A multinomial logit model was fitted in R using the nnet package with highest injury severity defined as the dependent variable and having three discrete (unordered) levels namely: no injury, minor injury and serious injury. The serious injury level included moderate injury observations for AV crashes and fatal injury observations for CV crashes, to minimize data imbalance across injury severity levels. No injury was defined as the reference level in the AV model while possible injury was defined as the reference level in the CV model.

The first fitted model included six (6) predictor variables namely cyclist pre-crash manoeuvre, AV pre-crash manoeuvre, AV-cyclist pre-crash trajectory, roadway type, incident time and vehicle model year.

iii. Defining reference/base levels for categorical predictors

Reference levels were defined for categorical variables having two or more levels. For the pre-crash trajectory and manoeuvres, the level “Other” was defined as the reference level while for roadway type, “Intersection” was set as the reference level.

iv. Model Fitting

A multinomial logit model was fitted in R using the nnet package. Several trial model fittings were made before a final model fit was settled on. The first model was fitted with six predictors. The subsequent models were fitted with predictors being dropped in a stepwise manner till the final set of 4 predictors was selected.

v. Model performance evaluation

To evaluate the performance of the model, data partitioning was done. With only 52 observations in the data set, a simple split of training versus testing data split would leave too few observations in the test set to reliably evaluate model performance. Further, some injury severity levels would be missing entirely in the test set, making the evaluation biased or invalid. The stratified k-fold cross-validation was therefore used. Instead of one split, the data is divided into k (5) equal parts (folds): The model trains on k-1 folds and tests on the remaining fold. This process repeats k (5) times, with each fold used once as the test set. Final performance is the average of all test results. Stratification ensures each fold has a similar distribution of injury severity categories as the whole dataset, preventing bias and ensuring more reliable model evaluation.

After the first run of the model, the categorical predictor variable levels with lower frequencies were further lumped together to form bigger levels with higher number of observations. This helped to lower and stabilize coefficient estimates.

vi. Rescaling of numeric predictors

Numeric predictors were rescaled by centering and standardizing to further stabilize coefficients and standard errors. Rescaling entailed performing the following to each value of a numeric predictor independently:

- i. Centering – subtracting its sample mean

- ii. Standardizing – dividing by its (sample) standard deviation

Afterwards every transformed variable has a mean that tends towards 0 and standard deviation that tends towards 1 . Rescaling serves to improve the numerical stability of the maximum-likelihood optimizer by placing all predictors on a common variance scale. The resulting coefficients are directly comparable as each reflects the change in log-odds produced by a one-standard-deviation increase in the predictor.

vii. Pooling of observations

AV and CV datasets were combined and a pooled model with an indicator for vehicle-type (AV or CV) was fitted. Combining the autonomous vehicle (AV) and conventional vehicle (CV) datasets into a single regression model with a binary indicator for vehicle type, enhances both statistical power and internal consistency. By pooling the AV-crash and CV-crash records, the overall sample size increases, yielding more precise coefficient estimates, particularly for predictors with relatively few observations.

Moreover, this approach also enables direct comparison of the effects of AVs and CVs in a crash. The coefficient associated with the AV versus CV indicator quantifies the change in odds of experiencing, for instance, a serious injury when the striking vehicle is autonomous rather than human driven.

3.5.5. Model Validation

The multinomial logit model is susceptible to correlation of unobserved effects across injury severity levels. Such a correlation causes violation of the model's independence of irrelevant alternatives (IIA assumption). Thus, when the severity levels share some unobserved effects, the MNL model may not be appropriate. To address this constraint, the Hausman–McFadden test was done to assess this property.

The serious injury level was dropped and the model estimates compared with the estimates obtained when all the three injury levels are included in the model.

4. Results

4.1. Imputation of Missing Data

Density plots comparing observed versus completed (imputed) values for AV pre-crash speed, posted speed limit and mileage were made to visualize the outcome of the imputation process. Density in this context implies the estimated probability distribution of speeds or mileage. The plots are shown in the figures below.

The density plot of observed and imputed values for AV pre-crash speed in figure 4 below shows strong alignment between distributions, indicating that the multiple imputation procedure effectively preserved the data's underlying structure. Minor deviations, including a slight overrepresentation of imputed values around 5–10 mph and around the distribution's peak, were observed but did not significantly alter the overall distribution. Both distributions peak near 0 mph, reflecting common low speed precrash conditions among autonomous vehicles. Beyond 15 mph, the alignment remains close with negligible differences. Overall, the imputations appear realistic and unbiased, supporting the reliability of the completed dataset for further analysis.

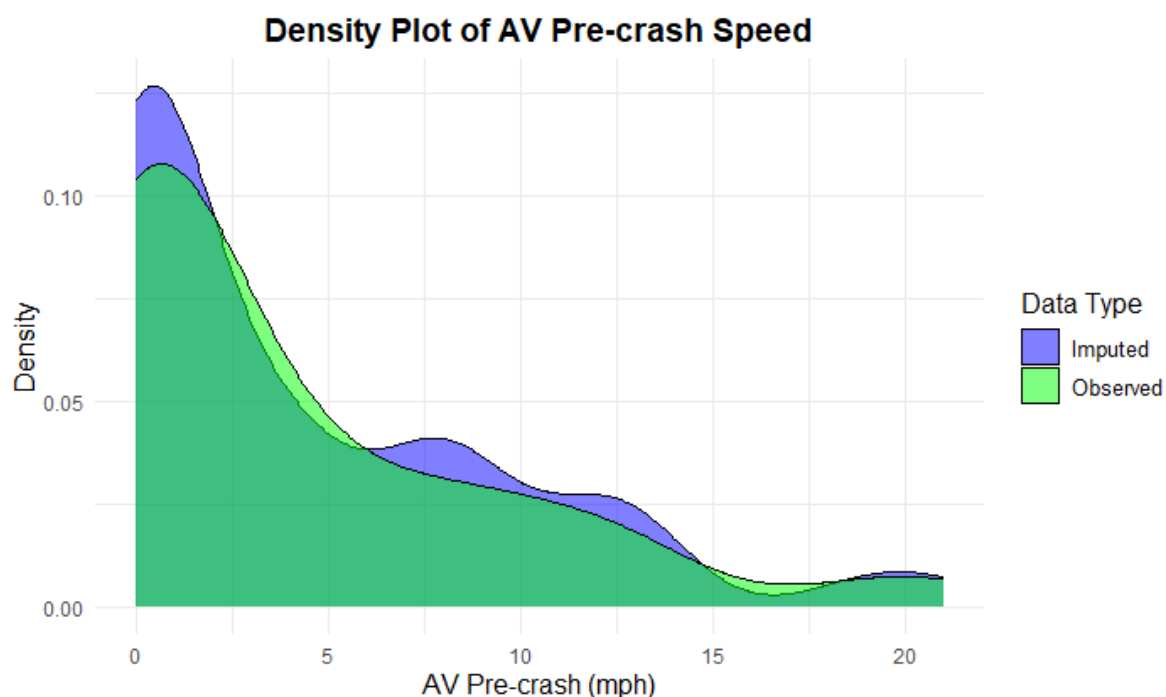


Figure 4 : Density Plot of Observed and Imputed Values for AV Pre-crash Speed (mph)

The density plot comparing observed and imputed values for posted speed limit (mph) in figure 5 below reveals a strong alignment between the two distributions. Both distributions exhibit a sharp peak around 25 mph, reflecting the concentration of posted speed limits in that range. Minor deviations were observed, with a slight overrepresentation of imputed values at lower speed ranges (5–10 mph) and near the distribution’s peak. Overall, the imputation process preserved the key characteristics of the observed data, supporting the reliability of the imputed values for further analysis.

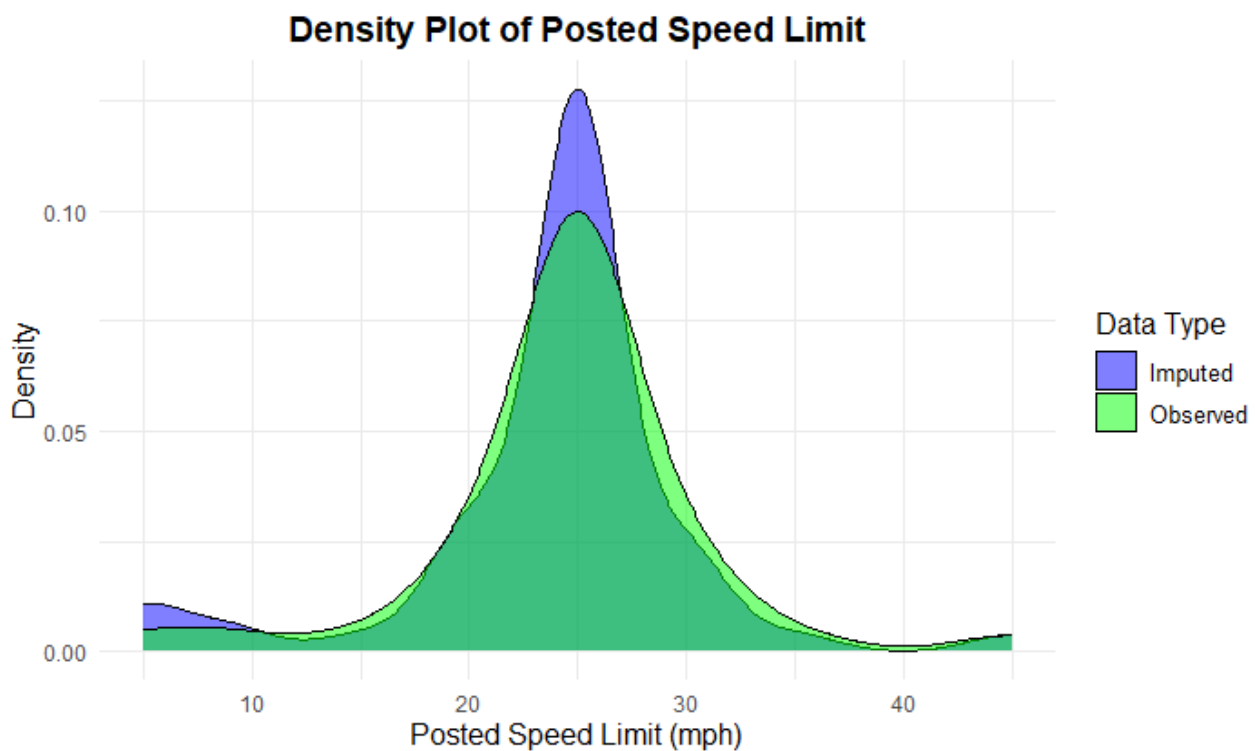


Figure 5 : Density Plot of Observed and Imputed Values for Posted Speed Limit (mph)

The density plot comparing observed and imputed values for mileage in figure 6 below indicates that the imputation model generally preserved the overall shape of the observed distribution. Both distributions follow a similar pattern, although imputed values show a higher density at lower mileage ranges and a slight underrepresentation at higher mileages. Despite these minor deviations, the imputed data adequately captures the main characteristics of the observed mileage distribution, supporting the reliability of the imputation process for subsequent analyses.

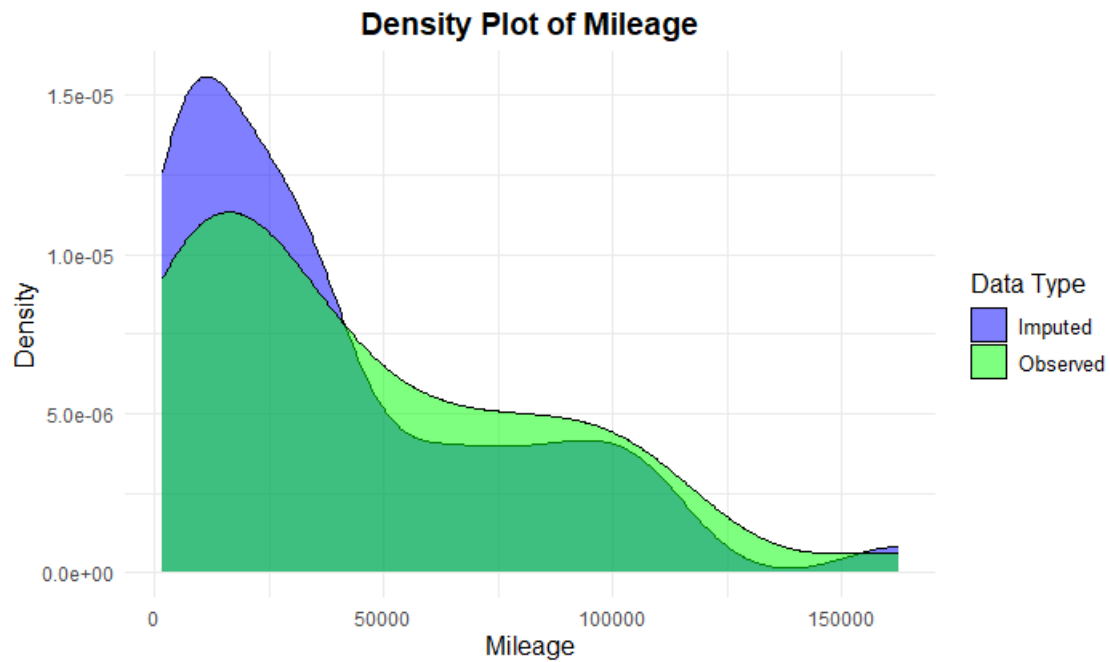


Figure 6 : Density Plot of Observed and Imputed Values for Mileage

Using the imputed/complete datasets, frequency bar charts were plotted for AV pre-crash speed, posted speed limit and vehicle mileage as presented in section 4.2.1.

4.2. Descriptive Analysis of Crashes

4.2.1. Autonomous Vehicle-to-Cyclist Crashes

The bar chart in figure 7 below shows that among all autonomous vehicle (AV) collisions with vulnerable road users (VRU), bicyclists account for the largest share of 71% of collisions. Micro-mobility users (scooterists and skateboarders) make up the next largest group at 19%, while pedestrians represent only 10% of AV-to-VRU crashes.

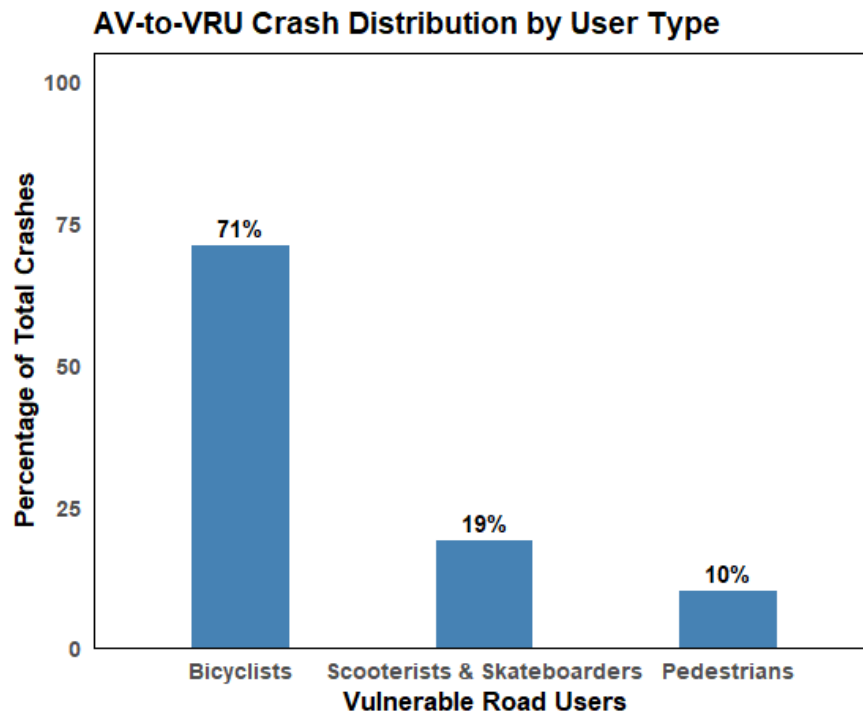


Figure 7 : Frequency distribution of crashes among vulnerable road users (VRU)

The bar chart in figure 8 below shows the outcomes of AV-to-cyclist collisions by the highest injury sustained in each crash. 54% of the crashes resulted in no bodily harm to the cyclist; 29% resulted in minor injury (bruises or cuts); 9% resulted in moderate injury requiring medical treatment, while 8% resulted in severe injury requiring hospitalization or emergency treatment. No fatal injury is observed.

