Electric Vehicles vs. Internal Combustion Engine Vehicles: A Comparative Study of Non-Motorist Crash Injury Severity

Jiahe Ling
School of Computer, Data & Information Sciences (CDIS)
University of Wisconsin Madison, Madison, WI, 53706
Tel: 608-556-4287; Email: jling9@wisc.edu

Xiaodong Qian, Ph.D. (Corresponding Author)
College of Engineering
Wayne State University, Detroit, MI 48202
Tel: 313-577-3854; Email: xdqian@wayne.edu

Konstantina Gkritza, Ph.D.
Lyles School of Civil Engineering
Purdue University, West Lafayette, IN, 47907
Tel: 765-494-4597; Email: nadia@purdue.edu

Word Count: 5459 words + 5 table (250 words per table) = 6,709 words

Submitted on July 28, 2023

ABSTRACT

Powered by electric engines, electric vehicles (EVs) exhibit unique dynamic characteristics that may lead to different crash characteristics and outcomes compared with traditional internal combustion engine vehicles (ICEVs). This paper focuses on non-motorist crashes and estimates crash characteristics and severity outcomes using statistical testing and regression analyses based on Chicago crash data from 2015 to 2022. Innovatively, this study supplements traditional police crash reports with Google Street View (GSV) images and employs computer vision neural network models to uncover previously unreported environmental variables at crash scenes. The results reveal both similarities and disparities in non-motorist crash characteristics between EV-involved and ICEV-involved incidents. The Likelihood Ratio Test suggests parameter transferability in injury severity models for both vehicle types. However, notable distinctions in factor distributions, such as non-motorist type, hit-and-run incidents, damage level, crash hour, crash weekday, weather conditions, and road surface conditions, along with the influence of season and road surface condition on injury severity, exist between EV and ICEV crashes. These distinctions may be attributed to driver demographics, vehicle design, and usage characteristics. These insights can guide the development of safety regulations for EVs and aid in devising specific safety measures and policies for non-motorists, including pedestrians and cyclists.

Keywords: *Non-motorist Crash, Electric Vehicles, Google Street View, Binary Probit Regression*

INTRODUCTION

Electric vehicles (EVs) refer to automobiles that utilize electricity as their primary or partial source of power. They can be classified into four main categories: battery electric vehicles (BEVs), plug-in hybrid electric vehicles (PHEVs), hybrid electric vehicles (HEVs), and fuel cell electric vehicles (FCEVs). Recent reductions in manufacturing costs and the progressive deployment of charging stations have significantly bolstered the appeal of EVs. Particularly, BEV registrations increased by 110% and PHEV registrations rose by 75% between 2016 and 2020 (*1*, *2*). This surge in EV numbers highlights the importance of addressing safety concerns, particularly pertaining to non-motorist safety. EVs display unique dynamic features, such as operating almost silently without the traditional engine noise, which calls for a reassessment of conventional safety measures.

The existing body of literature on non-motorist crashes involving EVs tends to focus predominantly on pedestrian incidents involving HEVs. However, there is a significant research gap when it comes to investigating non-motorist crashes involving BEVs and PHEVs. The distinct powertrain technology, weight distribution, and overall vehicle design inherent to these various types of EVs emphasize the importance of addressing this research gap. Given the extensive research available on HEVs due to their early introduction (*3*), and the limited crash records pertaining to FCEVs, this study primarily focuses on the analysis of BEVs and PHEVs (referred to as EVs later in this paper). In addition, the objective of this study is to employ an integrated approach that combines Google Street View (GSV), transformer models for image segmentation, and police reports to develop a comprehensive data description for non-motorist crashes involving EVs. This integrated approach contributes to the analysis of non-motorist crashes by providing comprehensive data integration, enhanced contextual understanding, precise object identification, and evidence-based insights for addressing the unique safety challenges associated with EVs. Subsequently, a binary probit model is estimated to identify the factors that influence EV-related non-motorist crashes.

The paper is structured as follows. Firstly, a comprehensive literature review on EV-related crash studies and use of GSV for crash analysis is presented. Subsequently, the data sources and statistical models employed in this study are described. The ensuing section presents the outcomes and significant findings derived from the statistical hypothesis testing and regression analyses. Finally, the paper concludes by discussing the research implications, acknowledging the study's limitations, and proposing potential avenues for future research.

LITERATURE REVIEW

With the increasing number of EVs on the road, some studies have shed light on the statistical analysis of EV crashes. Liu et al. (*4*) applied Pearson's chi-squared test to confirm that the distribution of severity levels for EV (mixing of BEV and PHEV) crashes is different from that of ICEV crashes. They used the logistic regression model to identify essential factors influencing crash severity. Based on their results, for EVs' crash data, the presence of medians has a negative effect on crash severity, and collisions with motorcycles have a positive effect on crash severity. However, this study only focused on environmental variables without taking human-related and vehicle-related variables into consideration. Moreover, several published government reports compared HEVs' crash data with ICEVs' crash data. Chen et al. directly compared the crash statistics between HEVs and ICEVs and noted that occupants of HEVs tended to be older than occupants of ICEVs, fire incidents were not common in both HEV and ICEVs, and occupants of HEVs were more likely to experience arm, wrist, and hand injuries but less likely to experience leg, ankle, and foot injuries when being compared with that of ICEVs (*5*). However, the study did not account for roadway and environmental factors into consideration and only descriptive analysis other than statistical models or testing is employed.

The former studies did not take heterogeneity into consideration, which may introduce biased estimation and inferences. Taking heterogeneity and heteroskedasticity into consideration, Huang et al. evaluated HEV crashes' severity through a hierarchical mixed logit model and concluded that higher occupant vehicles and older occupants were associated with higher injury counts, but crashes happen on the wet road surface and regional artery roads (not expressway) result in fewer injury counts (*6*). Also, Huang et al. pointed out that the statistical analysis results could support strong heterogeneity effects in crash data (*6*). Based on this conclusion, Seraneeprakarn et al. further validated the influence of unobserved heterogeneity by comparing estimation from the mixed logit model, mixed logit model with heterogeneity in means, and mixed logit models with heterogeneity in means and variance (*7*). These studies identify the importance of taking heterogeneity effects into consideration when analyzing EVs' crash data. Still, these studies only analyzed crash involving HEVs instead of other EV types.

