Identifying Highly Correlated Variables Relating to the Potential Causes of Reportable Wisconsin Traffic Crashes

Xiao Qin, Mohammad Razaur Rahman Shaon and Robert J. Schneider
University of Wisconsin-Milwaukee

David A. Noyce, Andrea Bill and Yashar Zeiynali Farid
University of Wisconsin-Madison

Dominique Lord
Texas A&M University

WisDOT ID no. 0092-16-11

July 2019

RESEARCH & LIBRARY UNIT

WISCONSIN DOT
PUTTING RESEARCH TO WORK

Technical Report Documentation Page

1. Report No.	2. Government Accession No	3. Recipient's Catalog No
4. Title and Subtitle Identifying Highly Correlated Variables Relating to the Potential Causes of Reportable Wisconsin Traffic Crashes		5. Report Date August 2019 6. Performing Organization Code
7. Authors Xiao Qin, Mohammad Razaur Rahman Shaon, Robert J. Schneider, David A. Noyce, Andrea Bill, Yashar Zeiynali Farid, and Dominique Lord		8. Performing Organization Report No.
9. Performing Organization Name and Address Civil and Environmental Engineering University of Wisconsin-Milwaukee Milwaukee, WI 53201-0784		10. Work Unit No. (TRAIS) 11. Contract or Grant No. WisDOT SPR# 0092-16-11
12. Sponsoring Agency Name and Address Wisconsin Department of Transportation Research & Library Unit 4822 Madison Yards Way Madison, WI 53705		13. Type of Report and Period Covered Final Report September 2016 - August 2019 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The goal of this research project is to provide the Wisconsin Department of Transportation (WisDOT) with the state of art methodologies for identifying the most pertinent behavioral and engineering variables relating to reportable crashes. In this regard, an extensive literature review has been performed to summarize the latest progress in crash modeling from a broad range of crash factors and analytical methodologies. A three-pronged approach has been taken to study the complexity of crash occurrence in diverse and varying contexts, including area-level modeling (census tract), site-specific modeling (roadway segment), and event-oriented modeling (crash events). More than 100 variables have been evaluated in over a dozen statistical models. New methodologies have been developed to effectively quantify the impact of behavioral variables on crash counts, or account for their absence when they are not available. Moreover, the effect of risky driving behaviors on traffic safety has been measured using Wisconsin traffic citation data. In addition, street corridor-based pedestrian and bike crash prediction models have been calibrated by data collected from various new sources throughout the state. The findings in the study will serve as an important and comprehensive reference for traffic safety professionals and researchers.

17. Key Words	18. Distribution Statement
Traffic citation, driver behavior, crash factors, crash prediction model, injury severity, driver error, pedestrian and bicycle crash, propensity score matching, cause-effect relationship	No restriction. This document is available to the public through the National Technical Information Service 5285 Port Royal Road Springfield VA 22161

19. Security Classif.(of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	178	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

This research was funded by the Wisconsin Department of Transportation and the Federal Highway Administration under Project 0092-16-11. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views of the Wisconsin Department of Transportation or the Federal Highway Administration at the time of publication.

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. This report does not constitute a standard, specification or regulation.

The United States Government does not endorse products or manufacturers. Trade and manufacturers' names appear in this report only because they are considered essential to the object of the document.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Driving is a major means of transportation in the United States (U.S.), providing an unparalleled degree of mobility. However, motor vehicle crashes in the U.S. remain one of the leading causes of deaths. In Wisconsin. Traffic fatality and injury rates have decreased steadily since 1950. In 2009, the fatality rate fell below 1.0 fatality per 100 million vehicle miles traveled (VMT) for the first time since the state started collecting crash records. The traffic fatality rate remained low in the past decade due to the collective efforts of law enforcement, driver education, highway engineering, and emergency medical services, as well as the advancements in vehicle safety technologies.

A comparison between nationwide and Wisconsin crash fatalities is provided in Figure 1-1. Following a steep decrease in mid-2000s, the total traffic fatalities in Wisconsin rose for a third straight year, increasing to 539 in 2017 (*1*). The recent upward trend warrants a careful and comprehensive review of the causes and factors that contribute to traffic fatalities crashes, especially for those that result in fatal and severe injuries. The review will be beneficial for transportation agencies to identify high priority safety issues and determine appropriate countermeasures in an effort to lower the rate of traffic fatalities in Wisconsin.

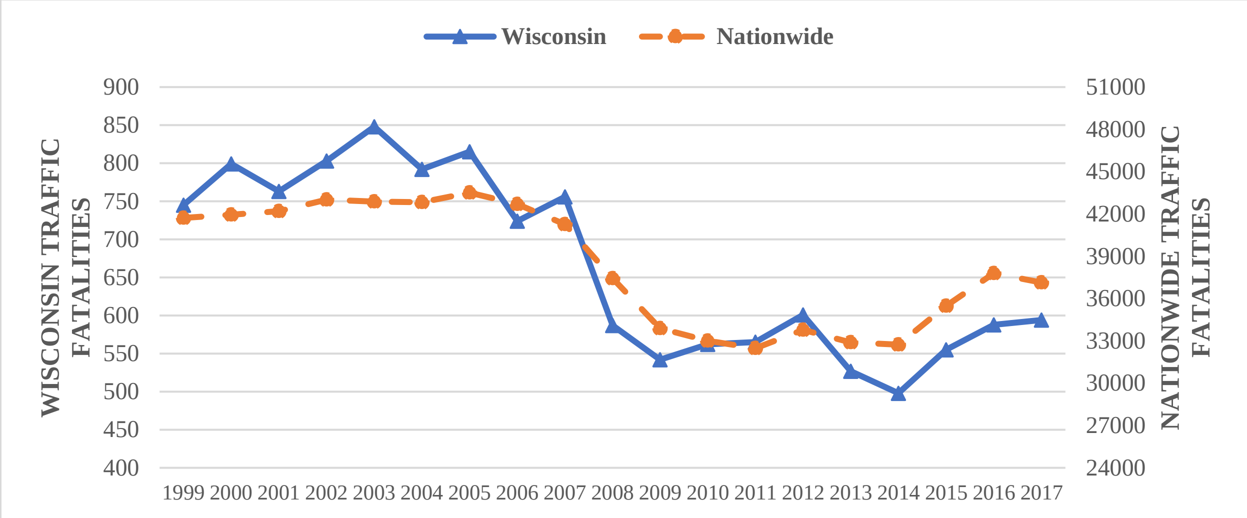


Figure 1-1 Trend in Traffic Crash Fatalities.

The notion that traffic crashes are not accidents, but avoidable events underscores the importance of identifying relevant, significant, and correctable factors leading to a crash. Traffic crash patterns react to changes in travel demand and patterns, driver demographics and behavior, highway design and traffic control, vehicle safety features, and broad economic trends. Crashes can also be affected by new or significant modifications to safety laws or policies that result in substantial increases in safety investments (e.g., national 55 mph maximum speed limit in 1974, mandatory seatbelt law in 1987, tougher Operating While Intoxicated (OWI) law in 1992, lowering the Blood Alcohol Limit (BAC) limit to 0.08 g/DL in 2003, the recent ban on cell phone texting in 2010). Global trends such as fluctuations in travel demand and vehicle registration, changes in travel preferences by age group, driver behavior, integration of new safety technologies into the vehicle fleet, and investments in safety also impact crashes. Furthermore, safety measures implemented by local agencies depend on factors such as weather conditions, roadway design treatments, local law enforcement activities, fuel prices and taxes, and unemployment rates. Thus, highway safety performance should be considered as an outcome of both global trends and local influences.

Crashes are usually caused by a confluence of risk factors. Among all possible contributing variables, driver error was the primary reason for over 90% of crashes involving drivers of all ages (*2*). Driver errors can be caused by a wide spectrum of elements, including environmental, roadway, and vehicle factors. Even for highways designed with good safety standards, the driving environment can be deteriorated by inclement weather, congested traffic conditions, and drivers' limited cognition, information processing, and decision-making skills. The complexity of causes with regard to crashes affirms that a single approach is inadequate to handle such a broad spectrum of data with diverse and varying characteristics.

To understand the effect of risky driving behaviors on crash occurrence, this study explored the characteristics of traffic citations and their relationships with crashes. The exploratory analysis with traffic violation data can provide valuable insight into the size, significance, and distribution of violations in Wisconsin. This study also took a unique, three-pronged approach to address safety issues by reviewing variables relating to crash causes at various levels. First, area-level models have been created to provide the linkage between total crash count of certain crash types and aggregated data by census tracts, augmented by roadway information from Wisconsin Information System for Local Roads (WISLR). This macroscopic analysis incorporates global trends such as socio-demographic-induced changes as well as infrastructure changes so that they can generate reliable estimates for the safety impacts of engineering and behavioral countermeasures under a multitude of growth scenarios. Second, novel statistical regression methods such as random parameters modeling have been applied to effectively account for local variations in the effects of highway and traffic characteristics. This microscopic view of specific roadway segments helps to identify the performance of key roadway design elements under the influence of human factors. Third, driver errors leading to a crash have been modeled in relation to their demographic, behavioral and personality factors. These granular details of driver mistakes can be identified when each crash is viewed individually.

With this three-pronged approach, the causes for crashes can be analyzed effectively and potential countermeasures can be recommended. The complex nature of modeling crash frequency and driver errors has led to a full-scale investigation of various data sources and hundreds of data elements. Although some results in this study require further investigation, the overall findings emphasize the importance of data quality, sharing, and distribution, as well as the reliable analytical methods for making Wisconsin's highways safer.

2. DATA COLLECTION AND DESCRIPTION

Evaluating the safety of a certain roadway element requires data elements for factors that may have safety effects. Accurate, accessible, timely, and standardized data allow decision-makers to identify the primary factors related to the source of crashes and their outcomes, develop and evaluate effective safety countermeasures, support traffic safety operations, measure progress in reducing crashes and their severity, design effective vehicle safety regulations, and target safety funding. State and national safety data are available for both traffic professionals and the public. The Wisconsin Department of Transportation (WisDOT) sponsored a study to develop the safety data resource guide for Wisconsin (*3*). The objectives of the resource guide are: 1) describe Wisconsin's "traffic records" or "highway safety data" system, 2) increase potential users' understanding of the quality and limitations of available data, and 3) streamline user access to the data. The data resource guide provided insight into available safety data items. Table 2-1 shows the data sources and variable types explored in this study. A detailed discussion of each data source is provided in the following sections.

Table 2-1 Safety Data Sources and Variable Categories.

Data source	Spatial Unit	Variable Categories
US Census Data	Census tract, Block group, County, etc.	Socioeconomic, demographic and travel behavior-related variables
State Trunk Network	Roadway corridor	Roadway geometry, pavement characteristics, mobility, safety, and other roadway-related data
Wisconsin Information System for Local Roads (WISLR)	Roadway corridor	Road geometry, pavement characteristics, and functional classification-related data
MV4000 (DT4000) Crash Data	Location-specific	Crash, Roadway, Vehicle, Weather, Occupants, Large and commercial vehicles, citation.
Wisconsin Department of Motor Vehicle (DMV)/ WisDOT; **Business Analyst**	City/Village/Town (C/V/T); Census tract, Block group, County, etc.	No. of registered vehicles, Vehicle Category; Job locations, Job types, Bar locations, etc.

2.1 U.S. Census Data

U.S. Census collects data on socioeconomic, demographic, and travel behavior-related variables for a defined geographical unit (e.g., census tract). These data attributes for Wisconsin Census Tracts are collected from TIGER/Line files, a product of U.S. Census. For this study, 2015 TIGER/Line files were collected from the U.S. Census. The 2015 TIGER/Line Shapefiles contain geography for the United States, the District of Columbia, Puerto Rico, and the Island

Areas. Geography in the 2015 TIGER/Line Shapefiles generally reflects the boundaries of governmental units in effect as of January 1, 2015, as well as other legal and statistical area boundaries that have been adjusted and/or corrected since the 2010 Census. This advantage includes boundaries of governmental units that match the data from the surveys that use 2015 geography, such as the 2015 Population Estimates and the 2015 American Community Survey. The 2015 TIGER/Line Shapefiles contain the geographic extent and boundaries of both legal and statistical entities. A legal entity is a geographic entity whose boundaries, name, origin, and area description result from charters, laws, treaties, or other administrative or governmental action. A statistical entity is any geographic entity or combination of entities identified and defined solely for the tabulation and presentation of data. Statistical entity boundaries are not legally defined and have no governmental standing.

2.2 Wisconsin State Trunk Network (STN)

Roadway geometry, pavement characteristics, mobility, safety, and other roadway-related data tables for the Wisconsin State Trunk Network (STN) are stored in Meta-Manager, a data management system developed by WisDOT. This system includes the state highway system of Wisconsin, the Interstate Highway System, and the United States Numbered Highway System, in addition to its other state trunk highways. The system integrates roadway and infrastructure data with geographical location and is updated every year. STN contains one GIS shapefile of all state highways and eight (8) separate database tables. The database tables represent different highway attributes related to each STN roadway segment. The tables can be integrated using a unique Meta-Manager segment ID (PDP ID). The Wisconsin STN is divided into five (5) zones, and highway attributes for each zone are separately maintained in the Meta-Manager database system. The Meta-Manager data from February 2017 were used in this study.

2.3 Wisconsin Information System for Local Roads (WISLR)

The roadway attributes for local roads in Wisconsin are collected from the Wisconsin Information System for Local Roads (WISLR). WISLR is an internet-accessible system that helps local governments and WisDOT manage local road data to improve decision-making and meet state requirements. WISLR combines local road data with interactive mapping functionality by using Geographic Information System (GIS) technology. The result is an innovative system that allows users to display their data in a tabular format, on a map, or both. WISLR provides a system for local governments to report local road information (i.e., width, surface type, surface year, shoulder, curb, road category, functional classification, pavement condition ratings) to WisDOT and serves as the highway inventory for local highways and streets.

2.4 Crash Data

Crashes that occurred on the Wisconsin roadway network are listed on the Wisconsin Motor Vehicle Accident Reporting Form 4000 (MV4000), now updated and renamed the Wisconsin

DT4000[1] Motor Vehicle Crash Report.[2] The database is stored and maintained in the WisTranPortal data hub which is developed through collaboration between the Wisconsin Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin-Madison and the WisDOT Bureau of Traffic Operations (BTO). The MV4000 crash database contains information on all police-reported crashes in Wisconsin since 1994. Information on the location of the crash, the vehicles involved, and general crash attributes are available in this dataset. Personal data have been removed from this dataset. A reportable crash is defined as a crash that resulted in injury or death, damage to government-owned non-vehicle property to an apparent extent of $200 or more, or total damage to property owned by any one person to an apparent extent of $1000 or more. An MV4000 crash report must have been completed by a police officer in order for a crash to be in the database. All police-reported crash data were collected from the MV4000 database for this study.

2.5 Traffic Citation Data

Enforcement officers issue traffic citations (tickets) to drivers for the violation of traffic laws. In Wisconsin, enforcement officers use the Wisconsin Uniform Traffic Citation (UTC) form and each citation issued must be resolved by a court action. A traffic conviction results from a guilty plea or court finding of guilty when a person is cited for a traffic violation. There are two types of courts in Wisconsin that resolve traffic citations: municipal court and circuit courts. Municipal courts resolve traffic citations issued by municipal police and circuit courts resolve traffic citations issued by state patrol or county sheriff. Both types of courts forward all citations to the Wisconsin Department of Motor Vehicle (DMV). The DMV is required to record convictions to establish a person's driving history. The DMV maintains this history of Wisconsin drivers to determine when license withdrawal is necessary. Some single convictions require that DMV withdraw a license. Other times a driver's accumulation of demerit points triggers an action. Traffic citations processed by both municipal and circuit courts in 2016 and 2017 were collected and used in this study.

2.6 Other Data Sources

A few proxy variables based on surrounding neighborhood characteristics, such as total number of jobs, retail jobs, bar locations, etc., were collected from Business Analyst, an ESRI product which contains job locations categorized by job types. Vehicle-related information such as the number of registered vehicles in a city/village/town were collected from WisDOT/ DMV to compensate for surrounding environment-related variables. Additionally, WisDOT publishes an annual "Facts and Figures" report from which information related to registered vehicles, vehicle type, license category, and conviction statistics were collected.

---

[1] MV4000 and DT4000 are used interchangeably in this report.

[2] As of January 1, 2017, the Wisconsin DT4000 crash report has replaced the MV4000 for all police reported motor vehicle crashes in Wisconsin. The DT4000 introduced a number of important changes to the overall set of crash data elements and attributes1, including adherence to the US DOT Model Minimum Uniform Crash Criteria (MMUCC) standard for crash data systems. Information about the DT4000 crash database modernization project is available on the TOPS Lab website: http://topslab.wisc.edu/research/cdi/.

3. METHODOLOGY

Statistical modeling is an effective approach to exploring the quantitative and statistically significant relationships between crash frequencies, traffic injury severities, driver mistakes, or other variables of interest. Once the relationship is established, the mean crash count or the probability of a crash-related error type can be estimated. Such a regression method assumes the error as random noise, and the mean can be represented as the true value around which observations fluctuate.

3.1 Crash Count Modeling

Crash count models focus on establishing a relationship between crash count and contributing factors based on the statistical significance unveiled from the data. Various methods have been developed and applied to handle data overdispersion, data heterogeneity, and variable selection. This section introduces the Negative Binomial (NB) regression model and more complicated NB-Lindley (NB-L) and random parameters NB-L generalized linear models (GLMs).

Negative Binomial Model

The NB model is one of the most notable models for crash frequency data which is a non-negative integer. The NB model accounts for data overdispersion, handles traffic exposure and offset variables, and has model parameters that are easy to estimate in commercial statistical software. The probability mass function (pmf) of the NB distribution can be written as:

$$\mathbf{P(Y = y; \phi, p) = \frac{\Gamma(\phi+y)}{\Gamma(\phi)\times y!}(1-p)^{\phi}(p)^{y}; \quad \phi > 0, 0 < p < 1} \tag{1}$$

Where, p = probability of success in each trial; Φ= Inverse dispersion parameter α (i.e. Φ=1/α);

The dispersion parameter ($\boldsymbol{\phi}$) measures the dispersion of the response variable. If the dispersion parameter equals zero, the NB model becomes the Poisson model, suggesting the Poisson distribution is a limiting case of the binomial distribution. If the dispersion parameter is greater than zero, it means that the response variable is over-dispersed. Using a log-link function, the mean response can be written as:

$$\boldsymbol{\ln(\mu) = \beta_0 + \sum_{i=1}^{q} \beta_i X} \tag{2}$$

Where, X= Covariates; β_i= Regression coefficient and q= Number of covariates.

NB-Lindley GLM

Although NB models can handle data dispersion, recent studies have pointed out that biased parameter estimates in the NB model can be found in a dataset with a long tail (*4; 5*). Hence, the NB-Lindley (NB-L) was introduced with the formulation as follows (*6; 7*):

$$\mathbf{P(Y = y|\ \mu, \phi, \theta) = \int NB(y; \phi, \varepsilon\mu) Lindley(\varepsilon; \theta)\ d\varepsilon} \quad (3)$$

where a general form of f (u; a, b) means that f is the distribution of the variable u, with parameters *a* and *b*. Following this explanation, the variable Y follows NB distribution with a mean and inverse-dispersion parameter of $\varepsilon\mu$ and ϕ (ϕ =1/α), respectively. The variable ε follows a Lindley distribution with parameter θ.

If the crash count is assumed to follow the NB-L(y; μ, ϕ, θ) distribution, the mean response function can be structured as follows (*6; 7*):

$$\mathbf{E(Y) = \left(e^{\beta_0 + \sum_{i=1}^{q} \beta_i X}\right) \times \frac{\theta+2}{\theta(\theta+1)} = e^{\left[\beta_0 + \log\left[\frac{\theta+2}{\theta(\theta+1)}\right]\right] + \sum_{i=1}^{q} \beta_i X}} \quad (4)$$

Where, $\beta_0^{'} = \beta_0 + \log\left[\dfrac{\theta+2}{\theta(\theta+1)}\right]$

The advantages of using NB-L GLM to model crash frequency data is its ability to account for extra overdispersion while maintaining the strength of the traditional NB model, especially for a dataset with a long crash tail.

Random Parameters GLM

In recent years, there has been great interest in developing data modeling alternatives that incorporate unobserved heterogeneity. Random parameters (RP) models can potentially capture unobserved heterogeneity by allowing parameters to vary across observations (such as a roadway segment) or be fixed within a group of observations but vary across groups that are specified by the analyst (such as roadway segments on the same highway route). Let x_{ij} denote the j-th covariate associated with i-th site. In an RP model, the coefficient β_{ij} is assumed to be random and is written as:

$$\boldsymbol{\beta_{ij} = b_j + w_{ij}} \quad (5)$$

where b_j denotes the fixed term (the mean parameter estimate), and w_{ij} denotes the random term. The random term is assumed to follow a predefined distribution such as a normal distribution with a zero mean and variance of σ^2. The random parameter β_{ij} should be used if the standard deviation of the random term w_{ij} is significantly different from 0; otherwise, a fixed parameter should be applied over all the individuals (*8; 9*). Considering the above parameterization, the pmf for RP NB model can be written as:

$$\mathbf{p(y_i) = \frac{\Gamma(\phi+y_i)}{\Gamma(\phi)\Gamma(y_i+1)} (1-p_i)^{y_i} {p_i}^{\phi};\quad \phi > 0, 0 < p_i < 1} \quad (6)$$

where, $p_i = \frac{\phi}{\mu_i+\phi}$; and the random coefficient definition in the model can be structured as:

$$\boldsymbol{\beta_{ij} \sim \mathrm{Normal}\left(\beta_j, \sigma_j^2\right);\ 1/\sigma_j^2 \sim \mathrm{Gamma}(0.01, 0.01)} \tag{7}$$

The advantages of using random parameters GLM to model crash frequency include handling (extra) unobserved heterogeneity in the target dataset due to the omission of important variables, and providing better inferences on the effect of explanatory variables on predicting the response variable.

3.2 Driver Error Modeling

Discrete choice models describe, explain, and predict choices between two or more discrete alternatives. Discrete or nominal scale data elements often appear in crash reports as crash types, manner of collisions, injury severities, traffic violations, and driver errors. A discrete choice model can be unordered or ordered. Unordered discrete outcome models do not consider the ordinal nature of the dependent variable (i.e., driver errors). The two most popular unordered discrete choice models are multinomial logit and multinomial probit, depending on the distributional assumption for the error term in the model.

Unordered Discrete Outcome Models — Multinomial Logit Model

The multinomial logit (MNL) model can be applied to predict the probability of different outcomes. The MNL model is formulated as:

$$\boldsymbol{p\left(y_i = j \middle| X_{i1}, X_{i2}, \ldots, X_{iJ}\right) = \frac{\exp(\beta_j X_{ij})}{\sum_{j=1}^{J} \exp(\beta_j X_{ij})}} \tag{8}$$

where $\boldsymbol{X_{ij}}$ is a vector of explanatory variables of observation i related with the jth driver error, mistake, or violation, and $\boldsymbol{\beta}_j$ is a vector of parameter estimates for $\boldsymbol{X_{ij}}$.

The MNL model allows the explanatory variables and parameter estimations related to one type of driver mistake or violation to vary. The MNL model should be appropriate for use when different driver errors are related to different contributing factors or are affected differently by the same factor. Yamamoto et al. (*10*) argued that non-ordinal models may offer unbiased parameter estimates, especially in situations of crash underreporting. MNL models, however, rely on the independence of irrelevant alternatives (IIA) assumption, meaning the odds of having one outcome over another do not depend on the presence or absence of other outcomes. The IIA assumption could lead to biased estimates when there is a correlation among various outcomes.

Unordered Discrete Outcome Models — Multinomial Probit Model

The multinomial probit (MNP) model relaxes the independence assumption built into the MNL model. The utility function of the MNP model that determines the preference or possible value of attaining the outcome i (i = 1, 2, . . . , I) for observation n can be written as (*11*):

$$\mathbf{U_{in}} = \boldsymbol{\beta}_\mathbf{i}\mathbf{X_{in}} + \boldsymbol{\varepsilon}_\mathbf{in} \text{ and } [\boldsymbol{\varepsilon}_\mathbf{1n}, \boldsymbol{\varepsilon}_\mathbf{2n}, \boldsymbol{\varepsilon}_\mathbf{3n}, \dots\dots, \boldsymbol{\varepsilon}_\mathbf{in}] \sim \boldsymbol{MVN}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{9}$$

Where, $\mathbf{X_{in}}$ = vector of independent variables for nth observation with i-th outcome, $\boldsymbol{\beta}_\mathbf{i}$ = vector of corresponding unknown coefficients, and $\boldsymbol{\varepsilon}_\mathbf{in}$ = disturbance term for unobserved effects.

The disturbance term $\boldsymbol{\varepsilon}_\mathbf{in}$ for i-th driver error type has a mean of zero, and errors can be correlated among different error types. Thus, the disturbance vector is defined by a multivariate normal distribution. In terms of log-likelihood which corresponds to the choice of i-th drive error, the choice of i-th driver error can be written as:

$$\boldsymbol{Prob}[\boldsymbol{Choice_{in}}] = \boldsymbol{Prob}[\mathbf{U_{in}} > \mathbf{U_{jn}},\ \ \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \dots\dots, \boldsymbol{I};\ \ \boldsymbol{i} \neq \boldsymbol{j}] \tag{10}$$

Ordered Discrete Outcome Models

The rationale behind using an ordered discrete choice model is the belief in a scale of severity among driver violations or errors from low to high (e.g., reckless driving such as disregarding traffic laws, speeding is probably the most risky behavior). The inclusion of the ordinal nature of the data in the statistical model defends the data integrity and preserves the information. The ordinal logit or probit model defines an unobserved variable, z_i, as a basis for modeling the ordinal nature of crash severity data. Z can be specified as:

$$\boldsymbol{z_i} = \boldsymbol{\beta X_i} + \boldsymbol{\varepsilon_i} \tag{11}$$

Where $\boldsymbol{X_i}$ is a vector of independent variables determining the discrete ordering for *i*th observation, $\boldsymbol{\beta}$ is a vector of estimable parameters, and ε_i is an error term.

The observed response variable, y_i, for *i*th observation is defined as:

$$\boldsymbol{y_i} = \boldsymbol{j}, \boldsymbol{if}\ \boldsymbol{\mu_{j-1}} < \boldsymbol{z_i} \leq \boldsymbol{\mu_j}\ , \boldsymbol{for}\ \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{J}; -\infty = \boldsymbol{\mu_0} < \boldsymbol{\mu_1} < \cdots < \boldsymbol{\mu_J} = +\infty$$

where j is the jth injury severity among all J levels in order, μ_j's are referred to as the threshold parameters.

The probability of y_i being the jth injury severity is given by:

$$\boldsymbol{p}(\boldsymbol{y_i} = \boldsymbol{j} | \boldsymbol{X_i}) = \boldsymbol{\Lambda}(\boldsymbol{\mu_j} - \boldsymbol{\beta X_i}) - \boldsymbol{\Lambda}(\boldsymbol{\mu_{j-1}} - \boldsymbol{\beta X_i}) \tag{12}$$

Where $\Lambda(.)$ represents the standard logistic cumulative distribution function (CDF) for the ordered logit model and the standard normal CDF for the ordered probit model. It is important to note that $\boldsymbol{\beta}$ is restricted to be the same across all levels.

4. ANALYSIS OF TRAFFIC VIOLATIONS

4.1 Introduction

Risky driver behaviors have long been identified as a major factor that contributes to crash occurrence. In the Wisconsin crash report MV4000 and DT4000, critical information is reported by police officers through interviews and crash investigations, such as road and traffic conditions occurring right before and during a crash. Table 4-1 lists 15 driver-related factors and 12 highway-related factors as possible crash contributing circumstances (PCC) reported in the MV4000 and DT4000 report.

Table 4-1 Possible Crash Contributing Circumstances Listed in MV4000 Database.

Driver-related Factors		Highway-related Factors	
• Driver condition • Physically disabled • Disregard traffic control • Following too close • Failure to yield • Failure to keep vehicle under control • In conflict • Inattentive driving	• Improper overtake • Improper turn • Left of center • Exceeding speed limit • Too fast for conditions • Unsafe braking • Others	• Snow/ Ice/ Wet • Narrow shoulder • Soft shoulder • Loose gravel • Rough pavement • Debris prior to accident	• Other debris • Sign obscured/ missed • Narrow bridge • Construction zone • Visibility obscured • Others

The driver-related PCCs consist of mistakes a driver can make such as speeding, becoming distracted, or violating traffic laws. Alcohol and drug use are listed separately in the MV4000[3]. If impaired driving from alcohol or drug use was treated as a driver-related PCC, over 71.5% of crashes in Wisconsin would be driver-related in 2016 based on 129,051 crashes. This is substantially lower than the national average, which is often more than 90% according to the National Highway Traffic Safety Administration (NHTSA)[4] (*2; 12*). The contribution of PCCs to crash injury severities varies. Take 2016 as an example, Figure 4-1 presents the proportion of different driver-related PCCs in both total crashes and injury crashes. Clearly, the contribution of driver-related PCCs is significantly higher in injury crashes compared with total crashes. Speeding, inattentive driving, and improper driving maneuver PCCs are the major contributors in both injury and total crashes.

[3] In MV4000, driver factors are about speeding, distractive and traffic violations, not including impairment. In DT4000, there are “suspected drug use” and “suspected alcohol use” checkboxes. Details of the violations are usually provided through traffic citation data.

[4] NHTSA broadly categorizes driver-related critical reasons into recognition errors, decision errors, performance errors and non-performance errors. Shaon et al. (2018) defined categories of driver-related critical reasons used by NHTSA using Wisconsin driver-related PCCs listed in MV4000. The percentage reported by NHTSA is based on collected on-scene information of nationally representative sample of 5,417 crashes on the events and associated factors leading up to crashes that involve light vehicles.

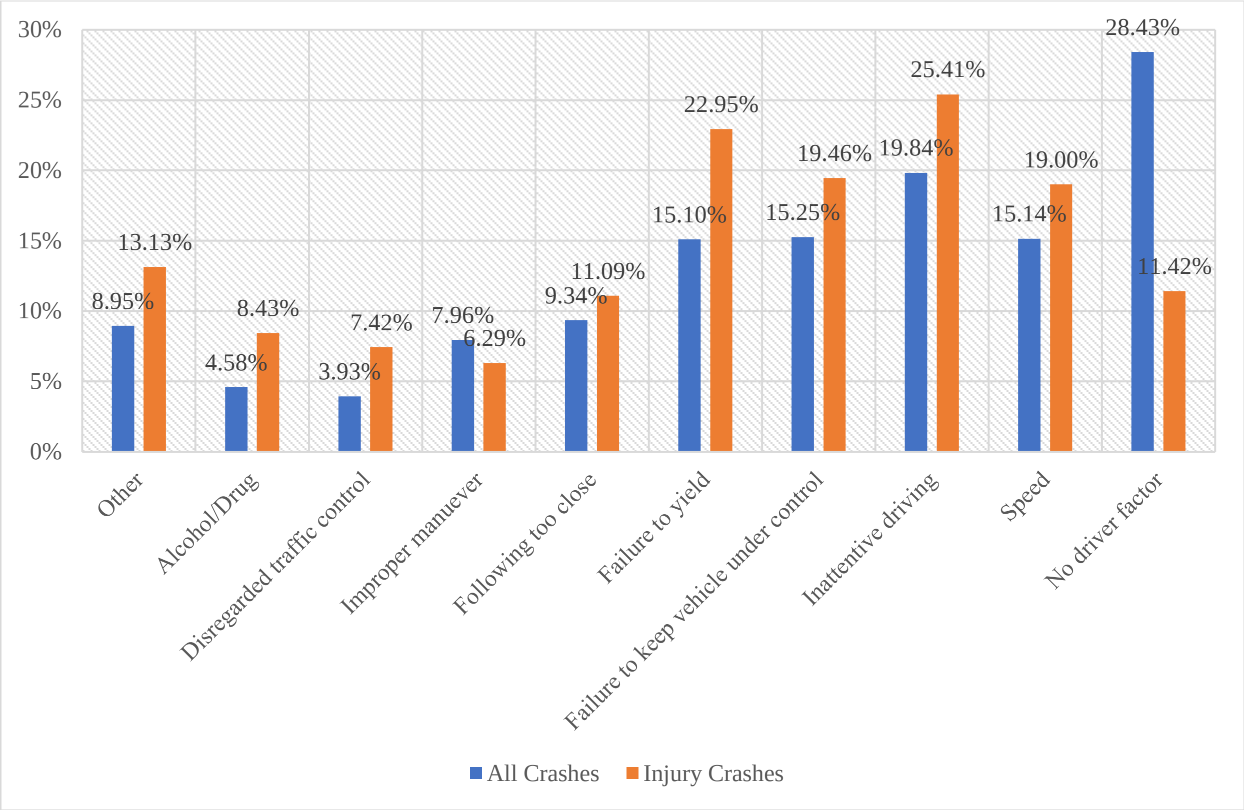


Figure 4-1 Driver related Possible Crash Contributing Factors.