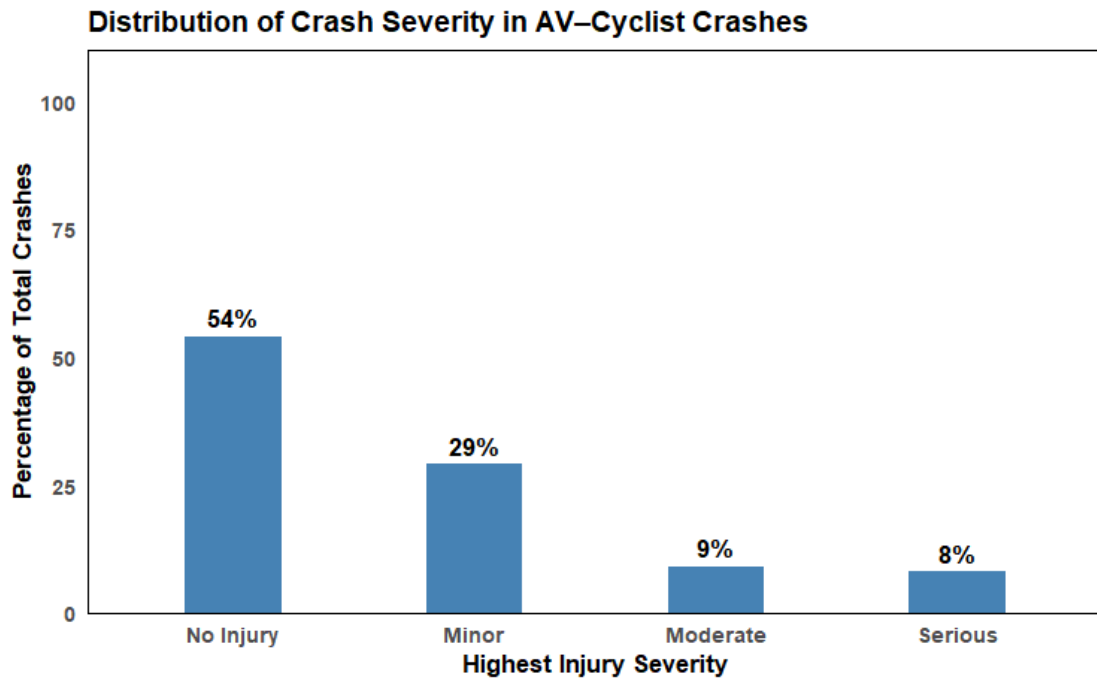


Figure 8 : Frequency distribution of crash severity in AV-to-cyclist crashes

The bar chart in figure 9 below shows how cyclists were moving just before collisions with autonomous vehicles. 46% of crashes occur when the cyclist was riding straight ahead, 17% when the cyclist was entering or traversing intersections, 11% when the cycling was moving across lanes, 10% when cyclist was overtaking another road user and 6% when the cyclist was turning left. Other low frequency pre-crash manoeuvres (each 2-4%) include riding on sidewalk (4%), crossing lanes (2%), traveling in a bike lane (2%) or turning right (2%).

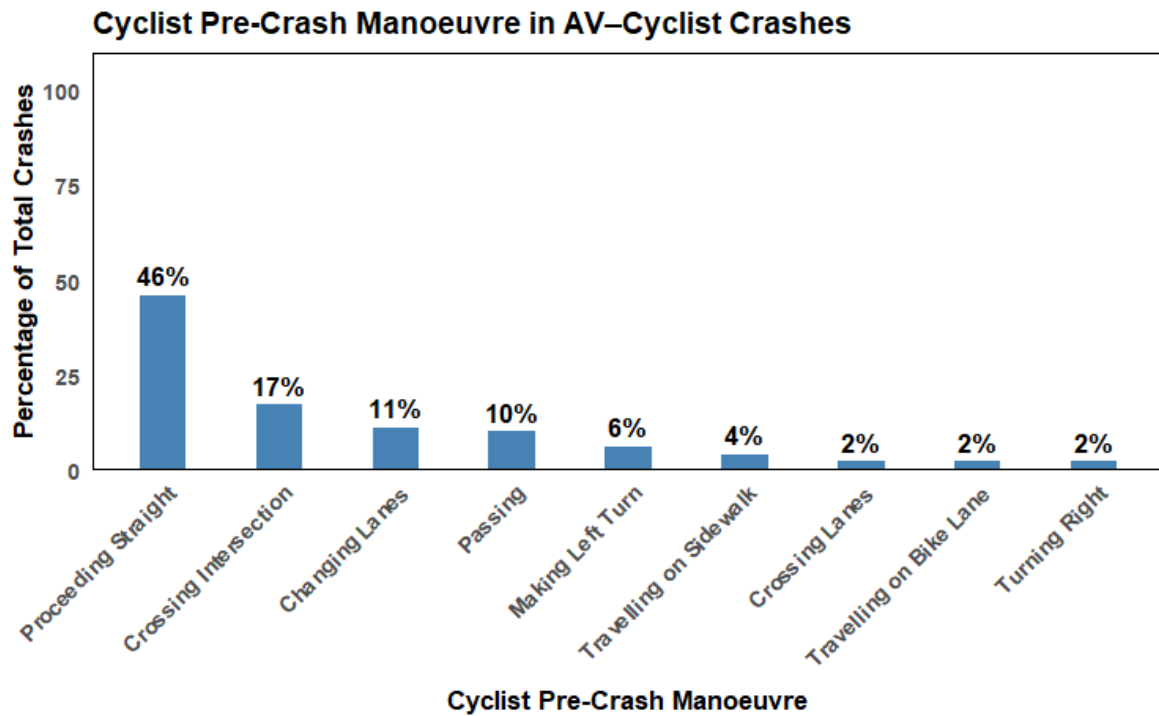


Figure 9 : Frequency distribution of cyclist pre-crash manoeuvre in AV-to-cyclist crashes

The bar chart in figure 10 below shows that AVs proceeding straight (36%) and stopped (32%) together account for 68% of AV-to-cyclist collisions. All the other AV manoeuvres including evasive manoeuvres, left turns, right turns (each 6%), backing, braking, parked (each 4%) and merging (2%) are comparatively less frequent.

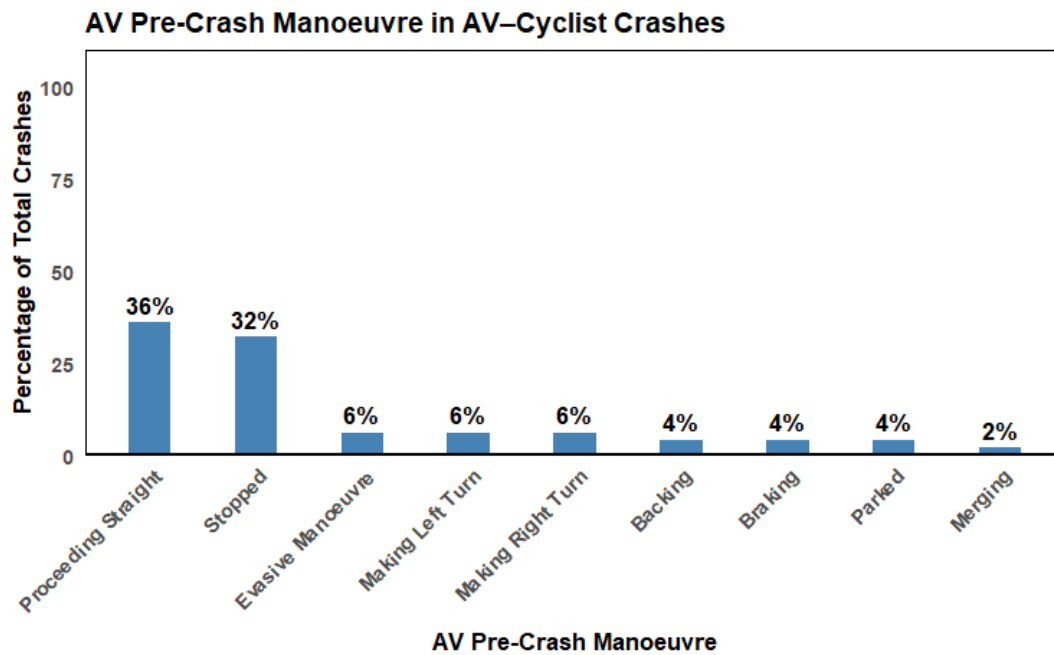


Figure 10 : Frequency distribution of AV pre-crash manoeuvre in AV-to-cyclist crashes

The bar chart in figure 11 below shows trajectories of cyclists relative to AVs in AV-to-cyclist crashes. The largest movement configuration is when a stationary AV is struck by a cyclist proceeding straight ahead (29%), followed by when both AV and cyclist are traveling straight ahead (27%) and when AVs moving straight collide with cyclists crossing their path at intersections (17%). All the other trajectories such as evasive manoeuvres, turning, backing and merging each account for $\leq 6\%$.

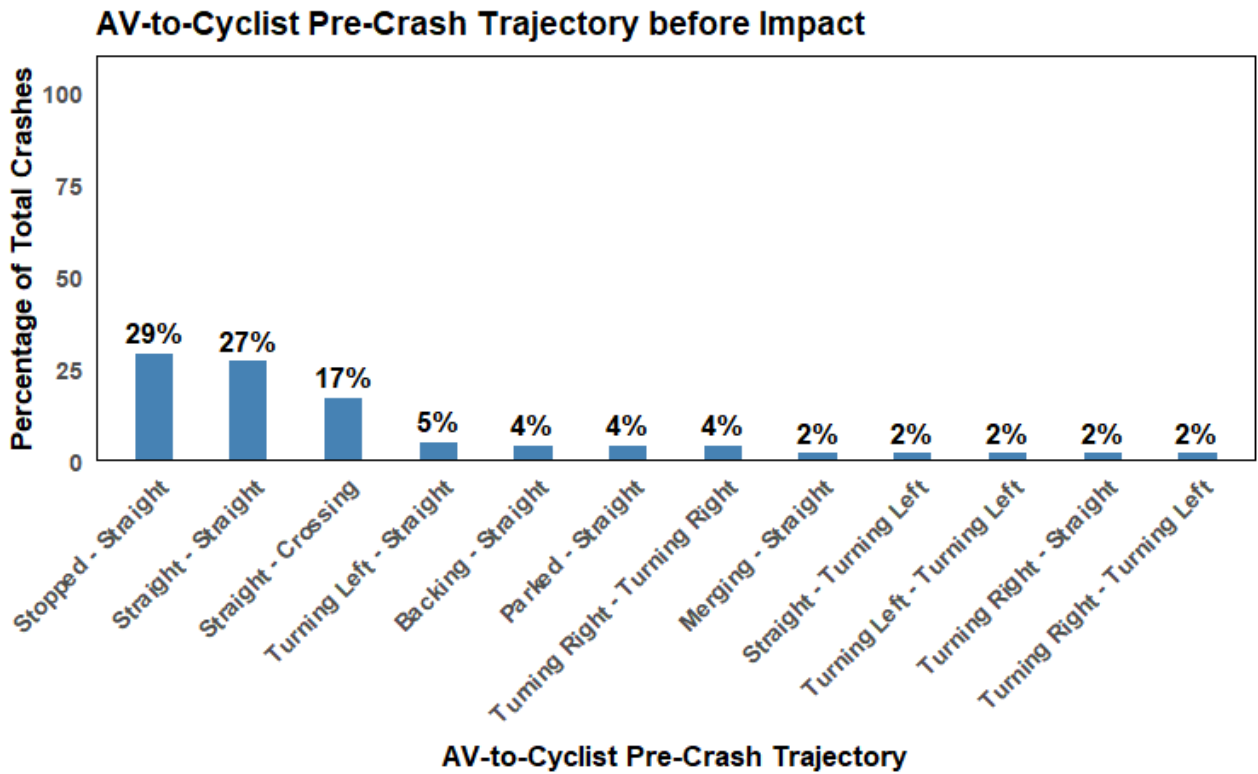


Figure 11 : Frequency distribution of AV-to-cyclist pre-crash trajectory

The bar chart in figure 12 below shows that 94% of AV-to-cyclist crashes occur in clear weather with only 6% occurring in cloudy weather. No crashes are recorded in precipitation.

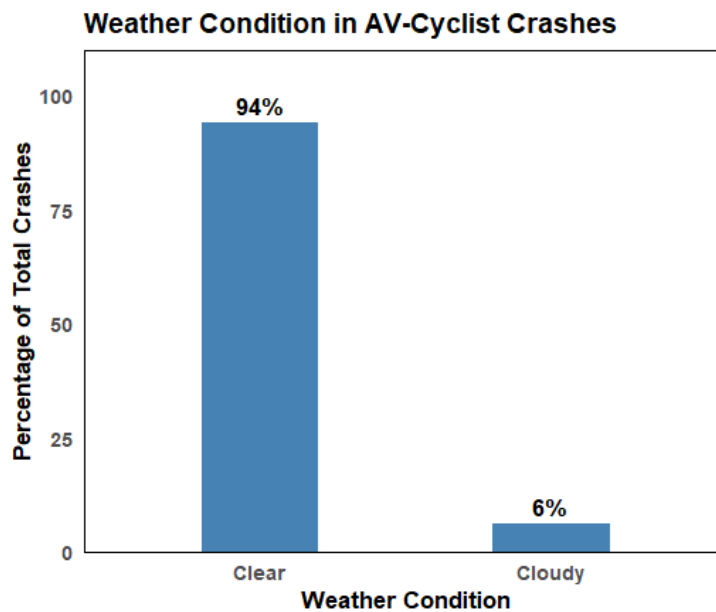


Figure 12 : Weather condition in AV-to-cyclist crashes

The bar chart in figure 13 below shows that 79% of AV-to-cyclist crashes occur during the day, 19% occur at night under artificial lighting and 2% occur in twilight.

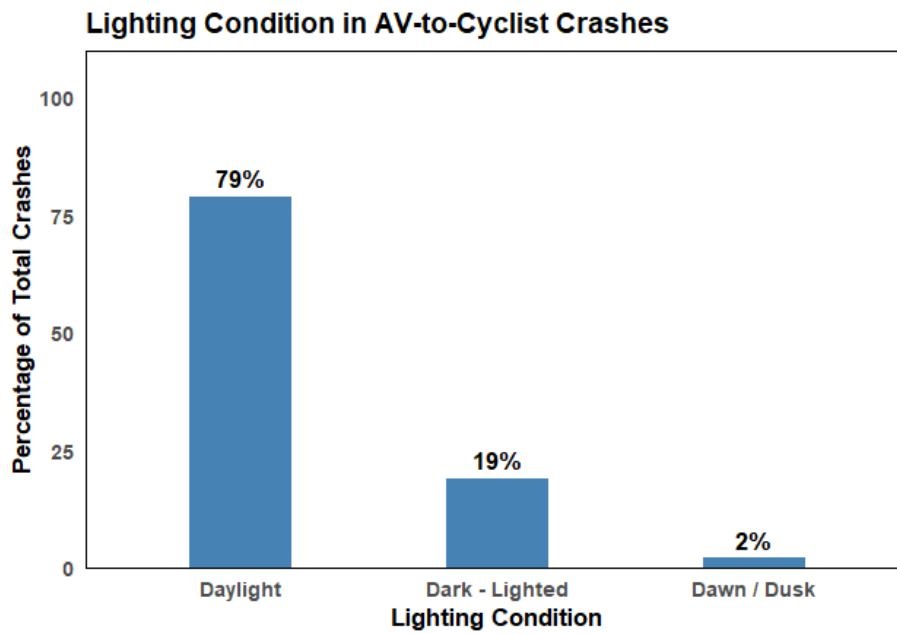


Figure 13 : Lighting condition in AV-to-cyclist crashes

The bar chart in figure 14 below shows that 50% of AV-to-cyclist crashes occur on streets, 48% on intersections and 2% in parking lots.

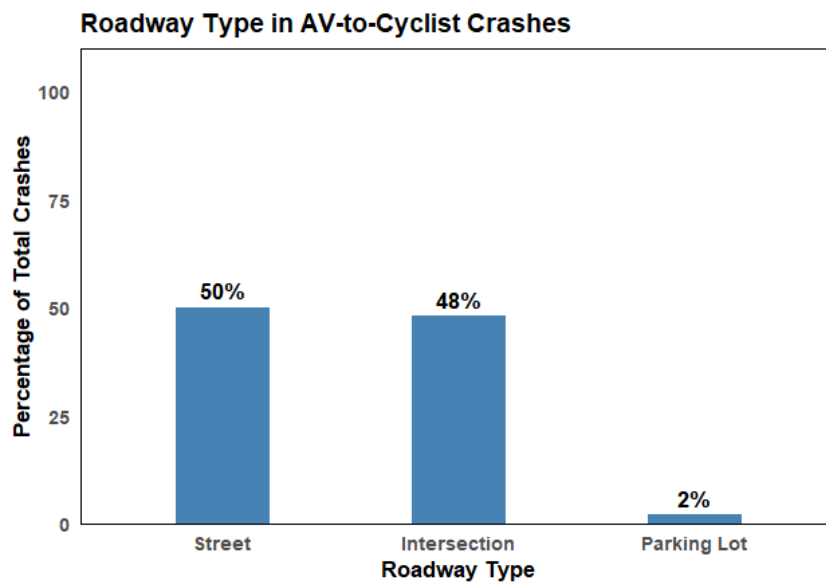


Figure 14 : Frequency distribution of type of roadway in AV-to-cyclist crashes

The bar chart in figure 15 below shows that 71% of AV-to-cyclist crashes occur when the AV has logged between 0–50 000 miles, 21% occur at 50 000–100 000 miles, 6% at 100 000–150 000 miles and 2% at 150 000–200 000 miles.