In addition to studies that use statistical modeling methods to analyze real crash data, Karaaslan et al. used agent-based modeling to conduct traffic simulation and showed that EVs have a greater potential of posing a threat to pedestrians than ICEVs by performing sensitivity analysis on the simulated crash data (*8*). Furthermore, Karaaslan et al. confirmed the simulation results by analyzing the crash data from the Fatal Analysis Reporting System (FARS) through a chi-squared test (*8*). This study further confirms the idea proposed in earlier studies that EVs have a higher possibility to hit non-motorists than ICEVs.

Other than focusing on driver injuries in EV-involved crashes, some studies focused on vulnerable road users (i.e., pedestrians or cyclists). Hanna studied pedestrian or cyclist crashes with HEVs and ICEVs (*3*). Based on hypothesis testing results, Hanna concluded that motor vehicle crashes involving pedestrians and cyclists usually happened on roads with low-speed limits under good lighting and weather conditions (*3*). More importantly, HEVs were more likely to collide with pedestrians and cyclists compared with ICEVs. Focusing on speed limits, vehicle actions, and crash locations, Wu et al. further verified this conclusion through statistical methods including a case-control approach, relative risk, and odds ratio (*9*). These studies suggest collision counterparts like pedestrians and bicycles are worth taking into consideration when analyzing EV crash data. However, a limited number of variables are explored in these studies. Such limitation makes it hard to identify potential differences and effects of various factors in crash data between EVs and ICEVs.

Besides, various types of data are used in vehicle crash analysis including police records, Event Data Recorder (EDR) data, and images of crash locations. Google Street View (GSV) has emerged as a valuable tool for various research domains, including transportation crash analysis and public health studies. Initially, it was primarily utilized for case study analysis, aiding in the interpretation of statistical models. For instance, Hanson, Noland, & Brown employed GSV to present case studies of crashes, which helped enhance the understanding of their statistical model results (*10*).

As the application of GSV expanded, researchers began utilizing trained auditors to extract features from the images, thereby enhancing data collection. Mooney et al., for example, employed five trained virtual street auditors, who collected data from the CANVAS system (*11*). They focused on factors such as crosswalk presence, billboards, road or sidewalk condition, bus stops, and pedestrian signals. Furthermore, with advancements in computer vision and neural networks, pre-trained Deep Convolutional Neural Networks (DCNNs) have facilitated the recognition of different spatial categories and scene types within Street View imagery. Kang et al. demonstrated the use of DCNNs for this purpose (*12*). For instance, Stiles, Li, & Miller Pyramid employed a Scene Parsing Network (PSPNet) to segment Street View imagery into up to 150 distinct object categories, enabling the identification of visual objects (*13*). It should be noted, however, that while pre-trained neural networks can extract certain elements from GSV images, some aspects, such as detailed road characteristics, may still require manual examination by researchers. This includes assessing factors like the presence of sidewalks, bike lanes, and other road conditions, as well as environmental conditions like blocked views or the presence of traffic signs. Still, GSV has proven instrumental in research for extracting valuable information that may not be included in traditional crash reports.

DATA DESCRIPTION AND PRELIMINARY ANALYSIS

This study utilizes crash data, which is based on police report, obtained from the electronic crash reporting system (E-Crash) maintained by the Chicago Police Department (CPD) (*14*). The dataset covers traffic crashes occurring on city streets within the jurisdiction of the City of Chicago, spanning the years 2015 to 2022. To ensure privacy, personally identifiable information like Vehicle Identification Number (VIN) is not included in the given data. Moreover, the GSV API is utilized to obtain panoramic 360-

degree views, which consists of 4 pictures, each covering a 90-degree field of vision, and a 90-degree pitch (horizontal visual angle) at each EV crash location, based on the given latitude and longitude coordinates of crashes.

Street view images provide human-like perspectives of urban streetscapes. Google Street View (GSV), launched in 2007, is the pioneering and leading provider in this domain (*12*). This study collected 360-degree GSV images from 2011 to 2023 with given latitude and longitude information of crash location. The primary distribution of image years is as follows: 41.67% from 2021, 19.05% from 2022, 16.67% from 2019, and 10.71% from 2018, 3.57% from 2017, 2.38% from 2016, and 1.19% from 2011, 2013, 2015, 2020, and 2023.

This study also incorporates the use of Segformer, a Transformer model designed for semantic segmentation of street view imagery, to aid the visual examination (*17*). Segformer consists of a hierarchical Transformer encoder and a lightweight all-MLP decode head. Pre-trained on CityScapes data, the model takes the input image and generates names of visual objects in the form of a list (*20*).

The distribution of crash locations of all victim types (including both motorist and non-motorists) for different types of vehicles are demonstrated in **Figure 1**.