Crash-based statistics provide extremely helpful, albeit limited information, because crash data are only collected after crashes happen. Moreover, data integrity and quality may be compromised by issues such as underreporting and location accuracy. One of the outstanding knowledge gaps is the lack of an unbiased estimate of the magnitude of risky driver behavior on Wisconsin highways and its influence on traffic safety. Drivers who operate vehicles in a reckless or risky manner do not always crash their vehicles, but they do have a greater risk of being involved in a crash. It is more effective to rectify behavior problems before a crash happens. Thus, knowing the size, severity, and significance of risky driver behaviors is essential for understanding challenges and exploring opportunities to improve highway safety on Wisconsin public roads.

Recognizing the limitation of crash data prompts the need to look beyond crash reports and finds sources that provide insight into risk factors not fully captured by crash statistics. Particularly of interest is traffic enforcement data (i.e., traffic arrests and citations) that may shed light on Wisconsin driver behavior and safety culture. Traffic violations have long been considered an important predictor of crashes. This study examines Wisconsin DMV data, traffic arrest data, citations, and licensed driver information. It was hypothesized that the enforcement data will present insightful statistics regarding risky driver behavior, establish the link between traffic violations and crashes, and identify top driver behavior problems that can be mitigated and corrected through engineering, enforcement, and education strategies.

4.2 The Trend of Traffic Violations in Wisconsin

Traffic violations occur when drivers violate laws that regulate vehicle operation on streets and highways. Traffic violations or judicial convictions of violations can provide valuable insight into driver behavior. Enforcement officers issue traffic citations to drivers who violate traffic laws. Most citations are written on the Wisconsin UTC form, and each citation must be resolved by a court action (*13*). When a court finds a driver guilty of a charge, the person usually pays a fine or forfeiture and is assessed demerit points on point-assessable offenses. Table 4-2 presents the summary statistics of crashes and crash-related citations in Wisconsin.

Table 4-2 Summary Statistics of Crashes and Citations in Wisconsin.

Year	Total Number of Crashes	Total number Crashes w/Citations	Total Citation Count	% Crashes w/Citations	% Change in Crashes w/Citation	% Change in Citations
2011	112,516	55,966	862,512	49.74%	1.66%	-1.70%
2012	109,385	53,842	886,899	49.22%	-3.80%	2.83%
2013	118,254	57,338	757,494	48.49%	6.49%	-14.59%
2014	119,734	57,410	713,185	47.95%	0.13%	-5.85%
2015	121,615	59,371	731,749	48.82%	3.42%	2.60%
2016	129,051	61,794	753,311	47.88%	4.08%	2.95%
2017	122,645	49,369	738,110	40.25%	-20.11%	-2.02%

As described in Table 4-2, citations are issued to fewer than 50% of crashes that occur as a result of driver mistakes, on an average. For example, 118,254 crashes occurred in 2013, whereas citations were issued in only 57,338 crashes (or 48.49% of all crashes) and in 2017, the figure is further down to 49,369 crashes (or 40.25% of all crashes). Another finding is that the percent change in crashes with a citation compared with the previous year does not seem to be in accordance with the percent change of citations issued. In 2013, a total of 757,494 citations were issued in Wisconsin with a negative 14.59% in total citation counts compared with 2012 while, there was a 6.49% increase in the number of crashes with citation in 2013. Although it is difficult to know exactly how many citations are issued to crashes, as a crash may have multiple citations, the out-of-sync trend between the two percent change measures suggests certain degree of citation discretion of police officers.

The eight-year trend of citation and crash count is plotted in Figure 4-2. Clearly, they do not share the same trend over time. The year 2013 seemed to be a watershed, as the total number of citations dropped by 14.59% from the previous year. The citation trend is relatively flat before and after 2013, but the total number of crashes has an overall ascending trend with a significant increase in 2013 and 2016. Unfortunately, there is no evidence to support the reduction in citations in 2013 could potentially be one of the reasons for more crashes in 2013-2016.

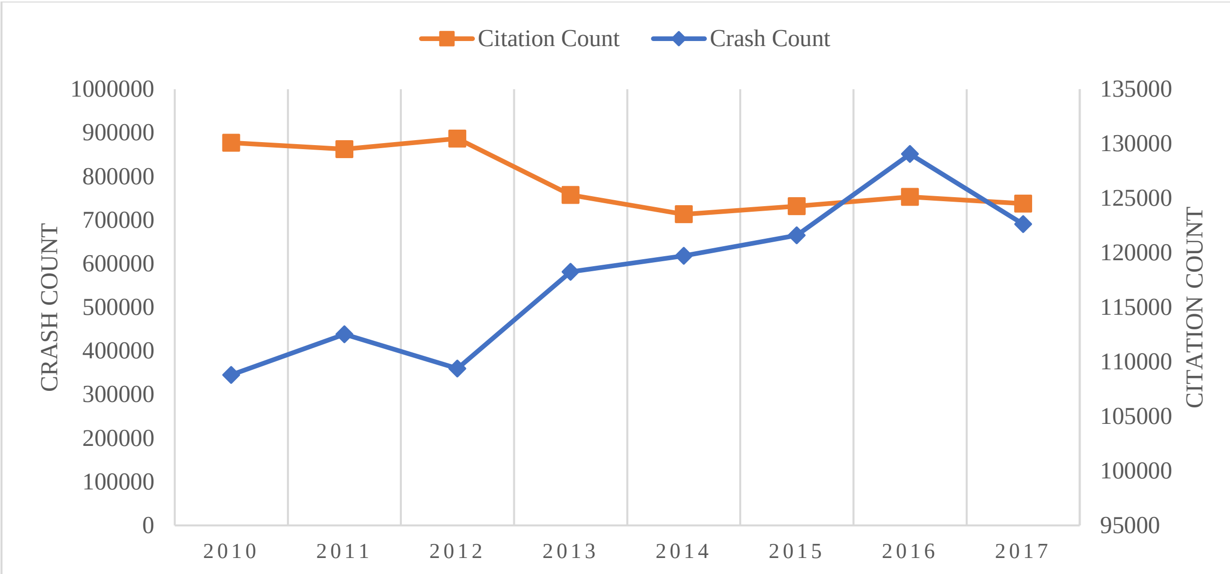


Figure 4-2 Crash and Citation Trend in Wisconsin.

4.3 Exploratory Data Analysis

All Violation Data

The Wisconsin DMV provided data on all violations processed in Wisconsin judicial courts for the year of 2016 and 2017, including violation charge code/description, processing court information, court decision, and the number of males and females involved in those corresponding violations. Table 4-3 represents the top 15 traffic citations out of 129 violation types in 2016 and 2017. The complete list of traffic violations and their corresponding percentage is included in Appendix A.

Table 4-3 Top 15 Traffic Citations and Conviction Rate in Wisconsin.

Code	Description	2016			2017		
		Count (%)	% Male	Conviction Rate	Count	% Male	Conviction Rate
SI	Speeding intermediate (11-19 over)	94,934 (12.01%)	57.42%	99.67%	86,722 (11.50%)	57.31%	99.65%
OWS	Operating while suspended	77,878 (9.85%)	63.15%	99.22%	78,456 (10.40%)	62.19%	99.12%
FFS	Failure to fasten seat belt	70,498 (8.92%)	70.19%	99.76%	58632 (7.77%)	70.47%	99.77%
CNI	Compulsory insurance - no insurance	63,878 (8.08%)	61.39%	98.59%	62,307 (8.26%)	61.21%	98.68%

CNP	Compulsory insurance - no proof	55,291 (6.99%)	59.95%	99.33%	50,838 (6.74%)	60.50%	99.31%
S	Speeding (1 - 10 over)	48,878 (6.18%)	56.94%	99.84%	43,500 (5.77%)	56.78%	99.84%
UV	Unregistered vehicle	35,131 (4.44%)	63.38%	98.35%	37,323 (4.95%)	62.99%	98.18%
OWL	Operating without driver license	30,079 (3.80%)	67.54%	99.24%	28,598 (3.79%)	67.32%	83.49%
OWI	Operating while intoxicated	29,673 (3.75%)	73.74%	83.30%	28,906 (3.83%)	73.55%	98.09%
SE	Speeding excess (20 or more over)	25,966 (3.28%)	65.04%	99.31%	26,254 (3.48%)	64.53%	99.20%
FOS	Failure to obey traffic sign or signal	22,926 (2.90%)	62.11%	99.21%	23,004 (3.05%)	62.62%	99.29%
PAC	Prohibited alcohol concentration	19,090 (2.41%)	74.15%	18.87%	18,868 (2.50%)	74.02%	18.07%
ORS	Operating while registration suspended	18,851 (2.38%)	58.63%	98.48%	19,138 (2.54%)	57.72%	98.73%
DS	Defective speedometer	15,714 (1.99%)	58.12%	99.94%	14,613 (1.94%)	58.46%	99.90%
OAR	Operating after revocation	14,727 (1.86%)	75.98%	99.46%	15,184 (2.01%)	74.72%	99.33%

After excluding non-behavior related traffic violations in Table 4-3 (i.e., operating while suspended or revoked license, unregistered vehicles or without insurance), violations due to speeding and alcohol were most prominent. Speeding-related citations made up 21.1% of all citations over two years, while OWI and PAC accounted for 6.2% of all citations. Interestingly, failure to fasten one's seat belt is listed as one of the top violations despite the steady improvement of safety belt use in Wisconsin[5]. Gender differences were not apparent for speeding-related citations but were starkly different for categories like seat belt usage and drunk-driving where male drivers consistently received more than 70% of those charges. Note that all statistics are based on citations instead of individuals, meaning a driver can receive multiple citations (e.g., SE and OWI).

Wisconsin has two separate charges for drunk driving – operating while intoxicated (OWI) and prohibited alcohol concentration (PAC) (*14*). With an OWI charge, the officer need only show that consuming alcohol and/or drugs affected your ability to competently operate a motor vehicle. With a PAC charge, however, the officer must have evidence that your blood alcohol content (BAC) was over the state's legal limit (0.08% or higher). In the case of PAC, the prosecution must establish that the driver's BAC was over the legal allowable limit—not that your ability to drive was compromised in any way. This can be done by introducing breathalyzer

[5] Wisconsin safety belt use is at an all-time high of 88 percent of drivers and passengers. https://wisconsindot.gov/Pages/safety/education/seat-belt/default.aspx

or blood test results that show your BAC was .08% or higher. Please note that the drivers could also be charged with OWI as well as PAC in certain Wisconsin counties. The conviction rate is near 100% in 2017 for OWI but the conviction rate for PAC was merely 18.07%, suggesting drivers fought hard to be exonerated from prohibited alcohol concentration or were convicted of a lesser charge such as OWI. By law, if someone is charged with both OWI and PAC and convicted of one of these charges, the other charge must be dismissed within a number of days, but the penalties, however, are the same (*15*). Moreover, the conviction rate for OWI was significantly different in 2016 compared with 2017. The conviction rate for OWI offense in 2016 was 83.30%, whereas the conviction rate is 98.09% in 2017 for the same offense type. It is probably worth exploring what factors help drive up the citation rate for OWI in 2017.

Based on NHTSA's definitions of risky driver behaviors (i.e., drunk driving, drug-impaired driving, distracted driving, speeding, seat belts, and drowsy driving), Wisconsin traffic violations can be categorized as: 1) speed-related, 2) impaired-driving-related, 3) traffic-rule-related, 4) inattentive and distracted driving, 5) license-related and 6) other, based on the description of each violation. The categories of violation data are presented in Table 4-4.

Table 4-4 Categorization of Violations.

Speed Related Violations	
• Speeding (1 - 10 over)	• Racing
• Speeding Intermediate (11-19 over)	• Commercial Speeding Intermediate (15-19 over)
• Speeding Excess (20 or mover over)	• Commercial Speeding Excess (20 or more over)
• Defective Speedometer	• Commercial Too Fast for Conditions
• Imprudent Speed	• Commercial Imprudent Speed
• Too Fast for Conditions	• Commercial Reckless Driving
• Reckless Driving	
Impaired Driving Related Violations	
• Operating While Intoxicated	• Juvenile Controlled Substance
• Prohibited Alcohol Concentration	• Underage ID
• Underage Alcohol	• Commercial Operating While Intoxicated
• Implied Consent	• Negligent Homicide Intoxicated
• Intoxicant in Vehicle-Operator	• Commercial Implied Consent
• Juvenile Alcohol	• Commercial Alcohol
• Underage Alcohol Operation	• Implied consent underage
• Intoxicant in Vehicle-Passenger	• HAZMAT commercial operating while intoxicated
• Operating While Intoxicated Causing Injury	• Commercial OWI causing injury
• Intoxicants in Vehicle Carrying Underage Person	• Juvenile ID
• Drugs	

Traffic Rules Related Violations	
• Failure to obey traffic sign or signal	• Parking on highway
• Failure to fasten seat belt	• Interfere w/ traffic sign/signal
• Failure to yield right of way	• Failure to give signal
• Failure to keep vehicle under control	• Commercial deviating from lane
• No or improper lights	• Commercial following too closely
• Following too closely	• Improper brakes
• Deviating from lane of traffic	• Unnecessary acceleration
• Driving on wrong side of highway	• Driving over walk
• Illegal turn	• Failure to dim lights
• Passing illegally	• Driving against traffic
• Signal violation	• Commercial passing illegally
• Child safety restraint	• Driving on divided highway improperly
• Backing illegally	• Flashing yellow signal violation
• Obstructing traffic	• Roadway violation
• Obstructed view or control	• Following emergency vehicle
• Failure to stop for school bus	• Failure to stop after accident
• Attempt to elude officer	
Inattentive and distracted Driving Related Violations	
• Inattentive driving	• Using cell phone while driving in work zone
• Texting while driving	• Commercial telephone use while driving
• Using Telephone While Driving w/Probationary or IP	
License Related Violations	
• Operating without driver license	• GDL passenger violation
• Operating while suspended	• Permit unauthorized person to operate
• Operating while registration suspended	• GDL curfew violation
• Operating after revocation	• GDL miscellaneous driving offense
• License not on person	• Operating with multiple licenses
• Improper plates	• Operating while disqualified
• Violation of restriction	• Illegal riding
Other	
• Failure to pay forfeiture	• Littering highway
• Compulsory insurance - no insurance	• Commercial failure to report accident
• Unregistered vehicle	• Commercial absolute sobriety
• Compulsory insurance - no proof	• Commercial duty upon striking property
• Failure to report accident	• Negligent homicide
• Failure to stop after accident - unattended vehicle	• Operating while out of service
• Duty upon striking property	• Falsified accident report
• Failure to pay forfeiture - juvenile	• Commercial possession of intoxicating beverage
• Improper muffler	• Miscellaneous
• Failure to notify of address or name change	• Falsified application

• Improper equipment	• Crossing fire hose
• Truancy	• Projecting loads on side of vehicle
• Non-trackable	• Transporting children in cargo areas of motor vehicle
• Failure to transfer title	• Illegal use of operator's license
• Vehicle used in commission of felony	• Compulsory insurance-fraud
• Special limitations on load	• Commercial failure to stop after accident-unattended
• Commercial unlawful operation	• Fraudulent application
• Ignition/immobilization device tampering	• Railroad failure to stop
• Restrictions on parking and stopping	• Failure to pay support
• Great bodily harm	• Haz commercial duty upon striking property
• Transport person or vehicle illegally	• Surrender of licenses and registration upon revocation or suspension
• Unnecessary noise	

The count of violations by gender for each category is presented in Table 4-5 along with the proportion of violations. The two-year statistics based on the categorized data shows that:

- The count of total violations in all categories is consistent across the time period.
- The distribution of violations between male and female is also consistent.
- Speed and traffic rule violations are the major sources of violations in Wisconsin. Around 24% of the total violations processed in the Wisconsin court system are related to speeding.
- A small fraction of total violations are related to impaired driving and inattentive and distracted driving.

The number of citations issued for risky driver behavior (i.e., the first four rows of Table 4-5: speed, impairment, violating traffic rules, inattentive and distracted) adds up to more than 419,000 per year, a noticeably high percentage considering there are approximately 4.23 million licensed drivers in Wisconsin.

Table 4-5 Violation Counts by Category.

Violation Category	**2016**			**2017**		
	Male	**Female**	**Total (%)**	**Male**	**Female**	**Total (%)**
Speed	114,804	79,397	194,201 (24.56%)	106,590	73,573	180,163 (23.59%)
Impairment	49,268	18,441	67,709 (8.56%)	47,977	18,092	66,069 (8.65%)
Traffic Rules	103,946	57,562	161,508 (20.42%)	95,318	52,927	148,245 (19.41%)

Inattentive/distracted	6,151	4,466	10,617 (1.34%)	5,611	3,965	9,576 (1.25%)
License	107,797	58,381	166,178 (21.01%)	106,025	59,036	165,061 (22.86%)
Others	119,731	70,718	190,449 (24.09%)	116,465	68,561	185,026 (24.23%)
Total	501,697 (63.45%)	288,965 (36.55%)	790,662	487,534 (63.38%)	276,154 (36.62%)	763,688

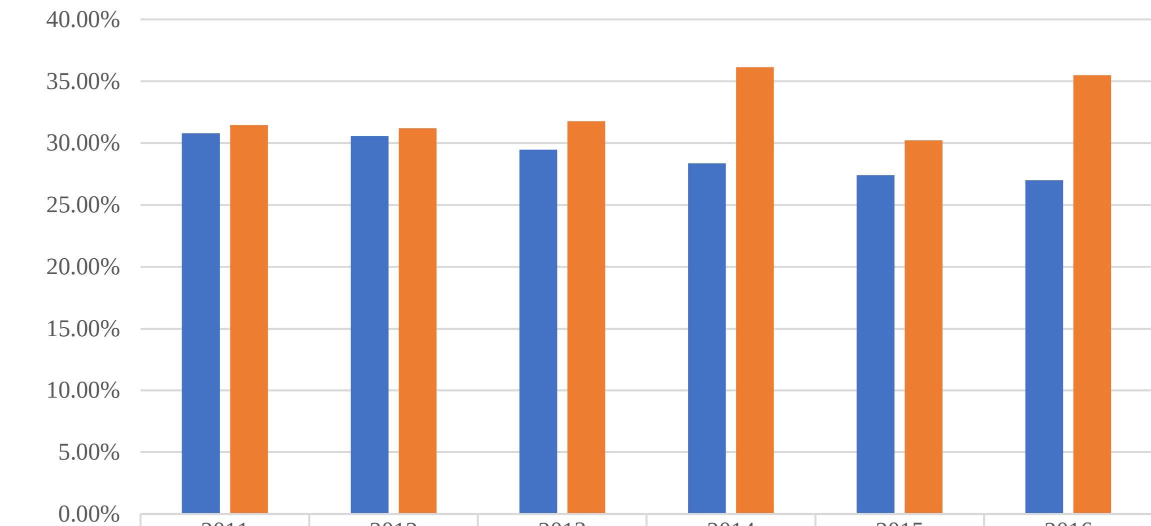
Each violation issued by a police officer is processed in the corresponding judicial court. The above-mentioned violation categories are subcategorized into general, juvenile or underage and commercial drivers to further gain the insight into the effect of driver age and driving experience on safety. Exploring the proportion of violations in different subcategories and the variation in conviction rates may support informed decision-making such as tougher laws against underage drinking and stricter regulation on commercial driver safety compliance. Table 4-6 describes the court decisions per each category of violation.

Table 4-6 Conviction Rate by Violation Category in 2016.

Category	Sub-Category	Sample Size	Guilty	Appealed	Dismissed	Error	Vacate
Speed	General	194,002	99.65%	0.08%	0.17%	0.07%	0.02%
	Commercial	199	97.31%	0.29%	2.00%	0.34%	0.05%
	Total	194,201	99.63%	0.08%	0.19%	0.08%	0.02%
Impairment	General	55,788	62.24%	0.21%	37.19%	0.27%	0.10%
	Juvenile	11,328	97.12%	0.22%	2.17%	0.33%	0.16%
	Commercial	55	80.00%	0.00%	20.00%	-	-
	Drugs	548	99.09%	0.36%	0.36%	0.18%	-
	Total	67,719	68.23%	0.21%	31.18%	0.27%	0.11%
Inattentive/ distracted	General	10,614	99.28%	0.08%	0.50%	0.11%	0.02%
	Commercial	3	100.00%	0.00%	0.00%	0.00%	0.00%
	Total	10,617	99.28%	0.08%	0.50%	0.11%	0.02%
Traffic Rules	General	161,212	99.30%	0.06%	0.46%	0.17%	0.01%
	Commercial	296	99.66%	0.34%	0.00%	0.00%	0.00%
	Total	161,508	99.31%	0.06%	0.46%	0.17%	0.01%
License	General	164,148	99.13%	0.08%	0.64%	0.13%	0.02%
	GDL	2,030	98.77%	0.15%	0.94%	-	0.15%
	Total	190,439	99.13%	0.08%	0.64%	0.12%	0.02%
Others		190,439	98.84%	0.07%	0.91%	0.15%	0.03%
Grand Total		790,662	96.57%	0.09%	3.17%	0.14%	0.03%

Speed-related Violation Data

According to NHTSA, a crash is considered to be speeding-related if any driver in the crash was charged with a speeding-related offense or if a police officer indicated that racing, driving too fast for the conditions, or exceeding the posted speed limit was a contributing factor in the crash. Wisconsin adopts the same definition as NHTSA to define speed-related crashes. In 2017, 35% of total fatalities in Wisconsin were speed-related, whereas the national average for speed-related fatalities is 27%. Figure 4-3 compares speed-related fatalities in Wisconsin and those for the United States from 2011 to 2016. Wisconsin's figure is consistently above the national average.



	2011	2012	2013	2014	2015	2016
National	30.79%	30.58%	29.48%	28.35%	27.40%	26.99%
Wisconsin	31.46%	31.21%	31.77%	36.14%	30.21%	35.50%

National Wisconsin

Figure 4-3 Comparison of Speed-Related Fatality Trend.

It is not easy to determine whether speed is indeed a crash contributing factor. The police officer assessment of speed as a contributing factor is more reliable for fatal crashes as extensive forensic investigations are required but not so for non-fatal crashes. On average, only 65% of speed-related crashes in Wisconsin receive citations. Table 4-7 provides the distribution of conviction rates for all speed-related citations issued in Wisconsin.

Table 4-7 Distribution and Conviction Rate of Speeding Violations.

	2016				2017			
Charge Description	**Male**	**Female**	**Total**	**Conviction Rate**	**Male**	**Female**	**Total**	**Conviction Rate**

General								
Speeding (1 - 10 over)	27,832	21,046	48,878 (25.17%)	99.84%	24,698	18,802	43,500 (24.14%)	99.84%
Speeding intermediate (11-19 over)	54,513	40,421	94,934 (48.88%)	99.67%	49,699	37,023	86,722 (48.14%)	99.65%
Speeding excess (20 or more over)	16,889	9,077	25,966 (13.37%)	99.31%	16,941	9,313	26,254 (14.57%)	99.20%
Defective speedometer	9,133	6,581	15,714 (8.09%)	99.94%	8,543	6,070	14,613 (8.11%)	99.90%
Imprudent speed	3,411	1,148	4,559 (2.35%)	98.75%	3,498	1,133	4,631 (2.57%)	98.79%
Too fast for conditions	1,363	740	2,103 (1.08%)	98.72%	1,466	845	2,311 (1.28%)	98.36%
Racing	108	7	115 (0.06%)	97.39%	127	7	134 (0.07%)	95.52%
Reckless driving	1,360	373	1,733 (0.89%)	97.23%	1,393	367	1,760 (0.98%)	96.42%
Commercial								
Commercial speeding intermediate (15-19 over)	149	2	151 (0.08%)	99.34%	159	10	169 (0.09%)	99.41%
Commercial speeding excess (20 or more over)	13	0	13 (0.01%)	100.00%	24	1	25 (0.01%)	96.00%
Commercial imprudent speed	7	2	9 (4.6E-3%)	100.00%	7	1	8 (4.4E-3%)	100.00%
Commercial too fast for conditions	21	0	21 (0.01%)	90.48%	29	1	30 (0.02%)	96.67%
Commercial reckless driving	5	0	5 (2.6E-3%)	80.00%	6	0	6 (3.3E-3%)	100.00%
Grand Total	**114,804 (59.11%)**	**79,397 (40.88%)**	**194,201**	**99.63%**	**106,590 (59.16%)**	**73,573 (40.84%)**	**180,163**	**99.57%**

It can be concluded from the speeding-related citation statistics that:

- Although the proportion of male drivers with a speed ticket is 63%, males and females had an almost similar proportion of speeding citations for the top two speeding citation types.
- The ratio of male to female drivers who were involved in reckless driving is 4:1.
- A very small proportion of speeding citations were issued to commercial drivers.
- The trend of different speeding citations is the same over the study period.

Drunk Driving

Drivers are considered to be alcohol-impaired when their blood alcohol concentrations (BACs) are .08 grams per deciliter (g/dL) or higher. Thus, any crash involving a driver with a BAC of .08 g/dL or higher is considered to be an alcohol-impaired-driving crash, and fatalities occurring in those crashes are considered to be alcohol-impaired-driving fatalities. In Wisconsin, for drivers with three or more prior Operating While Intoxicated (OWI) convictions, the limit is lower; they cannot operate a motor vehicle if their BAC is greater than 0.02. Figure 4-4 provides a comparison of alcohol-related fatalities between the national average and Wisconsin's average from 2011 to 2016. Compared to the pattern of speed-related fatality in Figure 4-4, Wisconsin's alcohol-related fatalities are very concerning. However, the year of 2016 is an exception in which 27% of traffic fatalities involving alcohol, the lowest record in the six-year span.

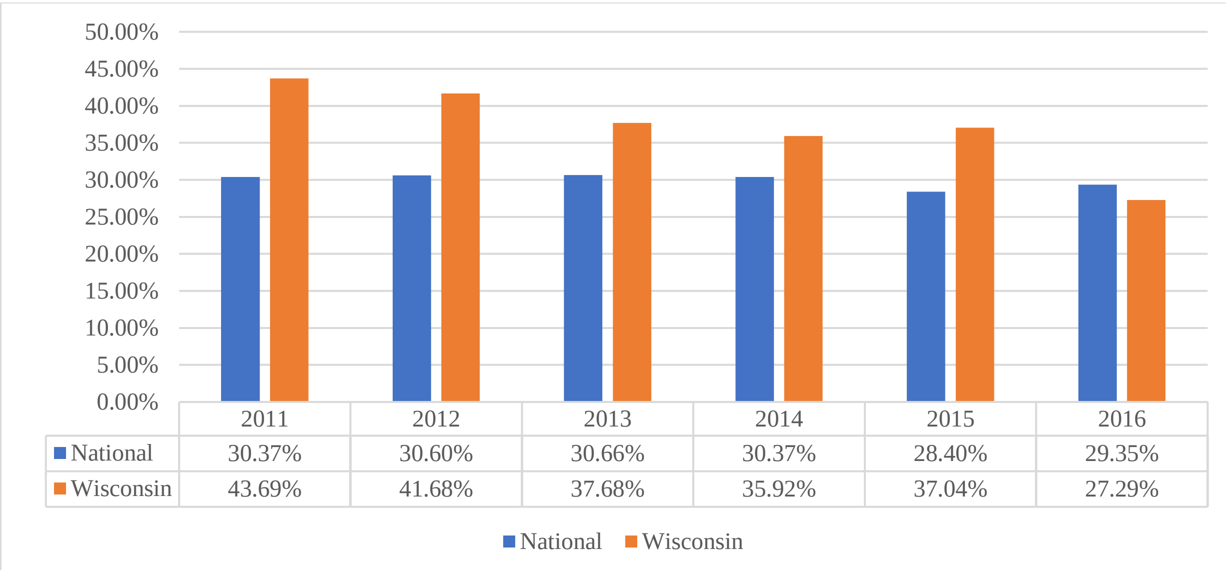


Figure 4-4 Comparison of Alcohol-Related Fatality Trend.

Drunk-driving citations are categorized as impairment-related citations in the database. To explore the impact of different age groups and driver types, the impairment-related citations are further sub-categorized into general, juvenile, commercial, and drug.

Table 4-8 Distribution and Conviction Rate of Impairment Related Citations.

Sub-Category	2016			2017		
	Total	% Male	Conviction Rate	Total	% Male	Conviction Rate
General	56,097 (82.85%)	74.22%	62.24%	55,323 (83.74%)	73.97%	61.59%
Juvenile	11,019 (16.27%)	65.13%	97.12%	10,213 (15.46%)	65.09%	96.15%
Commercial	45 (0.07%)	95.56%	80.00%	55 (0.08%)	100.00%	76.36%
Drug	548 (0.81%)	75.73%	99.09%	478 (0.72%)	73.22%	98.54%
Grand Total	67,709	72.76%	68.23%	66,069	72.62%	67.21%

Citation information for drunk-driving related driver behaviors is available in three different data tables: 1) OWI Citation, 2) OWI Arrests and, 3) Liquor Law Violations. Summary statistics and facts from each data table are provided in the following sections.

OWI Citation

The OWI Citation data table provides OWI-related traffic citation counts for each court in Wisconsin, including charges, offense count, and gender. The OWI Citation data table contains only citation counts for which the court decision was "Guilty." The distribution of OWI citations by offense type is presented in Figure 4-5.

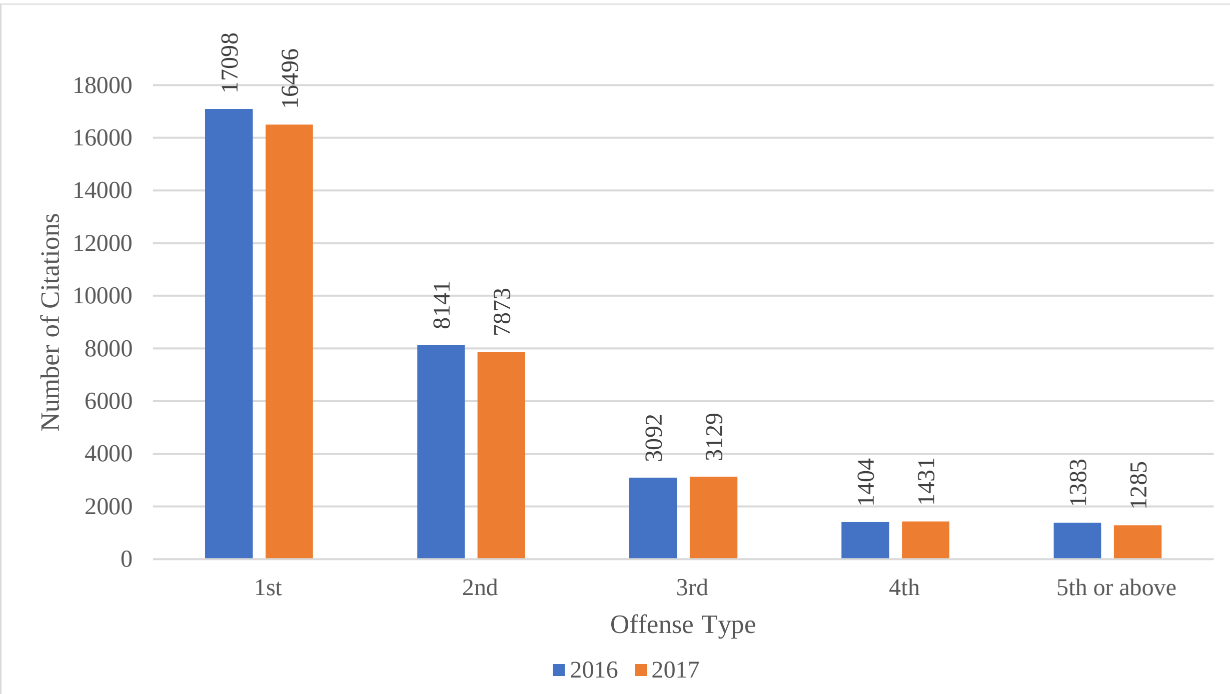


Figure 4-5 Distribution of OWI Citation by Offense Type.