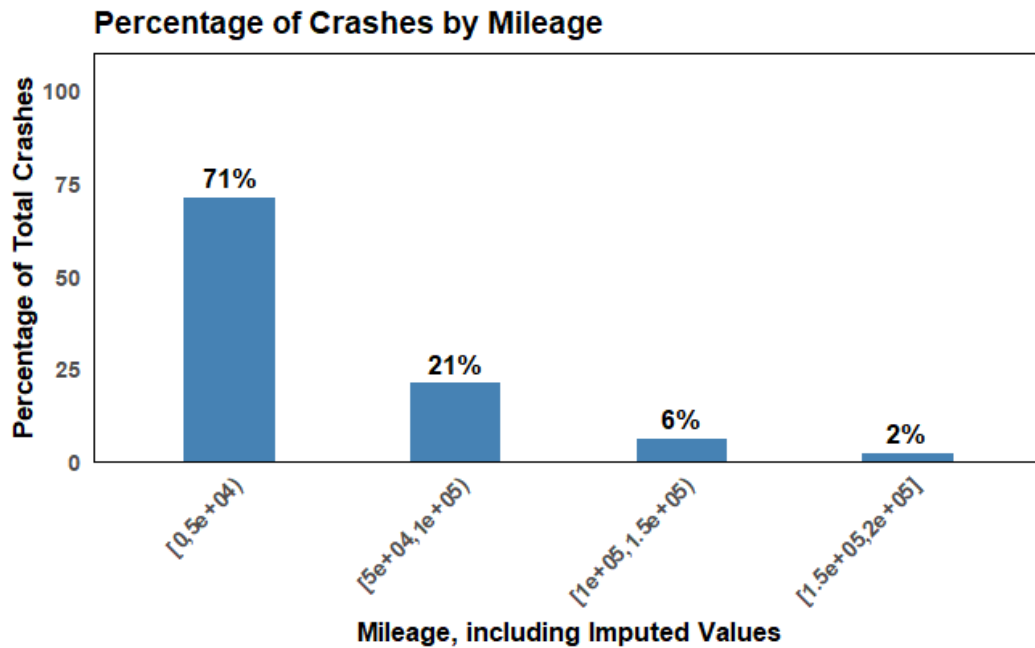


Figure 15 : Distribution of vehicle mileage in AV-to-cyclist crashes

The bar chart in figure 16 below shows that 65% of AV-to-cyclist crashes take place in 25–30 mph speed limit zones, 15% in 20–25 mph zones, 10% in 30–35 mph zones, 4% in 5–10 mph zones, 2% each in 15–20 mph, 35–40 mph and 45–50 mph zones.

Distribution of AV-Cyclist Crashes by Posted Speed Limit

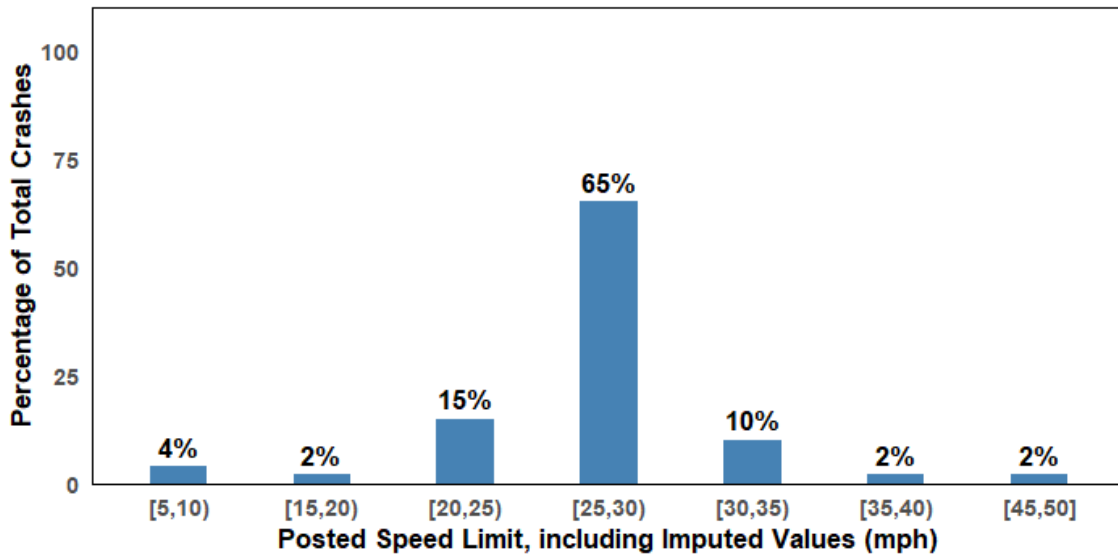


Figure 16 : Distribution of posted speed limit in AV-to-cyclist crashes

The bar chart in figure 17 below shows that 61% of AV-to-cyclist crashes occur when the AV was travelling at 0–5 mph, 25% at 5–10 mph, 10% at 10–15 mph, 2% at 15–20 mph and 2% at 20–25 mph.

Distribution of AV-Cyclist Crashes by AV Pre-Crash Speed

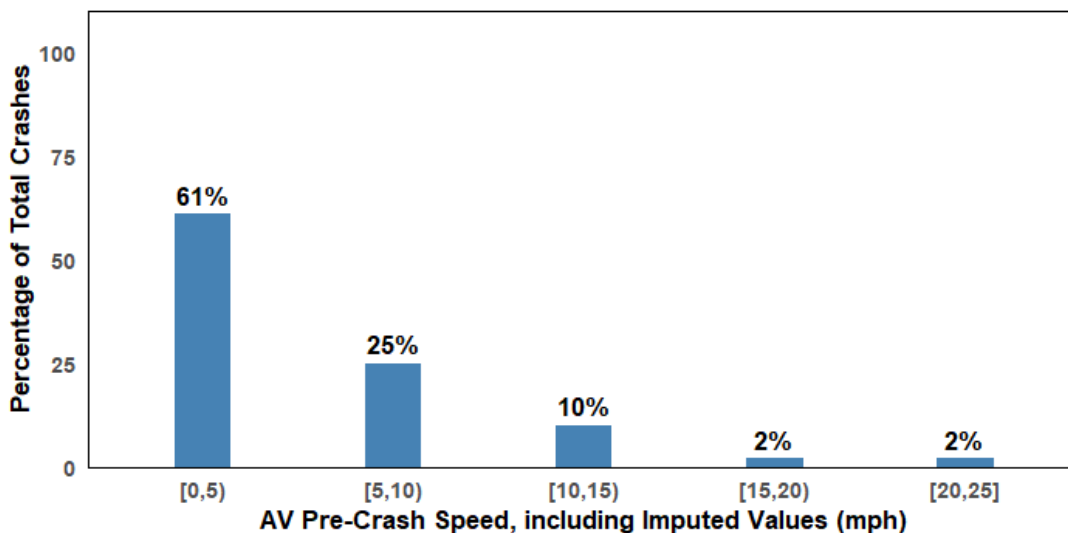


Figure 17 : Distribution of AV pre-crash speed in AV-to-cyclist crashes

Property Damage: 61% of AV-to-cyclist crashes result in property damage, 33% do not involve any damage to property and 6% are undetermined/unknown.

Vehicle Model Year: 50% of AV-to-cyclist crashes involved AVs manufactured in 2020 or 2021.

Cyclist Traffic Violations: Although in 61% of the crashes it is unknown whether the cyclist violated a rule, the most common cyclist traffic violations are traveling the wrong way (11%), running a stop sign (10%), running a red light (8%) and inattention (6%).

Time of crash: 57% of crashes occur between 12:00 and 18:00, 31% occur between 06:00 and 12:00, 8% occur in the evening (18:00–24:00) and 4% occur overnight (00:00–06:00).

Yearly Trend: Crashes have risen sharply from 2% in 2017 to 23% in 2024, with the period from 2022 to 2024 accounting for 65% of all recorded AV-cyclist incidents.

Collision Types: The most frequent types of collisions are broadside (29%), rear-end (23%), sideswipe (23%), dooring (6%) and head-on (6%). In 92% of the rear-end collisions, the cyclist struck the rear of the AV while in 88% of the broadside collisions, the cyclist struck the side of the AV.

Bar charts of the above-described AV-to-cyclist crash attributes are presented in *Appendix B: Frequency bar charts of non-key variables in AV-to cyclist and CV-to-cyclist crash datasets*.

4.2.2. Conventional Vehicle-to-Cyclist Crashes

The bar chart in figure 18 below shows the outcomes of CV-to-cyclist collisions by the highest injury sustained in each crash. 48.5% of the crashes are classified as “possible injury,” meaning the cyclist may have experienced aches, bruises, or minor injury that isn’t clearly documented. 43.5% resulted in confirmed minor injuries; 7.5% resulted in serious injury, while 0.5% resulted in fatal injuries.

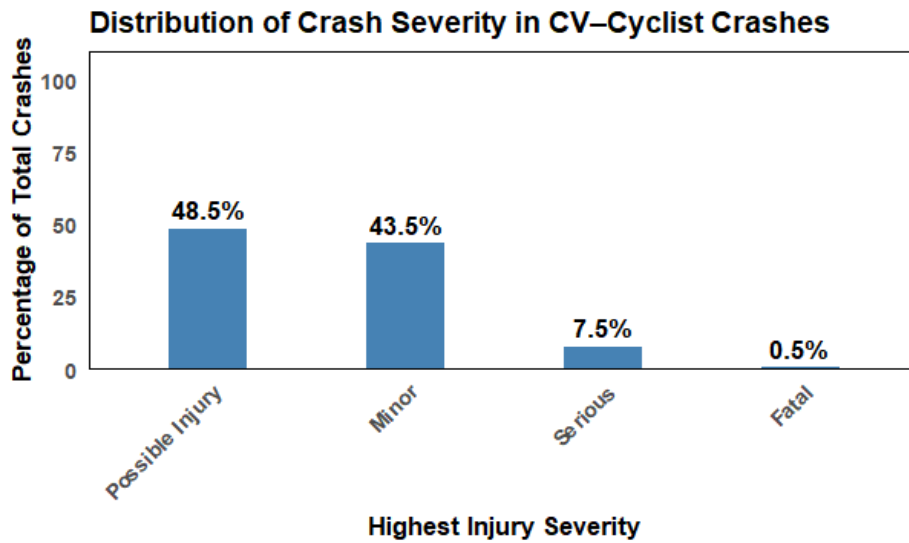


Figure 18 : Frequency distribution of crash severity in CV-to-cyclist crashes

The distribution in figure 19 below show that 81% of the collisions in CV-to-cyclist crashes occur when the cyclist is traveling straight ahead. Turning movements account for 8% of the crashes, while other manoeuvres such as changing lanes, entering traffic, passing another vehicle, stopped cyclists, riding the wrong way and other scenarios altogether account for only 11 % of crashes.

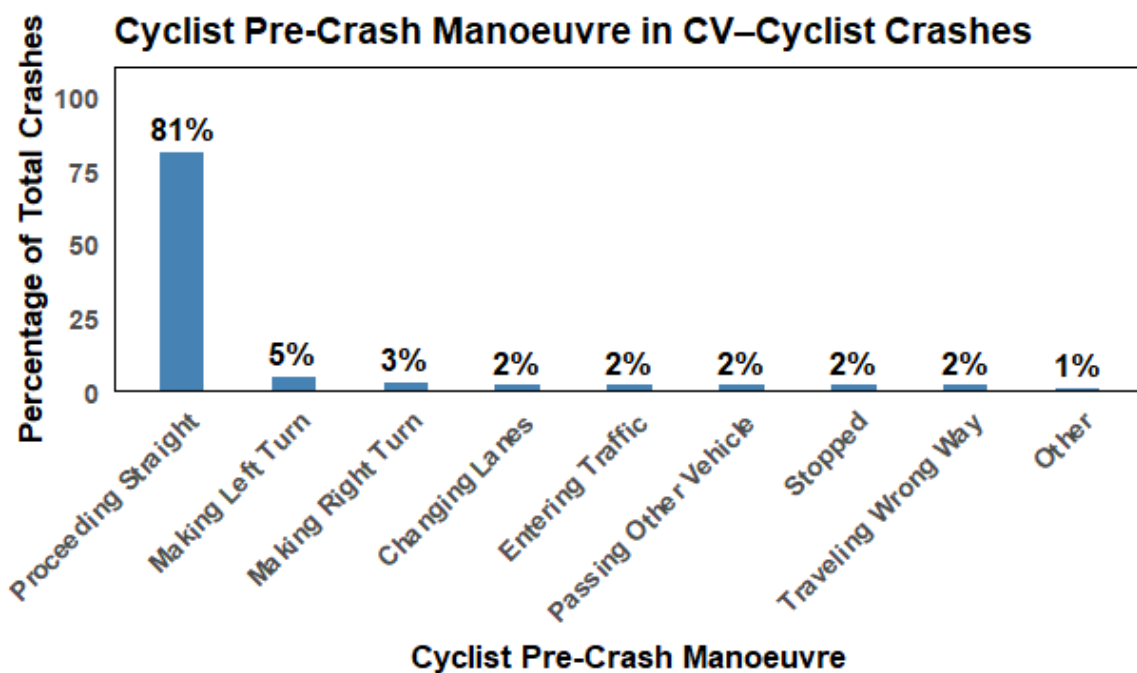


Figure 19 : Frequency distribution of cyclist pre-crash manoeuvre in CV-to-cyclist crashes

The distribution in figure 20 below show that when the driver is proceeding straight (35%), making left turns (18%) and making right turns (17%) together constitute 70% of CV-to-cyclist crashes. The remaining 30% are spread across stopped (8%), parked (6%), making U-turn (4%), entering traffic (3%), changing lanes, backing, parking, passing or stopping (1–2% each).

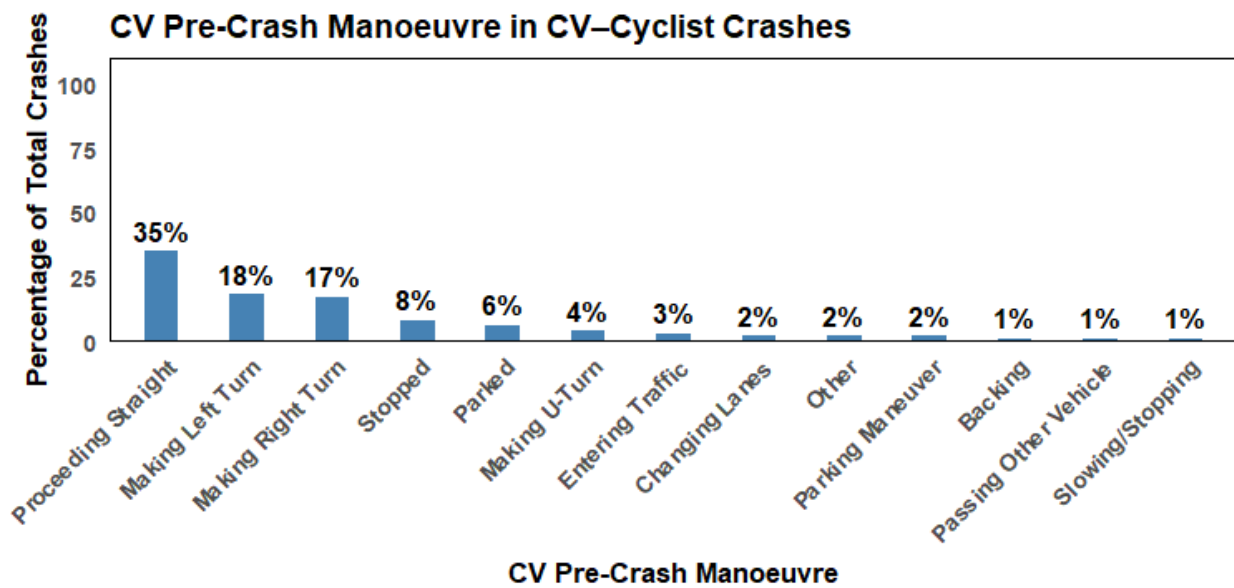


Figure 20 : Frequency distribution of driver pre-crash manoeuvre in CV-to-cyclist crashes

The distribution in figure 21 below show collisions where both driver and cyclist are traveling straight to be the most common single trajectory configuration (24%). Drivers turning left or right into straight-moving cyclists account for 31% of crashes. Other observed trajectories include parked–straight (6%), stopped–straight (6%), and making U-turn–straight (4%) and other low frequency trajectories (23%).