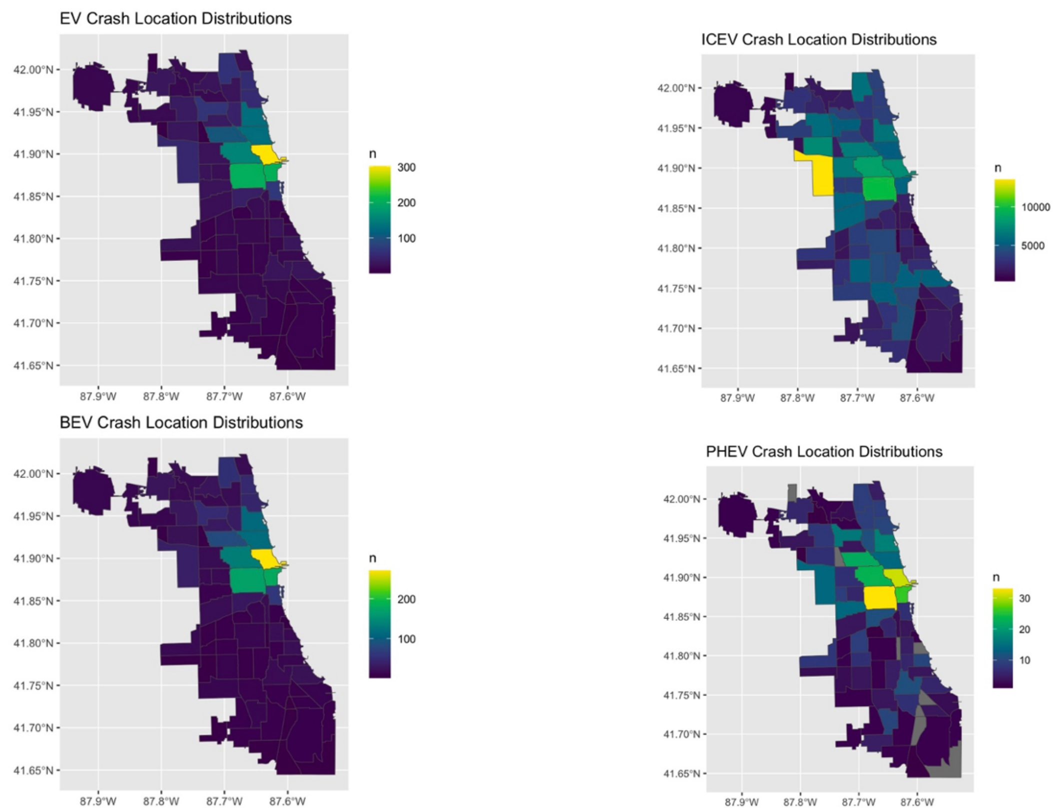

Figure 1 Crash Distribution for Difference Types of Vehicles

Police Report Data Processing

Due to the absence of VINs or other variables that explicitly identify EVs or ICEVs, the fuel type classification for each vehicle is determined based on the vehicle's make (brand), model, and year. This determination is made by referencing fuel economy information provided by the U.S. Department of Energy *(15)*.

Given the disparity in vehicle years and safety technologies between EVs and ICEVs in this study, a rigorous filtering process is implemented to ensure a fair comparison. Only ICEVs from model years 2010 to 2022 are considered to match the timeframe of the EVs under investigation.

Moreover, to address the critical aspects of accounting for geographical and temporal variations in vehicle crash analyses, established recommendations are followed, including the adoption of buffer analysis to control for spatial and temporal heterogeneity *(16)*. For this study, three types of buffers are employed to select relevant ICEV crashes that occurred in proximity to EV crashes. A geological buffer of 50 meters, a seasonal buffer, and a time-period buffer are utilized. The time-period buffer involves dividing each day into six distinct periods: 3:00-6:59, 7:00-10:59, 11:00-14:59, 15:00-18:59, 19:00-22:59, and 23:00-2:59. Only ICEV crashes that transpired within the same specific time-period, season, and geographical location as the corresponding EV crash are considered for further analysis.

In the end, there are 58 non-motorist crashes for BEVs, 17 non-motorist crashes for PHEVs, and 358 non-motorist crashes for ICEVs (327 from BEV crash buffer, 31 from PHEV crash buffer). The statistical summary of data is presented in **Table 1**.

TABLE 1 Descriptive Statistics of Selected Variables

	Electric Vehicles (EVs)		**Internal Combustion Engine Vehicles (ICEVs)**	
	Number	**Percentage %**	**Number**	**Percentage**
Injury Severity				
No Injury	21	28.0%	87	24.30%
Lightly Injured	48	64.0%	234	65.36%
Severely Injured/Fatal	6	8.0%	37	10.34%
Non-motorist Type				
Bike	42	56.0%	122	34.08%
Pedestrian	33	44.0%	236	65.92%
Gender				
Male	40	53.33%	206	57.54%
Female	35	46.67%	152	42.46%
Safety Equipment				
True	14	18.67%	47	13.13%
False	61	81.33%	311	86.87%
Location				
Bike Lane	14	18.67%	33	9.22%
Driveway Access	2	2.67%	10	2.79%
Crosswalk	On in 24	32.0%	113	31.56%
Roadway	29	38.67%	129	36.03%
Shoulder	0	0.0%	1	0.28%
Other	6	8.0%	72	20.11%
Driver Physical Condition				
Drug/Alcohol	1	1.33%	2	0.56%
Emotional	2	2.67%	0	0.0%
Normal	68	90.67%	338	94.41%
Other	4	5.33%	18	5.03%
Driver Vision				
Not Obscured	67	89.33%	321	89.66%
Obscured	8	10.67%	37	10.33%

Speed Limit				
Under 35 mph	75	100.0%	341	95.25%
35–50 mph	0	0.0%	17	4.75%
Above 50 mph	0	0.0%	0	0.0%
Traffic Control				
Present & Function	37	49.33%	174	48.60%
Not Present/Not Function	38	50.67%	184	51.40%
Intersection				
True	28	37.33%	132	36.87%
False	47	62.67%	226	63.12%
Not Right of Way				
True	1	1.33%	27	7.54%
False	74	98.67%	331	92.45%
Hit and Run				
True	16	21.33%	128	35.75%
False	59	78.67%	230	64.25%
Person Ejected				
True	4	5.33%	13	3.63%
False	71	94.67	345	96.37%
Damage				
$500 or less	30	40.0%	202	56.42%
$501 - $1,500	15	20.0%	51	14.25%
over $1,500	30	40.0%	105	29.33%
Hour				
3:00-6:59	0	0.0%	1	0.28%
7:00-10:59	18	24.0%	38	10.61%
11:00-14:59	15	20.0%	86	24.02%
15:00-18:59	28	37.33%	173	48.32%
19:00-22:59	13	17.33%	50	13.97%
23:00-2:59	1	1.33%	10	2.79%
Day of Week				
Monday	7	9.33%	29	8.10%
Tuesday	7	9.33%	55	15.36%
Wednesday	11	14.67%	57	15.92%
Thursday	13	17.33%	42	11.73%
Friday	15	20.0%	60	16.76%
Saturday	11	14.67%	80	22.35%
Sunday	11	14.67%	35	9.78%
Season				
Fall (9, 10, 11)	31	41.33%	129	36.03%
Spring (3, 4, 5)	10	13.33%	49	13.69%
Summer (6, 7, 8)	23	30.67%	108	30.17%
Winter (12, 1, 2)	11	14.67%	72	20.11%
Weather				
Clear	67	89.33%	293	81.84%
Cloudy	2	2.67%	15	4.19%
Rainy	6	8.0%	38	10.61%
Snowy	0	0.0%	12	3.35%

Surface Condition				
Dry	66	88.0%	292	81.56%
Icy	0	0.0%	16	4.47%
Wet	9	12.0%	50	13.97%

Table 1 contains the Safety Equipment variable, which denotes the presence of both regular helmets and bicycle helmets, Traffic Control variable, which indicates the presence and proper functionality of traffic signs, traffic signals, and police at the crash location, and Season variable, which is categorized into Spring (March, April, May), Summer (June, July, August), Fall (September, October, November), and Winter (December, January, February).