The distribution of OWI citation by gender and offense type is presented in Table 4-9. It is astonishing to see that nearly 45% of the OWI citations are for repeat offenders, 47.6% for male and 38.34% for female. These repeat offenders of OWI are serious threats to highway safety, as many studies have shown that drivers with prior OWI convictions are overrepresented in fatal crashes and are at greater risk of being involved in fatal crashes.

Table 4-9 Count of OWI Citation by Charge Type.

Offense Type	2016			2017		
	Male	Female	Total (%)	Male	Female	Total (%)
1st	12,043	5,055	17,098 (54.95%)	11,651	4,845	16,496 (54.60%)

2nd	6,109	2,032	8,141 (26.16%)	5,867	2,006	7,873 (26.06%)
3rd	2,449	643	3,092 (9.94%)	2,483	646	3,129 (10.36%)
4th	1,152	252	1,404 (4.51%)	1,187	244	1,431 (4.74%)
5th	652	105	757 (2.43%)	599	95	694 (2.30%)
6th	291	40	331 (1.06%)	277	46	323 (1.07%)
7th	159	11	170 (0.55%)	144	16	160 (0.53%)
8th	60	10	70 (0.22%)	64	2	66 (0.22%)
9th	36	4	40 (0.13%)	26	3	29 (0.10%)
10th	8	0	8 (0.03%)	7	0	7 (0.02%)
11th	5	0	5 (0.02%)	3	0	3 (0.01%)
12th	2	0	2	2	0	2
13th	0	0	0	1	0	1
Total	22,966 (73.80%)	8,152 (26.20%)	31,118	22,311 (73.84%)	7,903 (26.16%)	30,214

It can be concluded by OWI Citation statistics that:

- 45% of OWI citations were issued to repeated offenders, and nearly 19% of all OWI citations involved the 3rd or more offense.
- Around 74% of drivers involved in an OWI offense were male.
- Only male drivers were involved in more than the 9th OWI offense in both 2016 and 2017.
- The male to female ratio for 1st OWI offense is 2.38:1 and 2.40:1 in 2016 and 2017, respectively.
- The male to female ratio for repeat offenders is 3.53:1 and 3.49:1 in 2016 and 2017, respectively. This indicates more male drivers were involved in repeat OWI offenses.

OWI Arrest Data

As noted earlier, the court system determines whether drivers issued traffic citations are convicted of a violation. A traffic conviction results from either a guilty plea or the court finding the person guilty. The "Arrest" data shows the pre-adjudication OWI statistics. The OWI Arrest data table provides OWI-related arrest counts for each police department in Wisconsin. This data table contains the name of the police department, arrest type and count of males and females involved in the corresponding arrest type. Note that the OWI Arrest dataset does not contain any court decisions. Table 4-10 describes the count of OWI arrests in Wisconsin by gender for different arrest types.

Table 4-10 OWI Arrest Distribution.

Charge Category	**2016**			**2017**		
	Male	Female	Total	Male	Female	Total
Criminal or felony OWI	66	11	77	91	17	108

Traffic OWI	19,923	7,087	27,010	19,647	7,004	26,651
Total	19,989	7,098	27,087	19,738	7,021	26,759

The OWI Arrest data table also provides arrest count by offense type (i.e., 1st, 2nd, 3rd, etc.). The distribution of arrest counts by offense type is presented in Figure 4-6, and the trend is very similar to OWI Citation data.

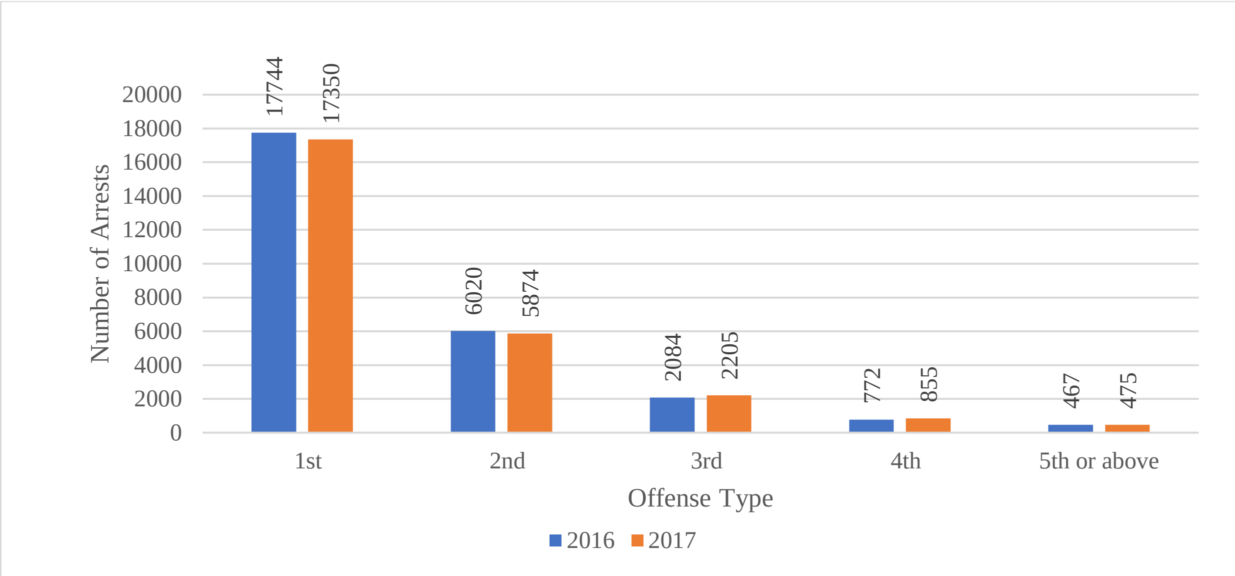


Figure 4-6 OWI Arrest Distribution by Offense Type.

It can be concluded by the OWI Arrest data statistics that:

- Less than 1% of arrests are categorized as criminal or felony OWI arrests in Wisconsin for both 2016 and 2017.
- Male drivers are more prone to drunk-driving, as more than 74% of OWI arrests involved male drivers.
- 35% of total OWI arrests in Wisconsin involved repeat offenders.
- The proportion of male drivers is higher for repeat offenses when compared with female drivers.

Liquor Law Violation

The Liquor law violation data table provides the count of liquor law violations (LLV) by gender for the Wisconsin court system. This data table contains the count for LLV-related traffic citations processed in Wisconsin courts by charges and gender for 2016 and 2017. The distribution of LLV by different charge type and by gender is presented in Table 4-11. The 2016-2017time period included 28,634 LLV violations: 60% involved underage alcohol, 8.3% were related to underage drinking, 3.3% were from underage alcohol operations and 2.1% were from intoxicants found in an underage person’s vehicle.

Table 4-11 Distribution of Liquor Law Violations by Charges.

Charge Code	Charge Description	2016			2017		
		Male	Female	Total (%)	Male	Female	Total (%)
IVO	Intoxicant in vehicle - operator	2069	697	2766 (18.76%)	2099	688	2787 (20.06%)
IVP	Intoxicant in vehicle - passenger	466	186	652 (4.42%)	410	186	596 (4.29%)
JA	Juvenile alcohol	676	523	1199 (8.13%)	679	505	1184 (8.52%)
JCS	Juvenile controlled substance	92	46	138 (0.94%)	79	30	109 (0.78%)
IIV	Intoxicants in vehicle carrying underage person	239	70	309 (2.10%)	213	87	300 (2.16%)
UAL	Underage alcohol	5903	3055	8958 (60.77%)	5428	2851	8279 (59.59%)
UAO	Underage alcohol operation	361	148	509 (3.45%)	317	121	438 (3.15%)
UID	Underage ID	142	68	210 (1.42%)	143	57	200 (1.44%)
Total		9948	4793	14741	9368	4525	13893

Inattentive/Distracted Driving

There is increasing evidence that driver inattention and distraction are major contributing factors in car and truck crashes. The problem is likely to escalate as more technologies are finding their ways into the vehicle. Driver inattention is defined as "*diminished attention to activities critical for safe driving in the absence of a competing activity*" (*16*). Distracted driving is any activity that diverts attention away from activities critical for safe driving towards a competing activity such as talking or texting, eating and drinking, talking to people in your vehicle, or fiddling with the stereo, entertainment, or navigation system. According to NHTSA, driver distraction is a specific type of driver inattention. Prior to 2010, Fatality Analysis Reporting System (FARS) was more inclusive of general inattentive behavior, whereas General Estimates System (GES) listed specific distracted-driving behaviors. In 2010, the two systems unified and coding in FARS was changed to "Yes-Distracted", "No-Not Distracted" or "Unknown if Distracted". Crashes involving distracted drivers during 2016 and 2017 in the United States killed 3,450 (9%) and 3,166 (8.53%) people, respectively (*17*).

In Wisconsin in 2016, 25,602 crashes - 19.84% of the total crashes - involved inattentive drivers and inattentive-related crashes caused 20.80% of all roadway fatalities – 109 lives lost – that year. At the national level, no state bans all cell phone use for all drivers. Seventeen states and Washington D.C., Puerto Rico, Guam, and the U.S. Virgin Islands prohibit all drivers from using hand-held cell phones while driving (*18*). In Wisconsin, total handheld device ban is applied while driving through work zone, on instruction permit, and with probationary driver license. Wisconsin state law also forbids driving "any motor vehicle while composing or sending an electronic text message or an electronic mail message" (*19*). Table 4-12 presents the distribution of inattentive driving and its subcategories. Note that the number of charges for “using cell phone while driving in work zone” increased 65 times from 2016 to 2017, a sign of probably strong work zone safety champions. The number of citations issued to commercial phone use while driving also is raised from 3 to 52 in a year. The total number of inattentive driving including all of its subcategories decreased by nearly 10% from 2016 to 2017.

Table 4-12 Distribution and Conviction Rate of Impairment related Citations.

Charge Description	**2016**			**2017**		
	Total	**% Male**	**Conviction Rate**	**Total**	**% Male**	**Conviction Rate**
Inattentive driving	10,122	58.18%	99.30%	8,897	58.94%	99.35%
Texting while driving	460	53.26%	98.91%	350	49.71%	99.14%
Using cell phone while driving in work zone	4	50.00%	100.00%	260	50.38%	99.23%
Using telephone while driving w/prob or IP	28	42.86%	100.00%	17	58.82%	100.00%
Commercial telephone use while driving	3	100.00%	100.00%	52	100.00%	100.00%
Grand Total	10,617	57.94%	99.28%	9,576	58.59%	99.34%

However, the small percentage made up by distracted driving such as texting or using a cell phone while driving can be misleading. In Wisconsin, inattentive driving (charge code: ID) is defined in Wisconsin State Legislature 346.89(1), (3)(a), (4), (5) as “*No person while driving a motor vehicle may be engaged or occupied with an activity, other than driving the vehicle, that interferes or reasonably appears to interfere with the person's ability to drive the vehicle safely*.” The texting while driving (charge code: ID), using cell phone while driving in work zone (charge code: IPW), using telephone while driving w/prob or IP (charge code: UTD) and commercial telephone use while driving (charge code: CTU) are defined as subsections of ID in Wisconsin State Legislature 346.89 in (3)(a), (4m), (4)(a) and (4)(b)2, respectively. As noted in Table 12, it is possible that the enforcement officer does not always differentiate between charge type (in subcategories under ID) as most of the citations were issued as ID in both 2016 and 2017.

Driver Data

The driver data tables contain a count of drivers by gender, class, and endorsement for 2,500 and 2,508 municipalities in 2016 and 2017, respectively. The data table contains driver’s license count by four license classes (“A”, “B”, “C” and “D”) and six license endorsements ("M", "H", "N", "P", "S" and "T"). According to Table 4-13, female drivers are more likely to have a regular license than male drivers. The table includes 831 municipalities but no county information. The data table does not contain city, town and village information, making it impossible to link driver counts with municipalities. Personal driver information (i.e., address, age) is also not available.

Table 4-13 Driver License Count by License Type in 2016.

PRODUCT	RGLR-REGULAR	PROB-PROBATIONARY	REGI-INSTRUCTION PERMIT	OCCL-OCCUPATIONAL	CYCI-INSTRUCTION PERMIT	JUVP-JUVENILE REST	MPDI-INSTRUCTION PERMIT	SPRI-INSTRUCTION PERMIT	SPRR-SPECIAL REST
MALE	2,020,480	81,937	15,886	471	5	1	1	2	7
FEMALE	2,035,838	77,236	21,282	90	0	0	1	1	2
"A"	199,181	0	0	0	0	0	0	0	0
"B"	299,144	0	0	0	0	0	0	0	0
"C"	305,388	0	0	0	0	0	0	0	0
"D"	4,056,284	159,166	37,168	561	0	1	2	3	9
"M"	536,273	1,644	0	77	5	0	0	0	0
"H"	86,331	0	0	0	0	0	0	0	0
"P"	65,755	0	0	0	0	0	0	0	0
"S"	48,686	0	0	0	0	0	0	0	0
"T"	49,073	0	0	0	0	0	0	0	0

4.4 Data Processing for Jurisdiction-Level Analysis

An in-depth crash analysis was conducted using collected citation attributes and crashes occurring in Wisconsin. Unlike roadway geometry and traffic variables, traffic violation data are available only at larger spatial units (e.g., court, police department). The availability of citation information for a large geographical unit is appropriate for safety analyses that occur on a macro level. The macro-level analysis of crashes evolved mainly to incorporate safety considerations within the transportation planning process and facilitate a proactive approach to assessing medium and long-term policy-based countermeasures.

The MV4000 crash database is used to identify crashes that occurred on the Wisconsin roadway network. Every record in the MV4000 database contains the geo-location of the crash along with its municipality and county name. To visualize the distribution of number of violations per capita and number of crashes per capita, a spatial distribution plot of total violations per capita and total crashes per capita at the county level is provided in Figure 4-7. Both total violations and total crashes per capita are estimated per 1,000 residents of each county.

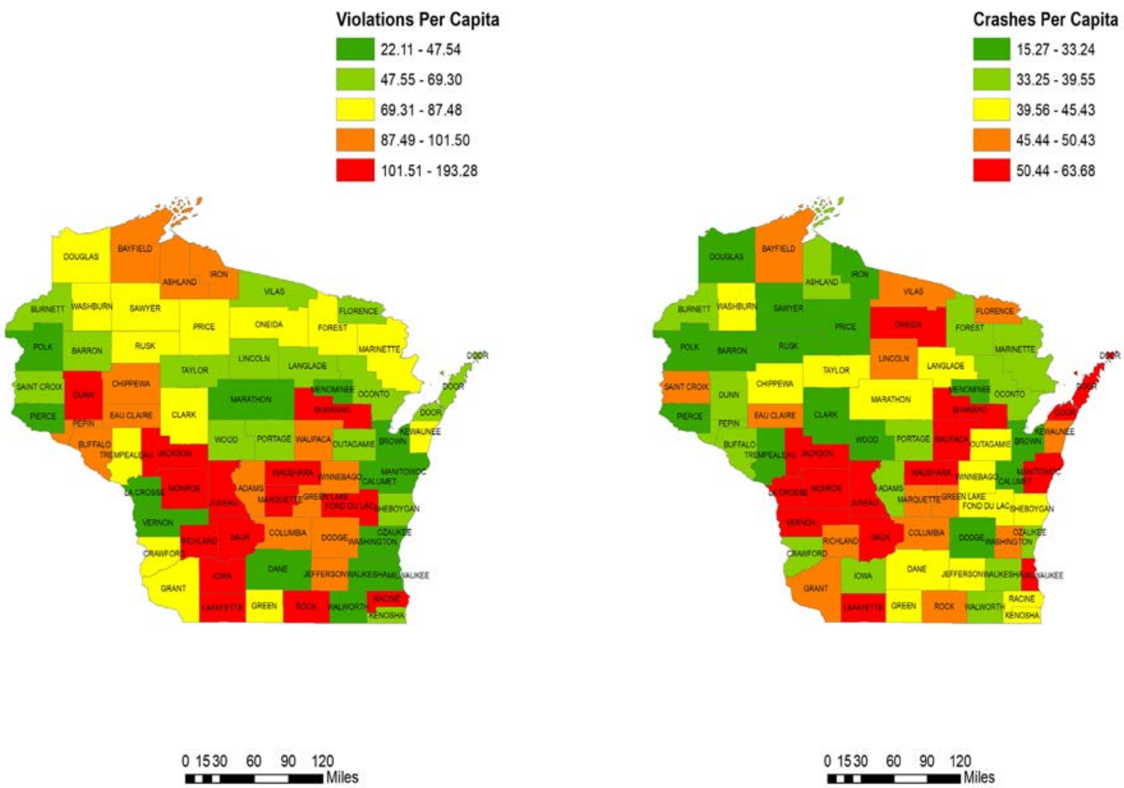


Figure 4-7 County-level Violations per Capita and Crashes per Capita.

In pursuit of a macro-level safety evaluation, "municipality" is used as a spatial unit to explore the effects of factors that contributed to crashes. The local government of a town, village, or city is defined as a municipality in Wis. Stats. 66.0621. A complete list of municipalities and their corresponding population estimates were collected from the Wisconsin demographic service center. Wisconsin has 1,854 municipalities; 190 units are "City", 412 units are "Village" and 1,292 units are designated as "Town". However, traffic citation information is available for each court/police department. The collected traffic citation data needs to be processed and linked to each municipality in order to conduct a macro-level safety evaluation. Table 4-14 provides an overview of the data and corresponding spatial units.

Table 4-14 Description of Collected Data Tables.

File Name	Designation	Spatial Unit	Available information
Final_Ests_MCD_2018	A	Municipality	Population
N1-5Drivers	B	Municipality	Count of different types of driver license
N6-Violations	C	Judicial Court	Adjudicated violations

N7-OWIArrests	D	Police department/ Sheriff's office	Count of OWI arrests
N8-LLViolations	E	Judicial Court	Adjudicated liquor law violations
N9a-Arrest	F	Police department/ Sheriff's office	Count of arrests by offense type (1st, 2nd, etc.)
N9b-OWICitations	G	Judicial Court	Adjudicated OWI citations

Several courts in Wisconsin process traffic citations issued by officers from different police departments: municipal court (MUN 242), circuit court (CIR 72), tribal court (TRI 3), federal court (FED 1) and administrative court (ADM 1). The figure in parenthesis represents the number of courts. Since municipal and circuit courts process most of the traffic citations in Wisconsin, tribal, federal and administrative courts were excluded for simplification and accuracy. Moreover, municipal courts have two different types in Wisconsin: those that serves only one municipality and those that serve multiple municipalities, also called a joint municipal court. Circuit courts are available for each county. Municipal courts administer most cases from urban or urbanized areas while circuit courts administer most cases from rural areas or municipalities with very low population. Considering very different driving behavior in urban and rural areas, the citation data for municipal courts and circuit courts were separately processed and analyzed. A flow-chart of data processing is provided in Figure 4-8.

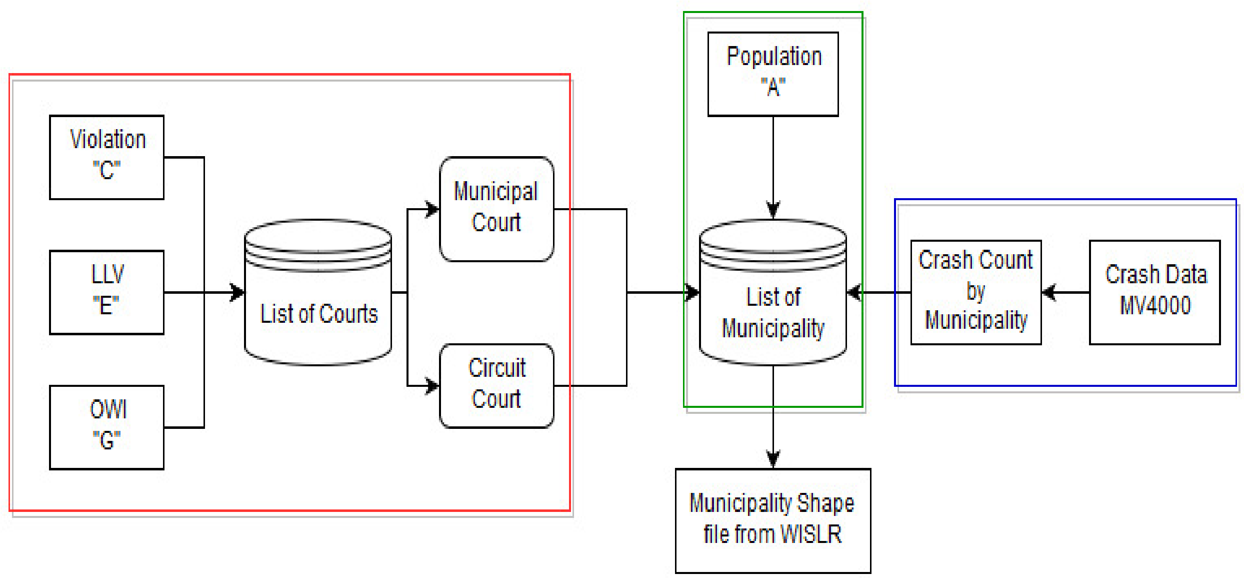


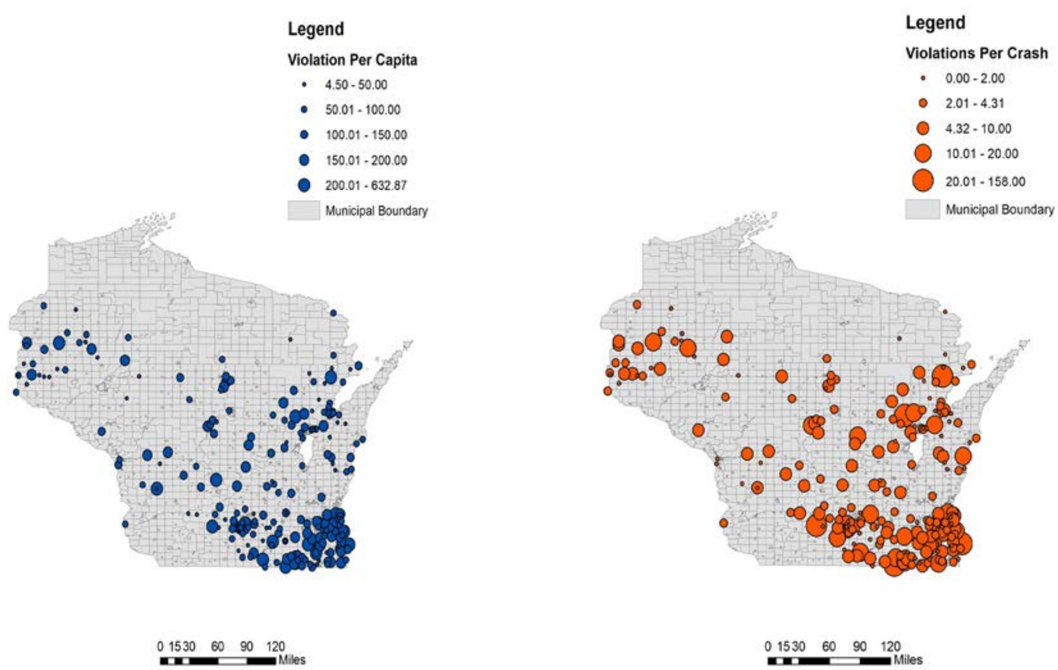
Figure 4-8 Data Processing Flowchart.

The list of joint municipal courts in Wisconsin was collected from the Wisconsin Municipal Court Directory 2016-17[6]. The steps for data processing are described below:

[6] https://www.wicourts.gov/contact/docs/muni.pdf

- Total Violation ("A"), Liquor Law Violation ("E") and OWI Citation ("G") data tables were processed jointly to generate a complete list of courts.
- All courts were then categorized as either municipal, joint municipal, or circuit.
- The population ("A") data table was used to generate a complete list of municipalities.
- Each municipal court was manually linked to its corresponding municipality based on municipality name, city/village/town information, and county.
- The total population of municipalities served by the corresponding joint municipal court was estimated. The violation counts were then distributed to each municipality based on the ratio of municipal population to the total population of the joint municipality.
- The total population of the corresponding county was estimated for circuit courts.

Although there are 242 municipal courts in Wisconsin, the above-mentioned data processing steps resulted in 418 municipalities with their own municipal courts or joint municipal courts where citations are processed. On the other hand, 72 circuit courts were joined to their corresponding county based on court names. Please note that traffic citations processed in municipal and circuit courts are issued by different police agencies. Municipal courts process traffic citations issued by municipal police (i.e., city, village or town police), whereas circuit courts process traffic citations issued by the highway patrol or county sheriff. Circuit courts may also process traffic citations issued by municipal police (or issued within the municipal police jurisdiction) if there is no municipal court in the nearby region. The violations per population and violations per crash for municipality and county are presented in Figure 4-9[7].



[7] A crash can be reported by either municipal police department, state patrol or county sheriff. Number of total crashes occurred within municipality is used to generate spatial graphs regardless to their reporting police department.

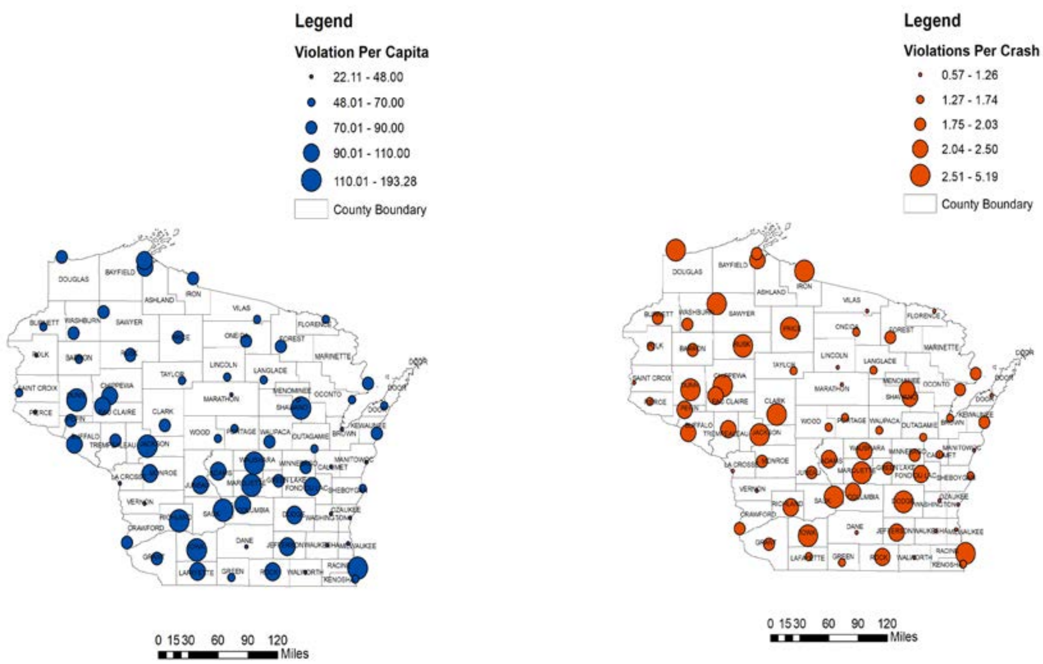


Figure 4-9 Violations based on Municipal and Circuit Court Locations.

The location distribution of municipal courts indicates that most of the municipal courts are located in the eastern region (Northeast and Southeast) and a portion of the Southwest region. The north region of Wisconsin has very few municipal courts. Both municipal and circuit courts show no visible spatial patterns for violation per capita. A similar conclusion can be drawn for violation per crash estimate. Thus, no spatial correlation was considered for further analysis. The summary statistics of collected variables are presented in Table 4-15.

Table 4-15 Variable Summary Statistics.

Variables	Municipality (418)				County (72)			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Population	9,426	33,394	175	595,555	80,778	135,130	4,258	950,381
City	19,702	56,658	996	595,555	-	-	-	-
Village	4,198	5,606	175	3,7574	-	-	-	-
Town	3,622	2,993	500	22,701	-	-	-	-
Violation Information								
Speed	209	517	0	6,761	1,325	1,490	40	8,412
Impaired	42	94	0	964	389	412	16	2,231
OWI	25	61	0	718	632	710	18	3,461
LLV	16	41	0	554	101	118	5	585
Inattentive	14	42	0	631	61	70	4	378

Traffic Rule	201	642	0	9,728	992	1,224	26	6,719
Male Count	557	1,730	0	29,882	3,242	3,621	106	19,823
Female Count	359	1,039	0	16,528	1,709	1,994	38	9,650
Total Violation	916	2,763	0	46,410	4,951	5,607	144	29,472
Crash Information								
Total Crash	363	1,771	0	33,221	3,496	6,505	65	49,123
Behavior	280	1,289	0	23,381	5,002	5,002	37	37,234
Speed	49	232	0	4,462	511	918	12	7,005
Impaired	16	61	0	1,068	164	258	5	1,861
Inattentive	58	227	0	4,024	502	937	11	6,796
Traffic Rule	154	715	0	12,611	1,213	2,780	8	20,394

4.5 Jurisdiction-Level Analysis

Relationship between Crashes and Citations – All Municipalities

Using processed municipal court data, a correlation matrix was generated between different traffic citation types and different crash types for 418 municipalities in Wisconsin. The correlation matrix in Table 4-16 shows that different types of violations are positively correlated and that different crash types are also highly correlated. The population of a municipality is strongly correlated with crashes and citations of different types except for liquor law violation.

Table 4-16 Correlation Matrix for Municipal Data.

	Population	**Speed Violation**	**OWI**	**LLV**	**ID Violation**	**Total Violation**	**Total Crash**	**Behavior Crash**	**Speed Crash**	**Impaired Crash**	**ID Crash**
Population	1.00										
Speed Violation	0.89	1.00									
OWI	0.81	0.80	1.00								
LLV	0.55	0.58	0.66	1.00							
ID Violation	0.68	0.80	0.71	0.74	1.00						
Total Violation	0.97	0.92	0.85	0.59	0.69	1.00					
Total Crash	0.98	0.84	0.77	0.50	0.60	0.95	1.00				
Behavior Crash	0.99	0.86	0.81	0.53	0.64	0.96	1.00	1.00			
Speed Crash	0.96	0.81	0.77	0.45	0.52	0.94	0.99	0.98	1.00		
Impaired Crash	0.99	0.88	0.85	0.56	0.68	0.96	0.98	0.99	0.97	1.00	
ID Crash	0.97	0.86	0.82	0.57	0.65	0.96	0.98	0.99	0.97	0.98	1.00

Apparently, at the municipal level, population for each municipality is a common contributor or exposure variable in both violation and crash counts. The relationship can be illustrated in the left side of Figure 4-10. As both traffic citations and crash counts have high correlation with population, both variables were normalized by population to further explore the relationship between violation and crash count. The actual confounding variables that underpin the high correlation between crashes and citations are driver's socioeconomic and demographic factors that are not available in the aggregated citation data, as illustrated in the right side of Figure 4-10. Ideally, these factors should be controlled in order to study the relationship between crashes and citations or included as the contributing factors to either crashes or citations.