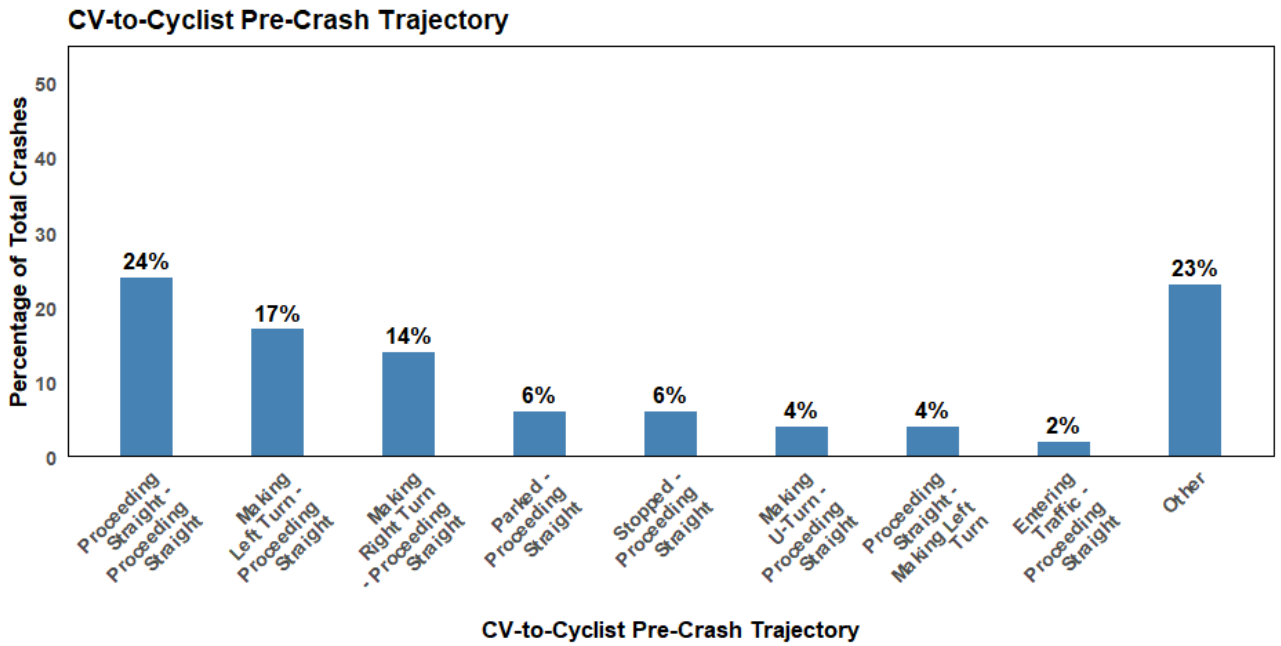


Figure 21 : Frequency distribution of CV-to-cyclist pre-crash trajectory

Figure 22 below show that 91% of CV-to-cyclist crashes occur in clear weather, 6% occur in cloudy weather and 3% occur during rain.

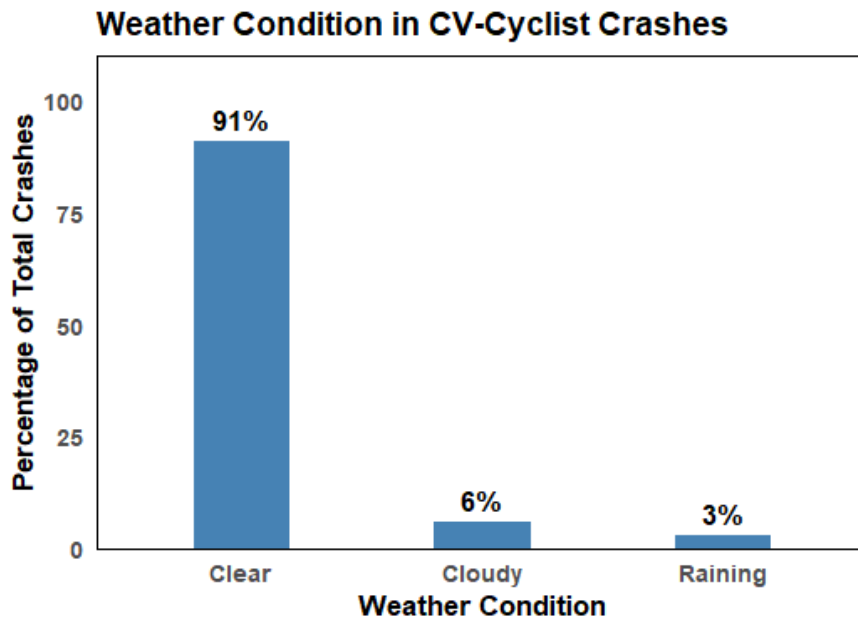


Figure 22 : Weather condition in CV-to-cyclist crashes

Figure 23 below show that 75% of CV-to-cyclist crashes occur in daylight, 20% occur at night under artificial illumination, 4% occurring in twilight and 1% occur on unlit roadways.

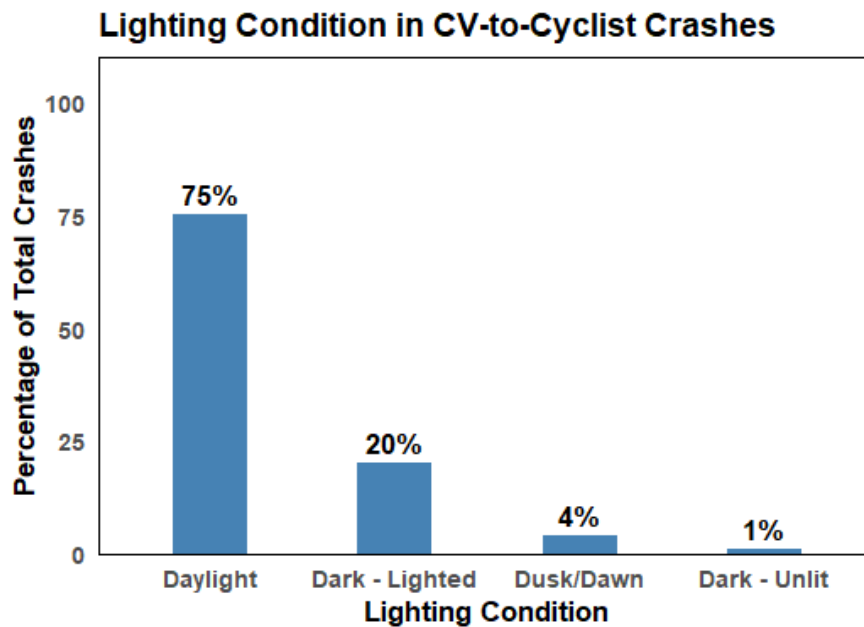


Figure 23 : Lighting conditions in CV-to-cyclist crashes

Figure 24 below show that that 63% of CV-to-cyclist crashes occur on intersections and 37% on non-intersections.

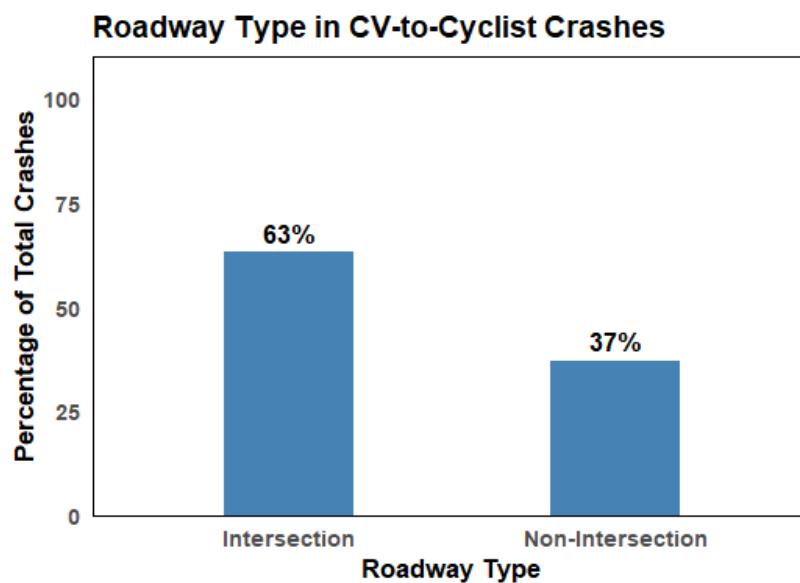


Figure 24 : Type of roadway in CV-to-cyclist crashes

Vehicle Model Year: 60% of CV-to-cyclist crashes involved CVs manufactured between 2010 and 2019.

Time of crash: 41% of crashes occur between 12:00 and 18:00, 30% occur between 06:00 and 12:00, 27% occur in the evening (18:00–24:00) and 2% occur overnight (00:00–06:00).

4.3. Parameter Estimation

The parameter estimates are assessed at 90% confidence interval where p value = 0.01

4.3.1. Multinomial Logit Model for AV-to-cyclist Crashes

The estimates from the multinomial logit models fitted for AV-to-cyclist crashes are presented in tables 9 and 10 below.

Minor Injury vs No Injury								
Predictor	Estimate	Std. Error	z value	p value	90% CI Lower	90% CI Upper	Odds Ratio	Odds Ratio (%)
(Intercept)	0.11225587	0.6690424	0.17	0.867	-0.988	1.213	1.12	+12 %
Vehicle Model Year	0.07878232	0.4463266	0.18	0.860	-0.655	0.813	1.08	+8 %
Cyclist Precrash Manoeuvre: Proceeding Straight	1.1097868	0.8338499	1.32	0.186	-0.269	2.489	3.03	+203%
AV Precrash Manoeuvre: Proceeding Straight	-1.38287903	0.8620806	-1.60	0.109	-2.801	0.035	0.25	-75%
AV Precrash Manoeuvre: Stopped	-2.777244386	1.1627241	-2.38	0.017	-4.685	-0.860	0.06	-94%
Roadway Type: Street vs Intersection	0.07083313	0.8661285	0.08	0.935	-1.354	1.495	1.07	+7.3

Table 9: Parameter Estimates for Minor Injury AV-to-cyclist crashes

When the autonomous vehicle was stationary prior to impact, the odds of the cyclist suffering a minor injury decrease by about 94% (OR ≈ 0.06, p = 0.017), compared to other AV manoeuvres.

Although marginally non-significant (p = 0.109), autonomous vehicles proceeding straight prior to impact decreases odds of minor injury by about 75% (OR ≈ 0.25).

Although statistically non-significant ($p = 0.186$), the estimate suggests that relative to other cyclist manoeuvres, cyclist proceeding straight prior to impact nearly triple the odds of minor injury ($OR \approx 3.03$; +203%).

AV model year and type of roadway show trends but do not have any significant effects on minor injuries (p values > 0.1).

Serious Injury vs No Injury								
Predictor	Estimate	Std. Error	z value	p value	90% CI Lower	90% CI Upper	Odds Ratio	Odds Ratio (%)
(Intercept)	-2.1050716	1.1558851	-1.82	0.069	-4.006	-0.204	0.12	-88%
Vehicle Model Year	0.2447703	0.4909969	0.50	0.618	-0.563	1.052	1.28	-28%
Cyclist Precrash Manoeuvre: Proceeding Straight	1.7586693	1.0002354	1.76	0.079	0.113	3.404	5.80	+480%
AV Precrash Manoeuvre: Proceeding Straight	0.3859855	1.1797706	0.33	0.744	-1.555	2.327	1.47	+47%
AV Precrash Manoeuvre: Stopped	-1.3934614	1.2720984	-1.10	0.273	-3.486	0.699	0.25	-75%
Roadway Type: Street vs Intersection	0.7248081	0.9812732	0.74	0.460	-0.889	2.339	2.06	+106%

Table 10: Parameter Estimates for Serious Injury AV-to-cyclist crashes

The intercept for serious injury is significantly negative (-2.1050716). Thus, the odds of serious injury are 88% less likely compared to no injury ($OR \approx 0.12$; $p = 0.069$).

Compared to other pre-crash manoeuvres, when the cyclist was proceeding straight prior to impact, odds of serious injury increase nearly six-fold (+480%; $OR \approx 5.8$; $p = 0.079$).

AV model year, type of roadway, AV proceeding straight or stopped show trends but do not have any significant effects on serious injuries (p values > 0.1).

Accuracy across 5 folds: 0.55, 0.3, 0.64, 0.56 and 0.73

Average cross validation accuracy:0.55

4.3.2. Multinomial Logit Model for CV-to-cyclist Crashes

The estimates from the multinomial logit models fitted for CV-to-cyclist crashes are presented in tables 11 and 12 below.

Minor Injury vs Possible Injury								
Predictor	Estimate	Std.Error	z value	p value	90% CI Lower	90% CI Upper	Odds Ratio	Odds Ratio (%)
(Intercept)	-0.04033317	0.14633303	-0.28	0.783	-0.281	0.200	0.96	-4%
CV Vehicle Model Year	0.01300979	0.04485943	0.29	0.772	-0.061	0.087	1.01	+1%
Cyclist Pre-crash Manoeuvre: Proceeding Straight	0.23780897	0.11695965	2.03	0.042	0.045	0.430	1.27	+27%
CV Pre-crash Manoeuvre: Proceeding Straight	-0.34430048	0.11948613	-2.88	0.004	-0.541	-0.148	0.71	-29%
CV Pre-crash Manoeuvre: Turning	-0.31132319	0.11950016	-2.61	0.009	-0.508	-0.115	0.73	-27%
Roadway Type: Non-intersection vs Intersection	-0.07143010	0.10125088	-0.71	0.481	-0.238	0.095	0.93	-7%

Table 11: Parameter Estimates for Minor Injury CV-to-cyclist crashes

Relative to other cyclist manoeuvres, cyclist proceeding straight prior to impact increases the odds of minor injury by about 27% (OR \approx 1.27; $p = 0.042$).

Compared to other pre-crash manoeuvres, when the human driven vehicle (CV) was proceeding straight or turning left or right prior to impact, odds of minor injury decrease by about 29% and 27% respectively (OR \approx 0.71 and 0.73; $p = 0.004$ and 0.009).

CV model year and type of roadway do not show any significant effects on minor injuries (p values > 0.1).

Serious Injury vs Possible Injury								
Predictor	Estimate	Std.Error	z value	p value	90% CI Lower	90% CI Upper	Odds Ratio	Odds Ratio (%)
(Intercept)	-2.18543148	0.28039523	-7.79	0.000	-2.647	-1.724	0.11	-89%
CV Vehicle Model Year	-0.09064268	0.07582107	-1.20	0.232	-0.215	0.034	0.91	-9%
Cyclist Pre-crash Manoeuvre: Proceeding Straight	0.42365782	0.21785779	1.94	0.052	0.065	0.782	1.53	+53%
CV Pre-crash Manoeuvre: Proceeding Straight	0.29043778	0.21231855	1.37	0.171	-0.059	0.640	1.34	+34%
CV Pre-crash Manoeuvre: Turning	-0.29989404	0.23236440	-1.29	0.197	-0.682	0.082	0.74	-26%
Roadway Type: Non-intersection vs Intersection	0.07262602	0.18144073	0.40	0.689	-0.226	0.371	1.08	+8%

Table 12: Parameter Estimates for Serious Injury CV-to-cyclist crashes

The intercept for serious injury is significantly negative (-2.18543148). Thus, the odds of serious injury are 89% less likely compared to possible injury (OR ≈ 0.11; p = 0.000).

Relative to other cyclist manoeuvres, cyclist proceeding straight prior to impact increases the odds of serious injury by about 53% (OR ≈ 1.53; p = 0.052).

CV model year, type of roadway, CV proceeding straight and CV turning right or left, do not show any significant effects on serious injuries (p values > 0.1).

Accuracy across 5 folds: 0.49, 0.48, 0.54, 0.51 and 0.51

Average cross validation accuracy:0.51

4.3.3. Multinomial Logit Model for pooled AV and CV datasets

The estimates from the pooled multinomial logit models fitted for the combined AV-to-cyclist and CV-to-cyclist crashes are presented in tables 13 and 14 below.

Minor Injury vs No Injury (Possible Injury)								
Predictor	Estimate	Std.Error	z value	p value	90% CI Lower	90% CI Upper	Odds Ratio	Odds Ratio (%)
(Intercept)	-0.04434639	0.14432698	-0.31	0.759	-0.282	0.193	0.96	-4%
Vehicle Model Year	0.01325588	0.04511278	0.29	0.769	-0.061	0.088	1.01	+1%
Cyclist Pre-crash Manoeuvre: Proceeding Straight	0.25963349	0.11544174	2.25	0.025	0.070	0.450	1.30	+30%
Vehicle Pre-crash Manoeuvre: Proceeding Straight	-0.36503802	0.11797606	-3.09	0.002	-0.559	-0.171	0.69	-31%
Vehicle Pre-crash Manoeuvre: Stopped	-1.81226556	0.85477815	-2.12	0.034	-3.218	-0.406	0.16	-84%
Vehicle Pre-crash Manoeuvre: Turning	-0.32615291	0.11884793	-2.74	0.006	-0.522	-0.131	0.72	-28%
Roadway Type: Non-intersection vs Intersection	-0.07398656	0.10025031	-0.74	0.461	-0.239	0.091	0.93	-7%
Vehicle Type: AV vs CV	-0.05672477	0.38539879	-0.15	0.883	-0.691	0.577	0.96	-6%

Table 13: Parameter Estimates for Minor Injury AV+CV-to-cyclist crashes

Relative to other cyclist manoeuvres, cyclist proceeding straight prior to impact increases the odds of minor injury by about 30% (OR \approx 1.30; $p=$ 0.025).

Vehicle proceeding straight or turning reduces the odds of the cyclist suffering a minor injury by about 31% and 28% respectively (OR \approx 0.69, $p =$ 0.02; OR \approx 0.72, $p =$ 0.06).

When a vehicle was already stopped prior to impact, the odds of the cyclist suffering a minor injury decrease by about 84% (OR \approx 0.16, $p =$ 0.034), compared to other vehicle manoeuvres.

Involvement of an autonomous vehicle (AV) in a crash as compared to a human-driven vehicle has no significant effect on the chances of minor injury.