Table 1 shows that the majority of categorical variables for EV-related non-motorist crashes closely resemble those of ICEV-related non-motorist crashes. Nonetheless, there are notable differences in variables such as non-motorist type, hit-and-run incidents, damage level, crash hour, crash weekday, weather conditions, and road surface conditions. The Chi-square tests confirm that all these differences are statistically significant.

Regarding non-motorist type, it appears that EVs are more likely to be involved in crashes with cyclists, while ICEVs are more frequently associated with pedestrian crashes. Additionally, non-motorist crashes involving EVs exhibit a lower proportion of hit-and-run incidents. This observation aligns with findings from a survey conducted by Jensen & Marbit, which indicated that households owning EVs tend to have higher education and income levels, potentially leading to a reduced likelihood of hit-and-run behavior (*17*). Furthermore, non-motorist crashes involving EVs result in higher financial damage levels, as supported by a study conducted by Mersky et al., highlighting the generally higher cost of EVs compared to traditional vehicles (*19*).

Notably, non-motorist crashes involving EVs are more prevalent during rush hours (7:00-10:59) and on workdays compared to those involving ICEVs. Jensen & Marbit's study provides an explanation for this trend, suggesting that EVs are more commonly used for commuting purposes due to the ease of planning home-work trips and a lower level of flexibility often required for this scenario, making EVs well-suited for such usage patterns (*17*).

Finally, the proportion of non-motorist crashes involving EVs is lower during snowy weather and icy road surface conditions. This can be attributed to the comparatively reduced usage of EVs during winter, possibly due to concerns about reduced battery range and overall performance in cold weather conditions.

Google Street View Image

Initially, the acquired images are processed through a pre-trained Segformer model, which is a type of Transformer neural network, to identify visual objects within each image *(17)*. The frequency of occurrence and corresponding percentage for each identified object are then documented in **Table 2**.

During the visual examination of GSV images, 10 environment factors are considered: crosswalk traffic light, traffic sign indicating pedestrians or cyclists, crosswalk type, number of lanes, number of directions, near the intersection, lane divided, and area type. The visual examination data is presented in **Table 3**.

To determine area types, as the geological boundary of the Chicago urban area is not obtained, this study relies on GSV images. More specifically, the area types are deduced based on the building type, which allows for a rough estimation of the population density of that area and, consequently, its area type. **Figure 2** provides an example of a crash location categorized as an urban area, while **Figure 3** illustrates an example of a crash location categorized as a suburban area.

Figure 2 E Pearson St and N Michigan Ave, Chicago (Lat. 41.89763, Long. -87.62427)

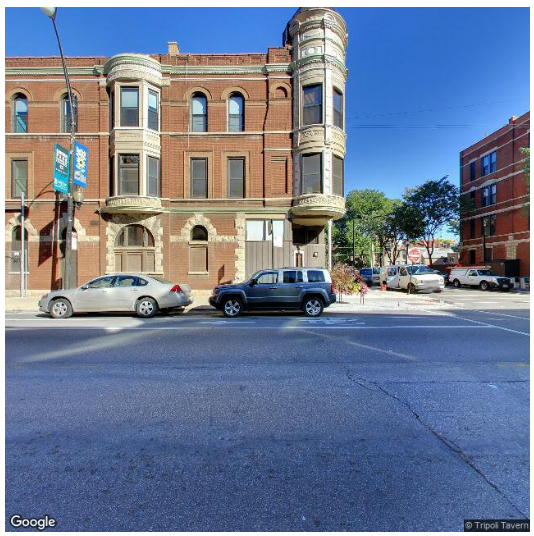

Figure 3 W Armitage Ave, Chicago (Lat. 41.91794, Long. -87.6573)

Excluding mobile objects such as bicycles, buses, and cars, notable distinctions between non-motorist crashes involving EVs and ICEV in **Table 2** are primarily attributed to the presence of terrain and traffic lights. The analysis of **Table 2** suggests that EV-related non-motorist crashes tend to occur more frequently in suburban or rural areas, since terrains are more prevalent there compared to urban areas. This observation finds support in the data presented in **Table 3**, which reveals that approximately 68% of non-motorist crashes involving EVs transpire in suburban locations. Moreover, **Table 2** underscores that non-motorist crashes involving EVs often take place in areas lacking traffic lights. This finding aligns with the reasonable expectation that urban areas, characterized by higher traffic volumes and intricate road networks, typically exhibit a higher density of traffic signals compared to suburban areas.

Table 3 indicates that non-motorist crashes involving EVs predominantly occur in proximity to intersections, two-way two-divided-lane trafficways, and areas lacking traffic control devices such as crosswalk lights and traffic signs.

TABLE 2 Portion of Visual Objectives by Pre-trained Segformer Model

	Electric Vehicles (EVs)		Internal Combustion Engine Vehicles (ICEVs)	
	Number	**Percentage**	**Number**	**Percentage**
Bicycle	46	61.33%	233	65.08%
Building	75	100%	358	100%
Bus	18	24%	159	44.41%
Car	75	100%	358	100%
Fence	75	100%	358	100%
Motorcycle	34	45.33%	186	51.96%
Person	71	94.67%	353	98.60%
Pole	73	97.33%	356	99.44%
Rider	23	30.67%	145	40.50%
Road	75	100%	358	100%
Sidewalk	75	100%	357	99.72%
Sky	72	96.0%	358	100%
Terrain	71	94.67%	314	87.71%
Traffic light	57	76.0%	337	94.13%
Traffic sign	75	100%	357	99.72%
Train	34	45.33%	229	63.97%
Truck	39	52.0%	240	67.04%
Vegetation	72	96.0%	358	100%
Wall	74	98.67%	354	98.88%