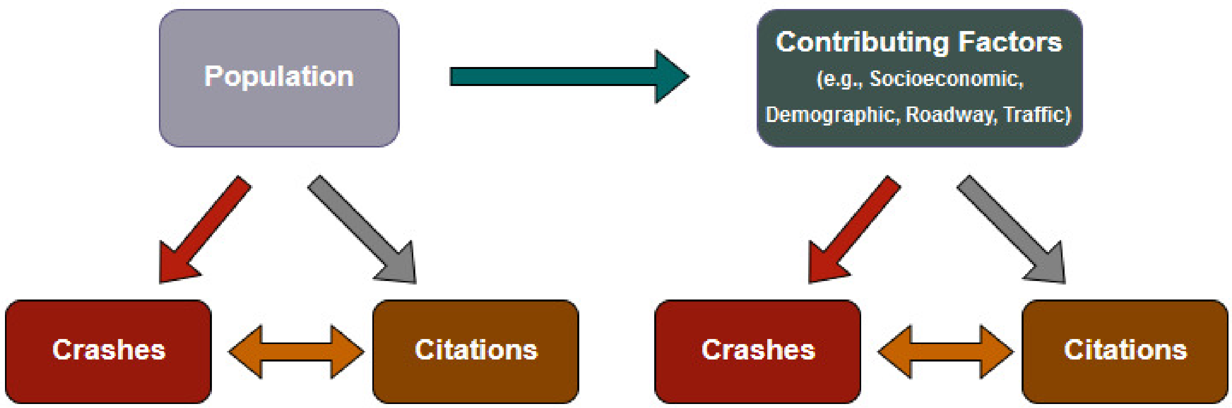


Figure 4-10 Confounding Effect between Population, Crashes and Citations

Figure 4-11 provides the scatterplot of total violation per capita and total crashes per capita for 418 municipalities in Wisconsin that served by a municipal court. In Figure 4-11, the scatterplot was generated using four quartiles of municipal population and each quartile was represented with different shape and color of point. The scatterplot helps visually explore whether municipalities in different quantiles of population can better explain the relationship between the estimates of total violations and total crashes per capita. Note that there is no apparent visual relationship between violations and total crashes per capita. Generating the scatterplot with different population quantiles also shows no specific trend between violations and total crashes per capita in different population quartiles.

Two hypotheses were used to further explore the relationship between total violations and total crashes after normalizing with population:

- **Hypothesis A:** A municipality's traffic violation per capita is directly proportional to its total crashes per capita; an increase in violation per capita will increase crash per capita. This hypothesis implies that violation per capita is a truthful reflection of safety culture.
- **Hypothesis B:** A municipality's traffic violation per capita is inversely proportional to its crash per capita; an increase in violation per capita will decrease the crash per capita estimate. This hypothesis implies that tougher and stricter enforcement (measured by the violation per capita) leads to better safety records, or a lower crash per capita.

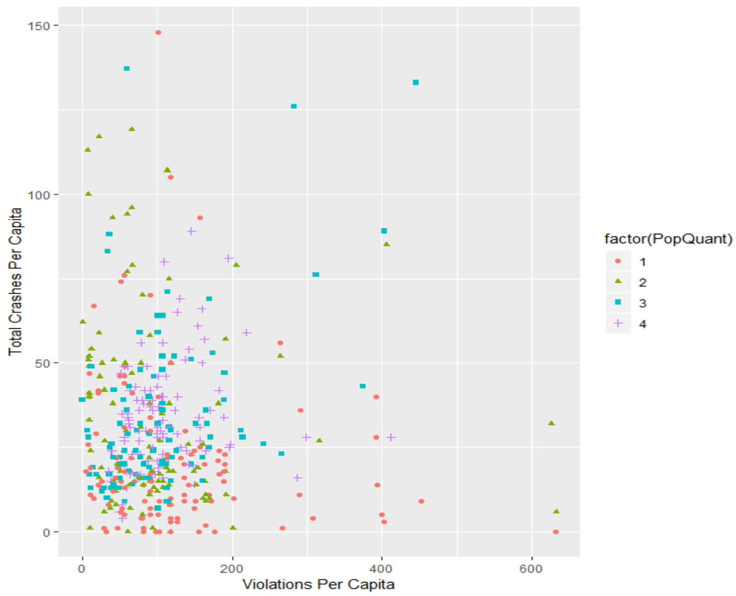


Figure 4-11 Scatterplot of Violation Per Capita and Crashes Per Capita for Municipalities.

Figure 4-12 shows how a mean-approach was used to evaluate the above hypotheses. The complete municipal dataset was categorized into four groups (A, B, C and D). The mean violations per capita and mean crashes per capita were estimated from the statewide average of municipalities which are served by a municipal court. Groups A and D represent Hypothesis A. Group A is comprised of municipalities that have both a violation and crash per capita that are less than the state mean. Group D is comprised of municipalities with a violation and crash per capita that are higher than the state mean. Groups B and C represent Hypothesis B. Group B municipalities have a violation per capita that is more than the state mean, while Group C's crash per capita is less than the state mean.

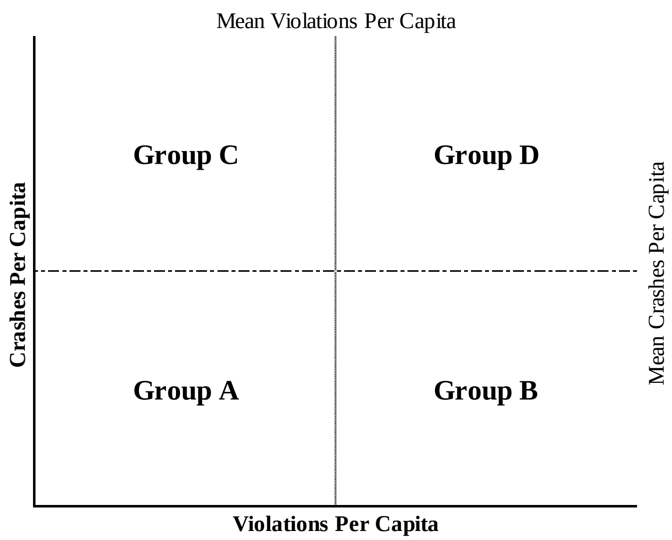
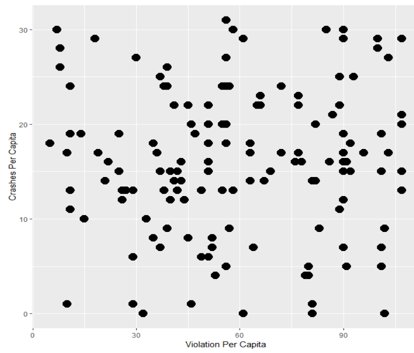
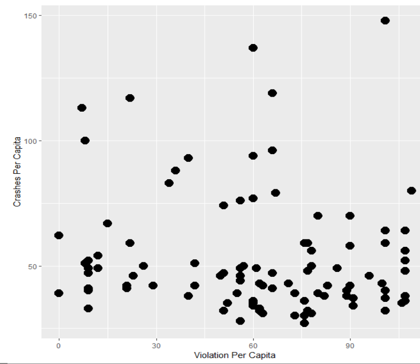
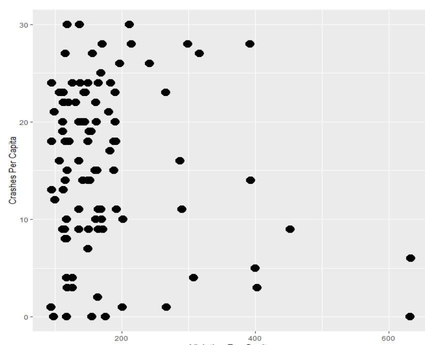
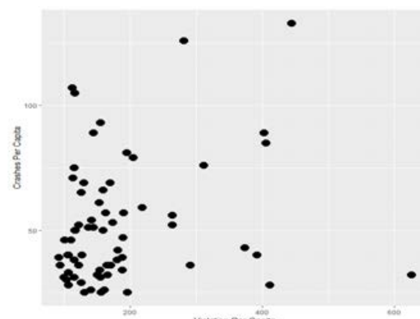


Figure 4-12 Grouping of Municipalities.

A cross-classification table was generated based on the above categories of municipalities. The statewide average is used as the mean to divide the space into four quadrants. The cross-classification table is presented in Table 4-17. Scatterplots of crashes per capita and violations per capita are created for each quadrant.

Table 4-17 Municipal-Level Violation Per Capita and Crash Per Capita.

		Crashes per capita		Total
		Less than Mean	More than Mean	
Violations per capita	Less than Mean	Count (%): 159 (38.04%) Mean population: 5,738 **A** C/V/T: 57/85/17 	Count (%): 105 (25.12%) Mean population: 17,492 **C** C/V/T: 31/17/57 	**264**
	More than Mean	Count (%): 101 (24.16%) Mean population: 3,671 **B** C/V/T: 27/64/10 	Count (%): 53 (12.68%) Mean population: 12,965 **D** C/V/T: 21/18/14 	**154**
Total		**260**	**158**	**418**

The count of municipalities described in Table 4-17 shows that the mean population is significantly different among the groups. However, counts of municipalities are almost equally divided between groups representing Hypothesis A and B. The equal dividend suggests that violation per capita may have a mixed effect on crash occurrences. The number of municipality count by C/V/T does not support hypothesis A or B. The scatterplot between violation per capita and crash per capita for each group also indicates that there is no significant relationship between violations and crash occurrences.

Relationship between Crashes and Citations – All Counties

Following a similar approach used for municipal court data, a correlation matrix was generated in Table 4-18 with the violation data collected from circuit courts. Similar to the correlation matrix in Table 4-16, different types of violations are positively correlated and that different crash types are also highly correlated. The population of each county, again is highly correlated with both crashes and citations except for liquor law violations.

Table 4-18 Correlation Matrix for Circuit Court Data.

	Population	**Speed Violation**	**OWI**	**LLV**	**ID**	**Total Violation**	**Speed Crash**	**Impaired Crash**	**ID Crash**	**Total Crash**
Population	1.00									
Speed Violation	0.84	1.00								
OWI	0.77	0.88	1.00							
LLV	0.38	0.55	0.64	1.00						
ID	0.79	0.80	0.86	0.67	1.00					
Total Violation	0.85	0.95	0.98	0.64	0.88	1.00				
Speed Crash	0.98	0.81	0.75	0.34	0.79	0.81	1.00			
Impaired Crash	0.99	0.86	0.81	0.41	0.82	0.87	0.98	1.00		
ID Crash	0.99	0.83	0.78	0.39	0.81	0.84	0.99	0.99	1.00	
Total Crash	0.99	0.83	0.76	0.35	0.79	0.83	0.99	0.99	0.99	1.00

Table 4-18 shows that population is a major contributor to both county-level violation counts and crash counts. A cross-classification table was generated for Wisconsin counties based on the categories described in Figure 4-12 and is presented in Table 4-19. The statewide average is used as the mean to divide the space of counties into four quadrants.

Similar to the conclusion drawn from the municipal court data exploration, Table 4-19 shows that the counts of counties are almost equally divided between groups representing Hypothesis A and B. The equal dividend suggests that the violation rate may have a mixed effect on crash occurrences. The only difference between different municipality groups was found in the mean population estimate for each group. The mean population is significantly different between groups, which indicates that population may be more correlated with both crash and violation counts. Following a similar trend, it can also be concluded that the violation per capita estimate by behavior type also does not have any significant relationship with the estimate of corresponding crashes per capita.

Table 4-19 County-Level Violation Per Capita and Crash Per Capita.

		Crashes per capita		Total
		Less than Mean	More than Mean	
Violations per capita	Less than Mean	A Count (%): 20 (27.78%) Mean population: 70634	C Count (%): 16 (22.22%) Mean population: 165320	36
	More than Mean	B Count (%): 14 (19.44%) Mean population: 34413	D Count (%): 22 (30.56%) Mean population: 58021	36
	Total	34	38	72

Behavior-based Crashes and Citations Per capita - All Municipalities

Although no distinct pattern exists between total violations per capita and total crashes per capita, it is important to explore the relationship between violations and crashes by behavioral attributes. As noted earlier, traffic violations are categorized into speed, impairment and inattentive driving violations based on the description of each violation. On the other hand, the crash database contains driver-related attributes that can be used to categorize crashes as being related either to speed, impairment or inattentive driving. Figure 4-13 shows a scatterplot of behavior-based crashes per capita (i.e., speed-related, impairment-related, inattentive and distracted driving (ID)-related) with corresponding violation type which was generated using violation data from 418 municipalities.

Figure 4-13, which illustrates the distribution of behavioral crashes per capita and their corresponding violation per capita, shows there is no significant relationship between behavioral crashes per capita and violations per capita. The distribution of behavioral crashes per capita with respect to violations by behavioral types per capita also varies by behavior type (i.e., speeding, impairment, inattentive driving). Thus, violations per capita by behavioral type cannot be used to effectively quantify or predict crashes per capita by behavioral type.

The above examination of municipal and circuit court data shows that violation per capita and crash per capita estimates are not statistically correlated. Further exploration using behavior-based traffic citation and crash count also indicated that violation and crash counts do not have any statistical correlation after controlling for population. According to a meta-analysis of 99 road safety studies conducted by Barraclough et al., the average correlation between crashes and traffic violations is only 0.18, suggesting traffic violations are a limited safety proxy (*20*). There are many factors that can influence traffic violations in terms of driving behavior and police enforcement. Unfortunately, police enforcement information was not available to the research team, thus the effect of police enforcement of traffic violations and in-turns on crash occurrence was not explored. Shawky et al. noted that driver's socioeconomic and demographic characteristics such as age, gender, education level, poverty level, nationality, vehicle ownership can influence number of citations received by driver and in-turns has a significant impact on the value of crash rate using data collected in Emirates of Abu-Dhabi, UAE (*21*). In current data collection, driver's personal information such as physical address, points on driver record, socioeconomic and demographic characteristics, etc., were not available. Thus, further exploration using driver-specific information is needed to define a relationship between traffic violation and crash occurrence.

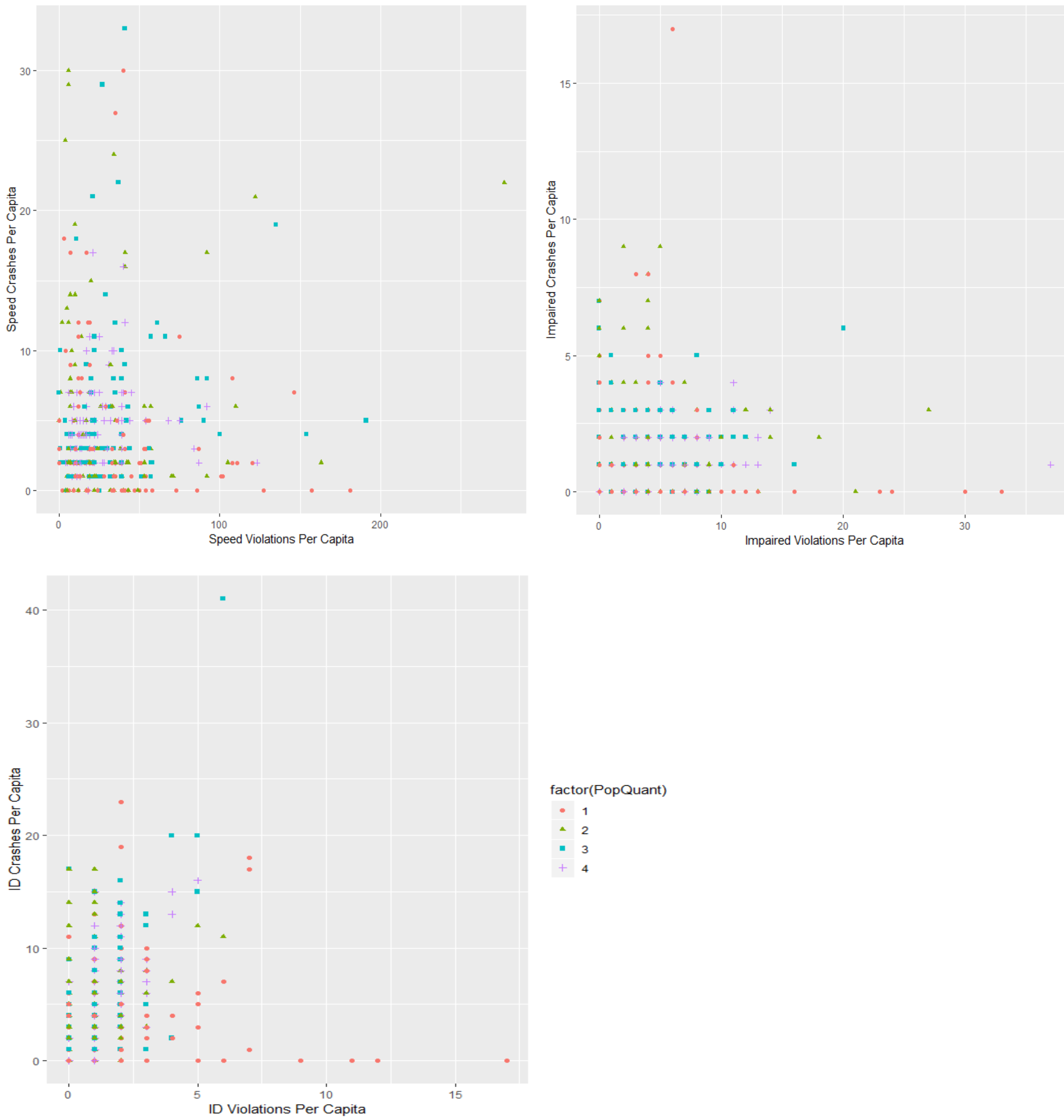


Figure 4-13 Behavior related Violations/Capita and Behavior Related Crashes/Capita.

A Tale of Four Cities

Among 418 municipalities, four municipalities with high population in different geographic locations – City of Milwaukee, City of Madison, City of Green Bay and City of La Crosse – were compared to examine how different attributes of municipalities and violations are associated with crashes. Law enforcement data for all four municipalities were collected from the FBI's 2015 Crime Statistics Report (*22*). The site comparison is described in Table 4-20.

Table 4-20 Site Comparison with Highly Populated Municipalities in Wisconsin (2016-2017).

Particulars	Unit	Milwaukee	Madison	Green Bay	La Crosse
C/V/T		C	C	C	C
Population	Person	595,555	252,546	105,477	52,282
Area	Sq. Miles	96.48	93.50	46.09	22.61
Population Density	Persons/Sq. Miles	6,173	2,701	2,288	2,312
Drivers Per Capita	Per 1000 Population	0.59	0.46	0.68	0.50
Total Crashes	Count	33,221	10,743	1,902	3,400
Municipal Crashes [1]	Count	27,325	10,220	1,837	3,363
County Crashes [1]	Count	5,896	523	65	37
Behavior-related Crashes	Count	23,381 (70.38%)	8,735 (81.31%)	1,673 (87.96%)	2,957 (86.97%)
Speed	Count	4,462	928	286	321
Impairment	Count	1,068	454	186	135
Inattentive	Count	4,024	1,462	336	834
Traffic rule	Count	12,611	5,393	904	1,723
Total Citations	Count	46,410	17,844	10,031	6,575
Speed	Count	6,761	5,953	2,054	894
Impairment	Count	964	636	636	758
OWI	Count	718	356	380	204
LLV	Count	246	278	257	554
ID	Count	226	631	300	221
Traffic	Count	9,728	5,992	1,888	1,249
Crashes Per Capita	Per 1000 Population	55.78	42.54	18.03	65.03
Citations Per Capita	Per 1000 Population	78	71	95	126

Citations/Crash	Value	1.40	1.66	5.27	1.93
Speed Citations/Crash	Value	1.52	6.41	7.18	2.79
Impairment Citations/Crash	Value	0.90	1.40	3.42	5.61
ID Citations/Crash	Value	0.06	0.43	0.89	0.26
Police Officer	Count	2518	589	227	107
Police Officers Per Capita	Value	4.23	2.33	2.15	2.05
Citations/Police Officer	Value	18.43	30.30	44.19	61.45
Crashes/Police Officer	Value	13.19	18.24	8.83	31.78

Note: 1) Crashes were reported by city/village/town police departments, county sheriff's office and state patrol.

According to Table 4-20, La Crosse has the highest number of crashes per capita, and Green Bay has the lowest number. Citations per capita also is highest in La Crosse and lowest for Green Bay. Law enforcement officers per capita is highest in Milwaukee and lowest in La Crosse. The number of citations per police officer is highest for La Crosse and lowest for Milwaukee. The number of crashes per police officer is the highest in La Crosse and lowest in Green Bay. The number of citations per crash is highest for Green Bay and lowest for Milwaukee. The higher value of violation-to-crash ratio indicates that more citations are issued for non-crashes. The violation to crash ratio for Green Bay is 5.27, indicating that a significant number of citations were issued without a crash event to strictly enforce traffic laws. Hence, a higher violation to crash ratio led to the lowest crash per capita estimate in the city of Green Bay. La Crosse, on the other hand, issues the most citations per police officer, which is the most citations per capita, but it also has the most crashes per capita. The differences between Green Bay and La Crosse suggest that connections between citations, police officers and crashes are complex and that additional analysis is needed to explore possible causal relations.

4.6 Summary and Recommendations

Traffic violations have long been considered to be a proxy variable for unsafe driving behavior that is directly responsible for or associated with traffic crashes. The number of traffic citations issued also depends on the intensity and frequency of enforcement activities. Hence, the relationship between traffic violations and crashes can be complicated. In this regard, a comprehensive analysis of citation data and its association with reportable crashes at the municipality level and county level has been performed for the most common risky driving behaviors in Wisconsin.

Following the NHTSA's definition of risky driver behavior (e.g., drunk driving, drug-impaired driving, distracted driving, speeding, seat belts, and drowsy driving), Wisconsin's citations were re-categorized and summarized. The numbers of citations issued for speed, impairment, violating traffic rules, inattentive which are the first four rows of Table 4-3 add up to more than 419,000 per year when there are approximately 4.23 million licensed drivers. The high number of traffic violation corroborates the fact that the percentage of fatal crashes in Wisconsin is consistently higher than the national average in almost all categories of risky driving behavior.

Among all violations, nearly 45% of the OWI citations are for repeat offenders and nearly 19% of all OWI citations are for the 3rd or more offense. Similarly, 35% of OWI arrests are repeat offenders. It is interesting to find that the court decision on an OWI and a PAC conviction can be very different. The conviction rate is near 100% for OWIs in 2017 while the conviction rate for PACs is merely 18.5%. The large difference in conviction rate suggests that a driver either fought to be acquitted from PAC or is convicted of a lesser charge.

A comparison of speed-related fatalities in Wisconsin and those of the U.S. from 2011 to 2016 shows that Wisconsin's figure is consistently above the national average. In 2017, 35% of the fatalities in Wisconsin are speed-related, whereas the national average is 27%. Only 65% of the speed-related crashes in Wisconsin are issued with a citation. A further comparison between crashes with citation(s) and different types of behavior-related crashes indicates that traffic citations may be under-issued for all types of behavior-related crashes. From 2011 to 2017, driver-related primary contributing factors (PCC) are cited in more than 71% crashes, whereas less than 50% of the crashes involving driver-related PCCs have citations.

Because of different coding systems for inattentive driving and distracted driving between Wisconsin and NHTSA, they cannot be directly compared. In Wisconsin, any activity while driving a motor vehicle "interferes" with the person's ability to safely operate the vehicle is treated as "inattentive driving". That said, the Wisconsin percentage of inattentive-related fatal crashes is 21%. The national percentage of distracted-related fatal crashes is 10% or lower.

According to the eight-year trend of citation and crash count, the yearly percent change in crashes does not seem to be in accordance with the percent change of total citations issued over time. The total number of crashes has an overall ascending trend with a significant increase in 2013 and 2016. The total number of citations has a 14.59% reduction in 2013, and the trend is relatively flat before and after that. This may imply some external crash contributing factors such as enforcement discretion.

The association of traffic violation counts and crash counts at the municipality level was explored in an attempt to quantify the effect of traffic violation on crash occurrence. The correlation matrix indicates that the population is a common exposure contributor to both violations and crashes. Next, violation and crash data were normalized by population and the analysis was repeated. The result shows that traffic violations per capita is not statistically correlated with crashes per capita. A further investigation for different risky driver behaviors also returned weak association between violations per capita and crashes per capita. This is not a surprise because at an aggregated level such as municipality, both traffic violation counts and crash counts can be influenced by a series of factors such as driver's age, socioeconomic and demographic characteristics, police enforcement in different manners. Certain at-risk driver population groups may exhibit more explicit patterns than others. Without driver-specific information, the relationship between violations and crashes cannot be clearly defined and measured.

Crashes are preventable. Recognizing "to err is human", there is a need to continue to improve the systems, processes, and conditions to deliver a safer traveling environment and lead people to make fewer mistakes. This is in accordance with the four principles underpinning the Safe System (http://www.towardszerofoundation.org/thesafesystem/):

- human fallibility,
- human vulnerability,
- road safety is a shared responsibility, and
- building a safe and forgiving road system.

Hence, the following recommendations are made from the analysis of traffic citation data:

1) Provide better access for safety professionals and researchers to traffic citation data. Like crash data, traffic citations contain valuable spatial, temporal and environmental factors pertaining to risky driving. Citations also contain driver information and offense types. Analyzing detailed citation data can help us to characterize the attributes of at-risk driver population (not personal identity), discover spatial and temporal patterns of violations, and identify potential highway design and operational issues that discourage driver compliance.
2) Link citation data with other data sources. By linking data from different sources such as crashes, citations, census, roadway and traffic, weather, and work zones, we can exploit synergistic factors of crashes, support effective and targeted enforcement and act proactively and appropriately. A linkage can be established based on the increasingly available location information and technologies at WisDOT.
3) Explore innovative ways to address resource shortage for law enforcement. The tightening resource, fewer law enforcement officers, and continuously expanding public roads attribute in part to inconsistent enforcement and limited coverage. New approaches such as alternative (or automated) enforcement, grant application assistance, coordinated enforcement, and internal training can be actively explored.
4) Raise the safety performance standards and expectations. Maintaining a high standard is vital for continued safety improvement. This can be done through the regular updates and modifications of SHSP, and through the leadership and authoritative recommendations from the Traffic Safety Council as the safety oversight committee at WisDOT.
5) Continue to use data, research and evaluation to understand crashes and risks. We need to promote research to study best practices and create tools for identifying and ranking safety hot spots; developing and evaluating countermeasures; and allocating resources that are essential for safety investment.

6) Of course, the driver must be held responsible for their actions. Hence, we should continue to consider and evaluate legislations and interventions with proven success in targeting risky driving behavior-related crashes, such as stricter driving laws for OWI repeat offenders.

4.7 Limitations

While the exploration of traffic violation in Wisconsin has provided an in-depth insight into risky driving behavior, this study has some data constraints. The limitations of this study are as follows:

1) While total yearly statewide traffic citation count is available in the WisDOT “Crash Fact sheet”, querying citation database for specific data items takes time and effort. Hence, only two-year (2016 and 2017) traffic violation count data disaggregated by gender and adjudicating court was used in the study.
2) As noted in the data description, the court decision (e.g. guilty, dismissed, error, appealed, etc.) for each violation is only available in the “All Violation” data table.
3) Location information for citation is not available and location-specific analysis cannot be performed.
4) Driver specific information is not available in any of the five traffic violation data tables. Thus, an in-depth investigation on repeated offenders or other risky driving-related violations cannot be conducted.
5) Detailed enforcement information by municipality or county is not available. The impact of enforcement activities on citations and crashes cannot be evaluated.

5. AREA-BASED CRASH PREDICTION MODEL (CPM)

5.1 Introduction

Based on the exploratory analysis with traffic citation data, it was noted that the population variable has a better correlation with crash data compared with traffic citation data. This finding, however, is not enlightening because as an exposure measure, population does not provide theoretical support to explain why crashes happened; and thus, yields no meaningful solutions to reduce them. In fact, it is the socioeconomic and demographic characteristics of the residents that contribute to crash occurrence. Traffic citation data analysis needs be extended to a multivariable crash analyses in which crash prediction models (CPMs) are used to relate traffic crashes aggregated by a specific spatial scale to area-level factors such as socioeconomic status, demographic characteristics, land use, and traffic patterns.

CPMs can help agencies to incorporate safety considerations in the long term transportation planning process (*23*) but the selection of a spatial unit is an important element of developing a useful CPM. A wide array of spatial units have been employed, such as regions (*24*), counties (*25*), zip codes (*26*), census tracts (*27*), block groups (*28*), and traffic analysis zones (*29*). Studies related to macro-level CPMs most commonly involve aggregate CPMs that have been developed to relate roadway crashes to a variety of explanatory factors, including road network composition, traffic patterns, and area-level demographic and socioeconomic characteristics. Among spatial units explored in the literature, variables related to socioeconomic, demographic, and traffic patterns are readily available at the census tract level from the U.S. Census. For this pilot macro-level model development, census tract was chosen as the spatial unit.

The objective of this section is to investigate the key contributors and their effects on various driver behavior-related crashes as well as all other crashes. The negative binomial modeling approach was used to develop CPMs for all crashes, for speed-related crashes, and for alcohol-related crashes. The parameter estimates of the developed models will shed light on the effects of covariates on different crash types based on driver behavior. The model outputs can be used to identify communities with a higher crash risk and help agencies develop more informative and cost-effective countermeasures.

5.2 Data Processing and Exploratory Analysis

The literature shows that driver error is one major type of factor contributing to crashes (*30*). In macro-level CPMs, the crash data and covariates are aggregated to a spatial unit, which in turn makes it nearly impossible to incorporate specific driver factors into CPMs. The covariates may have different effects on different driver behaviors. The development of separate CPMs for only driver behavior-related crashes is one way of exploring the effects of driver behavior on area-level CPMs. Crash reports note that speeding is one of the main driver-related factors that contribute to crashes. In 2014, 9,262 of the total 32,675 driving-related fatalities in the United States were due to speeding (*31*). Alcohol-impaired driving is another driver-related error that causes many crashes. According to the 2015 Wisconsin Fatal Crash Trend analysis, 34 percent of all fatal crashes that occurred in Wisconsin involved alcohol-impaired driving (*32*). The development of a behavior-based CPM along with a CPM for all crashes can test the hypothesis that covariates may have different effects on different crash types based on driver behavior.

Census tract has been used as a spatial unit for developing area-based crash frequency prediction models. The census tract information from Wisconsin has been collected by the U.S. Census. The analysis in this study used 2015 TIGER/Line Shapefiles data from the U.S. Census. The 2015 TIGER/Line dataset for census tract contains 28 separate data tables with 17,812 attributes total. All data tables can be integrated using a unique census tract identification number. Based on the literature, a series of attributes were selected to explore as covariates in a crash frequency prediction model for census tract. ArcMap was used to join selected attributes to each census tract. Data were processed further based on variable definitions in order to normalize the covariates. The final dataset was represented in percentage format.

The roadway network-related attributes in each census tract were obtained from WISLR since they were not available in the TIGER/Line database. The WISLR database contains roadway and traffic-related information, including the geographical location of the roadway, for all roadway networks in Wisconsin. The WISLR dataset was spatially joined with census tract using ArcMap to obtain the total roadway length, AADT, and the total number of intersections within each census tract.