Vehicle model year and type of roadway do not have any significant effects on odds of minor injuries (p values $>$ 0.1).

Serious Injury vs No Injury (Possible Injury)								
Predictor	Estimate	Std.Error	z value	p value	90% CI Lower	90% CI Upper	Odds Ratio	Odds Ratio (%)
(Intercept)	-2.27207318	0.27511263	-8.26	0.000	-2.725	-1.820	0.10	-90%
Vehicle Model Year	-0.08793031	0.07604085	-1.16	0.248	-0.213	0.037	0.92	-8%
Cyclist Pre-crash Manoeuvre: Proceeding Straight	0.49285292	0.21300540	2.31	0.021	0.143	0.843	1.64	+64%
Vehicle Pre-crash Manoeuvre: Proceeding Straight	0.31334832	0.20730254	1.51	0.131	-0.028	0.654	1.37	+37%
Vehicle Pre-crash Manoeuvre: Stopped	-0.39671782	0.81881567	-0.48	0.628	-1.744	0.950	0.67	-33%
Vehicle Pre-crash Manoeuvre: Turning	-0.28685779	0.23010417	-1.25	0.213	-0.665	0.092	0.75	-25%
Roadway Type: Non-intersection vs Intersection	0.10786617	0.17689562	0.61	0.542	-0.183	0.399	1.11	+11%
Vehicle Type: AV vs CV	0.96954735	0.50411958	1.92	0.054	0.140	1.799	2.64	+164%

Table 14: Parameter Estimates for Serious Injury AV+CV-to-cyclist crashes

Relative to other cyclist manoeuvres, cyclist proceeding straight prior to impact increases the odds of serious injury by about 64% (OR \approx 1.64; $p = 0.021$).

A crash involving an autonomous vehicle increases the likelihood of serious injury by about 164% (OR \approx 2.64; $p = 0.054$), compared to a crash involving a conventional vehicle.

Although marginally non-significant ($p = 0.131$; CI includes zero), vehicles proceeding straight prior to impact increases odds of serious injury by about 37% (OR \approx 1.37).

It is observed that the results in the pooled model mirror and therefore reinforce the findings in the separate AV and CV models.

Accuracy across 5 folds: 0.49, 0.48, 0.54, 0.51 and 0.51

Average cross validation accuracy:0.51

4.4. Model Validation Results

The Hausman–McFadden test was used to evaluate the Independence of Irrelevant Alternatives (IIA) assumption by comparing two multinomial logit specifications:

- i. Full model: comprising “No Injury,” “Minor,” and “Serious” injury severity levels
- ii. Restricted model: comprising “No Injury” and “Minor” injury severity levels only

The test yielded:

$$\chi^2 = -0.0134 \text{ on } 12 \text{ df, with } p = 1.00$$

5. Discussion

5.1. Descriptive Analysis of Crashes

5.1.1. Autonomous Vehicle-to-Cyclist Crashes

Cyclists are the predominant vulnerable road users (VRU) in AV collisions, accounting for 71% of the collisions. This reflects both their high exposure in traffic and specific detection or prediction challenges that AV systems face with bicycles.

Most AV-to-cyclist crashes are low in severity. 83% of the crashes result in either no injuries or only minor injuries, implying that even when collisions occur, AV systems may be reducing impact forces or speeds. Although serious injuries are least common, an 8% rate is non-negligible and highlights the need for continued refinement of AV braking and avoidance systems around cyclists.

The most frequent cyclist movement immediately prior to collision with an AV is the straight-ahead movement. 46% of crashes occur while the cyclist is riding straight ahead while 17% occur when cyclists are traversing intersections. Lateral movements including lane changes (11%) and passing maneuvers (10%) account for 21% of crashes, with left turns (6%) and other scenarios (riding on sidewalk or bike lane, right turns and crossing lanes) comprising the remainder. This implies that AV systems face challenges in detecting straight-ahead cyclist motion, perceiving lateral cyclist behavior, predicting cyclist intention at intersections and indicates the diverse scenarios that AV systems ought to recognize in a cyclist path to reduce collision risk.

AVs most often collide with cyclists either when moving straight ahead (36%) or while stationary (32%). The high proportion of crashes occurring while AVs are stopped suggests a substantial risk during start/stop phases, when cyclists fail to notice a stopping or stationary AV or when AV sensors struggle to detect approaching cyclists while stopping.

The most frequent AV-to-cyclist pre-crash trajectory is Stopped–Straight (29%), followed by Straight–Straight (27%) and Straight–Crossing (17%).

Nearly all AV-to-cyclist collisions occur in clear weather (94%), suggesting that either AV operations and exposure are largely confined to clear conditions or that adverse weather testing is limited.

Although 79% of AV-to-cyclist collisions occur in daylight, a notable 19% of crashes occur at night under artificial illumination. Very few collisions take place in twilight, suggesting lower AV exposure at those times.

AVs face nearly equal crash risk on streets (50%) and at intersections(48%), while off-road environments such as parking lots represent a negligible share of crashes (2%).

Most AV-to-cyclist crashes (71%) involve vehicles with low mileage (0–50,000 miles), suggesting they tend to happen early in deployment when sensor calibration and software tuning are still being refined.

Most AV-to-cyclist crashes (65%) occur in standard urban speed zones, particularly 25–30 mph, indicating where AVs operate most frequently and cyclist exposure is highest.

Most AV-to-cyclist crashes (61%) occur at very low speeds, suggesting that although AVs typically detect cyclists and brake, they cannot fully avoid impact in confined urban environments and further highlights the vulnerability of AVs to collisions that occur when they are stationary.

61% of AV-to-cyclist crashes result in property damage indicating that even when collisions result in no injury, there is always an extent of property damage to the bicycle, vehicle or roadway infrastructure.

Although violation of traffic rules by cyclists contribute to 39% of the crashes, most crash reports fail to clearly assign fault.

Most crashes occur in the afternoon between 12:00 and 18:00, an indication of when both bicycle and AV traffic peak.

50% of collisions involve AVs manufactured in 2020 or 2021, reflecting an increase in AV deployment during this period.

Crashes have increased rapidly from 2% in 2017 to 23% in 2024, with 65% of crashes occurring between 2022 and 2024, highlighting a rapid year-by-year rise in AV-cyclist collisions as AV deployment increase.

Broadside (29%), rear-end (23%) and sideswipe (23%) impacts, alongside dooring (6%) and head-on impacts (6%) account for most AV-to-cyclist crashes, with cyclists striking the AV's

side in 88% of broadside cases and its rear in 92% of rear-end cases. This highlights shortcomings in AV braking and lateral detection systems in predicting cyclist trajectories.

5.1.2. Conventional Vehicle-to-Cyclist Crashes

Most CV-to-cyclist crashes are low in severity. 92% of CV-to-cyclist collisions result in possible or minor injuries, indicating that while these crashes occur frequently, the most severe outcomes are relatively rare. Although only 7.5% and 0.5% of crashes result in serious injuries and fatalities respectively, these rates are significant and highlight the prevailing safety challenges faced by cyclists.

The dominant cyclist pre-crash movement in CV-to-cyclist crashes is the straight-ahead movement which accounts for 81% of the crashes indicating that drivers often fail to notice or yield to cyclists in this motion. Only a small percentage (8%) of crashes occur when cyclists are turning, suggesting that although intersections present turning conflicts, they only account for a small proportion of total crashes.

The main driver pre-crash movements in CV-to-cyclist crashes are the straight-ahead movement which also accounts for 35% of the crashes and the turning manoeuvre which also accounts for 35% of the crashes (left turning: 18% and right turning: 17%). Human-driven vehicles most often collide with cyclists when proceeding straight or executing turning manoeuvres, reflecting known driver difficulties in yielding and judging cyclist trajectories at intersections.

The most frequent CV-to-cyclist pre-crash trajectory is 'Turning–Straight' (31%), followed by 'Straight–Straight' (24%) and Straight–Crossing (17%).

CV-to-cyclist crashes occur mostly in clear weather conditions (91%), with 6% occurring during cloudy weather and 3% occurring during rain.

Human drivers crash with cyclists mostly during the day (75%) but show a slightly higher vulnerability in low-light and unlit conditions compared to AVs.

Intersections (63%) dominate CV-to-cyclist crashes, with 37% occurring on non-intersection locations.

Most CV-to-cyclist crashes (41%) occur in the afternoon between 12:00 and 18:00.

5.1.3. Comparison of AV-to-Cyclist and CV-to-Cyclist Crashes

In both AV-to-cyclist and CV-to-cyclist crashes, the majority result in low severity injuries. The similar serious injury rates for AVs (8%) and CVs (7.5%), the lower minor injury rate in AV crashes (29% for AV; 43.5% for CV) and the fact that fatalities (0.5%) are observed only in CV crashes, suggest that AVs may reduce both the most and the least severe crash outcomes.

Only 46% of AV-to-cyclist crashes occur in the straight-ahead cyclist movement compared to 81% for human drivers, implying that AVs are faced with a wider range of cyclist movements that they need to detect and predict.

Straight-ahead vehicle movements are common to both AV-to-cyclist (36%) and CV-to-cyclist (35%) crashes. AVs exhibit a unique vulnerability when stationary with 32% of crashes occurring while AVs are stopped, versus only 8% for CVs. This points to a shortcoming in AV design around stationary hazards and cyclist approach detection. 35% of CV collisions happen during turning manoeuvres, compared to just 12% for AVs. This highlights the difficulty human drivers face at intersections, an aspect in which AVs demonstrate a comparative advantage.

The most frequent AV-to-cyclist pre-crash trajectory is Stopped–Straight (29%), followed by Straight–Straight (27%) and Straight–Crossing (17%). On the other hand, the most frequent CV-to-cyclist pre-crash trajectory is ‘Turning–Straight’ (31%), followed by ‘Straight–Straight’ (24%) and Straight–Crossing (17%). This further emphasizes the points raised in the previous paragraph.

Both AV and CV collisions occur almost exclusively in clear weather conditions (94% versus 91%). Given the small proportions outside clear weather, neither vehicle type shows strong weather-driven differences in crash frequency. While CV crashes are likewise concentrated in clear weather conditions, a small number of CV crashes occur during rain, reflecting real-world exposure to wet weather that is absent from the AV sample. This suggests that either AV operations and exposure are largely confined to clear conditions or that adverse-weather testing is limited.

In both AV-to-cyclist and CV-to-cyclist crashes, most crashes occur in daylight (79% for AVs; 75% for CVs). Lighted nighttime scenarios are equally challenging (19% for AVs; 20% for CVs), indicating both AV sensors and human drivers can misjudge cyclists under artificial illumination.

CV-to-cyclist crashes are concentrated on intersections (63%), whereas AV-to-cyclist crashes are distributed almost evenly between intersections (48%) and streets (50%). AVs experience a higher proportion of crashes on non-intersections/streets (50%) compared to CVs (37%), suggesting AV forward-travel detection on road segments remains a key challenge.

In both AV-to-cyclist and CV-to-cyclist crashes, the majority occur in the afternoon between 12:00 and 18:00 (57% for AVs; 41% for CVs), an indication of when both bicycle and vehicle traffic peak.

5.2. Parameter Estimation

5.2.1. Comparison of AV-to-Cyclist and CV-to-Cyclist Crashes

Cyclist “Proceeding Straight”: In both scenarios, a cyclist riding straight ahead raises the odds of more severe crash outcomes. However, AV-to-Cyclist crash scenarios show a larger effect (OR \approx 5.8; +480%), compared to CV-to-Cyclist crash scenarios (OR \approx 1.53; +53%). This implies that AV systems face challenges in detecting straight-ahead cyclist motion.

Although “**AV Stopped**” is a protective factor against minor injury (–94% odds), nearly one-third of AV crashes occur while AV is stationary. This suggests a substantial risk during start/stop phases, when cyclists fail to notice a stopping or stationary AV or when AV sensors fail to detect approaching cyclists while stopping.

Vehicle “Proceeding Straight”: A human-driven vehicle proceeding straight significantly reduces odds of minor injury, whereas AV proceeding straight tend towards reducing odds of minor injury but do not reach significance (OR \approx 0.71; -29%).

Vehicle “Turning”: A human-driven vehicle turning right or left significantly reduces odds of minor injury. This can be attributed to low speeds necessitated by turning movements which consequently mitigate collisions. Autonomous vehicles turning do not show any clear impact on injury severity (OR \approx 0.73; -27%).

Vehicle model year and type of roadway: Newer vehicle model years and intersections/streets show no evidence of effect on injury severity in both scenarios, suggesting these factors have no substantial effect on injury once manoeuvres and vehicle types are accounted for.

The **intercept for serious injury** is significantly negative (≈ -2), indicating that both autonomous and human-driven vehicle crashes with cyclists tend towards low injury severities.

AVs compared to CVs increase the risk of serious injuries: Autonomous vehicle crashes increase the likelihood of serious injury by nearly 2.6 times (164%), compared to human-driven vehicle crashes. There is no evidence of comparable effect of AVs on minor injuries.

The average **cross validation accuracy** across the three model fits ranges from 51% to 55% demonstrating performance well above random chance. However, the wide range of fold-specific accuracies (30% to 73%) indicates that the models' predictive performance is likely highly sensitive to limited sample size and class imbalance.

5.3. Model Validation

The results of the Hausman–McFadden test which was used to evaluate the Independence of Irrelevant Alternatives (IIA) assumption by comparing two multinomial logit specifications yielded results with a statistical significance of $p = 1$. Given that this p value is far above any conventional thresholds, the study fails to reject the null hypothesis of IIA. This implies that omitting the serious injury category has no significant impact on the estimated coefficients, and the odds of minor injury versus no injury remain unaffected by the presence or absence of the serious injury level in the model.

6. Conclusion

6.1. Main Findings

The study reveals straight-ahead cyclist movements to be a key predictor of injury severity for both autonomous and conventional vehicles and highlights this effect to be over three times larger in AV crashes ($OR \approx 5.8$ vs. 1.53). This suggests limitations in AV detection of straight-ahead cyclist trajectories. Although a stationary AV significantly reduces minor-injury odds by about 94%, one-third of AV-cyclist collisions occur at zero speed, revealing a major vulnerability during start/stop phases of AVs. In contrast, human-driven vehicles achieve significant minor-injury reductions when proceeding straight or turning, effects that fail to reach significance for AV manoeuvres reflecting operational differences between the two vehicle types. Vehicle model year and roadway type show no evidence of impact on crash severity and both AVs and CVs predominantly result in low-severity crashes. AV involvement carries a near 2.6-fold increase in serious-injury risk compared with CVs, with no comparable effect on minor injuries, indicating a need for AV systems to refine their severe-outcome avoidance capabilities.

6.2. Study Limitations and Future Research

This study is subject to limitations that should inform both the interpretation of the results and the design of follow-on studies:

- The relatively small number of AV-to-cyclist collisions constrained the statistical power of the multinomial logit models, making it challenging to detect all the effects at the 90% confidence level. Future work should leverage larger AV crash datasets to improve precision and uncover subtler risk factors.
- Severe class imbalances in both injury severity levels and key categorical predictors (e.g., weather, lighting, road surface, pre-crash manoeuvre) necessitated the aggregation of low frequency categories and the exclusion of overly dominant predictors. While this approach stabilized the models, it may have obscured influences of low-frequency scenarios.
- The AV and conventional-vehicle crash datasets differed in structure and level of detail. AV crash records included narrative descriptions that allowed reconstruction of pre-impact trajectories, whereas the CV data were pre-tabulated, limiting deeper crash event reconstruction.

- Like most crash datasets, the CV crash dataset, lacked records of “no injury”. “Possible injury” was therefore used as the baseline severity level for CV crashes, whereas the AV analysis used “no injury”.
- Cyclist perspective of crash events is not adequately captured in the AV crash records as reports are submitted by AV operators and manufacturers. Although police investigations especially in high severity crashes is noted, their findings particularly on cyclists intentions are not detailed in the crash records.

Building on the insights and limitations drawn from this study, the following opportunities for further research were identified:

- Integration of driver and cyclist factors in pre-crash scenario typology for AV-to-cyclist crashes
- Assembling of larger AV crash datasets
- Assessing AV performance under adverse weather such as rain and poor lighting conditions
- Cyclist-centred crash collection

Appendix

Appendix A: Sample excerpt from a crash report obtained from California DMV Database

SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

Autonomous Mode Conventional Mode

A Waymo Autonomous Vehicle ("Waymo AV") was stopped in autonomous mode on southbound N. Shoreline Boulevard at W. Middlefield Road in Mountain View when a bicyclist made contact with the side of the Waymo AV. The Waymo AV was traveling in the far right lane on N. Shoreline Boulevard and slowing for a red light at W. Middlefield Road, when a bicyclist began to cross all lanes of traffic from the east side of N. Shoreline Boulevard to the west side of the boulevard, in a diagonal manner, against the flow of traffic. As the bicyclist approached the Waymo AV, the Waymo AV began braking for the bicyclist. As the Waymo AV came to a stop, the bicyclist then made contact with the driver's side rear quarter panel of the Waymo AV at approximately 8 MPH. The Waymo AV sustained minor scratches to the driver's side rear taillight, and the bicyclist, remaining upright, appeared to sustain no damage or injuries as they cycled away from the scene without exchanging information.

Additional information attached.