TABLE 3 Data by Visual Examination of GSV Image for EV-related Non-motorist Crash

	Number	**Percentage**
Crosswalk Light		
False	44	58.67%
True	31	41.33%
Traffic Sign		
False	62	82.67%
True	13	17.33%
Crosswalk Type		
Block	48	64.0%
Enhanced (Green)	2	2.67%
Line	7	9.33%
No Crosswalk	18	24.0%
Sidewalk Exists		
False	5	6.67%
True	70	93.33%
Lane Count		
0	1	1.33%
1	11	14.67%
2	41	54.67%
3	12	16.0%
4	9	12.0%
6	1	1.33%

Number of Direction		
0 (Parking)	1	1.33
1	25	33.33
2	49	64.33
Near Intersection		
False	24	32.0%
True	51	68.0%
Lane Divided		
False	21	28.0%
True	54	72.0%
Area/Location		
Highway	2	2.67%
Parking	4	5.33%
Rural	1	1.33%
Suburban	51	68.0%
Urban	17	22.67%

METHODS

This study analyzes GSV images of non-motorist crash locations through visual examination and pre-trained Transformer model, as described in the previous section. Additionally, the relationship between non-motorist injury levels and vehicle type is explored using the Pearson's chi-squared test and Cramer's V statistics. The crash severity level is further modeled using Binary Probit Regression model.

Statistical Analysis

Through statistical analysis, whether there is a significant relationship between non-motorist injuries and vehicle type and evaluate the magnitude of that association could be determined. Additionally, a Binary Probit Regression model is employed to understand how various factors impact the severity of non-motorist crashes involving EVs.

Hypothesis Testing

The Pearson's Chi-squared test is used to examine whether there is a significant association between non-motorists' injury counts and vehicle type (BEV, PHEV, and ICEV). This non-parametric statistical test analyzes categorical data and compares the observed frequencies of non-motorist injuries across different vehicle types with the expected frequencies under the assumption of independence *(21)*. Additionally, Cramer's V statistic is employed to measure the strength of the association between the variables. Suppose the χ^2 is already obtained, k is the number of vehicle types ($k = 2$), and r is the number of injury levels ($r = 2$), then Cramer V statistic is calculated through **Equation 1**:

$$\sqrt{\frac{\chi^2/n}{\min\{k-1, r-1\}}} \qquad (1)$$

Statistical Models

Due to the limited sample size of 75 EV-related non-motorist crashes and 358 ICEV-related non-motorist crashes, probit regression models are preferred over other models, such as multinomial logit and mixed logit, as they better mitigate the bias caused by the small sample *(22)*. Additionally, due to the scarcity of severe and fatal injury observations, the study combines injuries into two categories: injury and no injury. Consequently, a binary probit regression is deemed a suitable and efficient approach for this analysis.

The following specification is used that Y_i^* is the latent continuous measure of injury severity faced by non-motorist i in a crash, x_i is a vector of explanatory variables describing the non-motorists, driver, traffic condition, and environmental condition, β is a vector of parameters to be estimated, and ϵ_i is a random error term which is assumed to follow standard normal distribution, **Equation 2**:

$$Y_i^* = \beta' x_i + \epsilon_i \qquad (2)$$

Suppose μ_k represent unknown thresholds which need to be estimated along with β, the observed discrete injury severity Y_i is coded in **Equation 3**:

$$Y_i = \begin{cases} 0 & \text{if } Y_i^* \leq 0 \quad \text{(No Injury)} \\ 1 & \text{if } Y_i^* > 0 \quad \text{(Injury or Fatal)} \end{cases} \qquad (3)$$

Since ϵ_n is assumed to follow the standard normal distribution, and $\phi(\cdot)$ is the standard normal cumulative distribution function, the probability that ith non-motorist experiences k level of injury ($k = 0,1$) could be expressed in **Equation 4-1** and **Equation 4-2**:

$$\mathbb{P}(Y_i = 0 \mid x_i) = \mathbb{P}(Y_i^* \leq 0) = \mathbb{P}(\beta' x_i + \epsilon_i \leq 0) = \mathbb{P}(\epsilon_i \leq -\beta' x_i) = 1 - \phi(\beta' x_i) \qquad (4\text{-}1)$$
$$\mathbb{P}(Y_i = 1 \mid x_i) = 1 - \mathbb{P}(Y_i^* \leq 0) = 1 - \mathbb{P}(\beta' x_i + \epsilon_i \leq 0) = 1 - \phi(-\beta' x_i) = \phi(\beta' x_i) \qquad (4\text{-}2)$$

In binary probit model, the binary classes display an increasing trend (0 for no injury, 1 for injury). Positive signs in the associated variables indicate a higher likelihood of driver injury severity as their values increase. Conversely, negative signs suggest a lower likelihood of driver injury severity.

Model Transferability

A Likelihood Ratio Test is proposed to test whether estimated parameters are transferable spatially or temporally for a model *(23).* In this study, the same Likelihood Ratio Test is applied to test examine whether estimated parameters can be transferred between EV-related and ICEV-related non-motorist injury severity models.

In this context, $LL(\beta_T)$, $LL(\beta_E)$, and $LL(\beta_I)$ represent the log-likelihoods at convergence of models estimated using data from both EV-related and ICEV-related non-motorist crash, EV-related non-motorist crash only, and ICEV-related non-motorist crash only, respectively.

The test statistics χ^2 would follow a chi-square distribution with the degree of freedom equal to the sum of the number of estimated parameters in both EV-non-motorist and ICEV-non-motorist models minus the number of estimated parameters in the overall model, if the null hypothesis (the parameters are the same) holds. The calculation of the test statistic is performed using **Equation 5**:

$$\chi^2 = -2[LL(\beta_T) - LL(\beta_E) - LL(\beta_I)] \qquad (5)$$

To ensure the accuracy of the transferability test, consistent variables are employed across all models. However, due to the absence of visual examination of GSV images for ICEV-related non-motorist crashes, an EV-related non-motorist injury severity model is developed solely using police crash report data. This model shares the same variables as the ICEV-related non-motorist injury severity model, but its purpose is solely for testing rather than analysis and interpretation.