Most analyses in this report use police-reported crash data from 2011 to 2015. Crashes that occurred on Wisconsin roadways from 2011 to 2015 were collected from the MV4000 dataset and were processed to develop area-level CPMs. The effect of human behavior was explored by extracting two subsets of crash data – speed-related and alcohol-related – from the all crash dataset based on the human factor related to each crash occurrence. The MV4000 dataset contains flags for each crash type. A speed flag and alcohol flag was used to extract the subset of each crash type. Once crash data were collected, the crashes need to be linked with the census tract based on location information; however, crashes may not always occur within the defined census tract boundary. When existing roadways are used for defining government boundaries, a portion of crashes occurred on these census tract boundaries. A major challenge is joining crashes that occur on the census tract boundary, or the “boundary collision issue”. Researchers have developed several methods for properly distributing boundary crashes among corresponding census tracts. A list of available methods is provided below:

- Equal proportion: Proportioning the crash based on the number of adjacent spatial unit
- Geo-processing methods: Data attributes aggregated for each spatial unit as they were geo-coded in ArcGIS (Wei, 2010)
- One-to-one method: Each spatial unit forming the boundary is assigned one whole collision) (Wei, 2010)
- Vehicle Kilometer Traveled (VKT) Proportion: Proportioning the boundary crashes based on the value of VKT of corresponding spatial units

- Total Lane Kilometers (TLK) Proportion: Same as above, but with measured “total lane kilometers”
- Density Probability: Aggregation of boundary collisions by density probability ratio (Cui et al. 2015)

The equal proportion method was used in this study for joining boundary crashes with the census tract. The following steps were completed in ArcMap to filter boundary crashes and join them with the census tract using equal proportion method:

1. Convert polygon shapefile of census tract to line features.
2. Create point shapefile of all crashes using location attributes (longitude and latitude) available in the MV4000 database.
3. Use “Select by location” tool to select crashes occurred within a specified distance from census tract line shapefile. In this study, a distance of 30 meters was used.
4. Based on selected crashes from “Select by location” tool, create two separate shapefiles for crashes by splitting them: “Crashes on Boundary” and “Crashes within Boundary”.
5. Use “Spatial Join” tool with join option as “One-to-One” to count number of crashes occurred with each census tract from “Crashes within Boundary” file.
6. Use “Spatial Join” tool with join option as “One-to-Many” with “Crashes on Boundary” file to count the number of zones related to each crash occurred on census boundary.
7. If a crash occurred on the boundary of “n” census tracts, split the crash value to “1/n” in each census tract.
8. Sum joining results from step 5 and step 7 to obtain total crashes in a census tract.

Table 5-1 Summary Statistics of Wisconsin Census Tract Data.

Variable	Unit	Mean	Standard Deviation	Minimum	Maximum
Total Crash	Count	351.505	259.586	4.500	2193.167
Speed-related Crash	Count	63.01	53.24	0	451.33
Alcohol-related Crash	Count	16.88	10.97	0	102.5
Area	Sq. Mile	41.1	76.471	0.068	799.8
Roadway Length	Miles	166.4	218.288	0.112	2242.551
VMT	Veh-mile	80972.3	61319	5.600	553114.024
Number of Intersections	Count	248.3	189.04	0.000	1664
Population Density	Count/Sq. Mile	2919.1	4657.4	0.000	50428.739
Male	%	0.495	0.042	0.000	1.00
White	%	0.841	0.228	0.000	1.00
Proportion w/Age <18	%	0.226	0.064	0	0.49
Proportion w/Age >64	%	0.150	0.063	0	0.53
Median Age	Years	39.52	7.646	0.000	67.5

Enrolled in school	%	0.264	0.091	0.000	0.970
Primary work commute mode (Car)	%	0.878	0.107	0.000	1.000
Primary work commute mode (Public Transit)	%	0.027	0.056	0.000	0.653
Primary work commute mode (Bicycle)	%	0.008	0.018	0.000	0.212
Primary work commute mode (Walk)	%	0.035	0.057	0.000	0.613
Median Income	USD	53744.2	19695	0.000	156250
Below Poverty	%	0.146	0.126	0.000	0.864
Less_High School	%	0.083	0.077	0.000	0.540
High School Degree	%	0.297	0.103	0.000	0.571
College degree	%	0.335	0.070	0.000	0.558
Bachelor degree	%	0.284	0.168	0.000	0.928
Unemployment	%	0.337	0.086	0.105	1.000
Number of vehicles	Count	4501.393	2403.197	0.000	19880
Number of Bars	Count	2.187	2.572	0.000	23.000

Table 5-1 shows the summary statistics of the processed dataset used to develop area-level CPMs. The explanatory variables presented as percentages were calculated from information provided by the TIGER/Line dataset. Vehicle Miles Travelled (VMT) and number of intersections are considered to be roadway information in the dataset. All other variables are extracted and calculated from the TIGER/Line dataset. Please note that all explanatory variables in the final dataset are continuous variables. Categorical variables were not generated for this pilot run analysis.

Three NB models were developed with the processed dataset to quantify the effects of explanatory variables on total crashes, speed-related crashes, and alcohol-related crashes in a census tract. The Variance Inflation Factor (VIF) was estimated for each model to check multicollinearity. Any covariate with a VIF value greater than 5 were excluded from final models. Table 5-2 provides the summary of model coefficient estimates of area-level crash prediction models. The negative binomial model uses the following equation:

$$Y_i = exp\,(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_j X_{ji})$$

where:
Y_i = total number of reported crashes in census tract i from 2011 to 2015
X_{ji} = quantitative measure of each characteristic j associated with census tract i
β_j = coefficient corresponding to X_{ji} to be determined by negative binomial regression
β_0 = constant to be determined by negative binomial regression

Table 5-2 Parameter Estimate Summary for Area-level Crash Prediction Model.

Variable Category	All Crashes	Speed-related Crashes	Alcohol-related Crashes
Intercept	Intercept (0.179)	Intercept (-3.044)	Intercept (-3.001)
Traffic and Trip	Log(VMT) (0.544) No. of intersections (1.285E-03) Car Trips (-0.454)	Log(VMT) (0.601) No. of intersections (1.41E-03)	Log(VMT) (0.434) No. of intersections (9.692E-04) Car Trips (-0.385)
Demographic Variables	Area (-2.815E-03) Percent Male (0.978) Male w/age <18 (-0.849) Percent White (-0.449) Median Age (-8.427E-03)	Area (-2.065E-03) Percent Land (0.427) Population Density (-3.081E-05) Percent Male (1.039) Per. Male <18 yrs (-1.458) Percent White (-0.563) Median Age (-8.409E-03)	Area (0.001496) Percent Land (-0.316) Population Density (-9.108E-06) Percent Male (1.556) Per. Male <18 yrs (-7.723E-03) Percent White (0.408) Median Age (-7.718E-03)
Socioeconomic Variables	Median Income (-5.207E-06) In Labor Force (0.846)	Less_High_School (0.724) In labor force (0.56)	Median Income (-5.998E-06) Less_High_School (1.915) In labor force (1.234) Housing (7.154E-05)

Table 5-3 illustrates the effects and comparisons of crash contributing factors between different crash types. The sign provided in the parenthesis indicates a positive or negative effect of the contributing factors on different types of crash occurrences.

Table 5-3 Potential Contributing Factors in Area-level Crash Occurrences.

Category	Total Crashes	Speed-related	Alcohol-related
Traffic Variable and Trip Pattern	• VMT (+) • No. of Intersections (+)	• VMT (+) • No. of Intersections (+)	• VMT (+) • No. of Intersections (+)
Travel Pattern	• Car Trip (-)	• Car Trip (-)	
Demographic Variable	• Area (-) • Percent White (-) • Age<18 (-) • Number of Vehicles (+) • Bar Count (+)	• Area (-) • Population Density (-) • Percent Male (+) • Percent White (-) • Age<18 (-)	• Area (-) • Population Density (-) • Percent White (+) • Age<18 (-) • Age>64 (-) • Number of Vehicles (+) • Bar Count (+)
Socioeconomic Variable	• Median Income (-) • Education less than High School Percentage (+) • Unemployment rate (-)	• Education less than High School (+) • Unemployment rate (-)	• Median Income (-) • Education less than High School (+)

5.3 Findings

The CPM results show that roadway, travel pattern, socioeconomic, and demographic variables were statistically significant in predicting total crashes and behavior-related crashes in a census tract. VMT and intersection density were both statistically significant in predicting total crashes and behavior-related crashes. The parameter estimates of the roadway network-related variables are positive, meaning that an increase in any variable will increase the total number of crashes in a census tract overall, as well as the total number of speed-related crashes and alcohol-related crashes, specifically. The percentage of car trips is statistically significant in predicting all crashes and speed-related crashes but is not significant in predicting alcohol-related crashes; this indicates that alcohol-related crashes are not dependent on the number of car trips made within a census tract. A higher percentage of car trips also indicates a more uniform traffic mix within a census tract. The coefficient estimate of car trips is negative with regard to total crashes and speed-related crashes; therefore, a more uniform traffic mix will decrease these types of crashes.

Among demographic features, total area, population density, number of cars, gender, race-related variables, and bar counts are associated statistically with crash frequency. The negative sign of the area variable in the models for total crashes, speed- and alcohol-related crashes indicates that number of crashes decreases as area size increases. This relationship indicates that rural areas may have lower crash density because the size of rural census tract is usually larger. Population density can also be considered as a surrogate measure for area type, as more densely populated areas represent urban areas. Population density is statistically significant in predicting both types of behavior-related crashes (alcohol-related and speed-related). The negative sign of the population density variable means fewer crashes may occur in more populated areas. This relationship indicates that both speed- and alcohol-related crashes occur more frequently in rural areas compared with urban areas. The positive coefficient of the percent male variable indicates that male drivers are statistically more prone to speed-related crashes than female drivers. The population composition of a census tract is represented by exploring the coefficient estimates of people less than 18 years of age and people more than 64 years of age. The negative coefficient estimate of the percentage of people less than 18 years of age means that fewer crashes occurred in areas with more people who are younger than 18. This relationship is reasonable because the percentage of licensed drivers or vehicle owners is the lowest among young people (<18) compared with other age groups. The percentage of people older than 64 years of age is only significant in predicting alcohol-related crashes. A census tract with older people usually has less alcohol-related crashes.

Among socioeconomic variables, median income, education status, and employment status were found to be statistically significant. Higher income indicates the community is more educated. Model parameter estimates also show that both total crashes and alcohol-related crashes in a census tract decrease with an increase in median income. Interestingly, median income was not statistically significant in predicting speed-related crashes, implying that a person’s income or socioeconomic status does not predict speeding behaviors. A similar conclusion can be made with regard to educational status. The coefficient estimates for the “less than high school degree” variable implies that all crashes, speed-related crashes, and alcohol-related crashes increase when the percentage of uneducated people within a census tract increases. The estimated coefficient sign is consistent for all modeled crash types. Similarly, employment status (percent unemployed) reflects the status of household median income and education levels. The total number of vehicles in a census tract can be used to present trips generated from a census tract. The estimated coefficients for the total number of vehicles is positive for both total crashes and alcohol-related crashes, indicating that a census tract's crash count increases with an increase in traffic. Speed-related crashes, however, do not depend on the number of vehicles in a census tract.

5.4 Summary and Recommendations

The area-level crash frequency modeling results can help transportation agencies monitor area-level safety, identify major crash determinants, and evaluate safety programs and investment decisions. These results can be used to identify communities with a high risk of crashes and develop effective countermeasures to increase safety.

The area-level CPM analysis provides an opportunity to collect new data items for more rigorous crash analysis. The "bar count" variable collected from Business Analyst was available only for southern Wisconsin, but even with this limitation, the bar count within a census tract was found to be statistically significant in predicting total crashes and alcohol-related crashes. The CPM results also indicate that socioeconomic status and demographic variables are related to all types of crashes. Exploration and incorporation of these variables could provide a better understanding of safety issues within a census tract and help to develop effective safety countermeasures.

6. SITE-SPECIFIC CRASH PREDICTION MODEL

6.1 Introduction

Site-specific (e.g. roadway segment, intersection) crash data is often characterized by a large sample variance compared with the sample mean[8] (*33; 34*). Extensive research has been devoted to modeling and analyzing this type of crash dataset (*35-37*). One of the most notable accomplishments is the application of Negative Binomial (NB) models in crash frequency data. NB models can handle data over-dispersion by assuming a gamma distribution for the exponential function of the disturbance term in the Poisson mean. However, recent studies have pointed out that biased parameter estimates in the NB model can be found in dataset with a long tail (*4; 5*). A heavy tail is a statistical phenomenon that occurs when sample observations have a few very high crash counts or have preponderant zero observations which shift the overall sample mean to near zero (*5*). Failure to account for data over-dispersion leads to biased and inconsistent parameter estimates, which in turn causes erroneous inferences from models and inaccurate crash prediction information.

The mixed model is a well-known methodology used to incorporate heterogeneity into statistical analysis. In safety literature, mixed distribution NB models expanded the linear mixed model for continuous responses to discrete responses (e.g., crash count) by incorporating correlated non-normally distributed outcomes. Several mixed NB models have been proposed, including NB-Lindley (NB-L), NB-Generalized Exponential (NB-GE), and NB-Dirichlet process (NB-DP) generalized linear models (GLMs)-(*5-7; 38*). The advantage of the mixed model is that it provides flexibility by adding a mixed distribution to account for extra variance in the crash data which is caused by preponderant zero crash responses and a long tail. The underlying hypothesis is that the crash datasets are comprised of distinct subpopulations which have different probabilistic distributions. On the other hand, accessing all data items associated with the likelihood of crash occurrence and or injury severity is impossible. Omitting important variables such as driver-related factors causes data heterogeneity. Random parameters (RP) models can account for unobserved heterogeneity by allowing the parameter of variables to vary from one observation to the next and by estimating the unbiased mean effect of explanatory variables (*36*). Therefore, incorporating both random parameters and mixed probabilistic distributions within a single model can be a viable alternative for handling crash data with high over-dispersion and unobserved heterogeneity.

The objective of this section was to develop an RP NB model with Lindley mixed effect for heterogeneous count data due to unavailability of human factors, featuring an excess number of zero responses and a long tail. The proposed RP NB-L model was developed in a Bayesian hierarchical framework that is expanded from fixed coefficients NB-L GLM (*6; 7*). The parameters in RP NB-L GLM were calibrated with crash data from the Meta-Manager maintained by WisDOT and characterized by data over-dispersion with a high percentage of zero

[8] In a statistical term, the sample data is over-dispersed when the variance is greater than the mean. Data over-dispersion is often caused by unobserved data heterogeneity due to unobserved, unavailable, or unmeasurable variables that are important to explain model responses.

responses and a long tail. The model fitting and the modeling results were compared with the traditional NB, RP NB, and NB-L models.

6.2 Data Processing and Exploratory Analysis

Roadway geometry, pavement characteristics, mobility, safety, and other roadway-related data tables for Wisconsin are stored in a data management system developed by WisDOT called Meta-Manager. The most recent (February 2017) Meta-Manager data were used in this study along with the segment-related crash data from 2011 to 2015 for all crash types. As one of the largest highway facility types, rural two-lane two-way (RT) roadway was selected for developing crash prediction models. The rural RT segment database contained 8,287 observations after cleaning for null values in explanatory variables. 18 percent of the rural RT segments in this dataset did not experience any crashes from 2011 to 2015. Table 6-1 provides the descriptive summary statistics for the Wisconsin data.

Table 6-1 Summary Statistics of Wisconsin Segment-level Data.

Variable	Definition	Mean	Std. Dev.	Minimum	Maximum
Crash Count	Count of crashes	3.08	3.26	0	62
Segment Length	Length of segment in miles	0.962	0.39	0.01	2.79
AADT	Annual Average Daily Traffic	3367.37	2306.389	80	17238
Truck Percent	Percent of Truck in AADT	11.287	4.27	0	35
Lane Width	Lane width in feet	11.97	0.75	9	18
Shoulder Width	Shoulder width in feet	6.78	2.77	0	20
Percent No Passing	Percentage of no passing zone in segment	48.32	25.6	0	100
Horizontal Curve	Yes	660 (7.96%)			
	No	7627 (92.04%)			
Bridge Flag	Yes	1505 (18.16%)			
	No	6782 (81.84%)			

6.3 Findings

The performance of the RP NB-L model was compared with that of the NB, RP NB, and NB-L GLMs models. Table 6-2 summarizes the results in which coefficients for log(AADT), truck percentage, lane width, shoulder width, and passing zone were found to be random from site to site. The first part of Table 6-2 provides the estimates of parameter means, and the second part of the table provides the estimated standard error of random parameters.

Table 6-2 Parameter Estimates for Wisconsin RT Segments.

Parameters	NB		RPNB		NB-L		RPNBL	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.

Fixed Effect								
Intercept	-5.798	0.009	-5.658	0.011	-5.921	0.030	-5.778	0.031
Segment_Length	1.047	0.025	1.054	0.026	1.083	0.036	1.084	0.043
log(AADT)	0.818	0.015	0.826	0.016	0.82	0.017	0.814	0.019
Truck Percent	-0.01	0.002	-0.01	0.002	-0.01	0.003	-0.008	0.003
Lane Wid	-0.045	0.013	-0.066	0.016	-0.044	0.019	-0.051	0.027
Shoulder Width	-0.044	0.004	-0.042	0.004	-0.046	0.005	-0.045	0.004
Horizontal Curve	0.338	0.040	0.308	0.041	0.337	0.058	0.343	0.050
Percent Passing	0.004	0.0004	0.004	0.0003	0.004	0.001	0.004	0.0003
Bridge Flag	0.104	0.023	0.107	0.023	0.116	0.031	0.11	0.033
Alp	0.268	0.010	0.21	0.014	0.101	0.0001	0.1007	0.0001
t					1.326	0.031	1.334	0.033
Random Effects								
Std. Error Of AADT	NA		0.220	0.032	NA		0.070	0.010
Std. Error Of Truck Percentage			0.027	0.005			0.011	0.003
Std. Error Of Lane Width			0.108	0.016			0.055	0.013
Std. Error Of Shoulder Width			0.030	0.008			0.015	0.003
Std. Error Of Percent Passing			0.003	0.0007			0.001	0.0002
Model Performance								
Dbar	33630		32890		29260		29240	
Dhat	33620		32230		24990		24950	
pD	9.597		664.3		4266		4285	
DIC	33640		33560		33530		33520	

The NB model results in Table 6-2 show that all variables were statistically significant at a 10 percent significance level. The parameter-mean estimates of all explanatory variables have the same directions but not necessarily the same effects. Parameter estimates for all covariates were compared, finding that the parameter means were statistically significant at a 10 percent significance level. The random parameters between RP NB and RP NB-L were compared, finding that all continuous explanatory variables were statistically significant at a 10 percent significance level. The NB-L model has smaller standard deviation estimates for all model coefficients (random parameters). The smaller the standard deviation for random parameter estimates, the more the normal distribution of a covariate parameter is centered to the mean value when the RP NB-L model is used. This may be a result of the site-specific frailty term used in the NB-L formulation which accounts for a portion of data variation.

The segment length and AADT variables in the RP NB-L model were positively correlated with crash count for almost all segments, but with varying magnitude. For the truck percentage variable, more than 65% of the sites had parameter estimates less than zero. As the standard deviation estimate of the random coefficient for percent passing is very small, it can be considered to have a fixed effect in the model. Similar observations can be made for the lane width variable, where 79% of the random parameter estimates have a value of less than 0. Shaon and Qin used the same dataset and made similar observations (*7*). The authors noted that lane width may have mixed safety effects, and an increasing lane width may not always bring additional safety benefits. Further research should look into whether increased lane width does lead to increased safety.

The estimate of the dispersion parameter is smallest in the RP NB-L model, suggesting that the mean estimates in the RP NB-L model are less affected by data dispersion. In other words, the RP NB-L model is capable of accounting for more variations in the data than the other three models.

Table 6-2 also provides model performance estimates based on the Deviance Information Criterion (DIC). Geedipally et al. explained that the model parameterization can influence the estimation of the DIC value, and comparisons with DIC should be made only between models that have similar parameterizations (*39*). As both the NB-L and RP NB-L models are developed based on NB model parameterization, all developed models can be adequately compared using the DIC measure. The DIC consists of two components: (a) measures of how well the model fits the data (Dbar) and (b) a measure of model complexity (pD). A comparison of DIC between models illustrated that the RP NB-L model performed better than NB-L and RP NB. Table 6-2 shows that the DIC value is highest in the traditional NB model, but the pD value illustrates that estimation is easiest with NB. According to the estimated pD value, the RP NB-L model is the most complex of the models due to its mixed distribution and random components in explanatory variables. The point estimate of deviance illustrated by Dhat represents that the RP NB-L model has the smallest deviance. Dbar represents almost the same information as Dhat except that it represents the posterior mean of deviance rather than a point estimate. Although it does have the highest penalty value of pD, the RP NB-L model has a significant improvement in DIC values when compared with the RP NB model and fixed-coefficient NB-L model, respectively.

6.4 Summary and Recommendations

It is challenging to understand the underlying crash generating process and to produce reliable model coefficients and statistical inferences from crash data. Although it is well established that human factors play a very important role in crash occurrence, human-related variables may not always be readily available.

As discussed earlier, a lack of critical data can cause unobserved heterogeneity and excess overdispersion in the crash dataset. Without exclusively addressing these issues in modeling techniques, a crash prediction model may provide biased parameter estimates. The above section proposed the application of an RP NB-L GLM for analyzing crash data by implementing an NB-L model with coefficients that varied from site to site using a two-lane, two-way rural highway crash database from Wisconsin. The model results were compared with NB, RP NB, and fixed coefficient NB-L models. Results showed that both the fixed coefficient NB-L and newly developed RP NB-L GLMs performed best. According to the standard deviation of random parameters, the estimated effects of covariates using RP NB-L were less dispersed compared with the RP NB model. The RP NB-L model’s highly dispersed data led to its ability to achieve a significant improvement in DIC when compared to the RP NB model. Therefore, it is safe to conclude that both the fixed and random parameters of NB-L GLMs offer a viable alternative to the traditionally fixed and random parameters NB GLMs when analyzing over-dispersed crash datasets.

Although the flexibility in the model structure of an RP NB-L GLM allows unobserved data heterogeneity to be properly treated in the model estimation, it does not help to explain the effect of missing variables. As data and modeling go hand-in-hand, more variables such as driver behavior-related factors need to be collected to minimize the effects of unobserved heterogeneity in the crash dataset if resources are permitted.

7. DRIVER ERROR PREDICTION MODEL

7.1 Introduction

State and federal highway administrators in the United States have been working together to achieve the goal of Vision Zero, an injury severity study in which engineering-related variables such as roadway design and traffic conditions are examined in depth. Despite the well-known knowledge that driver errors partially or mainly contribute to over 90 percent of all crashes (*30*), these variables are rarely and explicitly considered in crash prediction models. The primary reason for not including driver behavior variables in these models is the scarcity of such information, especially at the site level. Crash model results could be biased if driver factors are not considered, and the effect of engineering variables may be overestimated. Thus, understanding why drivers make mistakes and how to incorporate human factors into crash prediction models has become an increasingly important topic.

It is of particular interest to know when, where, and how drivers make mistakes that contribute to a crash. During the driver information processing period of seeing and reacting to a hazard: perception, intellection, emotion, and volition, or "PIEV", errors can happen during any of the four phases. Driver decision-making is only as good as the information drivers can glean from their environment. Situations with conflicting or confusing information can make the situation worse. Increased workload and driver distraction may come from a comprehensive list of factors ranging from roadway geometry, traffic conditions, weather, lighting conditions, unexpected events like construction zones, debris on the roadway, as well as in-vehicle distractions. The National Motor Vehicle Crash Causation Study (NMVCCS) adopted a way to categorize the errors associated with information processing: recognition, decision, performance, and non-performance errors on a roadway segment. Explanatory variables identified through statistical models can be different for each unique error type, which helps to design custom treatments and countermeasures.

Driving becomes more challenging where two or more roads meet. Situations occurring at intersections can be complex and overwhelming under certain traffic controls. Certain circumstances can elevate an error from a traffic infraction to an intentional traffic violation. A better understanding of the underlying relationships between driver errors and contributing factors can lead to the development and deployment of more effective safety strategies.

The goal of developing driver error prediction models is to investigate key contributors of various driver errors that are attributable to a crash. Specifically, the models will 1) examine common driver errors committed during information processing; 2) explore the critical factors affecting different severity levels of driver errors and identify the effects of these factors by intersection type; and 3) recommend cost-effective countermeasures to mitigate driver errors. New insights on the circumstances that lead to driver error will shed light on the development of tangible, practical, targeted, cost-effective enforcement strategies, driver education and training programs, engineering solutions, and vehicle technologies.

7.2 Data Processing and Exploratory Analysis

Wisconsin crash data were collected from 2013 to 2015, excluding deer-related crashes (*40*). 48,441 rural crashes and 46,221 urban crashes were retrieved from the Wisconsin state highway segments after excluding crashes that lacked good location information. Segment crashes were limited to state highways so that crash data could be combined with highway inventory information that is available only for state highways. Fourteen types of driver errors were extracted from the Wisconsin Motor Vehicle Accident Reporting Form 4000 (MV4000), in which the investigating police officers documented detailed accident information (*40; 41*). The police investigation guides which factor is listed as the most severe when there are multiple factors involved. The NMVCCS study guided the classification of driver-related errors as either recognition, decision, performance, or nonperformance-based (*30*). Recognition error includes driver inattention, internal and external distraction, inadequate surveillance; decision error includes aggressive driving behavior, driving too fast, etc.; performance error includes overcompensation, poor directional controls; sleep and physical impairment are considered as nonperformance errors. Table 7-1 shows the NMVCCS driver error types and corresponding Wisconsin driver factors along with summary statistics for each category.

Table 7-1 Categorization and Distribution of Driver Error.

Error Type	NMVCCS Criteria	Wisconsin Criteria	Rural	Urban
Recognition Error	• Inadequate surveillance • Internal distraction • External distraction • Inattention	• Inattentive driving	8659 (17.88%)	9044 (19.57%)
Decision Error	• Too fast for conditions • Too fast for curve • False assumption of other's action • Illegal maneuver • Misjudgment of gap or other's action • Following too closely • Aggressive driving behavior	• Too Fast for condition • Exceed Speed Limit • Disregard traffic control • Following too close • Improper overtake • Improper turn	17139 (35.38%)	17662 (38.21%)
Performance Error	• Overcompensation • Poor directional control • Panic/Freezing • Other performance error	• Failure to keep vehicle under control • Left of center • Unsafe backing • Failure to yield	10288 (21.24%)	9867 (21.35%)
Non-Performance Error	• Sleep • Heart attack • Other non-perf. error	• Disability • Driver Condition • Others	2402 (4.96%)	3030 (6.55%)
No Error			9953 (20.55%)	6620 (14.32%)

As shown in Table 7-1, 18-20 percent of total crashes that occurred between 2013 to 2015 in Wisconsin were due to inattentive driving; 35-38 percent of crashes occurred were due to decision error; approximately 21 percent of total crashes were due to performance errors, and the remaining 5-7 percent were due to non-performance errors made between rural and urban areas. The crash dataset does not contain roadway geometric information at the crash location. Roadway geometry, pavement characteristics, mobility, safety and other roadway-related data tables stored in Meta-Manager at the Wisconsin Department of Transportation (WisDOT) were linked with crash data using spatial join in ArcGIS. The joined dataset contains all information collected by the investigating police officer, roadway geometry, and traffic information for each crash. Table 7-2 provides summary statistics of explanatory variables.

Table 7-2 Summary Statistics of Explanatory Variables.

Variable	Description	Type	Rural		Urban	
			Mean	Std. Dev.	Mean	Std. Dev.
AADT	Annual Average Daily Traffic (In thousand unit)	Continuous	21610.37	26554.21	55450.19	48566.43
Truck	Truck Percentage (%)	Continuous	11.44	4.59	7.735	2.86
Speed	Posted Speed Limit (MPH)	Continuous	57.49	11.31	46.95	14.17
Lane	Number of lanes (Count)	Continuous	2.13	0.43	2.59	0.715
LW	Lane width (feet)	Continuous	12.10	0.83	12.34	1.08
SW	Shoulder width (feet)	Continuous	8.60	3.87	5.58	5.49
Rut	Pavement rutting (inch)	Continuous	0.088	0.08	0.07	0.07
Percent Passing	Passing percentage (%)	Continuous	26	31.90	3.25	15.85
Highway Type	Interstate	Categorical with 3 levels	9840 (20.31%)		12012 (25.99%)	
	State Highway		37377 (77.16%)		30062 (65%)	
	Other state roadway		1224 (2.53%)		4147 (9%)	
Roadway Type	Undivided	Categorical with 3 levels	24177 (49.91%)		8225 (17.79%)	
	Divided		23889 (49.32%)		36324 (78.59%)	
	One Way		375 (0.77%)		1672 (3.62%)	
Presence of Median	No	Categorical with 2 levels	31530 (65.09%)		20170 (43.64%)	
	Yes		16911 (34.91%)		26051 (56.36%)	
Roadway Condition	Dry	Categorical with 4 levels	27830 (57.45%)		31740 (68.67%)	
	Wet		5255 (10.85%)		7517 (16.26%)	
	Snow		10281 (21.22%)		5307 (11.48%)	
	Ice		5075 (10.48%)		1657 (3.58%)	
Weather Condition	Clear	Categorical with 5 levels	20591 (42.51%)		22378 (48.42%)	
	Fog/Cloudy		13619 (28.11%)		14671 (31.74%)	
	Wind		1041 (2.15%)		140 (0.3%)	
	Rain		3057 (6.31%)		4157 (8.99%)	
	Snow/Sleet		10133 (20.92%)		4875 (10.55%)	
Lighting Condition	Day	Categorical with 3 levels	33065 (68.26%)		30046 (73.66%)	
	Night-Unlit		13477 (27.82%)		3026 (6.55%)	
	Night-Lit		1899 (3.92%)		9149 (19.79%)	
Horizontal Curve	No	Categorical with 2 levels	39390 (81.32%)		41750 (90.33%)	
	Yes		9051 (18.68%)		4471 (9.67%)	
Vertical Curve	No	Categorical with 2 levels	38865 (80.23%)		40542 (87.71%)	
	Yes		9576 (19.77%)		5679 (12.29%)	
Age group	Adolescent (<18 years)	Categorical with 5 levels	2363 (4.88%)		1789 (3.87%)	
	Young Adults (18-25 years)		11206 (23.13%)		11271 (24.39%)	

	Adults (26-35 years)		10309 (21.28%)	11406 (24.68%)
	Middle Age (36-65 years)		20223 (41.47%)	18294 (39.58%)
	Old (>65 years)		4340 (8.96%)	3461 (7.49%)
Gender	Male	Categorical with 2 levels	30090 (62.12%)	26989 (58.39%)
	Female		18351 (37.88%)	19232 (41.61%)
Vehicle	Passenger car	Categorical with 4 levels	35498 (73.28%)	38051 (82.32%)
	Motorcycle		888 (1.83%)	545 (1.18%)
	Light truck		8096 (16.71%)	4898 (10.6%)
	Heavy truck		3959 (8.17%)	1727 (5.9%)
Alcohol	No	Categorical with 2 levels	45725 (94.39%)	44470 (96.21%)
	Yes		2716 (5.61%)	1751 (3.79%)
Drug	No	Categorical with 2 levels	47920 (98.92%)	45881 (99.26%)
	Yes		521 (1.08%)	340 (0.74%)
Visibility Obscured	No	Categorical with 2 levels	48078 (99.25%)	46013 (99.55%)
	Yes		363 (0.75%)	208 (0.45%)
Work Zone	No	Categorical with 2 levels	47625 (98.32%)	45339 (98.09%)
	Yes		816 (1.68%)	882 (1.91%)
Debris on road	No	Categorical with 2 levels	47695 (98.46%)	45860 (99.22%)
	Yes		746 (1.54%)	361 (0.78%)

Intersection-related crashes are unevenly distributed between rural and urban areas, with 24,774 in rural areas and 76,583 in urban areas. Data in this study includes 7,203 intersection-related crashes that occurred in Madison, WI between 2008 and 2010. Approximately 90 percent of intersection-related crashes were related to driver errors. Crashes were further categorized by the intersection's traffic control strategy (*i.e.* uncontrolled, sign-controlled and signalized) as suggested by Devlin *et al.(42)*. Roundabouts were omitted due to their low count and the fact that a very limited number of crashes occurred in roundabouts.

Driver errors were extracted from MV4000 where investigating police officers documented detailed accident information. Similarly, it is not unusual for one crash to be associated with multiple violations. The "NA" bubble in the traffic accident report was marked if no driver errors applied (*43*). According to the citation documentation for traffic violations in Wisconsin (*44*), driver errors are classified as either improper, careless, or reckless driving, with an increasing ordinal nature to account for the severity of violation. Improper overtaking, improper turning, or driving too fast for the road conditions are traffic infractions that are punishable by a fine of no more than $500, for example. Careless driving incidents such as following too close, failure to keep the vehicle under control, inattentive driving, left of center, or unsafe backing, are often defined as operating a motor vehicle in an offensive and negligent manner, but doing so unintentionally. These offenses shall be punishable by a fine that is higher than the improper driving fine. Reckless driving is usually defined as a mental state in which the driver intentionally breaks traffic rules. Reckless driving often causes severe accidents or other damages and is punishable by fines, imprisonment, and/or driver license suspension or revocation (*45; 46*). Reckless driving violations include a disregard for traffic control, failure to yield, and exceeding the speed limit. The upper panel of Table 7-3 includes all driver errors and describes the distributions of specific driver errors by intersection type. The lower panel of Table 7-3 describes the distributions of specific driver error severities by intersection type. Only the most severe driver errors were considered for this study.