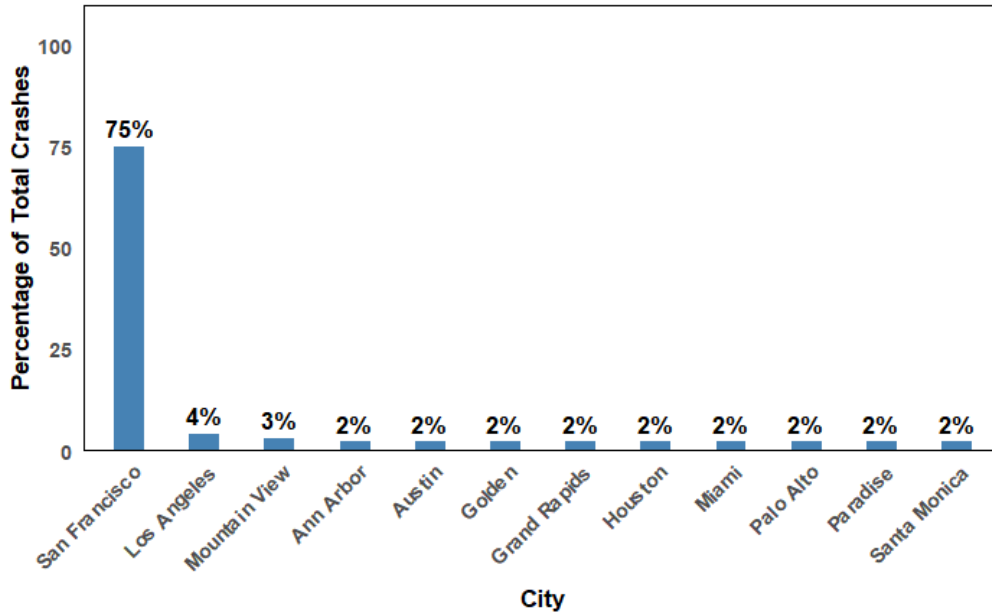
Go to Page 3

DL 316 (REV. 2/2017) WWW

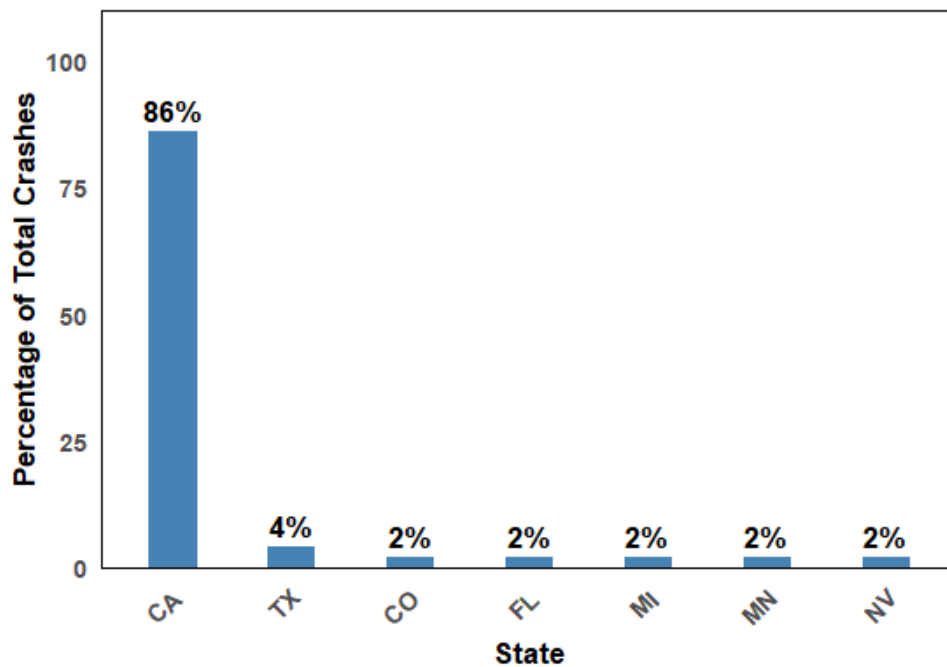
ITEMS MARKED BELOW FOLLOWED BY AN ASTERISK (*) SHOULD BE EXPLAINED IN THE NARRATIVE							
WEATHER (MARK 1 to 2 ITEMS)	VEH 1	VEH 2	MOVEMENT PRECEDING COLLISION	VEH 1	VEH 2	OTHER ASSOCIATED FACTOR(s) (MARK ALL APPLICABLE)	
A. CLEAR	X	X	A. STOPPED			A. CVC SECTIONS VIOLATED	
B. CLOUDY			B. PROCEEDING STRAIGHT			CITED <input type="checkbox"/> YES <input type="checkbox"/> NO	
C. RAINING			C. RAN OFF ROAD				
D. SNOWING			D. MAKING RIGHT TURN				
E. FOG/VISIBILITY			E. MAKING LEFT TURN				
F. OTHER			F. MAKING U TURN			B. VISION OBSCUREMENT <input type="checkbox"/>	
G. WIND			G. BACKING			C. INATTENTION* <input type="checkbox"/>	
LIGHTING			H. SLOWING/STOPPING	X		D. STOP & GO TRAFFIC <input type="checkbox"/>	
A. DAYLIGHT	X	X	I. PASSING OTHER VEHICLE			E. ENTERING/LEAVING RAMP <input type="checkbox"/>	
B. DUSK - DAWN			J. CHANGING LANES			F. PREVIOUS COLLISION <input type="checkbox"/>	
C. DARK - STREET LIGHTS			K. PARKING MANUEVER			G. UNFAMILIAR WITH ROAD <input type="checkbox"/>	
D. DARK - NO STREET LIGHTS			L. ENTERING TRAFFIC			H. DEFECTIVE WEH EQUIP	
E. DARK - STREET LIGHTS NOT FUNCTIONING*			M. OTHER UNSAFE TURNING		X	CITED <input type="checkbox"/> YES <input type="checkbox"/> NO	
ROADWAY SURFACE			N. XING INTO OPPOSING LANE				
A. DRY	X	X	O. PARKED			I. UNINVOLVED VEHICLE <input type="checkbox"/>	
B. WET			P. MERGING			J. OTHER* <input type="checkbox"/>	
C. SNOWY - ICY			Q. TRAVELING WRONG WAY			K. NONE APPARENT <input type="checkbox"/>	
D. SLIPPERY (MUDDY, OILY, ETC.)			R. OTHER*			L. RUNAWAY VEHICLE <input type="checkbox"/>	
ROADWAY CONDITIONS (MARK 1 TO 2 ITEMS)			TYPE OF COLLISION				
A. HOLES, DEEP RUT*			A. HEAD-ON				
B. LOOSE MATERIAL ON ROADWAY			B. SIDE SWIPE				
C. OBSTRUCTION ON							

Appendix B: Frequency bar charts of non-key variables in AV-to cyclist and CV-to-cyclist crash datasets

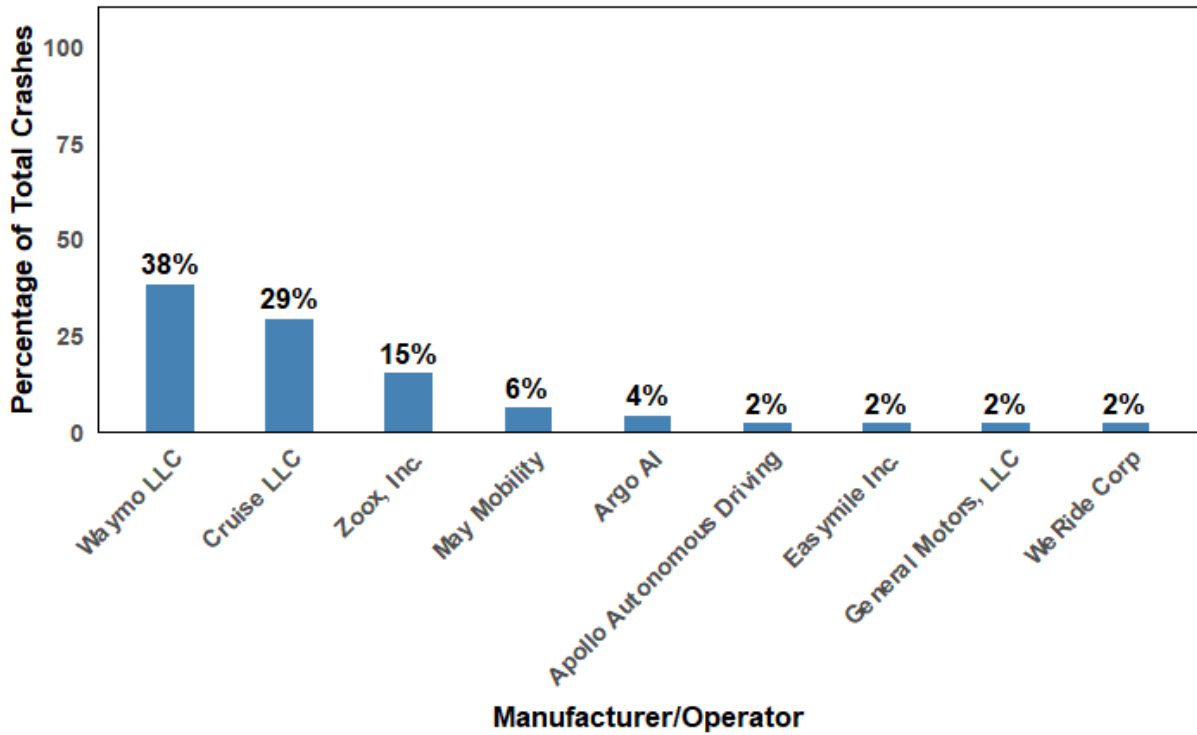
Distribution of AV-to-Cyclist Crashes across US Cities



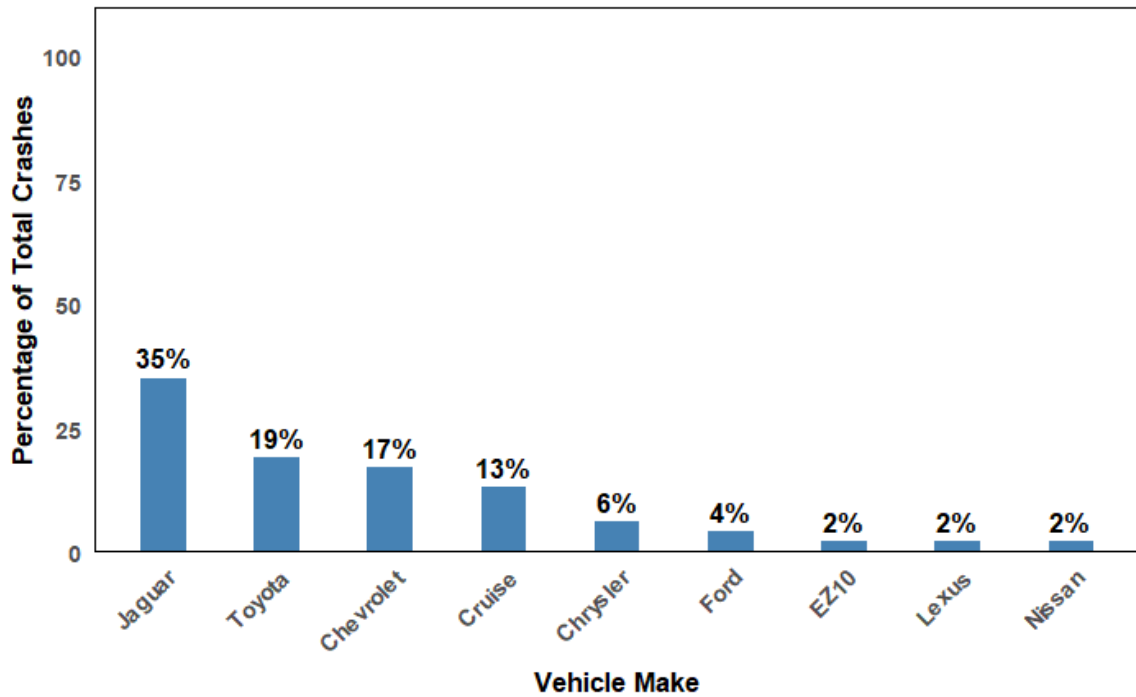
Distribution of AV-to-Cyclist Crashes across States



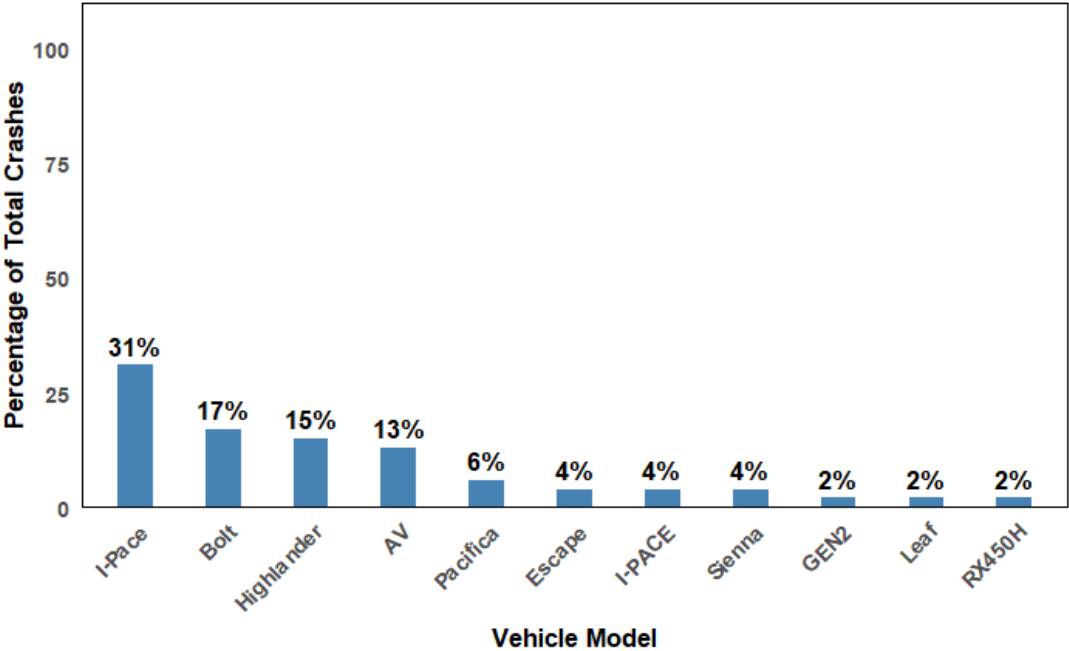
Distribution of AV-to-Cyclist Crashes among Manufacturers/Operators



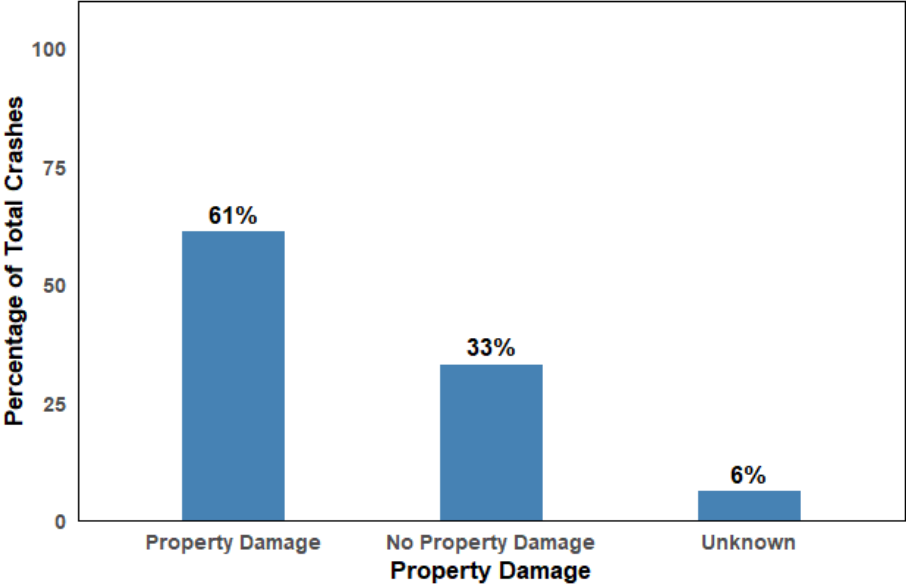
Distribution of AV-to-Cyclist Crashes across Vehicle Make