RESULTS

In this section, the results of our analysis are presented. The hypothesis testing section includes the contingency table with Chi-square test statistics, P-values, and Cramer's V statistics. Additionally, the outcomes of the Likelihood Ratio Test for model transferability are provided. In the statistical model section, the reference group for each categorical variable are discussed, and a table displaying estimated coefficients and standard errors is presented.

Hypothesis Testing

In **Table 4**, contingency tables display the frequency distribution of vehicle type and person type and non-motorist type. This facilitates the examination of potential associations between vehicle type and person type or non-motorist type. The **Table 4** also provides the respective chi-square statistics and p-values for each contingency table.

TABLE 4 Contingency Table and Corresponding Chi-square Test Results

	EV	ICEV	χ^2	d.f.	P-value	Cramer V
Non-Motorists	75	358	0.2826	1	0.595	0.003
Motorists	6,192	27,397				
Cyclists	42	122	11.75	1	0.001	0.171
Pedestrians	33	236				

The Likelihood Ratio Test for model transferability between EV-related and ICEV-related non-motorist injury severity models yields a test statistic $\chi^2 = 10.767$ with 15 degrees of freedom. The corresponding P-value for the null hypothesis (assuming the parameters are the same) is 0.769.

Statistical Model

The research model encompasses five variable groups: person, driver, traffic, environment, and age. We have conducted Binary Probit Regression, using "No Injury" as the baseline severity level, and the outcomes are presented in Table 5.

In this model, age is the only numerical variable. As for categorical variables, “female” represents the reference group for Gender, “cyclist” for Non-motorist Type, “not obscured” for Driver Vision, “false” for Hit and Run, “sidewalk” for location, “no control” for Traffic Control, “false” for Lane Divided, “false” for Intersection, “one-way” for Number of Directions, “Fall” for season, “dry” for Road Surface Condition, and “false” for Urban Area.

Additionally, Table 5 provides the corresponding standard errors and significance levels of coefficient estimates.

TABLE 5 Binary Probit Model Estimation of Crash Severity Outcome

		Electric Vehicles (EVs)		Internal Combustion Engine Vehicles (ICEVs)	
Categories	**Variable**	**Coef. Est.**	**Std. Err.**	**Coef. Est.**	**Std. Err.**
	Intercept	0.3222	1.2204	-0.4270	0.3472
Non-motorist Characteristics	**Gender**				
	Male	-1.1500	0.5211 *	-0.4471	0.1901 *
	Age	0.0035	0.0183	0.0127	0.0062 *
	Non-motorist Type				
	Pedestrian	0.9435	0.6468	0.8964	0.2111 ***
Driver Characteristics	**Driver Vision**				
	Obscured	-1.2674	0.7370	0.0723	0.2891
	Hit and Run				
	True	0.1970	0.5079	0.3024	0.1874
Traffic Characteristics	**Location**				
	Crosswalk	1.1836	1.0795	0.3383	0.3394
	Roadway	0.5247	0.7650	0.1582	0.2295
	Bike lane	1.2166	0.8476	1.5440	0.3931 ***

	Traffic Control				
	Present & Function	-0.2390	0.5636	-0.2832	0.2322
	Lane Divided				
	True	1.1823	0.6769	\	\
	Intersection				
	True	0.4533	0.6639	0.4229	0.3051
	Num of Directions				
	Two-way	-0.6938	0.5158	\	\
Environmental Characteristics	**Season**				
	Spring	0.3295	0.7112	-0.4782	0.2559
	Summer	0.2158	0.5500	0.1611	0.2164
	Winter	-0.5741	0.7363	0.2263	0.2623
	Surface Condition				
	Wet	1.6550	0.7363	-0.0497	0.2509
	Icy	\	\	-0.7975	0.4279
	Urban Area				
	True	-0.3065	0.5569	\	\

Significance Level Codes: < 0.001 (***), < 0.01 (**), < 0.05 (*)
Reference Severity Level: No Injury

DISCUSSION

Hypothesis Testing

Based on the Chi-square test with a p-value of 0.595 > 0.05, there is insufficient evidence to reject the null hypothesis, indicating no significant association between vehicle type (EV, ICEV) and person type (non-motorist, motorist). However, since exposure rates for non-motorist are not obtained, we could only conclude that, based on police report based crash data, there is no sufficient evidence to support the claim that there is an association between vehicle type (EV, ICEV) and victim type (non-motorist or not).

Nonetheless, EV-related and ICEV-related non-motorist crashes are not entirely identical, as evidenced by the Chi-square test with a p-value of 0.0006 < 0.05, indicating a statistically significant association between vehicle type (EV, ICEV) and non-motorist type (pedestrian, cyclist). However, the Cramer V statistic suggests such association is not strong. The higher likelihood of crashes involving EVs and bicycles in Chicago may be attributed to the city's substantial EV adoption and efforts to enhance bike-friendliness. With a wide array of public charging stations, bike lanes, and bike-sharing programs, there is increased interaction between EVs and bicycles, which potentially leads to a higher crash risk.

The Likelihood Ratio Test for model transferability yields a p-value of 0.769 > 0.05, indicating no sufficient evidence to reject the claim of parameter transferability between EV-related and ICEV-related non-motorist injury severity models. This suggests that the factors influencing non-motorist injury severity in crashes involving EVs and ICEVs are similar and consistent. Thus, a unified set of variables and coefficients can be used to model non-motorist injury severity in both EV and ICEV crashes.

Statistical Model

Given the limited observations, the study identified a limited set of significant variables for predicting the severity of EV-related non-motorist crashes. Despite these limitations, the estimated coefficients provide valuable insights into the impact of each factor on non-motorist injury severity in

EV-related crashes. Additionally, comparing the estimates between EV-related and ICEV-related non-motorist injury severity models can offer valuable insights into potential differences between the two types of crashes. Findings for different variable groups are discussed separately below.

Non-motorist Characteristics

The estimated coefficients for non-motorist characteristics, including gender, age, and type, do not show a significant difference between EV-related and ICEV-related non-motorist injury severity models.

For non-motorist victims, males have a lower likelihood of injury than female, aligning with prior studies showing females' increased vulnerability due to biomechanical differences *(24)*. Moreover, injury likelihood increases with age, consistent with previous research on non-motorist crashes *(25)*. This can be attributed to age-related factors such as decreased bone density, muscle strength, slower reaction time, and medical conditions, making older non-motorists more susceptible during crashes.