The number and severity of driver errors vary by intersection type and by area type. As shown in Table 7-3, sign-controlled intersections have the highest percentage of driver errors and the highest percentage of reckless driving violations; this is almost two times the percentage of violations occurring at uncontrolled or signalized intersections. The severity distributions of signalized intersections and uncontrolled sites are very close. Specific driver errors vary by intersection type within each severity type. “Fail to yield” is fairly prevalent across all intersection types with regard to reckless driving, but it is a dominating violation for sign-controlled intersections. “Inattentive driving” is the most frequently made mistake when looking at careless driving factors, followed by “follow too close”. “Improper turning” and “too fast for the condition” are commonly observed at uncontrolled intersections when improper driving violations are examined. The percentage of careless driving is higher in rural areas than in urban areas when all intersection types are considered, whereas the percentage of reckless driving is higher in urban areas. Recognizable patterns may exist in factors other than traffic control. Crashes with driver errors were classified into four categories: driver characteristics, highway and traffic characteristics, environmental factors, and vehicle type. The corresponding distribution is presented in Table 7-4.

The distribution of drivers by age and by gender is consistent irrespective of traffic controls. A very high percentage of driver errors took place at intersections with a low speed limit (<35mph), especially intersections on urban highways. Posted speed limit is the speed limit of the street on which a crash occurred. The percentage of driver mistakes made in the morning peak hours is considerably higher than other time periods when all intersection types are considered. It is worth noting that when observing nighttime errors for all intersection types, a higher percentage of errors occurred when street lights were present as opposed to when they were not present, especially on urban highways. The percentage of driver errors involving only passenger cars is markedly higher than that of trucks.

Table 7-3 Distribution of Driver Error by Intersection Type.

Category	Rural Highways (24774)			Urban Highways (76583)		
	Frequency by All Crashes (%)			Frequency by All Crashes (%)		
	Uncontrolled (6403)	Sign-controlled (12925)	Signal-controlled (5446)	Uncontrolled (15570)	Sign-controlled (23671)	Signal-controlled (37342)
No Errors	667 (10%)	930 (7%)	568 (10%)	1967 (13%)	2120 (9%)	4841 (13%)
Improper Driving						
Improper overtaking	247 (4%)	16 (0%)	23 (0%)	311 (2%)	52 (0%)	242 (1%)
Improper turning	235 (4%)	271 (2%)	152 (3%)	820 (5%)	500 (2%)	1308 (4%)
Too fast for the condition	277 (4%)	871 (7%)	308 (6%)	594 (4%)	1018 (4%)	1915 (5%)
Careless Driving						
Follow too close	639 (10%)	496 (4%)	674 (12%)	2022 (13%)	978 (4%)	4059 (11%)
Fail to keep vehicle under control	356 (6%)	857 (7%)	476 (9%)	594 (4%)	733 (3%)	2125 (6%)
Inattentive driving	2063 (32%)	1956 (15%)	1214 (22%)	3494 (22%)	2662 (11%)	7182 (19%)
Left of center	166 (3%)	104 (1%)	13 (0%)	152 (1%)	80 (0%)	95 (0%)
Unsafe backing	177 (3%)	240 (2%)	56 (1%)	332 (2%)	246 (1%)	251 (1%)
Reckless Driving						
Disregard traffic control	10 (0%)	806 (6%)	547 (10%)	60 (0%)	1487 (6%)	5592 (15%)
Fail to yield	1484 (23%)	6155 (48%)	1385 (25%)	4954 (32%)	13422 (57%)	9219 (25%)
Exceed speed limit	82 (1%)	223 (2%)	30 (1%)	270 (2%)	373 (2%)	513 (1%)
Category	Frequency by All Crashes (%)			Frequency by All Crashes (%)		
	Uncontrolled (6403)	Sign-controlled (12925)	Signal-controlled (5446)	Uncontrolled (15570)	Sign-controlled (23671)	Signal-controlled (37342)
No Errors	667 (10%)	930 (7%)	568 (10%)	1967 (13%)	2120 (9%)	4841 (13%)
Improper Driving	759 (12%)	1158 (9%)	483 (9%)	1725 (11%)	1570 (7%)	3465 (9%)
Careless Driving	3401 (53%)	3653 (28%)	2433 (45%)	6594 (42%)	4699 (20%)	13712 (37%)
Reckless Driving	1576 (25%)	7184 (56%)	1962 (36%)	5284 (34%)	15282 (65%)	15324 (41%)

Table 7-4 Driver Error Distribution by Contributing Factors.

				Rural Highways (24774)			Urban Highways (76583)		
				Frequency by Driver Errors (%)			Frequency by Driver Errors (%)		
Category	**Variable**	**Type and Value**	**Description**	**Uncontrolled**	**Sign**	**Signal**	**Uncontrolled**	**Sign**	**Signal**
Driver	Aggend	**Categorical**	**Age and gender**						
		younger male	Male driver (age<25)	1174 (18%)	2122 (16%)	817 (15%)	2525 (16%)	3448 (15%)	5743 (15%)
		middle-aged male	Male driver (age 25-55)	1660 (26%)	3362 (26%)	1535 (18%)	4187 (27%)	5665 (24%)	10466 (28%)
		older male	Male driver (age>55)	889 (14%)	2098 (16%)	685 (13%)	1636 (11%)	3153 (13%)	4131 (11%)
		younger female	Female driver (age<25)	937 (15%)	1452 (11%)	723 (13%)	2401 (15%)	3195 (14%)	5040 (14%)
		middle-aged female	Female driver (age 25-55)	1123 (18%)	2431 (19%)	1136 (21%)	3390 (22%)	5248 (22%)	8392 (22%)
		older female	Female driver (age>55)	620 (9%)	1460 (12%)	550 (10%)	1431 (9%)	2962 (12%)	3570 (10%)
	DUI	**Dummy**	**Drugs or alcohol**	175 (3%)	565 (4%)	113 (2%)	299 (2%)	476 (2%)	905 (2%)
Highway and traffic	Hor	**Dummy**	**Horizontal curve**	728 (11%)	1597 (12%)	259 (5%)	1000 (6%)	2475 (10%)	1366 (4%)
	Vert	**Dummy**	**Vertical curve**	909 (14%)	1862 (14%)	471 (9%)	1404 (9%)	2302 (10%)	2569 (7%)
	Postspd	**Categorical**	**Posted speed limit**						
		low	Low (<35 mph)	2586 (40%)	5329 (41%)	2314 (42%)	11409 (73%)	18856 (80%)	23332 (62%)
		middle	Middle (35 mph-55 mph)	3505 (55%)	6892 (53%)	2934 (54%)	3888 (25%)	4239 (18%)	13427 (36%)
		high	High (>55 mph)	312 (5%)	704 (6%)	198 (4%)	273 (2%)	576 (2%)	583 (2%)
	Acctime	**Categorical**	**Accident time**						
		morning peak	7:00am-9:59am	3025 (47%)	5861 (45%)	2492 (46%)	7040 (45%)	11100 (47%)	16604 (44%)
		day time	10:00am-3:59pm	1058 (17%)	2424 (19%)	954 (18%)	2448 (16%)	4387 (19%)	6124 (16%)
		afternoon peak	4:00pm-6:59pm	1770 (28%)	3049 (24%)	1364 (25%)	4393 (28%)	5804 (25%)	9238 (25%)
		nigh time	7:00pm-6:59am	550 (8%)	1591 (12%)	636 (11%)	1689 (11%)	2380 (9%)	5376 (15%)
Environmental	Wthr	**Categorical**	**Weather condition**						
		clear	Clear	3582 (56%)	6682 (52%)	2640 (48%)	8508 (55%)	12453 (53%)	19407 (52%)
		cloudy	Cloudy	1962 (31%)	4350 (34%)	1913 (35%)	4969 (32%)	7822 (33%)	11762 (32%)
		rain	Rain	382 (6%)	820 (6%)	431 (8%)	1137 (7%)	1791 (8%)	3333 (9%)
		snow	Snow/hail	477 (7%)	1073 (8%)	462 (9%)	956 (6%)	1605 (6%)	2840 (7%)
	Lgt	**Categorical**	**Light condition**						
		day	Day	5390 (84%)	10319 (80%)	4340 (80%)	12670 (81%)	19692 (83%)	28446 (76%)
		without	Night (without street light)	565 (9%)	1593 (12%)	72 (1%)	332 (2%)	408 (2%)	277 (1%)
		light	Night (with street light)	448 (7%)	1013 (8%)	1034 (19%)	2568 (17%)	3571 (15%)	8619 (23%)
	Road	**Categorical**	**Road condition**						
		dry	Dry	4666 (73%)	8926 (69%)	3825 (70%)	11186 (72%)	16579 (70%)	26352 (71%)
		wet	Wet	741 (12%)	1602 (12%)	871 (16%)	2217 (14%)	3507 (15%)	6415 (17%)
		snow	Snow/slush	781 (12%)	1787 (14%)	590 (11%)	1826 (12%)	2987 (13%)	3945 (11%)
		ice	Ice	215 (4%)	610 (5%)	160 (3%)	341 (2%)	598 (2%)	630 (1%)
Vehicle	Vehtype	**Categorical**	**The vehicle type**						
		pc	Passenger car	4851 (76%)	9760 (76%)	4313 (79%)	13377 (86%)	20365 (86%)	31689 (85%)
		light truck	Light truck	1165 (18%)	2425 (19%)	798 (15%)	1620 (10%)	2448 (10%)	3855 (10%)
		heaver truck	Heaver truck	387 (6%)	740 (5%)	335 (6%)	573 (4%)	858 (4%)	1798 (5%)

7.3 Findings

The coefficient estimates of the MNP regression models for rural and urban crashes are presented in Table 7-5 and Table 7-6, respectively. In both tables, the coefficient estimates represent the log-odds ratio between the probabilities of defined driver error type and no error category with a positive sign for increase and a negative sign for decrease. The "no error" category was considered as the base outcome in the MNP model.

The modeling results for "Non-performance error" were excluded because this error category does not include driver behavioral factors. The middle-aged and old-aged groups are more prone to non-performance error. Alcohol and drug consumption increases the probability of non-performance error compared to no error.

Table 7-5 Coefficient Estimates for MNP Model for Driver Errors in Rural Crashes.

Variable		**Recognition Error**		**Decision Error**		**Performance Error**	
		Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
AADT		**0.002**	**0.000**	**0.010**	**0.000**	**0.004**	**0.000**
Truck		**-0.005**	**0.003**	**-0.006**	**0.002**	**-0.011**	**0.002**
Speed		**-0.02**	**0.002**	**-0.019**	**0.002**	**-0.01**	**0.002**
Lanes		-0.046	0.032	**-0.084**	**0.03**	**-0.12**	**0.032**
Shoulder Wid		**0.012**	**0.004**	0.001	0.004	**-0.014**	**0.004**
Pavement Rutting		**-0.462**	**0.166**	**-0.515**	**0.154**	**-0.518**	**0.158**
Highway Type	Interstate			Base Condition			
	State Highway	-0.033	0.037	**-0.102**	**0.031**	**0.08**	**0.034**
	Other state roadway	-0.114	0.077	**-0.186**	**0.072**	**0.338**	**0.072**
Roadway Type	Undivided	0.07	0.047	**-0.146**	**0.043**	0.00004	0.044
	Divided			Base Condition			
	One Way	**-0.36**	**0.143**	0.036	0.123	-0.026	0.13
Horizontal Curve	No			Base Condition			
	Yes	**0.153**	**0.031**	**0.255**	**0.027**	**0.356**	**0.028**
Vertical Curve	No			Base Condition			
	Yes	-0.029	0.029	**0.065**	**0.025**	**0.054**	**0.026**
Roadway Condition	Dry			Base Condition			
	Wet	-0.026	0.048	**0.413**	**0.045**	**0.208**	**0.046**
	Snow	**-0.892**	**0.057**	**1.072**	**0.039**	**0.284**	**0.042**
	Ice	**-1.56**	**0.076**	**0.947**	**0.038**	**0.18**	**0.041**
Weather Condition	Clear			Base Condition			
	Fog/Cloudy	**0.157**	**0.026**	**0.16**	**0.025**	**0.19**	**0.026**
	Wind	**-0.895**	**0.169**	0.054	0.070	-0.068	0.077
	Rain	**-0.163**	**0.065**	**0.269**	**0.057**	0.019	0.06
	Snow/Sleet	**-0.335**	**0.063**	**0.17**	**0.041**	0.085	0.044
	Day			Base Condition			

Lighting Condition	Night-Unlit	**-0.207**	**0.026**	**-0.341**	**0.023**	**-0.116**	**0.024**
	Night-Lit	-0.020	0.059	**-0.281**	**0.056**	**-0.194**	**0.058**
Visibility	No			Base Condition			
	Yes	**-0.364**	**0.132**	-0.046	0.108	-0.105	0.114
Work Zone	No			Base Condition			
	Yes	**0.210**	**0.082**	**0.565**	**0.076**	-0.074	0.089
Debris on road	No			Base Condition			
	Yes	**-2.136**	**0.117**	**-1.863**	**0.094**	**-1.762**	**0.101**
Age group	Adolescent			Base Condition			
	Young Adults	**-0.181**	**0.056**	**-0.174**	**0.052**	**-0.178**	**0.054**
	Adults	**-0.420**	**0.056**	**-0.341**	**0.052**	**-0.278**	**0.054**
	Middle Age	**-0.561**	**0.053**	**-0.527**	**0.050**	**-0.44**	**0.051**
	Old	**-0.339**	**0.061**	**-0.582**	**0.058**	**-0.211**	**0.059**
Gender	Male	0.022	0.024	0.006	0.021	**-0.056**	**0.022**
	Female			Base Condition			
Vehicle	Passenger car	**0.194**	**0.040**	**0.332**	**0.037**	**0.216**	**0.039**
	Motorcycle	**-0.564**	**0.091**	**0.225**	**0.081**	**0.488**	**0.078**
	Light truck	**0.193**	**0.046**	**0.325**	**0.041**	**0.237**	**0.044**
	Heavy truck			Base Condition			
Alcohol	No			Base Condition			
	Yes	**1.120**	**0.067**	**1.282**	**0.066**	**1.459**	**0.065**
Drug	No			Base Condition			
	Yes	**0.754**	**0.129**	**0.740**	**0.129**	**0.875**	**0.128**
Intercept		**1.277**	**0.255**	**1.324**	**0.244**	**1.138**	**0.252**

[Note: Variables that are statistically significant at 90% confidence interval are presented in bold font]

Driver age, gender, vehicle type, alcohol, and drug impairment were found to be statistically significant in predicting all driver error categories in Table 7-5. Adolescents are more prone to driver errors compared with all other age groups. The probability of decision error gradually reduces with the increase in age. Older drivers make more performance and recognition errors compared with younger and middle-aged drivers. Decision and recognition errors do not depend on a driver's gender, whereas female drivers were found to have a higher probability of performance error. Motorcycle drivers are least likely to make recognition errors, but they are most likely to commit a performance error. Alcohol or drug impairment increases the probability of all errors with a maximum increase in performance error.

Traffic variables such as AADT and truck percentage are statistically significant in predicting all types of driver error. A unit change (in thousand vehicles) in AADT results in 1.002 ($e^{0.002}$) times of a recognition error, 1.01 ($e^{0.010}$) times of a decision error and, 1.004 ($e^{0.004}$) times of a performance error compared to no error situation, respectively. The signs of estimated coefficients of truck percentage, speed, number of lanes, shoulder width, and pavement rutting represent the reduction in the probability of an error compared to no error because of a unit increase in the independent variable.

Roadway classification results show that highway type is significantly related to both decision and performance errors. Decision errors occur mainly on interstate highways, but performance errors are least likely to occur on the interstate. The change in the probability of performance error is the highest in highways such as rural city or town roads. One-way roads reduce recognition errors are reduced on one-way roads, and decision errors are reduced on undivided highways. Horizontal and vertical curves significantly increase the probability of all error categories with a maximum increase in performance error for horizontal curves and a maximum increase in decision error for vertical curves.

Roadway events have a significant effect on driver errors. A comparison between roadway and weather condition variables illustrates a few important observations. For example, snowy pavement increases decision error by 4.13 times, while snow precipitation increases decision error by 1.24 times. Another important observation is that snowy pavement has a higher increase in probability than icy pavement. Drivers tend to be more cautious during adverse weather events because of the negative impact on recognition error. A construction zone increases the probability of decision and recognition error but is not statistically significant for performance error. The negative impact of roadway debris on all types of errors suggests drivers may be more vigilant when there are unusual objects on the roadway.

Table 7-6 provides the coefficient estimates for the MNP model with urban crash data. All explanatory variables except for the median variable were statistically significant at a 10% level in predicting driver error categories in urban crashes.

Table 7-6 Coefficient Estimates for MNP Model for Driver Errors in Urban Crashes.

Variable		**Recognition Error**		**Decision Error**		**Performance Error**	
		Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
AADT (In thousand)		-3E-05	0.000	**0.002**	**0.000**	-0.0002	0.000
Truck		**-0.015**	**0.004**	**-0.015**	**0.004**	**-0.01**	**0.004**
Speed		**-0.006**	**0.001**	-0.001	0.001	**-0.011**	**0.001**
Lanes		**0.057**	**0.019**	**0.049**	**0.018**	**0.095**	**0.019**
Lane Wid		**0.047**	**0.011**	**0.045**	**0.011**	**0.036**	**0.011**
Shoulder Wid		0.003	0.003	**0.006**	**0.003**	**0.017**	**0.003**
Pavement Rutting		**-0.482**	**0.162**	**-0.393**	**0.152**	**0.387**	**0.159**
Percent Passing		**0.005**	**0.001**	**0.003**	**0.001**	**0.002**	**0.001**
Highway Type	Interstate			Base Condition			
	State Highway	0.001	0.033	-0.017	0.029	**-0.244**	**0.032**
	Other state roadway	**-0.12**	**0.049**	**-0.335**	**0.045**	**-0.146**	**0.047**
Roadway Type	Undivided	0.043	0.043	**-0.105**	**0.04**	**-0.069**	**0.042**
	Divided			Base Condition			
	One Way	-0.056	0.063	**-0.171**	**0.06**	**-0.248**	**0.064**
Horizontal Curve	No			Base Condition			
	Yes	**-0.196**	**0.043**	0.053	0.037	**0.283**	**0.038**
Vertical Curve	No			Base Condition			
	Yes	**-0.082**	**0.036**	-0.029	0.032	**-0.12**	**0.034**

Roadway Condition	Dry			Base Condition			
	Wet	-0.069	0.045	**0.157**	**0.042**	**0.203**	**0.044**
	Snow	**-0.802**	**0.063**	**0.323**	**0.049**	**0.104**	**0.052**
	Ice	**-1.991**	**0.123**	0.061	0.055	**-0.408**	**0.063**
Weather Condition	Clear			Base Condition			
	Fog/Cloudy	**0.145**	**0.026**	**0.168**	**0.025**	**0.198**	**0.026**
	Wind	-0.456	0.303	0.119	0.173	-0.074	0.2
	Rain	**-0.153**	**0.06**	**0.219**	**0.054**	0.077	0.057
	Snow/Sleet	**-0.296**	**0.07**	**0.181**	**0.052**	**0.12**	**0.056**
Lighting Condition	Day			Base Condition			
	Night-Unlit	**-0.19**	**0.05**	**-0.321**	**0.043**	**-0.159**	**0.047**
	Night-Lit	**-0.243**	**0.029**	**-0.37**	**0.027**	**-0.225**	**0.028**
Visibility	No			Base Condition			
	Yes	**-0.588**	**0.166**	**-0.682**	**0.153**	0.22	0.142
Construction Zone	No			Base Condition			
	Yes	-0.043	0.083	**0.289**	**0.075**	-0.052	0.084
Debris on road	No			Base Condition			
	Yes	**-2.04**	**0.149**	**-1.895**	**0.113**	**-1.925**	**0.14**
Age group	Adolescent			Base Condition			
	Young Adults	**-0.202**	**0.065**	**-0.213**	**0.061**	**-0.188**	**0.065**
	Adults	**-0.376**	**0.065**	**-0.431**	**0.061**	**-0.296**	**0.065**
	Middle Age	**-0.393**	**0.064**	**-0.5**	**0.059**	**-0.307**	**0.063**
	Old	**-0.271**	**0.073**	**-0.531**	**0.069**	**0.001**	**0.072**
Gender	Male	**-0.069**	**0.024**	-0.011	0.022	-0.035	0.023
	Female			Base Condition			
Vehicle	Passenger car	**0.413**	**0.046**	**0.512**	**0.042**	**0.44**	**0.046**
	Motorcycle	**-0.536**	**0.118**	0.02	0.099	**0.436**	**0.099**
	Light truck	**0.44**	**0.056**	**0.512**	**0.051**	**0.434**	**0.054**
	Heavy truck			Base Condition			
Alcohol	No			Base Condition			
	Yes	**0.917**	**0.086**	**0.849**	**0.083**	**1.226**	**0.083**
Drug	No			Base Condition			
	Yes	**0.667**	**0.18**	**0.527**	**0.176**	**0.783**	**0.176**
Intercept		0.035	0.195	-0.011	0.184	0.036	0.194

[Note: Variables that are statistically significant at 90% confidence interval are presented in bold font]

Dissimilarities exist between the urban crash analysis and the rural crash analysis. AADT is significant only in predicting decision error, meaning the probability of making performance or recognition errors in an urban setting does not vary by AADT. Posted speed limit does not affect decision error, which is counterintuitive because one of the major driver errors in this category is "Exceeding Speed Limit". Plausibly, crashes related to speed violations may occur at any posted speed limit. Numbers of lanes, lane width, shoulder width, and passing percent all have positive effects on driver errors.

Decision and performance errors occur mainly on interstate highways in urban areas. However, in other highway types, the highest increase in probability has been observed for recognition error compared with no error. The roadway type variable is not significant in predicting recognition error, but it is significant for both decision and performance error at all levels. Divided highways increase the probability of both decision and performance error compared to no error. In urban areas, drivers are least likely to make performance mistakes when ice is on the roadway. Other explanatory variables show trends similar to rural crashes.

7.4 Discussion: Driver Error Contributing Factors

Driver errors related to a crash are caused by a confluence of factors. Table 7-7 provides a review of marginal effects of variables relating to driver errors with a statistically significant confidence interval of 90% for rural highway crashes. The marginal effect has varying definitions based on the variable type. The marginal effect for a continuous variable is the difference in the probability at each level following a one-unit change in the independent variables. The marginal effect for a categorical variable is calculated as the changes in the probabilities for each level caused by a change in the value from its base level. The variables can be viewed individually or collectively to examine their effects on driver errors.

Recognition errors are more likely to happen on interstate highways, followed by undivided highways and one-way streets, as a higher level of traffic volume increases the probability of these errors. Even if an interstate highway is well designed, situations such as high traffic, high speed, complex interchanges, and exits can increase the potential for human error. The driving situation can become even more challenging if vertical and/or horizontal curves are present. Weather can also create issues, as foggy/cloudy and/or windy conditions lead to more recognition mistakes. Recognition errors also have a higher probability of happening at night when (street) lights are present.

However, drivers are less likely to commit a recognition error when the pavement is either wet or covered in snow, when snow/sleet/rain is occurring, or during nighttime when lights are not present. This suggests that drivers typically adjust their behavior in response to the perceived level of risk, becoming more careful when they sense greater risk. Similarly, when visibility is low or roadway debris is present, risk compensation is found again, as the probability of recognition error is low.

Decision errors stem from misjudgment, especially when it comes to vehicle operating speed. Decision errors are more likely to take place on interstate highways or on one-way streets. The presence of roadway alignments worsens the situation. A high level of traffic volume or a high percentage of truck traffic may perpetuate decision errors such as following too close or improper overtake. The only physical deterrent for decision errors was found to be poor pavement condition (i.e., large rutting value). Adverse weather (e.g., fog/cloudy, snow/sleet, rain) and/or slippery pavement (snow, ice, and wet), as well as nighttime conditions irrespective of the availability of street lighting, will increase the probability of decision errors that are related to speed, aggressive driving behavior, and disregarding traffic controls. Another noteworthy finding is that work zones may see a higher probability of decision error.

Table 7-7 Review of Marginal Effects for Rural Crashes.

Variable	Recognition Error	Decision Error	Performance Error
Traffic Variables	-	Truck (-0.0001)	Truck (-0.0001)
	AADT (0.0002)	AADT (0.0001)	-
Roadway Geometry	-	-	Lanes (-0.015)
	Speed (-0.002)	Speed (-0.003)	Speed (0.001)
Highway Type (base: Interstate)	Other highways (-0.028)	Other highways (-0.076)	Other highways (0.014)
	-	State highways (-0.032)	State highways (0.006)
Roadway Type (base: Divided)	Undivided (0.026)	Undivided (-0.042)	Undivided (0.015)
	One-way (-0.043)	One-way (0.041)	-
Alignment	Ver.Curve = Yes (0.010)	Hor.Curve = Yes (0.0177)	Hor.Curve = Yes (0.048)
	Hor.Curve = Yes (0.009)	Ver.Curve = Yes (0.0129)	Ver.Curve = Yes (0.010)
Pavement	-	Rutting (-0.047)	-
Roadway Condition (base: Dry)	Snow (-0.197)	Snow (0.342)	
	Wet (-0.048)	Wet (0.094)	Wet (0.015)
		Ice (0.340)	
Weather Condition (base: Clear)	Fog/Cloudy (0.007)	Fog/Cloudy (0.011)	Fog/Cloudy (0.017)
	Snow/Sleet (-0.085)	Snow/Sleet (0.061)	Snow/Sleet (0.025)
	Rain (-0.045)	Rain (0.082)	-
	Wind (0.137)	-	-
Lighting Condition (base: Day)	Night-Unlit (-0.012)	Night-Unlit (0.061)	Night-Unlit (0.026)
	Night-Lit (0.021)	Night-Lit (0.056)	Night-Lit (0.020)
Events	Debris = Yes (-0.145)	Debris = Yes (-0.226)	Debris = Yes (-0.131)
	-	Work Zone = Yes (0.145)	Work Zone = Yes (0.084)
	Visibility: Yes (-0.058)	-	-
Impairment	Alcohol (-0.028)	Alcohol (0.022)	Alcohol (0.073)
	-	Drug (0.021)	Drug (0.060)
Age (base: Adolescent)	Old (-0.022)	Old (-0.112)	Old (0.023)
	Adult (-0.040)	Adult (-0.033)	-
	Middle age (-0.046)	Middle age (-0.058)	Middle age (-0.016)
Gender (base: Female)	-	-	Male (-0.003)
Vehicle type (base: Heavy truck)	Motorcycle (-0.104)	Motorcycle (0.032)	Motorcycle (0.133)
	-	Passenger car (0.053)	Passenger car (0.014)
	-	Light truck (0.047)	-

[Note: Marginal effect presented with "-" is not significant at 90% confidence interval]

Performance error is most probable with changes in roadway geometry and driving environment compared with the other error types. The probability of performance error is high for non-interstate highways, narrow lanes, high posted speed limit, horizontal or vertical alignments, adverse weather, wet pavement surfaces, and night driving. Performance error, not unlike decision error, is more likely to occur in a work zone.

Human factors are a critical factor for all error types. Decision and performance errors are increased with the use of alcohol and drugs. The negative effect of alcohol on recognition errors requires further investigation. Adult drivers make fewer mistakes than teen drivers, however, drivers older than 65 seem to have the highest chance of performance error. Gender differences did not exist for recognition or decision errors, but male drivers seem to make fewer performance mistakes. Truck drivers have a low probability overall for errors of all types, possibly because they are following more safety regulations. Motorcycle riders have the lowest probability of recognition errors and the highest probability of performance errors.

Table 7-8 provides estimates of the marginal effects of covariates for urban crashes. Similar to rural crashes, the estimated marginal effects that were statistically significant at a 90% confidence interval were shown in the table.

Table 7-8 Review of Marginal Effects for Urban Crashes.

Variable	Recognition Error	Decision Error	Performance Error
Traffic Variables	-	Truck (-0.002)	Truck (0.0001)
	AADT (-1.45E-7)	AADT (4.89E-7)	-
Roadway Geometry	-	Speed (0.002)	Speed (-0.002)
	-	-	Lanes (0.013)
	Lane Wid (0.004)	Lane Wid (0.006)	-
	-	-	Shoulder Wid (0.003)
	Percent Passing (0.001)	-	-
Highway Type (base: Interstate)	Other highways (0.011)	Other highways (-0.070)	-
	State highways (-0.015)	State highways (-0.019)	State highways (-0.057)
Roadway Type (base: Divided)	Undivided (0.022)	Undivided (-0.029)	-
	-	One-way (-0.026)	One-way (-0.039)
Alignment	Hor.Curve = Yes (-0.060)	-	Hor.Curve = Yes (0.075)
	Ver.Curve = Yes (-0.006)	Ver.Curve = Yes (0.014)	Ver.Curve = Yes (-0.017)
Pavement	Rutting (-0.084)	Rutting (-0.105)	Rutting (0.164)
Roadway Condition (base: Dry)	Snow (-0.159)	Snow (0.157)	Snow (0.031)
	Wet (-0.040)	Wet (0.037)	Wet (0.041)
	Ice (-0.210)	Ice (0.200)	-
Weather Condition (base: Clear)	-	Fog/Cloudy (0.011)	Fog/Cloudy (0.016)
	Snow/Sleet (-0.074)	Snow/Sleet (0.073)	Snow/Sleet (0.030)
	Rain (-0.056)	Rain (0.067)	-
	-	Wind (0.091)	-
Lighting Condition (base: Day)	Night-Unlit (-0.004)	Night-Unlit (0.061)	Night-Unlit (0.008)
	Night-Lit (-0.002)	Night-Lit (0.059)	Night-Lit (0.002)
Events	Debris = Yes (-0.142)	Debris = Yes (-0.230)	Debris = Yes (-0.142)

	Work Zone = Yes (-0.003)	Work Zone = Yes (0.096)	Work Zone = Yes (-0.040)
	Visibility: Yes (-0.075)	Visibility: Yes (-0.162)	Visibility: Yes (0.183)
Impairment	Alcohol (-0.033)	Alcohol (-0.074)	Alcohol (0.064)
		Drug (-0.051)	-
Age (base: Adolescent)	Young Adult (-0.006)	-	-
	Adult (-0.052)	Adult (-0.053)	-
	Middle age (-0.015)	Middle age (-0.075)	Middle age (-0.011)
	-	Old (-0.134)	Old (0.075)
Gender (base: Female)	Male (-0.013)	-	-
Vehicle type (base: Heavy truck)	Passenger car (0.023)	Passenger car (0.079)	Passenger car (0.031)
	Motorcycle (-0.095)	-	Motorcycle (0.151)
	Light truck (0.028)	Light truck (0.074)	Light truck (0.027)

[Note: Marginal effect presented with "-" is not significant at 90% confidence interval]

Similar to rural crashes, traffic characteristics, roadway geometric design, weather, and pavement conditions significantly affect the probability of all types of errors on urban roads. However, some variables seem to have strikingly different effects across different error types. The highest probability of recognition errors occurs on undivided other highways. Higher traffic volume seems to be associated with a lower probability of recognition errors, while wider lane width and higher passing percentage show the opposite. The presence of horizontal or vertical curves, poor pavement conditions, nighttime driving, adverse weather events, slippery pavement, as well as challenges such as debris, work zones, and obstructed visibility, all lower the probability of recognition errors on urban highways. This finding supports the theory of risk compensation.

The effects of variables relating to decision errors are very consistent between rural and urban highway crashes. One exception is that a higher percentage of trucks may increase the probability of decision errors. In addition, lane width is found to be statistically significant in increasing the chance of recognition errors.