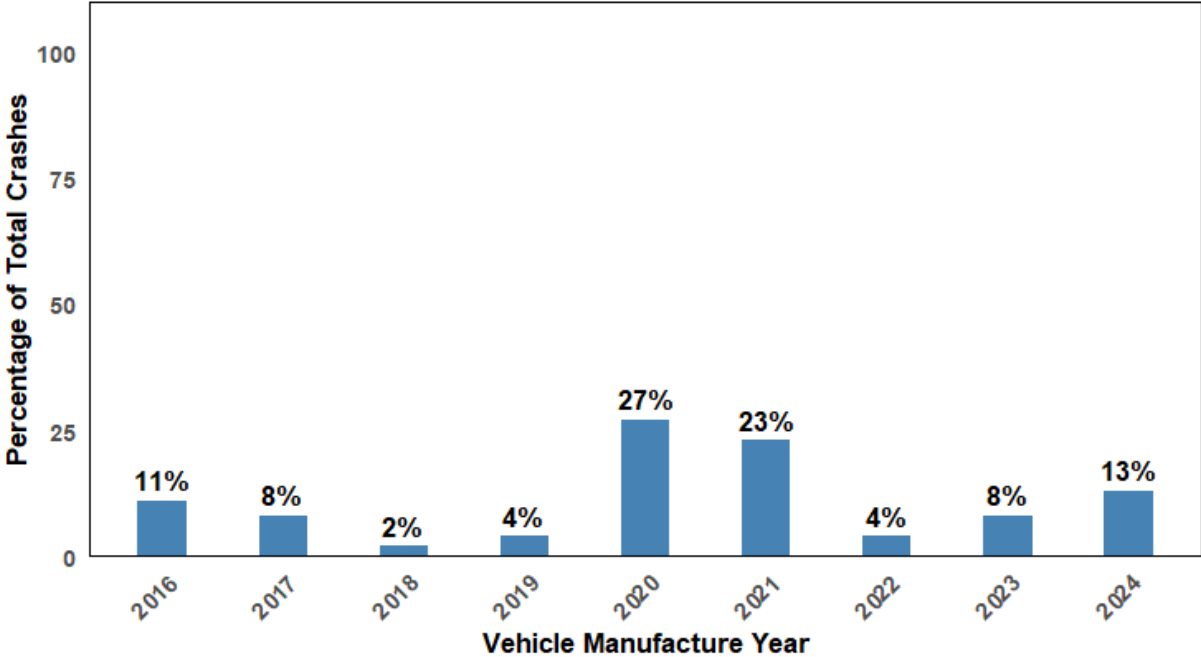
Distribution of AV-to-Cyclist Crashes across Vehicle Model



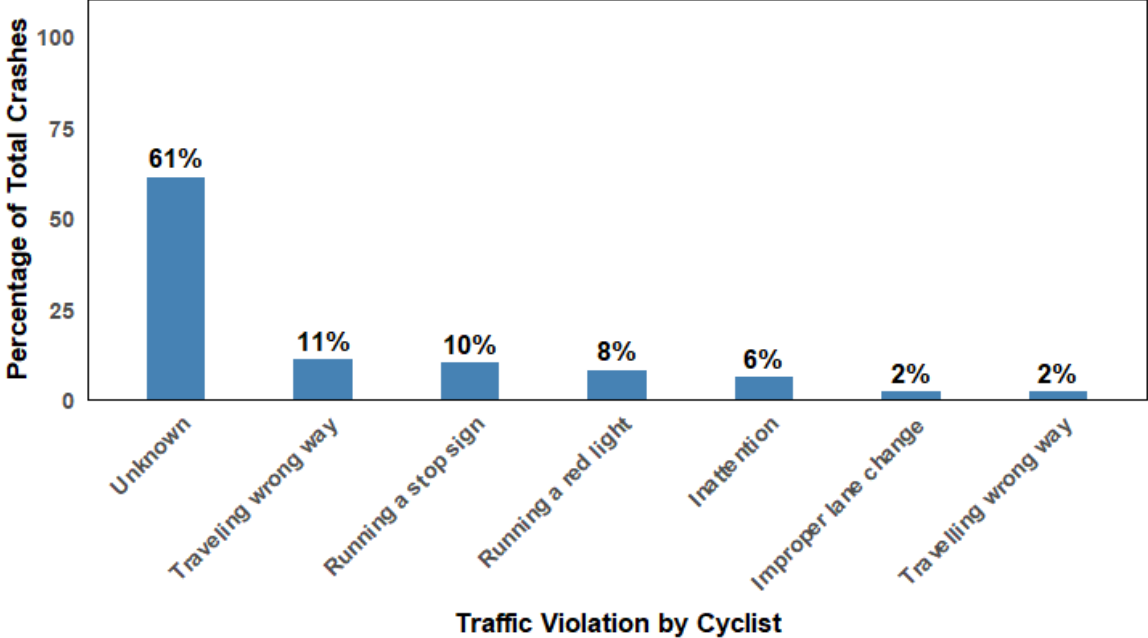
Property Damage in AV-to-Cyclist Crashes



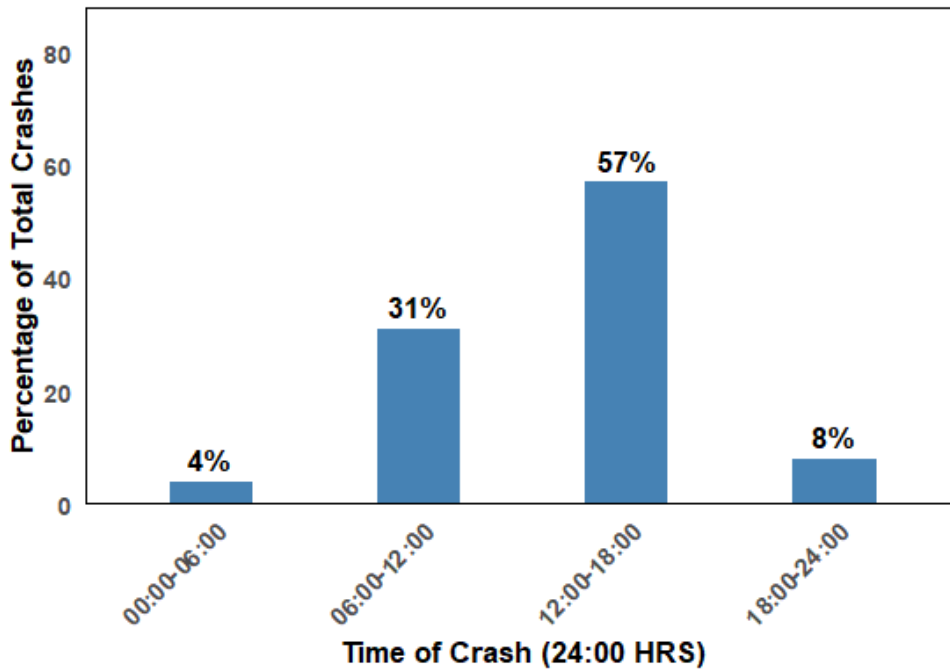
Distribution of AV-to-Cyclist Crashes by Year of Vehicle Manufacture



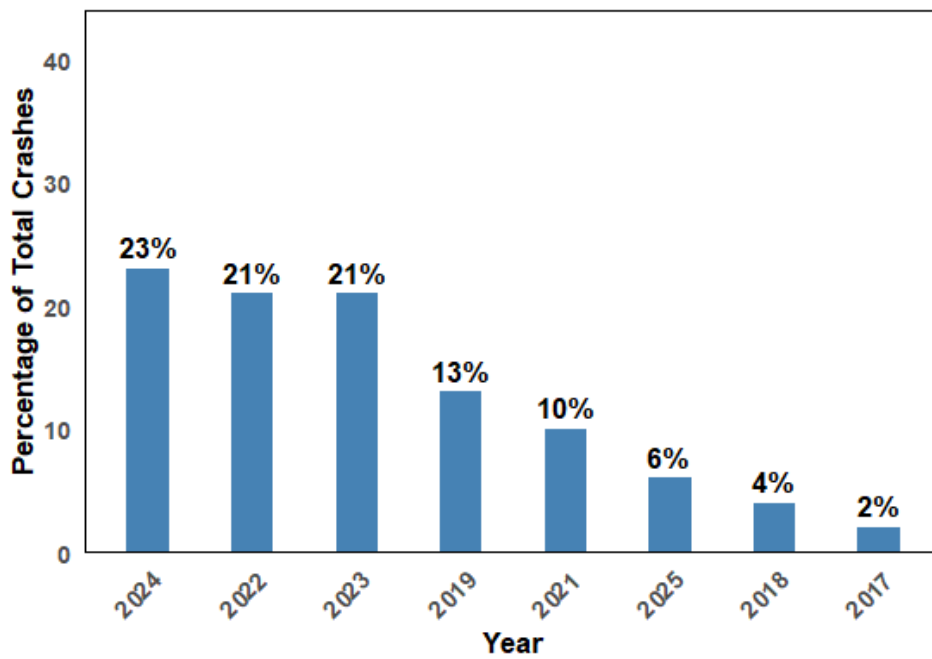
Distribution of AV-Cyclist Crashes by Cyclist Traffic Violation



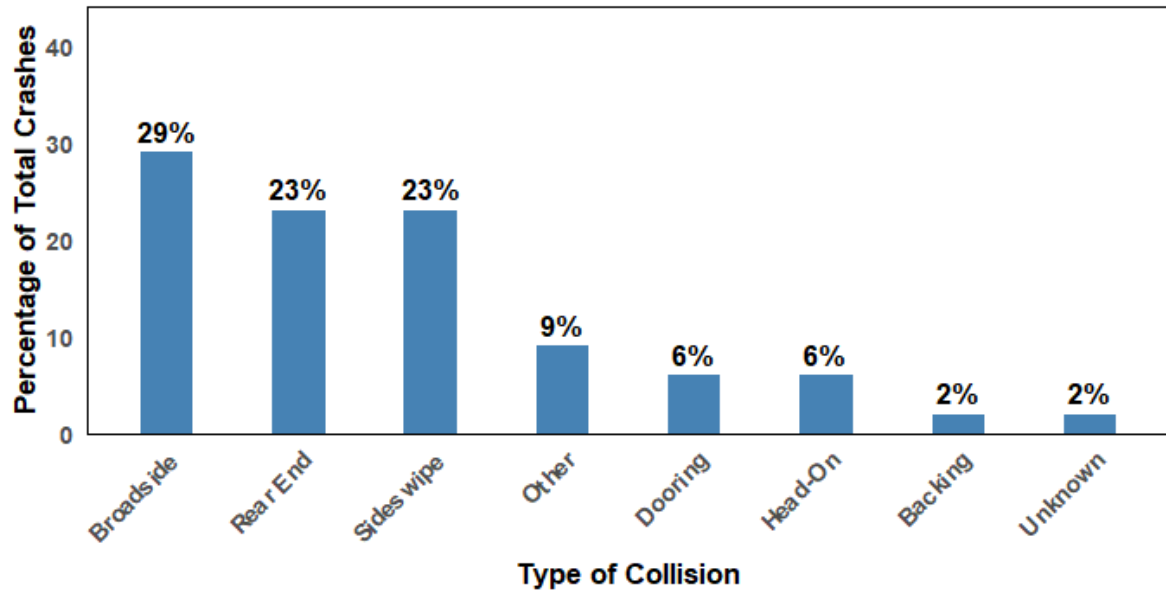
Distribution of AV-Cyclist Crashes by Time of Crash



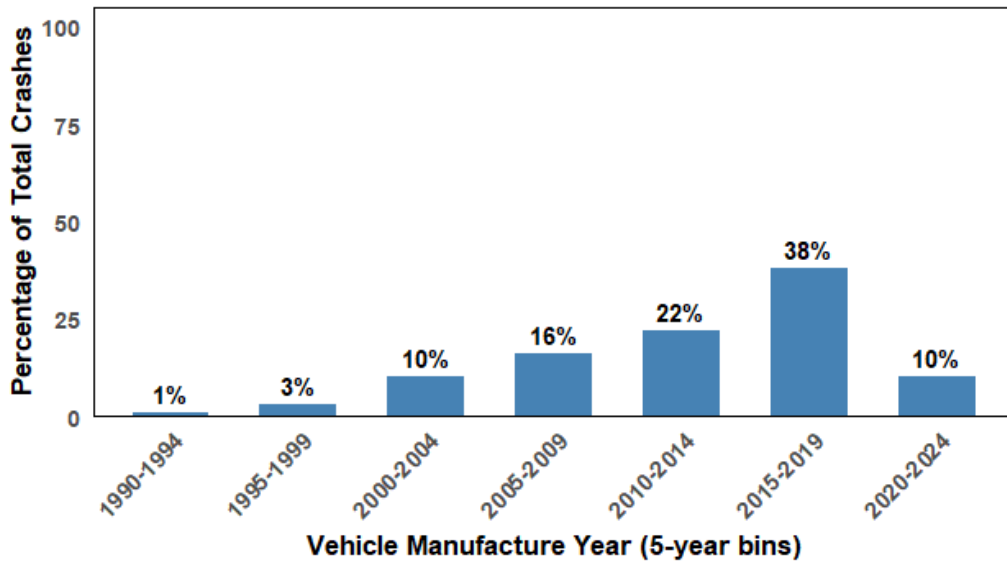
Distribution of AV-to-Cyclist Crashes across Years



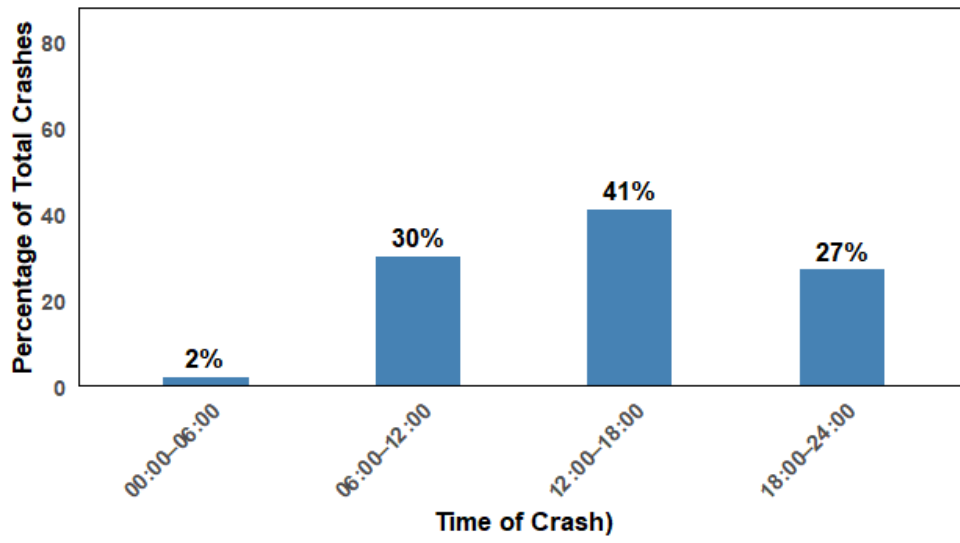
Distribution of AV-to-Cyclist Crashes by Type of Collision



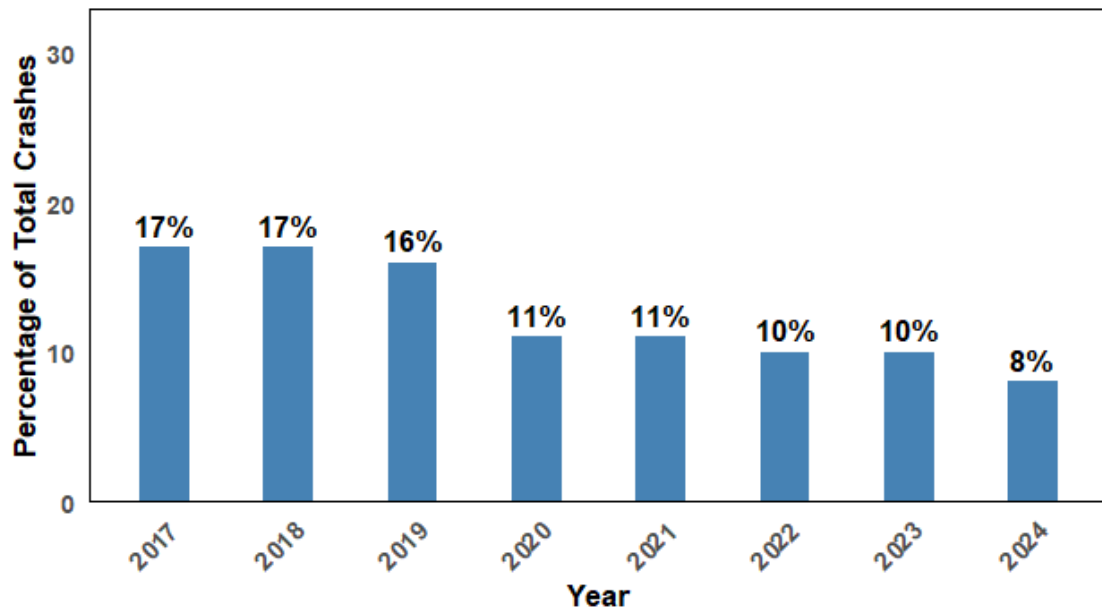
Distribution of CV-Cyclist Crashes by 5-Year Vehicle Manufact



Distribution of CV–Cyclist Crashes by Time of Crash



Distribution of CV-to-Cyclist Crashes across Years



Declaration concerning the Master's Thesis

I hereby confirm that the presented thesis work has been done independently and using only the listed sources and resources. This thesis has not been submitted elsewhere for purposes of assessment.

Munich, June 10, 2025

A handwritten signature in blue ink, consisting of several overlapping loops and strokes, positioned above a horizontal line.