Pedestrians are at a greater risk of injury compared to cyclists, which can be attributed to factors such as visibility, speed, and protection. Cyclists' greater visibility to drivers due to their height and reflective clothing, their ability to travel faster and maneuver easily, and their use of protective gear like helmets contribute to their lower vulnerability in crashes.

Driver Characteristics

The estimated coefficients for driver vision differ between EV-related and ICEV-related non-motorist injury severity models. Specifically, when driver vision is obscured, the likelihood of injury decreases for EVs but increases for ICEVs. This could be attributed to reduced noise and vibration in EVs, making it easier for drivers to detect hazards even with obscured vision. Additionally, EV drivers' heightened awareness of their quieter vehicles may contribute to exercising greater caution in the presence of pedestrians or cyclists.

In both EV-related and ICEV-related non-motorist injury severity models, hit-and-run incidents exhibit a higher likelihood of injury. This can be attributed to the delayed medical attention, potentially worsening injuries, and association with severe injury or reckless driving behaviors such as driving under the influence of alcohol or drugs *(26)*.

Traffic Characteristics

For both EV-related and ICEV-related non-motorist crashes, injuries are more likely at crosswalks, roadways, bike lanes, and intersections compared to sidewalks. This can be attributed to higher traffic speeds and volumes at these locations, increasing the injury risk. Additionally, the presence of traffic control devices, such as pedestrian signs and traffic lights, effectively reduces the likelihood of injury, emphasizing their importance in reducing non-motorist injuries in both types of crashes.

Furthermore, utilizing GSV images, this study examines the impact of divided lanes and the number of directions on the injury severity of EV-related non-motorist crashes. The estimated coefficients suggest a higher likelihood of injury in EV-related non-motorist crashes on One-Way roads with divided lanes (where lanes are used to separate parking spaces or bike lanes), potentially due to limited escape routes and obstructed visibility caused by roadside vehicles for both drivers and non-motorists.

Environmental Characteristics

For EV-related non-motorist crashes, there is a higher likelihood of injuries in Spring and Summer, but lower in Winter compared to Fall. Conversely, for ICEV-related non-motorist crashes, injuries are more likely in Summer and Winter, but less likely in Spring compared to Fall. The precise explanation for these seasonal differences remains challenging, but they could be attributed to the interplay of seasonal patterns of non-motorist activities and the use of EVs.

The estimated coefficients for road surface condition show differences between EV-related and ICEV-related non-motorist injury severity models, with EVs more likely to have injuries on wet road surfaces, while ICEVs are less likely. The contrasting responses of EVs and ICEVs to wet road surfaces could be attributed to specific vehicle characteristics, although this study lacks relevant data for further validation. EVs' regenerative braking may be less effective on wet roads, impacting braking performance, and their weight distribution might reduce traction. In contrast, ICEVs with ABS, traction control, and front-wheel or all-wheel-drive configurations tend to handle wet and icy conditions better, particularly with the use of winter tires. Additionally, EVs' instant torque delivery can lead to sudden acceleration and potential loss of control on wet surfaces.

This study investigates the influence of area type on the injury severity of EV-related non-motorist crashes using GSV images. The estimated coefficient reveals a lower likelihood of injuries in urban areas compared to suburban and rural areas, possibly due to factors such as lower vehicle speeds, shorter distances between intersections, better road infrastructure, and sufficient traffic control devices.

CONCLUSIONS

This study aims to bridge critical research gaps in the investigation of non-motorist crashes involving EVs (BEVs and PHEVs) by conducting a comprehensive analysis of factors of crash and crash injury severity levels in both EV and ICEV collisions with non-motorists. The research focuses on the city of Chicago and employs various statistical methodologies to explore specific aspects of the crashes. Through contingency table and Chi-square testing, the study investigates whether EVs are more prone to colliding with non-motorists and examines which type of non-motorists is more likely to be involved in EV crashes. Additionally, a likelihood ratio test is applied to assess the transferability of factors between EV-related and ICEV-related non-motorist crash severity models. Furthermore, the research utilizes Binary Probit Regression analyses to explore and compare three main groups of crash factors: human, traffic, and environment, between EV-related and ICEV-related non-motorist crash injury severity models.

The results reveal a significant scale of similarities between EV-related and ICEV-related non-motorist crashes. The Likelihood Ratio Test suggests parameter transferability between the injury severity models for the two vehicle types, indicating that the factors influencing non-motorist injury severity apply consistently across both EV and ICEV crashes. Also, the Chi-square test suggests there is not enough evidence to claim that EVs are more likely to crash with non-motorists compared to ICEVs. Moreover, the estimated coefficients for various factors, including non-motorist gender, age, type, hit and run incidents, crash location, and presence of traffic control devices, do not exhibit significant differences between the two vehicle types of injury severity models.

Nevertheless, significant variations in factor distributions, including non-motorist type, hit-and-run incidents, damage level, crash hour, crash weekday, weather conditions, and road surface conditions, as well as the impact of season and road surface condition on injury severity, are observed between EV and ICEV crashes. These disparities can be attributed to the demographic characteristics of drivers, vehicle structure and design, and usage patterns unique to each vehicle type.

The study's contributions extend beyond the statistical analyses. Through combining GSV images with traditional police crash report data, this study reveals and uses previously unreported environmental variables for analysis. This study also highlights the potential of machine learning models for computer vision, such as Transformer Neural Network (Segformer), in enhancing crash analysis in certain perspectives. Moreover, this study addresses a critical concern regarding the rise of EVs, contributing to targeted safety measures and policies for non-motorists, including pedestrians and bicyclists.

Despite its strengths, this study does have several limitations. The data processing involves using vehicle make, model, and year to filter EV-related crashes since VINs are not provided. Consequently, some EV-related crashes might be overlooked, leading to potential underestimation of their numbers. Non-motorist crashes involving ICEVs in locations similar to those where EV-related non-motorist crashes occur, but not covered by EV-crash buffers, may also be disregarded. Additionally, due to a limited number of observations, cyclists and pedestrian data are combined in one model, and only a restricted set of variables is considered, with the usage of a simple regression model. Lastly, it is important to note that the information extracted from GSV images may not accurately reflect the circumstances at the time of the crashes.