The comparison of variables affecting performance errors between rural and urban highway crashes are mixed. Findings are inconsistent with regard to some variables being associated with a higher probability of performance errors (interstate highways, high percentage of trucks, wider lane width and/or shoulder width, and poor pavement conditions) and others being associated with a lower probability of performance errors (including posted speed limit and presence of a vertical curve). One consistent finding is with regard to a higher probability of performance error when a horizontal curve, adverse weather, wet pavement surfaces, night conditions, or work zone is present.

Human factors such as age, gender, type of vehicle driven, and driver impairment affect the chances of driver errors that lead to crashes. However, alcohol and/or drugs are found to negatively affect decision errors, which goes against conventional wisdom. Another counterintuitive observation is that alcohol negatively affects recognition errors made in both rural and urban highway crashes. These questionable findings deserve further investigation. A discrepancy was also found when looking at the effect of gender on error types. Males have a lower probability of recognition error for urban highway crashes, but no statistically significant difference is found for performance error.

7.5 Discussion: Driver Error by Severity Scale

The coefficient estimates of the OPM regression models for rural and urban crashes are presented in Table 7-9 and Table 7-10, respectively. Driver characteristics and behavior appear to have a large influence on the error severity outcome. Younger male drivers have a significantly higher probability of making severe mistakes when compared to other age groups, except older females. This finding is consistent with a previous study where reckless driving was most prevalent among male drivers under the age of 25 (*47*) . Younger drivers are more likely to violate red signals due to their more aggressive and risky driving behavior (*7; 47; 48*). Old drivers behave consistently irrespective of area types. Older females are the most vulnerable drivers at intersections, where they are more likely than any other driver group to make severe mistakes at all types of intersections. Older male drivers are more prone to errors at uncontrolled intersections in rural areas and at sign-controlled intersections in rural areas. Traffic signs can be a challenge for older drivers due to deteriorating vision, slower recovery from glare, and misjudgment of gap or speed of other vehicles. These challenges may contribute to an increased likelihood of making more severe mistakes at sign-controlled intersections (*49*). Findings indicate that drivers over 60 years of age are more likely to fail to stop or fail to yield the right-of-way at an intersection (*50*-*52*). From a behavior perspective, the probability of making severe mistakes while under the influence drastically increases at intersections with all kinds of traffic controls, regardless of the age or gender of a driver. Compared to passenger car drivers, truck drivers are more likely to avoid severe mistakes. This may be because truck operators have more driving experience in general. Bonneson *et al.* suggested that heavy vehicles are more likely to be involved in running a red light (*53*).

In terms of highway design factors, Devlin *et al.* discovered the increased probability of failing to notice traffic signs at sign-controlled intersections at vertical and horizontal curves (*42*). However, this study analysis shows that the coefficient of roadway alignment seems to suggest that driver error severity decreases at horizontal and/or vertical curves at intersections of all types of traffic control. It is noted that the coefficients of the vertical curve are positive at uncontrolled intersections, but they are not statistically significant. Drivers have a higher chance of committing more severe mistakes when their vision is obscured. The chances of limited stop sight distance affected by vertical curves may be low in Wisconsin, due to the state’s mostly flat terrain.

Drivers are more likely to commit serious mistakes at rural sign-controlled intersections with intermediate posted speed limit (35-55mph) or at urban uncontrolled intersections with a high posted speed limit (>55mph). High-speed uncontrolled intersections prompt the question of potential coding errors by investigating law enforcement officers. A follow-up review shows that there are 273 crashes coded at urban intersections with no traffic controls and speed limits higher than 55mph.

All inclement weather conditions and adverse roadway surface conditions seem to be associated with lower severity of driver errors, which might be due to the fact that most drivers compensate for risk by reducing speed in these conditions (*54*). Nighttime driving seems to be associated with low driver error severity at all intersections except for signalized intersections with street lights. Initially, this finding seems to be counterintuitive, but as the study is focused on factors contributing to driver error severity for any intersection-related crashes, this could indicate that drivers are prone to disregarding traffic signals at night and that street lights may increase the risk for violation.

The thresholds of the ordered probit model can offer a clear hierarchy for driver errors distributed among three types of intersections. It is obvious that the distributions of driver errors vary among the intersection types. The probability of making mistakes at sign-controlled intersections is the highest because the threshold from no error to any error is the lowest ($\mu > -1.804$ in rural areas and $\mu > -1.683$ in urban areas); signal-controlled intersections are the next highest ($\mu > -1.575$ in rural areas and $\mu > -1.372$ in urban areas), and uncontrolled intersections ($\mu > -1.450$ in rural areas and $\mu > -1. 261$ in urban areas) are last. The probability of making reckless driver errors at sign-controlled intersections is highest because the threshold from careless driving to reckless driving is the lowest ($\mu > -0.342$ in rural areas and $\mu > -0.645$ in urban areas), followed by signal-controlled intersections ($\mu > 0.136$ in rural areas and $\mu > 0.047$ in urban areas) and then uncontrolled intersections ($\mu > 0.545$ in rural areas and $\mu > 0.330$ in urban areas).

Table 7-9 Coefficient Estimates of OPM for Rural Crashes.

Variable	Value	Uncontrolled	Sign-controlled	Signal-controlled
Age and Gender	male driver (age<25)	Base level		
	male driver (age 25-55)	-0.174 (0.00)	-0.066 (0.04)	-0.165 (0.00)
	male driver (age>55)	0.029 (0.57)	0.124 (0.00)	-0.074 (0.21)
	female driver (age<25)	-0.130 (0.01)	0.042 (0.30)	-0.087 (0.13)
	female driver (age 25-55)	-0.228 (0.00)	-0.039 (0.26)	-0.259 (0.00)
	female driver (age>55)	0.018 (0.75)	0.189 (0.00)	0.118 (0.06)
DUI (alcohol or drugs)		0.430 (0.00)	0.301 (0.00)	-0.014 (0.90)
Horizontal curve		-0.127 (0.00)	-0.221 (0.00)	-0.293 (0.00)
Vertical curve		0.035 (0.37)	-0.089 (0.00)	-0.169 (0.00)
Posted speed limit	low (<35mph)	Base level		
	middle (35mph-55mph)	0.045 (0.12)	0.059 (0.01)	-0.020 (0.52)
	high (>55mph)	0.029 (0.66)	-0.174 (0.00)	-0.169 (0.04)
Accident Time	AM peak (7:00am-9:59am)	Base level		
	day time (10:00am-3:59pm)	-0.01 (0.85)	-0.062 (0.03)	0.053 (0.21)
	PM peak (4:00pm-6:59pm)	0.035 (0.32)	-0.045 (0.10)	-0.001 (0.99)
	night time (7:00pm-6:59am)	-0.048 (0.44)	-0.080 (0.06)	-0.059 (0.36)
Weather	clear	Base level		
	cloudy	0.053 (0.10)	0.059 (0.01)	-0.010 (0.77)
	rain	-0.065 (0.42)	-0.044 (0.46)	-0.087 (0.27)
	snow	-0.11 (0.11)	-0.216 (0.00)	-0.133 (0.11)
Light	day	Base level		
	night without street light	-0.147 (0.01)	-0.087 (0.03)	-0.113 (0.41)
	night with street light	-0.027 (0.65)	-0.022 (0.61)	0.289 (0.00)
Road	dry	Base level		
	wet	-0.012 (0.84)	-0.027 (0.54)	-0.01 (0.90)
	snow	-0.255 (0.00)	-0.659 (0.00)	-0.440 (0.00)
	ice	-0.482 (0.00)	-1.074 (0.00)	-0.877 (0.00)
Vehicle type	pc	Base level		
	light truck	-0.067 (0.07)	-0.045 (0.11)	-0.073 (0.10)
	heaver truck	-0.355 (0.00)	-0.372 (0.00)	-0.615 (0.00)
$\mu 1$ (No error -> Improper Driving)		-1.450	-1.804	-1.575
$\mu 2$ (Improper Driving -> Careless Driving)		-0.692	-1.289	-1.166
$\mu 3$ (Careless Drving -> Reckless Driving)		0.545	-0.342	0.136

Table 7-10 Coefficient Estimates of OPM for Urban Crashes.

Variable	Value	Uncontrolled	Sign-controlled	Signal-controlled
Age and Gender	male driver (age<25)	Base level		
	male driver (age 25-55)	-0.022 (0.43)	**-0.135 (0.00)**	**-0.122 (0.00)**
	male driver (age>55)	**0.084 (0.02)**	0.012 (0.70)	-0.024 (0.30)
	female driver (age<25)	0.001 (0.98)	0.014 (0.65)	**-0.044 (0.04)**
	female driver (age 25-55)	**-0.066 (0.02)**	**-0.091 (0.00)**	**-0.112 (0.00)**
	female driver (age>55)	**0.144 (0.00)**	**0.090 (0.01)**	**0.102 (0.00)**
DUI (alcohol or drugs)		**0.324 (0.00)**	**0.121 (0.04)**	**0.200 (0.00)**
Horizontal curve		**-0.214 (0.00)**	**-0.361 (0.00)**	**-0.294 (0.00)**
Vertical curve		0.002 (0.94)	**-0.150 (0.00)**	**-0.070 (0.00)**
Posted speed limit	low (<35mph)	Base level		
	middle (35mph-55mph)	0.007 (0.73)	**-0.190 (0.00)**	**-0.025 (0.04)**
	high (>55mph)	**0.290 (0.00)**	**-0.272 (0.00)**	**-0.153 (0.00)**
Accident Time	AM peak (7:00am-9:59am)	Base level		
	day time (10:00am-3:59pm)	0.020 (0.45)	0.002 (0.93)	-0.001 (0.99)
	PM peak (4:00pm-6:59pm)	**-0.061 (0.00)**	**-0.070 (0.00)**	**-0.090 (0.00)**
	night time (7:00pm-6:59am)	**-0.154 (0.00)**	**-0.082 (0.02)**	**-0.107 (0.00)**
Weather	clear	Base level		
	cloudy	**0.064 (0.00)**	**0.077 (0.00)**	**0.074 (0.00)**
	rain	-0.038 (0.42)	**0.075 (0.08)**	**-0.063 (0.03)**
	snow	**-0.207 (0.00)**	**-0.204 (0.00)**	**-0.107 (0.00)**
Light	day			
	night without street light	-0.065 (0.31)	-0.092 (0.14)	-0.092 (0.17)
	night with street light	**-0.070 (0.02)**	**-0.110 (0.00)**	**0.062 (0.00)**
Road	dry	Base level		
	wet	-0.024 (0.502)	0.005 (0.88)	0.003 (0.88)
	snow	**-0.099 (0.01)**	**-0.369 (0.00)**	**-0.421 (0.00)**
	ice	**-0.439 (0.00)**	**-1.057 (0.00)**	**-0.939 (0.00)**
Vehicle type	pc	Base level		
	light truck	**-0.123 (0.00)**	**-0.065 (0.01)**	**-0.057 (0.00)**
	heaver truck	**-0.497 (0.00)**	**-0.590 (0.00)**	**-0.671 (0.00)**
μ1 (No error -> Improper Driving)		-1.261	-1.683	-1.372
μ2 (Improper Driving -> Careless Driving)		-0.823	-1.334	-0.993
μ3 (Careless Drving -> Reckless Driving)		0.330	-0.645	0.047

7.6 Summary and Recommendations

Driver error was involved in more than 90 percent of crashes on roadway segments. Driver error can be categorized as recognition, decision, performance, or non-performance depending on the physical definition of each error category introduced in NMVCCS. The factors contributing to these errors can be complex, and can include highway and traffic characteristics, environmental factors, roadway events, driver characteristics, and the type of vehicle.

A statistical relationship was established between driver errors and a series of factors including roadway, traffic, and crash data elements. MNP models were applied to quantify the effects of each explanatory variable. The model results suggest that many of the variables (roadway geometry, highway classification, traffic characteristics, roadway events, driver-related) are statistically correlated with different driver error categories in both rural and urban areas. Dissimilarities were found by comparing results between rural and urban crashes; the differences in safety culture between these areas seems to be a large contributor to these dissimilarities.

A review was conducted for rural crashes using marginal effects from the MNP model in order to better understand the impact of contributing factors of driver errors. The marginal effects of each explanatory variable represent the quantity of increase or decrease in the probability of a specific driver error type. Thus, each error category can be characterized by a combination of unique variables that help to differentiate future safety treatments.

Variables relating to different levels of error severity were studied intently with regard to intersection-related crashes. The OPM model was employed to account for the ordinal nature of driver errors, from minor infractions to reckless violations. The results unambiguously state that human factors strongly influence the number and type of driving errors.

Younger drivers are more likely to be reckless on the road compared to other drivers. Older drivers, especially females, have a higher probability of making severe mistakes at all three types of intersections. Accordingly, driver education and training about proper vehicle control and crash risk are recommended for teenagers and older drivers (*42*). This study recommends increasing the conspicuity of traffic signs and signals for older drivers by increasing the size of signs and installing additional warning signs as drivers approach intersections (*42; 55; 56*). Alcohol and drug use dramatically increases the probability of severe driver errors; hence, measures such as random breathing tests and public anti-drug and alcohol campaigns are recommended (*42; 57*).

Uncontrolled intersections have a very high percentage of both careless driving and reckless driving errors. Relevant countermeasures, such as installing “intersection ahead” warning signs, can help raise drivers’ awareness (*58*). Additionally, setting appropriate speed limits to improve drivers’ gap selection is also recommended (*56*). In particular, visibility obstruction is found to significantly increase driver error severity at uncontrolled intersections.

The elimination of objects that obstruct the driver's vision might help the driver maintain visibility from all directions (*42*).

Sign-controlled intersections have the highest percentage of driver errors and the highest percentage of reckless driving errors. Increasing the visibility and conspicuity of stop signs by installing them on both the left and right-hand sides of the road may help lower this percentage (*59*). Additionally, installing rumble strips across the lane may prompt drivers to slow down when approaching intersections (*59*). Careless driver errors are closely associated with driving during the night, which is likely due to driver fatigue. Drivers should take turns driving and use rest areas to prevent driver fatigue (*60*). FHWA recommends raising pavement markers and installing reflective strips on traffic signs to improve nighttime driving safety at unlighted or dark sign-controlled intersections (*56*). FHWA also recommended setting a lower speed limit to improve traffic safety at sign-controlled intersections, especially for trucks, which have much longer stopping distances (*56*).

The probability of severe driver errors occurring at signalized intersections increases with visibility obstruction and high posted speed limit. Devlin *et al.* suggested increasing the visibility of traffic lights by removing objects that obscure the driver's vision (*42*). FHWA recommended setting appropriate speed limits to account for roadway design, traffic, and environmental conditions (*56*). Additionally, increasing the length of yellow light time (not to exceed 5.5 seconds) or installing "no turn on red" signage have also been proven to efficiently prevent crashes due to running red lights (*42; 56*).

8. VARIABLES ASSOCIATED WITH PEDESTRIAN AND BICYCLE CRASHES

The *Wisconsin Strategic Highway Safety Plan, 2017-2020* includes “Improve Non-Motorist Safety” as one of ten “Highest Priority Issue Areas” (*61*). Between 2012 and 2016, pedestrians and bicyclists accounted for 9.9% of all traffic fatalities (55 per year), 9.8% of all incapacitating injuries (311 per year), and 5.3% of all injuries (2,171 per year) (*61*). Reducing the impact of pedestrian and bicyclist crashes is important for the health and vitality of individuals, families, and communities throughout Wisconsin.

This section of the report describes an exploratory analysis of roadway and surrounding environment variables associated with pedestrian and bicyclist crashes in Wisconsin. Our primary analysis focused on a random sample of 200 one-mile corridors along the Wisconsin State Highway System in areas with at least 100 residents per square mile. Ideally, we hoped to develop a model to predict pedestrian and bicyclist crashes along state highway corridors in these areas.

8.1 Previous Studies

There is extensive research on individual pedestrian and bicycle safety treatments. For example, the Federal Highway Administration summarizes the results of many studies in guidance such as PedSAFE (*62*), BikeSAFE (*63*), and Proven Safety Countermeasures (*64*). However, there relatively few pedestrian or bicyclist crash models that simultaneously control for multiple variables and can be used to predict pedestrian or bicyclist crashes in a particular location (*65-70*). The most common variables associated with pedestrian or bicyclist crashes in these models are pedestrian or bicyclist volume and motor vehicle volume (sometimes differentiated by turning movement or intersection approach leg). Higher levels of pedestrian, bicyclist, and motor vehicle activity tend to be associated with more pedestrian and bicyclist crashes. However, the risk of a pedestrian or bicyclist crash per pedestrian or bicyclist generally decreases when there are more pedestrian and bicyclist activity (*66; 68*). After controlling for these exposure variables, several studies have identified specific pedestrian facilities to be negatively associated with pedestrian crashes, such as median refuge islands (*66; 70*) and rectangular rapid flashing beacons (*70*).

None of the existing models could be applied to the Wisconsin State Highway System. Several of the models are based on data sets from a single community and require intensive data collection to gather pedestrian and bicyclist volume data and explanatory variables. Most models applied to intersections rather than an entire roadway corridor. To overcome these issues, other studies have used proxy variables to represent pedestrian and bicyclist activity levels, which can create difficulty describing the relationship with other explanatory variables accurately (*71; 72*). Some researchers have used conflict-prediction models rather than crash prediction models to increase the data available to quantify the safety outcome (*73*).

8.2 Selection of Wisconsin State Highway Study Corridors

We selected 200 one-mile study corridors so that they were located in areas with more than 100 people per square mile. This generally included cities, suburbs, and villages but excluded rural areas. We focused on more urbanized areas because they tend to have higher volumes of pedestrians and bicyclists and more pedestrian and bicyclist crashes.

Our selection process involved identifying all census tracts with more than 100 people per square mile and imposing an imaginary 200m by 200m grid on the selected census tracts using GIS. Then we selected the 200m by 200m cells that contained a state highway. This set of cells was assigned an identification number, and an initial set of 200 identification numbers were selected randomly. These geographic center of these cells were considered as the starting points for one-mile corridors along the state highway. For each of the 200 starting points, we then selected a direction randomly (north or south; east or west) and measured one mile in that direction to define the corridor.

Note that some of the corridors that were selected initially were not used. Some initial selection points were less than one mile from the end of a state highway, so it was not possible to identify a full mile corridor. Other corridors contained a roundabout, which was a feature excluded from this analysis. In these cases, the initially-selected starting point was replaced with a new starting point that produced a one-mile corridor that was eligible to study. Figure 1 shows the locations of the 200 study corridors throughout the state.

8.3 Crash Data

The pedestrian and bicycle crash analyses are based on police-reported crashes that occurred along 200 one-mile corridors within the Wisconsin state highway system. These crashes include all reported collisions between a motor vehicle and either a pedestrian and bicyclist on public or private property within 100 feet on either side of the roadway centerline during the ten-year period from 2006 to 2015. They only include crashes that are contained in the WisTransPortal Database and have latitude and longitude coordinates allowing them to be mapped to the corresponding corridors. Crashes that occurred within the corridors but were not reported to police, not contained in WisTransPortal Database, or not geocoded were not analyzed. Note that the ten-year time period provides more crash data for analysis, but it increases the chances that a particular corridor had different characteristics when the earliest crashes occurred.

Ten-year crash totals within the 200 corridors range from zero to 45 for pedestrians and zero to 30 for bicyclists. Of the 200 corridors, 102 had at least one reported pedestrian crash and 91 had at least one reported bicycle crash.

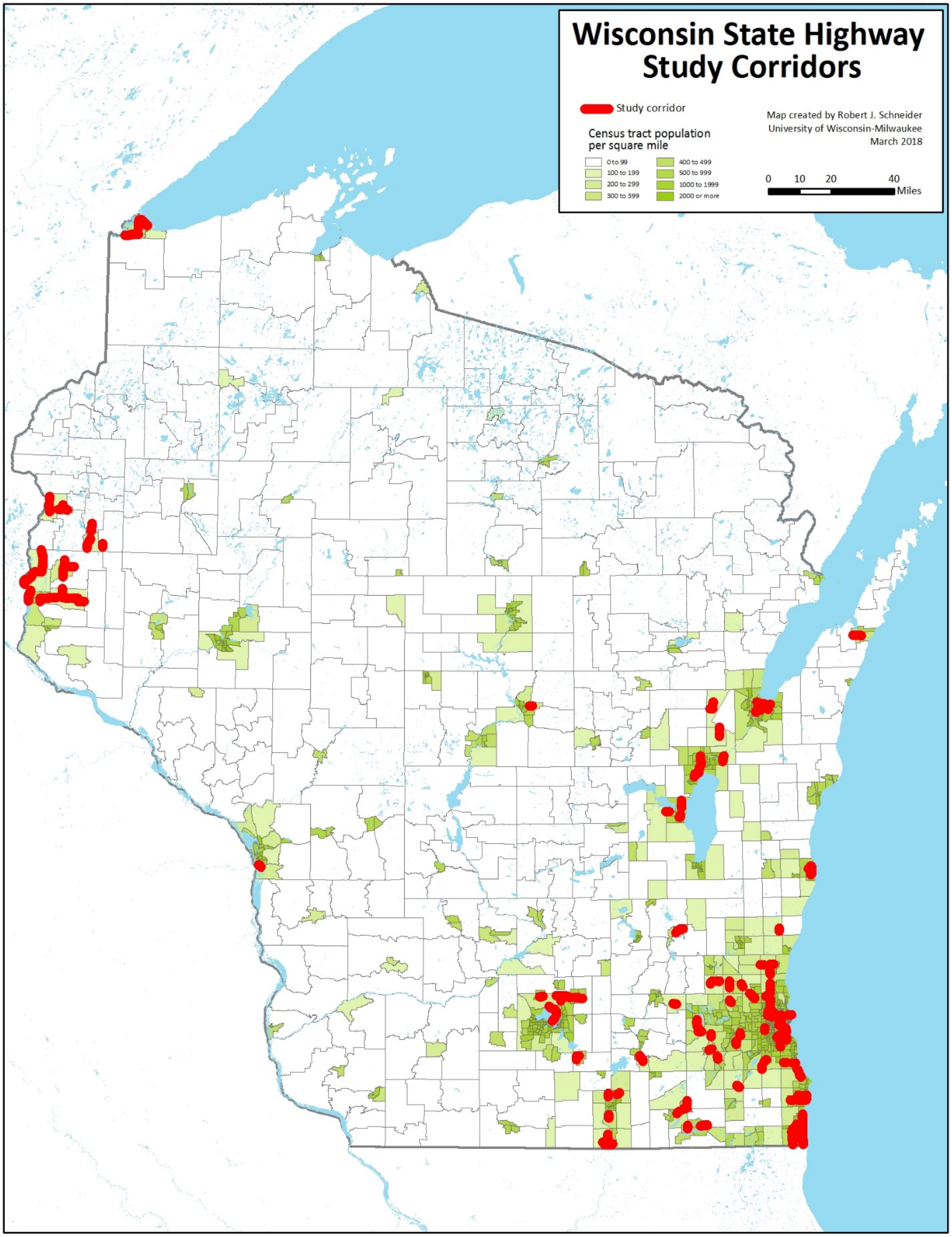


Figure 8-1 State Highway Study Corridors used for Pedestrian and Bicycle Crash Models.

8.4 Explanatory Variables

Several categories of explanatory variables were collected for consideration during the modeling process. These included exposure, roadway segment characteristics, and roadway intersection characteristics. The explanatory variables are summarized in Table 8-1.

Exposure Variables

Exposure variables represent walking and bicycling activity within the corridors. All else equal, more activity is typically associated with more reported crashes, so these variables are important to control in order to identify the impact of various roadway design features on pedestrian and bicyclist safety.

Ideally, annualized pedestrian and bicyclist volumes would be available for each state highway corridor. However, the pedestrian and bicyclist traffic monitoring field is still in its infancy, so none of the study corridors had pedestrian or bicyclist counts. Like several other previous studies, we used proxy variables based on surrounding neighborhood characteristics to represent the level pedestrian and bicyclist activity within the corridors. These proxy variables were:

- Population density (people per acre)
- Natural log of population density (people per acre)
- Total job density (jobs per acre)
- Natural log of job density (jobs per acre)
- Retail job density (retail jobs per acre)
- Square root of retail job density (retail jobs per acre)
- Percentage of households with no automobile
- Square root of households with no automobile
- Number of bus stops
- Square root of number of bus stops

Each of these variables was created using data from census block groups containing the corridor. Block group variable values were averaged for corridors within multiple census block groups. These data were available statewide from the United States Environmental Protection Agency Smart Location Database (*74*). The square root of retail job density, households with no automobile, and bus stops were used rather than the natural log because some corridors had a value of zero for these variables.

Roadway Segment and Intersection Characteristics

Roadway geometric characteristics and pedestrian and bicycle facilities may be related to pedestrian and bicyclist safety. We summarized characteristics of roadway segments and intersections within each corridor. Roadway segment characteristics included:

- Average pavement width

- Average number of through lanes
- Percentage of corridor with a median (also dummy variable for median along at least 50% of corridor)
- Percentage of corridor with a two-way left-turn lane (also dummy variable for two-way left-turn lane along at least 50% of corridor)
- Percentage of corridor with a sidewalk (also dummy variable for sidewalk along at least 50% of corridor)
- Percentage of corridor with a paved shoulder at least 4 feet wide (also dummy variable for paved shoulder along at least 50% of corridor)
- Percentage of corridor with a designated bike lane (also dummy variable for designated bike lane along at least 50% of corridor)
- Percentage of corridor with a sidepath (also dummy variable for median along at least 50% of corridor)
- Posted speed limit
- Pedestrian facility factor (value between 0 and 3; one point each for at least 50% sidewalk coverage, at least 50% median coverage, and at least one curb extension in the corridor)
- Bicycle facility dummy variable (value of 1 if the corridor has a bike lane, paved shoulder, or sidepath along at least 50% of the corridor)
- Complete street dummy variable (value of 1 if the corridor has at least one of the following pedestrian facilities: at least 50% sidewalk coverage, at least 50% median coverage, or at least one curb extension and at least one of the following bicycle facilities: at least 50% bike lane coverage, at least 50% paved shoulder coverage, and at least 50% sidepath coverage)

Motor vehicle annualized average daily traffic (AADT) was also considered to be a roadway characteristic for modeling pedestrian and bicycle crashes. AADT was gathered for each state highway corridor from the WISLR database. Two variables were considered:

- State highway AADT
- Natural log of state highway AADT

When several traffic volume measurements were available along the corridor, these were averaged. Note that this traffic volume variable does not include the traffic volumes on streets intersecting the corridor. Roadway intersection characteristics included:

- Number of roadway intersections
- Number of signalized intersections
- Number of unsignalized intersections
- Number of turn lanes on state highway approaches to all intersections
- Number of left-turn lanes on state highway approaches to all intersections
- Number of right-turn lanes on state highway approaches to all intersections
- Number of marked crosswalks across the state highway
- Number of marked midblock crosswalks across the state highway

- Number of crosswalks with a curb extension on at least one end (also dummy variable for at least one curb extension in the corridor)
- Number of residential driveways
- Number of non-residential driveways

Most of these roadway corridor attributes were collected by students from Google Maps and Google Street View imagery. These images were from 2016 and 2017. Detailed instructions used to create these roadway variables are provided in Appendix A.

Table 8-1 Variables Used in State Highway Corridor Pedestrian and Bicycle Crash Analysis.

Variable	Description	Sample	Mean	Std Dev	Min	Max
Length_Mil	Length of corridor (miles)	200	0.99	0.04	0.83	1.23
Ped0615	Reported pedestrian crashes (2006-2015)	200	3.67	6.32	0.00	45.00
Bike0615	Reported bicycle crashes (2006-2015)	200	2.61	4.78	0.00	30.00
Ped1115	Reported pedestrian crashes (2011-2015)	200	1.80	3.23	0.00	19.00
Bike1115	Reported bicycle crashes (2011-2015)	200	1.18	2.32	0.00	16.00
PopAcre10	Avg. population per acre in surrounding Census BGs	200	4.17	4.61	0.06	22.62
LnPopAcre	Natural log of Avg. population per acre in surrounding Census BGs	200	0.52	1.58	-2.77	3.12
TotJobAcre	Avg. total jobs per acre in surrounding Census BGs	200	2.18	3.77	0.00	20.77
LnJobAcre	Natural log of Avg. jobs per acre in surrounding Census BGs	200	-0.85	2.17	-6.17	3.03
RetJobAcre	Avg. retail jobs per acre in surrounding Census BGs	200	0.18	0.35	0.00	2.89
RtRetAcre	Square root of Avg. retail jobs per acre in surrounding Census BGs	200	0.30	0.30	0.00	1.70
Pct0Veh	Avg. percentage of households with no vehicle in surrounding Census BGs	200	0.07	0.07	0.00	0.34
RtPct0Veh	Square root of Avg. percentage of households with no vehicle in surrounding Census BGs	200	0.22	0.13	0.00	0.58
BusStops	Number of bus stops within 100 feet of the state highway centerline	200	2.78	5.17	0.00	20.00
RtBusStops	Square root of the number of bus stops within 100 feet of the state highway centerline	200	0.86	1.43	0.00	4.47
CBGs	Number of surrounding Census BGs	200	3.17	1.80	1.00	8.00
NatWalkInd	EPA National Walking Index in surrounding Census BGs	200	8.52	3.71	2.92	17.17
EPAIntSqMi	EPA non-automobile oriented intersections per square mile in surrounding Census BGs	200	48.24	47.31	0.41	194.07
EPASegSqMi	EPA all roadway segments per square mile in surrounding Census BGs	200	11.01	7.99	0.99	28.55
AvgWidthFt	Average curb-to-curb width along corridor	200	59.93	26.96	8.00	145.00
AvgThruLns	Average number of through lanes along corridor	200	3.32	1.30	1.00	6.50
PctMedian	Percentage of corridor length with a median	200	0.40	0.42	0.00	1.00
PctTWLTL	Percentage of corridor length with a two-way left-turn lane	200	0.06	0.16	0.00	1.00
PctSW	Percentage of corridor covered by sidewalks on both sides (sidewalk on only one side for full length = 0.5)	200	0.39	0.44	0.00	1.00

Variable	Description	Sample	Mean	Std Dev	Min	Max
PctShld	Percentage of corridor covered by paved shoulders on both sides (shoulder on only one side for full length = 0.5)	200	0.31	0.43	0.00	1.00
PctBL	Percentage of corridor covered by designated bike lanes on both sides (bike lane on only one side for full length = 0.5)	200	0.09	0.26	0.00	1.00
PctSidePth	Percentage of corridor covered by sidepaths on both sides (sidepath on only one side for full length = 0.5)	200	0.05	0.19	0.00	1.00
SpeedLim	Posted speed limit (miles per hour)	200	38.43	9.43	20.00	65.00
TotInts	Total number of intersections along corridor	200	6.41	4.67	0.00	18.00
SignalInts	Signalized intersections along corridor	200	1.85	2.05	0.00	12.00
UnsigInts	Unsignalized intersections along corridor	200	4.56	3.64	0.00	15.00
LTurnLanes	Left-turn lanes on state highway approaches to intersections along corridor	200	4.33	5.01	0.00	29.00
RTurnLanes	Right-turn lanes on state highway approaches to intersections along corridor	200	2.73	3.67	0.00	21.00
TTurnLanes	Total turn lanes on state highway approaches to intersections along corridor	200	7.06	7.49	0.00	41.00
MarkedXWs	Total marked crosswalks across the state highway along corridor	200	2.80	3.69	0.00	18.00
MrkMidXWs	Marked midblock crosswalks across the state highway along corridor	200	0.05	0.27	0.00	3.00
XWCrbxts	Number of crosswalks across the state highway with a curb extension on either end along corridor	200	0.31	1.43	0.00	11.00
ResDvwys	Residential driveways along corridor	200	9.97	12.81	0.00	57.00
NonResDvwy	Non-residential driveways along corridor	200	11.39	12.32	0.00	58.00
TrailXings	Major trail crossings along corridor ***(To be determined)***	200	0.00	0.00	0.00	0.00
MedianDum	Dummy variable for median along at least 50% of corridor (1 = yes; 0 = no)	200	0.37	0.48	0.00	1.00
TWLTLDum	Dummy variable for two-way left-turn lane along at least 50% of corridor (1 = yes; 0 = no)	200	0.03	0.17	0.00	1.00
SWDum	Dummy variable for sidewalk coverage of at least 50% along the corridor (1 = yes; 0 = no)	200	0.36	0.48	0.00	1.00
ShldDum	Dummy variable for shoulder coverage of at least 50% along the corridor (1 = yes; 0 = no)	200	0.28	0.45	0.00	1.00
BLDum	Dummy variable for bike lane coverage of at least 50% along the corridor (1 = yes; 0 = no)	200	0.09	0.28	0.00	1.00
SidPthDum	Dummy variable for a sidepath coverage of at least 50% along the corridor (1 = yes; 0 = no)	200	0.05	0.21	0.00	1.00
AADT_STH	Average of all Annualized Average Daily Traffic (AADT) volume counts along corridor	200	14247	9379	2690	56440
LnAADT_STH	Natural log of average of all Annualized Average Daily Traffic (AADT) volume counts along corridor	200	9.36	0.66	7.90	10.94
CrbxtDum	Dummy variable for at least one curb extension in the corridor (1 = yes; 0 = no)	200	0.08	0.27	0.00	1.00
PedFacFac	Pedestrian facility factor (value between 0 and 3; one point each for at least 50% sidewalk coverage, at least 50% median coverage, and at least one curb extension in the corridor)	200	0.80	0.83	0.00	3.00
BikeFacDum	Bicycle facility dummy variable (value of 1 if the corridor has a bike lane, paved shoulder, or sidepath along at least 50% of the corridor)	200	0.37	0.48	0.00	1.00
CSFacDum	Complete street dummy variable (value of 1 if the corridor has at least one of the following pedestrian facilities: at least 50% sidewalk coverage, at least	200	0.75	0.44	0.00	1.00

Variable	Description	Sample	Mean	Std Dev	Min	Max
	50% median coverage, or at least one curb extension and at least one of the following bicycle facilities: at least 50% bike lane coverage, at least 50% paved shoulder coverage, and at least 50% sidepath coverage)					

8.5 Variables Correlated with Pedestrian and Bicycle Crashes

Many individual explanatory variables are correlated with pedestrian and bicycle crashes reported from 2006 to 2015 along the 200 study corridors. As a group, the highest correlations are for pedestrian and bicyclist exposure proxy variables (Table 8-2). Among the exposure proxy variables, population density and the percentage of households with no vehicle in surrounding census block groups have the highest correlations with both pedestrian and bicycle crashes.