Asca Anne Amondi

List of References

References

- Abdel-Aty, M., & Ding, S. (2024). A matched case-control analysis of autonomous vs human-driven vehicle accidents. *Nature Communications*, *15*(1), 4931. <https://doi.org/10.1038/s41467-024-48526-4>
- Aldred, R., & Goodman, A. (2018). Predictors of the frequency and subjective experience of cycling near misses: Findings from the first two years of the UK Near Miss Project. *Accident; Analysis and Prevention*, *110*, 161–170. <https://doi.org/10.1016/j.aap.2017.09.015>
- Almaskati, D., Kermanshachi, S., & Pamidimukkala, A. (2024). Investigating the impacts of autonomous vehicles on crash severity and traffic safety. *Frontiers in Built Environment*, *10*, Article 1383144. <https://doi.org/10.3389/fbuil.2024.1383144>
- Automated Vehicles for Safety*. NHTSA. (n.d.). Retrieved March 2025, from <https://www.nhtsa.gov/vehicle-safety/automated-vehicles-safety>
- Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*(1), 40–49. <https://doi.org/10.1002/mpr.329>
- Boggs, A. M., Wali, B., & Khattak, A. J. (2020). Exploratory analysis of automated vehicle crashes in California: A text analytics & hierarchical Bayesian heterogeneity-based approach. *Accident; Analysis and Prevention*, *135*, 105354. <https://doi.org/10.1016/j.aap.2019.105354>
- Buehler, R., & Pucher, J. (2016). Trends in Walking and Cycling Safety: Recent Evidence from High-Income Countries, With a Focus on the United States and Germany.
- Channamallu, S. S., Almaskati, D., Kermanshachi, S., & Pamidimukkala, A. (2024). Predictive analytics in autonomous vehicles safety: crash outcome modeling. *Urban, Planning and Transport Research*, *12*(1), Article 2409676. <https://doi.org/10.1080/21650020.2024.2409676>
- Char, F., & Serre, T. (2020). Analysis of pre-crash characteristics of passenger car to cyclist accidents for the development of advanced driver assistance systems. *Accident; Analysis and Prevention*, *136*, 105408. <https://doi.org/10.1016/j.aap.2019.105408>
- Chimba, D., Emaasit, D., & Kutela, B. (2012). Likelihood Parameterization of Bicycle Crash Injury Severities. *Journal of Transportation Technologies*, *02*(03), 213–219. <https://doi.org/10.4236/jtts.2012.23023>
- Cubbin, W., van Paridon, K., Keyes, H., & Timmis, M. (2024). Close passes caught on camera – How knowledge and behavioural norms relate to perceptions of liability when cars overtake cyclists. *Transportation Research Part F: Traffic Psychology and Behaviour*, *100*, 308–322. <https://doi.org/10.1016/j.trf.2023.12.002>
- Das, S. (2021). Autonomous vehicle safety: Understanding perceptions of pedestrians and bicyclists. *Transportation Research Part F: Traffic Psychology and Behaviour*, *81*, 41–54. <https://doi.org/10.1016/j.trf.2021.04.018>
- Dash, I., Abkowitz, M., & Philip, C. (2022). Factors impacting bike crash severity in urban areas. *Journal of Safety Research*, *83*, 128–138. <https://doi.org/10.1016/j.jsr.2022.08.010>
- Eluru, N., Bhat, C. R., & Hensher, D. A. (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. <https://doi.org/10.1016/j.aap.2007.11.010>
- Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part a: Policy and Practice*, *77*, 167–181. <https://doi.org/10.1016/j.tra.2015.04.003>

- Fernández, P. D., Lindman, M [M.], Isaksson-Hellman, I [I.], Jeppsson, H [H.], & Kovaceva, J [J.] (2022). Description of same-direction car-to-bicycle crash scenarios using real-world data from Sweden, Germany, and a global crash database. *Accident Analysis & Prevention*, *168*, 106587. <https://doi.org/10.1016/j.aap.2022.106587>
- Filipović, F., Mladenović, D., Lipovac, K., Das, D. K., & Todosijević, B. (2022). Determining Risk Factors that Influence Cycling Crash Severity, for the Purpose of Setting Sustainable Cycling Mobility. *Sustainability*, *14*(20), 13091. <https://doi.org/10.3390/su142013091>
- Fu, H., Ye, S., Fu, X., Chen, T., & Zhao, J. (2025). New insights into factors affecting the severity of autonomous vehicle crashes from two sources of AV incident records. *Travel Behaviour and Society*, *38*, 100934. <https://doi.org/10.1016/j.tbs.2024.100934>
- Ghoul, T., & Sayed, T. (2025). Cyclist safety assessment using autonomous vehicles. *Accident; Analysis and Prevention*, *212*, 107923. <https://doi.org/10.1016/j.aap.2025.107923>
- Hamann, C. J., Peek-Asa, C., Lynch, C. F., Ramirez, M., & Hanley, P. (2015). Epidemiology and spatial examination of bicycle-motor vehicle crashes in Iowa, 2001–2011. *Journal of Transport & Health*, *2*(2), 178–188. <https://doi.org/10.1016/j.jth.2014.08.006>
- Hansen, K. S., Engesaeter, L. B., & Viste, A. (2003). Protective effect of different types of bicycle helmets. *Traffic Injury Prevention*, *4*(4), 285–290. <https://doi.org/10.1080/714040486>
- Isaksson-Hellman, I [Irene], & Töreki, J. (2019). The effect of speed limit reductions in urban areas on cyclists' injuries in collisions with cars. *Traffic Injury Prevention*, *20*(sup3), 39–44. <https://doi.org/10.1080/15389588.2019.1680836>
- Johannsen, H., Otte, D [Dietmar], & Urban, M. (2015). Pre-Crash Analysis of accidents Involving Turning Trucks and Bicyclists.
- Kaplan, S., Vavatsoulas, K., & Prato, C. G. (2014). Aggravating and mitigating factors associated with cyclist injury severity in Denmark. *Journal of Safety Research*, *50*, 75–82. <https://doi.org/10.1016/j.jsr.2014.03.012>
- Klug, C., Leo, C., Schachner, M., Grumert, E., Rizzi, M., Lindner, A., Heinzl, P., Davidse, R., Iancovici, S., Wagstom, L., & Lindman, M [M.] (2020). Current and future accident and impact scenarios for pedestrians and cyclists.
- Kockelman, K., Loftus-Otway, L., Stewart, D., Nichols, A., Wagner, W., Li, J., Boyles, S., Levin, M., & Liu, J. (2016). Best Practices Guidebook for Preparing Texas for Connected and Automated Vehicles (0-6849-P1).
- Kohanpour, E., Davoodi, S. R., & Shaaban, K. (2024). Analyzing Autonomous Vehicle Collision Types to Support Sustainable Transportation Systems: A Machine Learning and Association Rules Approach. *Sustainability*, *16*(22), 9893. <https://doi.org/10.3390/su16229893>
- Kuehn, M., Hummel, T., & Lang, A. (2015). Cyclist-Car Accidents: Their Consequences for Cyclists and Typical Accident Scenarios (Paper Number 15-0243). German Insurers Accident Research.
- Kuo, P.-F., Hsu, W.-T., Lord, D., & Putra, I. G. B. (2024). Classification of autonomous vehicle crash severity: Solving the problems of imbalanced datasets and small sample size. *Accident; Analysis and Prevention*, *205*, 107666. <https://doi.org/10.1016/j.aap.2024.107666>
- Kutela, B., Das, S., & Dadashova, B. (2022). Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors. *Accident; Analysis and Prevention*, *165*, 106473. <https://doi.org/10.1016/j.aap.2021.106473>
- Lahrman, H., Madsen, T. K. O., Olesen, A. V., Madsen, J. C. O., & Hels, T. (2018). The effect of a yellow bicycle jacket on cyclist accidents. *Safety Science*, *108*, 209–217. <https://doi.org/10.1016/j.ssci.2017.08.001>
- League of American Bicyclists (2018). *Bicycling and Walking in the United States: 2018 Benchmarking Report, Sixth Edition*.
- Lindman, M [M.], Jonsson, S., Jakobsson, L., Karlsson, T., Gustafson, D., & Fredriksson, A. (2015). Cyclists interacting with passenger cars; a study of real world crashes.

- Liu, Q., Wang, X., Liu, S., Yu, C., & Glaser, Y. (2024). Analysis of pre-crash scenarios and contributing factors for autonomous vehicle crashes at intersections. *Accident; Analysis and Prevention*, 195, 107383. <https://doi.org/10.1016/j.aap.2023.107383>
- Lopez, D. S., Sunjaya, D. B., Chan, S., Dobbins, S., & Dicker, R. A. (2012). Using trauma center data to identify missed bicycle injuries and their associated costs. *The Journal of Trauma and Acute Care Surgery*, 73(6), 1602–1606. <https://doi.org/10.1097/TA.0b013e318265fc04>
- MacAlister, A., & Zubby, S. D. (2015). Cyclist Crash Scenarios and Factors Relevant to the Design of Cyclist Detection Systems.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, in *Frontiers in Econometrics*.
- McGuckin, N., & Fucci A. (2018). Summary of Travel Trends: 2017 National Household Travel Survey. U.S. Department of Transportation and Federal Highway Administration.
- Moore, D. N., Schneider, W. H., Savolainen, P. T., & Farzaneh, M. (2011). Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. *Accident; Analysis and Prevention*, 43(3), 621–630. <https://doi.org/10.1016/j.aap.2010.09.015>
- Najm G., W., Smith D., J., & Yanagisawa, M. (2007). Pre-Crash Scenario Typology for Crash Avoidance Research.
- National Center for Statistics and Analysis (2023). Traffic safety facts 2021: A compilation of motor vehicle traffic crash data (Report No. DOT HS 813 527). National Highway Traffic Safety Administration.
- National Center for Statistics and Analysis (2024). Bicyclists and other cyclists: 2022 data (Traffic Safety Facts. Report No. DOT HS 813 591). National Highway Traffic Safety Administration. NHTSA. *Standing General Order for incidents involving ADS and Level 2 ADAS*. <https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting#data>
- Niemann, M., Braun, K. F., Otto, E., Tiefenbrunner, M., Wüster, J., Stöckle, U., Ahmad, S. S., Märdian, S., & Graef, F. (2023). Dangers of e-mobility: A systematic review and meta-analysis of sustained injury patterns and injury severity. *Safety Science*, 167, 106283. <https://doi.org/10.1016/j.ssci.2023.106283>
- Nobis, C. (2019). Mobility in Germany -MiD analyses of cycling and walking. BMVI, infas, DLR, IVT, infas 360. Bonn, Berlin.
- Puller, N., Lucas, G., Leschke, A., Maier, O., Mönnich, J., & Rocco, V. (2023). The "Typical" Car-Cyclist Collision under the Microscope: A GIDAS-Based Analysis of the Prevalent Crash Scenario.
- Qawasmeh, B., Oh, J.-S., & Kwigizile, V. (2024). Micro-Mobility Safety Assessment: Analyzing Factors Influencing the Micro-Mobility Injuries in Michigan by Mining Crash Reports. *Future Transportation*, 4(4), 1580–1601. <https://doi.org/10.3390/futuretransp4040076>
- Rash-ha Wahi, R., Haworth, N [Narelle], Kumar Debnath, A., & King, M. (2018). Influence of Type of Traffic Control on Injury Severity in Bicycle–Motor Vehicle Crashes at Intersections. *Transportation Research Record*, 2672(38), 199–209. <https://doi.org/10.1177/0361198118773576>
- Reish, L. (2021). Comparing demographic trends in vulnerable road user fatalities and the U.S. population, 1980–2019 (Research Note. Report No. DOT HS 813 178). National Highway Traffic Safety Administration.
- Ren, W., Yu, B., Chen, Y., Gao, K., & Bao, S. (2022). Divergent Effects of Factors on Crashes under Autonomous and Conventional Driving Modes Using a Hierarchical Bayesian Approach.
- Reynolds, C. C. O., Harris, M. A., Teschke, K., Cripton, P. A., & Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. *Environmental Health*, 8(1), 47. <https://doi.org/10.1186/1476-069X-8-47>

- Robartes, E., & Chen, T. D. (2017). The effect of crash characteristics on cyclist injuries: An analysis of Virginia automobile-bicycle crash data. *Accident Analysis & Prevention, 104*, 165–173. <https://doi.org/10.1016/j.aap.2017.04.020>
- Scarano, A., Rella Riccardi, M., Mauriello, F., D'Agostino, C., Pasquino, N., & Montella, A. (2023). Injury severity prediction of cyclist crashes using random forests and random parameters logit models. *Accident; Analysis and Prevention, 192*, 107275. <https://doi.org/10.1016/j.aap.2023.107275>
- Schindler, R., & Jeppsson, H [Hanna] (2024). In-depth analysis of scenarios and injuries in crashes between cyclists and commercial vehicles in Germany. *Traffic Safety Research, 7*, e000067. <https://doi.org/10.55329/uoqc5084>
- Shinar, D., Valero-Mora, P., van Strijp-Houtenbos, M., Haworth, N [N.], Schramm, A., Bruyne, G. de, Cavallo, V., Chliaoutakis, J., Dias, J., Ferraro, O. E., Fyhri, A., Sajatovic, A. H., Kuklane, K., Ledesma, R., Mascarell, O., Morandi, A., Muser, M., Otte, D [D.], Papadakaki, M., . . . Tzamalouka, G. (2018). Under-reporting bicycle accidents to police in the COST TU1101 international survey: Cross-country comparisons and associated factors. *Accident; Analysis and Prevention, 110*, 177–186. <https://doi.org/10.1016/j.aap.2017.09.018>
- Shiran, G., Imaninasab, R., & Khayamim, R. (2021). Crash Severity Analysis of Highways Based on Multinomial Logistic Regression Model, Decision Tree Techniques, and Artificial Neural Network: A Modeling Comparison. *Sustainability, 13*(10), 5670. <https://doi.org/10.3390/su13105670>
- Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics, 8*(2), 39–42. <https://doi.org/10.12691/ajams-8-2-1>
- Singh, S. (2018). Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. (Traffic Safety Facts Crash•Stats. Report No. DOT HS 812 506). Washington, DC: National Highway Traffic Safety Administration.
- Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2015). Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. *Accident Analysis & Prevention, 83*, 132–142. <https://doi.org/10.1016/j.aap.2015.07.014>
- Tafidis, P., Pirdavani, A., Brijs, T., & Farah, H. (2019). Can Automated Vehicles Improve Cyclist Safety in Urban Areas? *Safety, 5*(3), 57. <https://doi.org/10.3390/safety5030057>
- Thompson, D. C., Nunn, M. E., Thompson, R. S., & Rivara, F. P. (1996). Effectiveness of bicycle safety helmets in preventing serious facial injury. *JAMA, 276*(24), 1974–1975.
- Train, K. (2009). *Discrete Choice Methods with Simulation*.
- U.S. Census Bureau (2014). *Modes Less Traveled—Bicycling and Walking to Work in the United States: 2008–2012*.
- U.S. Census Bureau (2023a). *Age and Sex Composition: 2020 Census Briefs*.
- U.S. Census Bureau. (2023b). *American Community Survey: DP03 | Selected Economic Characteristics*. <https://data.census.gov>
- Wang, C., Lu, L., & Lu, J. (2015). Statistical Analysis of Bicyclists' Injury Severity at Unsignalized Intersections. *Traffic Injury Prevention, 16*(5), 507–512. <https://doi.org/10.1080/15389588.2014.969802>
- WHO. (2019). *Global Status Report on Road Safety 2018*. World Health Organization. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5910092>
- WHO. (2020). *Cyclist Safety: An Information Resource for Decision-Makers and Practitioners* (1st ed.). World Health Organization. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=30478506>
- Wisch, M., Lerner, M., Kovaceva, J [Jordanka], Bálint, A., Gohl, I., Schneider, A., Juhasz, J., & Lindman, M [Magdalena] (2019). CAR-TO-CYCLIST CRASHES IN EUROPE AND DERIVATION OF

USE CASES AS BASIS FOR TEST SCENARIOS OF NEXT GENERATION ADVANCED DRIVER ASSISTANCE SYSTEMS-RESULTS FROM PROSPECT. In

- Wood, J., Tyrrell, R., Marszalek, R., Lacherez, P., Carberry, T., & Chu, B. (2012). Using reflective clothing to enhance the conspicuity of bicyclists at night. *Accident; Analysis and Prevention*, 45, 726–730. <https://doi.org/10.1016/j.aap.2011.09.038>
- Yan, X., Ma, M., Huang, H., Abdel-Aty, M., & Wu, C. (2011). Motor vehicle-bicycle crashes in Beijing: Irregular maneuvers, crash patterns, and injury severity. *Accident; Analysis and Prevention*, 43(5), 1751–1758. <https://doi.org/10.1016/j.aap.2011.04.006>
- Zhou, Y., Aryal, S., & Bouadjenek, M. R. (2024, April 7). *A Comprehensive Review of Handling Missing Data: Exploring Special Missing Mechanisms*. <http://arxiv.org/pdf/2404.04905v1>
- Zhuo Wang, Richard L. Neitzel, Wenlong Zheng, Dezheng Wang, Xiaodan Xue, & Guohong Jiang (2021). Road safety situation of electric bike riders: A cross-sectional study in courier and take-out food delivery population. *Traffic Injury Prevention*, 22(7), 564–569. <https://doi.org/10.1080/15389588.2021.1895129>