Future research should focus on investigating EV-related non-motorist crashes with improved data quality and models that account for random effects and unobserved heterogeneity. Furthermore, considering a wider range of variables, including driver characteristics (age, gender, physical condition), vehicle characteristics (year, weight), and traffic characteristics (road class, trafficway type, pavement type), could lead to a more comprehensive analysis of EV-related crash characteristics. Furthermore, future research may explore how EV driver characteristics, the impact of Advanced Driver Assistance Systems (ADAS), and some vehicle-related factors, such as weight and structure, influence EV-crash characteristics.

ACKNOWLEDGMENTS

This study is partially supported by the Advancing Sustainability through Powered Infrastructure for Roadway Electrification (ASPIRE) award, an Engineering Research Center program by the National Science Foundation (NSF), grant no. EEC-1941524. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of NSF, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

AUTHOR CONTRIBUTIONS

The authors confirm their contributions to the paper as follows: study conception and design: J. Ling, X. Qian, K. Gkritza; data collection: J. Ling, X. Qian; analysis and interpretation of results: J. Ling, X. Qian; draft manuscript preparation: J. Ling, X. Qian, K. Gkritza. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. U.S. Department of Energy. (2020a) U.S. Plug-in Electric Vehicle Sales by Model. https://afdc.energy.gov/data. Accessed June 14, 2023.

2. U.S. Department of Energy. (2020b) U.S. HEV Sales by Model. https://afdc.energy.gov/data. Accessed June 14, 2023.

3. Hanna, R. (2009). *Incidence of pedestrian and bicyclist crashes by hybrid electric passenger vehicles* (No. HS-811 204).

4. Liu, C., Zhao, L., & Lu, C. (2022). Exploration of the characteristics and trends of electric vehicle crashes: a case study in Norway. *European transport research review*, *14*(1), 1-11. https://doi.org/10.1186/s12544-022-00529-2

5. Chen, R., Choi, K. S., Daniello, A., & Gabler, H. (2015). An analysis of hybrid and electric vehicle crashes in the US. In *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*.

6. Huang, S., Seraneeprakarn, P., & Shankar, V. (2016). *A hierarchical mixed logit model of hybrid involved crash severities* (No. 16-5162).

7. Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., & Milton, J. (2017). Occupant injury severities in hybrid-vehicle involved crashes: A random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research*, *15*, 41-55. https://doi.org/10.1016/j.amar.2017.05.003

8. Karaaslan, E., Noori, M., Lee, J., Wang, L., Tatari, O., & Abdel-Aty, M. (2018). Modeling the effect of electric vehicle adoption on pedestrian traffic safety: An agent-based approach. *Transportation Research Part C: Emerging Technologies*, *93*, 198-210. https://doi.org/10.1016/j.trc.2018.05.026.

9. Wu, J., Austin, R., & Chen, C. L. (2011). *Incidence rates of pedestrian and bicyclist crashes by hybrid electric passenger vehicles: an update* (No. HS-811 526).

10. Hanson, C. S., Noland, R. B., & Brown, C. (2013). The severity of pedestrian crashes: an analysis using Google Street View imagery. *Journal of transport geography*, *33*, 42-53. https://doi.org/10.1016/j.jtrangeo.2013.09.002

11. Mooney, S. J., DiMaggio, C. J., Lovasi, G. S., Neckerman, K. M., Bader, M. D., Teitler, J. O., ... & Rundle, A. G. (2016). Use of Google Street View to assess environmental contributions to pedestrian injury. *American journal of public health*, *106*(3), 462-469.

12. Kang, Y., Zhang, F., Gao, S., Lin, H., & Liu, Y. (2020). A review of urban physical environment sensing using street view imagery in public health studies. *Annals of GIS*, *26*(3), 261-275. https://doi.org/10.1080/19475683.2020.1791954

13. Stiles, J., Li, Y., & Miller, H. J. (2022). How does street space influence crash frequency? An analysis using segmented street view imagery. *Environment and Planning B: Urban Analytics and City Science, 49*(9), 2467-2483. https://doi.org/10.1177/23998083221090962

14. Chicago Open Data web-portal (2022). Chicago Data Portal. https://data.cityofchicago.org/. Accessed September 10, 2022.

15. Oak Ridge National Laboratory. (2022). FuelEconomy.gov Web Services. https://www.fueleconomy.gov/feg/ws/index.shtml#vehicle. Accessed September 14, 2023.

16. Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. Anal Methods Accid Res 11, 1–16. https://doi.org/10.1016/j.amar.2016.04.001.

17. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, *34*, 12077-12090.

18. Jensen, A. F., & Mabit, S. L. (2017). The use of electric vehicles: A case study on adding an electric car to a household. *Transportation Research Part A: Policy and Practice*, *106*, 89-99. https://doi.org/10.1016/j.tra.2017.09.004

19. Mersky, A. C., Sprei, F., Samaras, C., & Qian, Z. S. (2016). Effectiveness of incentives on electric vehicle adoption in Norway. *Transportation Research Part D: Transport and Environment, 46*, 56-68. https://doi.org/10.1016/j.trd.2016.03.011

20. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3213-3223).

21. Singhal, R., & Rana. R. (2015). Chi-Square Test and Its Application in Hypothesis Testing. *Journal of the Practice of Cardiovascular Sciences*, Vol. 1, No. 1, 2015, p. 69. https://doi.org/10.4103/2395-5414.157577.

22. Ye, F., & Lord, D. (2014). Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Analytic methods in accident research, 1*, 72-85.

23. Washington, S., Karlaftis, M. G., Mannering, F., & Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*. CRC press.

24. Obeng, K. (2011). Gender differences in injury severity risks in crashes at signalized intersections. *Accident Analysis & Prevention*, *43*(4), 1521-1531. https://doi.org/10.1016/j.aap.2011.03.004

25. Park, S. H., & Bae, M. K. (2020). Exploring the determinants of the severity of pedestrian injuries by pedestrian age: a case study of Daegu Metropolitan City, South Korea. *International journal of environmental research and public health*, *17*(7), 2358. https://doi.org/10.3390/ijerph17072358

26. Benson, A., Arnold, L., Tefft, B., & Horrey, W. J. (2018). Hit-and-run crashes: Prevalence, contributing factors and countermeasures.