Table 8-2 Exposure Variables Correlated with Corridor Pedestrian and Bicycle Crashes.

Variable Name	Variable Definition	Correlation with Pedestrian Crashes	Correlation with Bicycle Crashes
PopAcre10	Avg. population per acre in surrounding Census BGs	0.74	0.61
LnPopAcre10	Natural log of population per acre	0.61	0.56
Pct0Veh	Avg. percentage of households with no vehicle in surrounding Census BGs	0.73	0.56
RtPct0Veh	Square root of percentage of households with no vehicle in surrounding Census BGs	0.67	0.55
BusStops	Number of bus stops within 100 feet of the state highway centerline	0.60	0.51
RtBusStops	Square root of the number of bus stops within 100 feet of the state highway centerline	0.59	0.50
TotJobAcre	Avg. total jobs per acre in surrounding Census BGs	0.44	0.53
LnTotJobAcre	Natural log of total jobs per acre	0.52	0.54
RetJobAcre	Avg. retail jobs per acre in surrounding Census BGs	0.28	0.40
RtRetAcre	Square root of retail jobs per acre in surrounding Census BGs	0.39	0.50

Roadway segment variables are also correlated with the number of corridor pedestrian and bicycle crashes (Table 8-3). However, it is essential to recognize that many of these are also highly correlated with pedestrian and bicyclist exposure (e.g., sidewalk coverage is more common in dense urban areas where people walk and bicycle more, posted speed limits are lower in dense urban areas where more people walk and bicycle more). This means that the direction of the relationship between individual explanatory variables and numbers of pedestrian and bicycle crashes do not necessarily represent their relationships with pedestrian and bicycle crash risk.

Table 8-3 Roadway Segment Variables Correlated with Corridor Pedestrian and Bicycle Crashes.

Variable Name	Variable Definition	Correlation with Pedestrian Crashes	Correlation with Bicycle Crashes
PctSW	Percentage of corridor covered by sidewalks on both sides	0.60	0.55
SWDum	Dummy variable for sidewalk coverage of at least 50% along the corridor	0.58	0.51
PedFacFac	Pedestrian facility factor	0.51	0.38
AADT_STH	AADT along corridor	0.41	0.28
LnAADT_STH	Ln of AADT along corridor	0.40	0.31
SpeedLim	Posted speed limit (miles per hour)	-0.39	-0.39
AvgThruLns	Avg. through lanes along corridor	0.34	0.27
PctShld	Percentage of corridor covered by paved shoulders on both sides	-0.17	-0.26
MedianDum	Dummy variable for median along at least 50% of corridor	0.16	0.05
AvgWidthFt	Average curb-to-curb width along corridor	0.16	0.04
PctMedian	% of corridor length with a median	0.14	0.06
BLDum	Dummy variable for bike lane coverage	0.14	0.17

Roadway intersection variables correlated with corridor pedestrian and bicycle crashes include the number of intersections (total, signalized, and unsignalized), marked crosswalks, left-turn lanes, and non-residential driveways (Table 8-4). Similar to the roadway segment variables, many of these intersection variables are correlated with pedestrian and bicyclist exposure (e.g., there are more intersections and marked crosswalks in dense urban areas where more people walk and bicycle more), so their signs may not represent their actual relationship with pedestrian and bicycle crash risk.

Table 8-4 Roadway Intersection Variables Correlated with Corridor Pedestrian and Bicycle Crashes.

Variable Name	Variable Definition	Correlation with Pedestrian Crashes	Correlation with Bicycle Crashes
SignalInts	Signalized intersections along corridor	0.56	0.56
TotInts	Total intersections along corridor	0.55	0.58
MarkedXWs	Total marked crosswalks across corridor	0.54	0.58
LTurnLanes	Left-turn lanes on state highway approaches to intersections along corridor	0.39	0.31
UnsigInts	Unsignalized intersections along corridor	0.39	0.43
NonResDvwy	Non-residential driveways along corridor	0.39	0.36
XWCrbxts	Number of crosswalks across the state highway with a curb extension	0.32	0.18
TTurnLanes	Total turn lanes on state hwy approaches	0.32	0.24
CrbxtDum	Dummy variable for >=1 curb extension	0.26	0.20
RTurnLanes	Right-turn lanes on state highway approaches to intersections along corridor	0.11	0.07

ResDvwys	Residential driveways along corridor	0.06	0.21
MrkMidXWs	Marked midblock crosswalks across the state highway along corridor	-0.06	0.00

8.6 Regression Model Structure

A negative binomial model form was chosen to represent the relationship between the total number of reported crashes in each corridor and the explanatory variables. Separate models were created for pedestrian crashes and bicycle crashes. The negative binomial model uses the following equation:

$$Y_i = exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_j X_{ji})$$

where:
Y_i = total number of reported crashes in corridor *i* from 2006 to 2015
X_{ji} = quantitative measure of each characteristic *j* associated with corridor *i*
β_j = coefficient corresponding to X_{ji} to be determined by negative binomial regression
β_0 = constant to be determined by negative binomial regression

Regression Modeling Process

We first developed core models with pedestrian and bicycle exposure variables only and then developed models with additional roadway segment and intersection explanatory variables. Since several of the explanatory variables were highly correlated ($|\rho| > 0.6$), we did not include pairs of correlated variables in the same model to avoid problems with collinearity. In general, we followed a stepwise process, removing the least significant explanatory variables and leaving variables that were significant at the 90% confidence level ($p < 0.10$).

8.7 Pedestrian Crash Model Results

The pedestrian crash model results shown in Table 8-5 should be interpreted with caution. The core exposure variables used in the pedestrian crash models were the natural logarithm of population density, the percentage of households with zero vehicles, and the natural logarithm of total job density for the census block groups that contained the roadway corridor (Table 8-5, Model P1). While the square root of retail job density also showed a strong statistical association with pedestrian crashes (Table 8-5, Model P2), the natural logarithm of total job density provided a better model fit in most models with additional explanatory variables. Therefore, the variables in Model P1 were used as the core exposure variables in the models that included roadway variables.

Several different models were developed to show statistically-significant relationships between roadway segment and intersection variables and pedestrian crashes. The models in Table 8-5 generally reflect consistent results throughout the series of models that were tested. Pedestrian crashes were positively related to natural log of the state highway AADT (Models P3, P4, and P5), non-residential driveways (Models P3 and P4), total intersections (Models P3 and

P5), percent sidewalk coverage (Model P5), percent sidepath coverage (Model P5), and negatively related to percentage of corridor with a median (Model P4).

It is critical to recognize that the sidewalk variable in Model P5 is likely to be positively related to pedestrian crashes because it is correlated with pedestrian activity levels. For example, percentage of sidewalk coverage is positively correlated with the natural log of population per acre ($\rho = 0.71$), the natural log of jobs per acre ($\rho = 0.65$), and the percentage of households with no vehicles ($\rho = 0.63$). Previous research and theory suggest that the presence of sidewalks should not increase pedestrian crashes; they should decrease pedestrian crashes. Since this result suggests that pedestrian exposure is not being fully captured in the exposure proxy variables, the other statistically significant variables in these models may not be accurate representations of their relationship with pedestrian crashes.

8.8 Bicycle Crash Model Results

The bicycle crash model results were generally more consistent with theory than the pedestrian crash model results, though they should still be interpreted cautiously. The core exposure variables used in the bicycle crash models were the natural logarithm of population density, the percentage of households with zero vehicles, and the square root of retail job density for the census block groups that contained the roadway corridor (Table 8-6, Model B1). The square root of retail job density generally performed better than the natural logarithm of total job density in the exposure-only models (Table 8-6, Model B2) and most other bicycle crash models. Therefore, the variables in Model B1 were used as the core exposure variables in the models that included roadway variables.

Several different models were developed to show statistically-significant relationships between roadway segment and intersection variables and bicycle crashes. The models in Table 8-6 generally reflect consistent results throughout the series of models that were tested. Bicycle crashes were positively related to the number of non-residential driveways (Model B4) and negatively related to paved shoulder coverage (Model B3) and median coverage (Model B3). While bicycle crashes were negatively related to bicycle lane coverage (Model B5), the parameter was not statistically significant in the series of models.

Despite more theoretically-consistent results, the bicycle crash models should also be viewed cautiously because the exposure proxy variables may not provide a true representation of bicycle activity levels in the study corridors.

Table 8-5 Wisconsin State Highway Corridor Preliminary Pedestrian Crash Models.

	Model P1			Model P2			Model P3			Model P4			Model P5		
Variable	**Beta Coeff.**	**Std. Error**	**p-value**	**Beta Coeff.**	**Std. Error**	**p-value**	**Beta Coeff.**	**Std. Error**	**p-value**	**Beta Coeff.**	**Std. Error**	**p-value**	**Beta Coeff.**	**Std. Error**	**p-value**
Constant	-0.695	0.192	0.000	-1.045	0.189	0.000	-7.356	1.519	0.000	-7.272	1.640	0.000	-7.941	1.496	0.000
LnPopAcre	0.921	0.129	0.000	0.947	0.112	0.000	0.524	0.135	0.000	0.709	0.128	0.000	0.399	0.136	0.003
Pct0Veh	4.673	1.481	0.002	5.673	1.510	0.000	5.506	1.302	0.000	5.712	1.357	0.000	4.556	1.266	0.000
LnJobAcre	0.132	0.078	0.089				0.136	0.074	0.067	0.111	0.075	0.140	0.123	0.073	0.090
RtRetAcre				0.601	0.277	0.030									
LnAADT_STH							0.633	0.155	0.000	0.679	0.176	0.000	0.684	0.152	0.000
TotInts							0.079	0.021	0.000				0.070	0.020	0.001
NonResDvwy							0.020	0.006	0.002	0.027	0.007	0.000			
PctMedian										-0.436	0.214	0.042			
PctSW													1.150	0.258	0.000
PctSidepth													0.782	0.385	0.042
Model Pseudo R^2	0.236			0.238			0.286			0.274			0.298		
Model Log Likelihood	-331			-330			-309			-314			-304		

Table 8-6 Wisconsin State Highway Corridor Preliminary Bicycle Crash Models.

	Model B1			Model B2			Model B3			Model B4			Model B5		
Variable	**Beta Coeff.**	**Std. Error**	**p-value**	**Beta Coeff.**	**Std. Error**	**p-value**	**Beta Coeff.**	**Std. Error**	**p-value**	**Beta Coeff.**	**Std. Error**	**p-value**	**Beta Coeff.**	**Std. Error**	**p-value**
Constant	-1.600	0.241	0.000	-1.063	0.257	0.000	-1.282	0.245	0.000	-1.740	0.253	0.000	-1.635	0.245	0.000
LnPopAcre	1.145	0.147	0.000	1.084	0.163	0.000	1.082	0.144	0.000	1.085	0.148	0.000	1.156	0.148	0.000
Pct0Veh	3.376	1.856	0.069	1.614	1.865	0.387	3.313	1.739	0.057	3.457	1.827	0.058	3.641	1.864	0.051
LnJobAcre				0.229	0.092	0.012									
RtRetAcre	0.864	0.321	0.007				1.146	0.313	0.000	0.810	0.318	0.011	0.952	0.327	0.004
PctShld							-0.806	0.294	0.006						
PctMedian							-0.504	0.226	0.026						
NonResDvwy										0.015	0.008	0.054			
PctBL													-0.488	0.334	0.144
Model Pseudo R^2		0.244			0.243			0.262			0.249			0.247	
Model Log Likelihood		-287			-287			-280			-285			-286	

8.9 Lessons Learned

While we identified relationships between exposure proxy variables and pedestrian and bicycle crashes, data limitations prevented us from producing conclusive pedestrian and bicycle crash model results on the 200 study corridors. Some of the roadway variables had significant and theoretically-consistent relationships with pedestrian and bicycle crashes, but others did not, especially in the pedestrian crash models. Nonetheless, the analysis provided useful insights for future exploration of variables associated with pedestrian and bicyclist crashes.

Overall, pedestrian and bicycle crash modeling is challenging for several reasons.

- Pedestrian and bicyclist crashes are relatively infrequent compared to motor vehicle crashes. Small numbers of crashes at any given location along a roadway network make it challenging to identify statistically significant relationships between crashes and roadway characteristics.
- Pedestrian and bicyclist crashes tend to be concentrated in locations with more walking and bicycling activity. Locations with more walking and bicycling tend to have more pedestrian and bicycle facilities and other treatments that are designed to improve pedestrian and bicyclist safety (e.g., sidewalks, median crossing islands, paved shoulders, slower speed limits). This produces counterintuitive results: treatments designed and proven to improve pedestrian and bicyclist safety tend to be positively correlated with pedestrian and bicyclist crashes. Therefore, it is crucial to accurately control for levels of walking and bicycling activity in order to identify relationships between roadway characteristics and pedestrian and bicycle crashes (*75*).
- Pedestrian and bicycle counts or other volume data are often unavailable. This makes it difficult to address the problem described in the previous bullet. Proxy variables are often used to represent walking and bicycling activity (e.g., population density, employment density), but these may not provide sufficient accuracy to produce reliable model results.
- Small-scale roadway features that are likely to be important for pedestrian and bicyclist safety are rarely available in standard roadway inventory databases. For example, few roadway databases include median islands, curb extensions, bicycle lanes, or pedestrian signal timing.

State highway corridors may be especially difficult to use as a unit of analysis for pedestrian and bicyclist crash models. One reason is that pedestrian and bicycle activity is difficult to define along a one-mile corridor. Some pedestrian and bicyclist movements that have a risk of crashes are across the state highway, while other movements are along the corridor. Further, many pedestrian and bicycle movements along the corridor are less than one mile. So corridor pedestrian and bicyclist volumes may be represented more accurately as the sum of crossing movements over all intersections and midblock sections.

A second reason that corridors are challenging to analyze is that the explanatory variables summarized at the corridor level may not necessarily relate to the primary causes of pedestrian and bicyclist crashes in the corridor. The set of reported pedestrian or bicyclist crashes in any given corridor is likely to include many different crash types. These may involve left-turning automobiles, right-turning automobiles, bicyclists riding in the opposite direction as adjacent traffic, or drivers failing to yield to pedestrians at uncontrolled crosswalks. These crash types may be more likely because of specific design characteristics of particular intersections or roadway segments, but the crash data and roadway characteristics are all aggregated at the corridor level, potentially blurring the direct causal connection.

8.10 Future Research Recommendations

We propose several strategies to develop more conclusive results showing how roadway and surrounding neighborhood characteristics are associated with pedestrian and bicycle crashes in Wisconsin. In the short to medium term, we suggest analyzing local data and testing other statistical modeling approaches.

- Analyze different subsets of the database. This could include developing pedestrian and bicycle crash models just for the most urbanized corridors (e.g., at least 1,000 residents per square mile). It could also involve examining specific crash types, severity levels, or times of day.
- Add variables that quantify the socioeconomic characteristics of the communities surrounding each study corridor. These socioeconomic characteristics could potentially be proxies that account for additional components of exposure or pedestrian, bicyclist, or driver behaviors that impact safety in the local area.
- Analyze the variables associated with pedestrian or bicycle crashes in a single city. For example, this could be done in the City of Milwaukee. The City of Milwaukee Pedestrian Master Plan, which is being completed in 2018 with grant support from WisDOT, included a pedestrian intersection crossing volume model. Based on counts from 66 intersections throughout Milwaukee, the City estimated annual pedestrian crossing volumes at more than 4,000 intersections along arterial and collector roadways. These pedestrian volumes could be used to control for exposure and analyze the risk of pedestrian crashes per million pedestrian crossings. The characteristics of intersections with high versus low pedestrian crash risk could be summarized.
- Explore different statistical modeling approaches to account for pedestrian and bicyclist activity levels. These approaches might include applying structural equation models (SEMs) to identify which roadway variables may be associated with both crashes and exposure and compare the strength of each of these relationships. Two-stage models could also be used, first developing a model of pedestrian or bicyclist exposure and then testing other roadway and surrounding environment characteristics.
- Conduct a historical analysis of roadway design changes along particular sections of the state highway system to identify impacts on pedestrian and bicycle crashes. These analyses could be designed to compare time periods before and after specific projects and could investigate the effect of the project as a whole or the effect of particular pedestrian or bicyclist safety treatments. Larger samples of historical projects would provide better results. Like other crash studies, these historical comparisons should account for changes in pedestrian and bicycle activity levels.

In the longer term, it will be valuable to collect more comprehensive pedestrian and bicycle volume and facility data that can be used to develop pedestrian and bicycle safety performance functions.

- Institutionalize consistent pedestrian and bicycle counting procedures as a part of traffic monitoring programs. WisDOT could expand its efforts to include pedestrian and bicyclists as a part of intersection counts and work with local, county, and regional agencies to create a state repository for consistently-formatted counts. Initial phases of pedestrian and bicycle counting could be done on the sample of 200 corridors used in this study.
- Develop expansion factors to estimate annual pedestrian and bicyclist volumes from short-term counts. Annual volumes are typically used to compare crashes (often summarized for each year) with pedestrian and bicyclist activity levels. WisDOT could develop a program to collect continuous pedestrian and bicycle counts to document daily, weekly, and seasonal patterns throughout the state. This will eventually make it possible to estimate annual pedestrian and bicyclist volumes on state and local roadway systems.
- Collect pedestrian and bicycle facility variables as a routine part of roadway facility inventories. WisDOT could work with local municipalities to add features that may be related to pedestrian and bicyclist safety to the WISLR roadway database. These characteristics could include sidewalks, median refuge islands, curb extensions, bike lanes, sidepaths, pedestrian signals, turning lanes, and midblock crossing facilities.

9. CAUSAL INFERENCE

Identifying the proper cause-effect relationship and quantifying its impact is the key to determining cost-effective countermeasures. The statistical mantra, "correlation does not imply causation", means that just because two things correlate does not necessarily mean that one causes the other. A correlation between two factors can be caused by a third factor: a confounder. For instance, if a statistically significant relationship is found between the number of crashes and median household income, the confounding factors may be that affluent families own safer vehicles, that those with higher income levels have better driver behavior, or that safer roads come with higher property values. Causal claims cannot be substantiated from associations alone and it is necessary to make causal assumptions that are not testable in observational studies (*76*).

The objective of the present study was to perform a causality analysis to assess the impact of various factors on fatal or incapacitating injury crashes. In addition to driver characteristics, potential confounders including driver behavior (i.e., speeding), collision type, site-specific characteristics (i.e., road alignment, posted speed limit), and environmental conditions (i.e., weather condition, light condition) were considered in the causal inference.

9.1 Methodology

The objective of this observational study was to evaluate the causal effect of some treatment (exposure) on some outcome. The following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if the treatment is involved in unit } i \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if the treatment was not involved in unit } i \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for unit } i \text{ without treatment} \end{cases}$$

Causal effect of the treatment on the outcome for unit i is the difference between its two potential outcomes ($\tau_i = Y_{1i} - Y_{0i}$). However, for each unit only one outcome (either with treatment or without treatment) can be observed. This so-called “fundamental problem of causal inference” (*77*) in observational studies can be addressed by making some assumptions. In this study, four standard assumptions about the observed data for the causal inference are made:

- The stable unit treatment value assumption (SUTVA): there is one version of treatment and there is no interference between the units (i.e., treatment assignment of one unit does not affect the outcomes of other units),

- Consistency: the potential outcome with a treatment is equal to the observed outcome of the actual treatment,
- Ignorability: treatment assignment is independent from the potential outcomes, and
- Positivity: every unit has some chance of receiving the treatment and for every set of covariates, treatment assignment is stochastic with a probability of greater than zero.

Since the treatment and control groups are not randomly drawn from the same population, a matching method is used to balance the covariates across the two groups. The observations in the event-oriented analysis are matched using Propensity Score Matching (PSM).

The distances between the observations in PSM are determined as the difference between their propensity scores. Propensity score is the probability that observation i receives treatment given the covariates (*78*):

$$\pi_i = \Pr(T_i = 1|X) \tag{13}$$

In the present study, the propensity scores are estimated by a logistic regression:

$$\pi_i = \frac{1}{1 + e^{\beta^T X_i}} \tag{14}$$

One-to-one nearest neighbor greedy matching without replacement is applied to match the observations. Treated observations that are unreasonably distant from the control observations are removed using calipers. Calipers are specified ranges (caliper c) for the maximum distance allowed in matching (*79*). In this study, caliper value is set to 0.2 which is commonly used in the literature (*79,80*).

After implementing the matching method, the quality of covariate balance is evaluated using both graphical and statistical methods. In the graphical method histogram and jitter plots are used to illustrate the distribution of propensity scores before and after matching. In the statistical method, standardized mean difference (SMD) is used as the statistical criterion (*81*):

$$SMD = 100 * \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{S_T^2 - S_C^2}{2}}} \tag{15}$$

where $\bar{X}_T$ and $\bar{X}_C$ are mean of covariates (X) in treatment and control units, respectively, while S_T^2 and S_C^2 are sample covariances of X in the two groups.

There is no universally agreed upon threshold for the SMDs; however, an SMD of less than 0.1 has been taken to indicate a balanced covariate (*81,82*). In this study, SMD threshold of 0.1 is used to evaluate the quality of covariate balance.

After matching, the treatment effect was estimated using the difference in means statistic. In large samples, even tiny effects can become statistically significant (*83*); therefore, McNemar's test was also used to assess the treatment effect.

9.2 Event-oriented

Data

Crashes occurred on Wisconsin roadway network are available in Wisconsin Motor Vehicle Accident Reporting Form 4000 (MV4000) and are stored and maintained at WisTranPortal data hub (*84*). WisTranPortal is developed through collaboration between the Wisconsin Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin-Madison and the Wisconsin Department of Transportation (WisDOT) Bureau of Traffic Operations (BTO). Crash data include detailed information such as weather conditions, manner of collision, crash severity, and road conditions. The severity of crashes is based on the KABCO type scale indicating fatal (K), incapacitating (A), non-incapacitating evident (B), possible injury (C) and no apparent injury (O) crashes.

Segment-related crashes from 2012 to 2016 (5 years) were collected and filtered based on the following criteria: at least one car involved in the crash, no pedestrian involvement, no bicycle involvement, no motorcycle involvement, and no missing data points. After pre-processing, 207,597 crashes were selected for the analysis. Descriptive summary of the dataset is shown in Table 9-1. Fatal or incapacitating crashes occurred in 2.47% of the observations (5,134 crashes).

Table 9-1 Descriptive Statistics of the Event-Oriented Dataset

Variable	Name	Definition	% value 1
Manner	Manner_Angle	1 = If manner (first harmful event) in which participants collided in the crash is "angle"; 0 = Otherwise	10.35
	Manner_Head	1 = If manner (first harmful event) in which participants collided in the crash is "head-on collision"; 0 = Otherwise	1.69
	Manner_NoCol	1 = If manner (first harmful event) in which participants collided in the crash is "no-collision with another vehicle"; 0 = Otherwise	40.37
	Base	All other manners including "rear end", "rear to rear", "sideswipe/opposite direction", "sideswipe/same direction"	47.60
Road Condition	ROADCOND_Dry	1 = If surface condition of the road is dry; 0 = Otherwise	63.88
Light Condition	LGTCOND_NightWLight	1 = If light condition is nighttime with street lights; 0 = Otherwise	13.07
	LGTCOND_Night	1 = If light condition is nighttime without street lights; 0 = Otherwise	15.96
	LGTCOND_Day	1 = If light condition is day light; 0 = Otherwise	66.88

	Base	All other light conditions including dawn and dusk	4.09
Speed Limit	HighSpeed	1 = If posted speed limit is 55 mph or greater; 0 = Otherwise	41.29
Age	Age_Adolescent	1 = If age of the driver is less than 18; 0 = Otherwise	19.03
	Age_Young	1 = If age of the driver is less than equal 25 and greater than equal 18; 0 = Otherwise	34.98
	Age_Old	1 = If age of the driver is greater than 65; 0 = Otherwise	12.35
	Base	If age of the driver is between 25 and 65 (adult and middle age)	33.64
Speed Flag	SpeedFlag	1 = If at least one driver involved in the crash received a citation for speeding	23.97
Horizontal Curve	HCurve	1 = If the horizontal road terrain at the point of impact is a curve; 0 = Otherwise	15.43
Train Flag	TrainFlag	1 = If a train is involved in a crash	0.04
Urban/Rural	Urban	1 = If a crash occurred in an urban area; 0 = Otherwise (Rural)	56.00
Roadway Class	RoadClass_Street	1 = If the type of road a crash took place on is street; 0 = Otherwise	34.17
	RoadClass_Interstate	1 = If the type of road a crash took place on is interstate highway; 0 = Otherwise	13.55
	Base	If the type of road a crash took place on is state highway	52.28
Gender	Female	1 = If at least one female driver is involved in a crash; 0 = Otherwise	58.61
Seatbelt	Safety_NoBelt	1 = If shoulder and/or lap belt was used by a driver involved in a crash; 0 = Otherwise	2.15
Impaired Driver Flag	ImpairedFlag	1 = If a driver was listed on the police report as drinking alcohol and/or using drugs before the crash; 0 = Otherwise	5.79
Observed Outcome	Outcome	1 = If injury severity for a crash, taken over all persons involved in a crash is killed and/or incapacitating (K and A crashes); 0 = Otherwise	2.47

Analysis

Impaired Driving

To evaluate the impact of impaired driving on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if an alcohol or drug impaired driver was involved in crash } i \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if an alcohol or drug impaired driver was not involved in crash } i \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression since only two treatment levels (crashes with impaired drivers versus crashes without impaired drivers) were considered. The developed propensity score model (Table 9-2) estimates the probability that an impaired driver was involved in a crash. All covariates are statistically significant in the estimation of the propensity scores. Negative estimated parameters indicate that the respective covariates reduce the probability of impaired driver involvement while positive estimates are associated with a higher probability of impaired driver involvement.

Table 9-2 Propensity Score Model

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.79013	0.063412	-59.7702	0.000
Manner_Angle	-0.36837	0.055567	-6.62936	0.000
Manner_Head	1.311131	0.057974	22.61586	0.000
Manner_NoCol	0.830462	0.026416	31.43743	0.000
ROADCOND_Dry	1.095856	0.024312	45.0746	0.000
LGTCOND_NightWLight	1.226484	0.055012	22.29489	0.000
LGTCOND_Night	0.856161	0.053187	16.09705	0.000
LGTCOND_Day	-0.72304	0.053797	-13.4401	0.000
HighSpeed	-0.25206	0.026923	-9.36237	0.000
Age_Adolescent	-0.3958	0.028502	-13.8867	0.000
Age_Young	-0.06088	0.021408	-2.8437	0.004
Age_Old	-0.66271	0.047883	-13.8402	0.000
SpeedFlag	0.388532	0.023977	16.20403	0.000
HCurve	0.171499	0.025238	6.795146	0.000
TrainFlag	0.885043	0.374363	2.364133	0.018
Urban	-0.2468	0.02943	-8.38593	0.000
RoadClass_Street	0.231557	0.030365	7.625899	0.000
RoadClass_Interstate	-0.34013	0.037368	-9.10225	0.000
Female	-0.49347	0.02106	-23.4317	0.000
Safety_NoBelt	1.368888	0.039734	34.45155	0.000

Results of the propensity score model show that impaired drivers are more likely to be involved in head-on and no-collision crashes while they are less likely to be involved in angle crashes in comparison to the base case (sideswipe, rear end, and rear to rear crashes). Impaired driving crashes are more likely to occur on dry roadways than on wet, snowy, or icy roadways. Adolescent, young and old drivers have lower probabilities of impaired driving comparing to adult and middle-age drivers. During nighttime, crashes on roadways with streetlights have higher probabilities of impaired-driver involvement than roadways without lights. Daytime crashes are less likely to have impaired drivers than nighttime crashes. With respect to roadway classes, streets have higher probabilities of impaired driving than state highways (base case) while interstate highways have lower chances. High-speed segments, urban areas, and female drivers are associated with lower probabilities of impaired driver involvement while speeding, horizontal curves, train involvement, and no-seatbelt are associated with higher probabilities of impaired driving.

Using the developed propensity model, one-to-one nearest neighbor greedy matching without replacement was applied and the caliper was set to 0.2. Out of 12,024 observations with impaired drivers (treatment units), 11,982 observations were matched with observation from the control units. The caliper assigned in the model cut out 42 observations with impaired drivers. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-3.

SMD was used as the statistical criteria to evaluate the quality of covariate balance after matching. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-3 and shown in Figure 9-1. As shown, the absolute standardized mean differences are much larger for the unmatched data than for the matched data. In the unmatched dataset, SMDs range between 0.007 and 0.998 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 (ranging between less than 0.001 and 0.035) and therefore the matched dataset is balanced.

Table 9-3 Descriptive Statistics of Unmatched and Matched Datasets and SMDs

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	195533	12024		11982	11982	
Manner_Angle (mean (sd))	0.11 (0.31)	0.03 (0.18)	0.294	0.03 (0.17)	0.03 (0.18)	0.017
Manner_Head (mean (sd))	0.02 (0.12)	0.04 (0.19)	0.146	0.04 (0.19)	0.04 (0.19)	0.008
Manner_NoCol (mean (sd))	0.39 (0.49)	0.69 (0.46)	0.644	0.71 (0.46)	0.69 (0.46)	0.032
ROADCOND_Dry (mean (sd))	0.63 (0.48)	0.75 (0.43)	0.267	0.75 (0.43)	0.75 (0.43)	0.003
LGTCOND_NightWLight (mean (sd))	0.12 (0.32)	0.32 (0.47)	0.507	0.33 (0.47)	0.32 (0.47)	0.02
LGTCOND_Night (mean (sd))	0.15 (0.35)	0.39 (0.49)	0.57	0.38 (0.48)	0.39 (0.49)	0.022
LGTCOND_Day (mean (sd))	0.69 (0.46)	0.25 (0.43)	0.988	0.25 (0.43)	0.25 (0.43)	0.006
HighSpeed (mean (sd))	0.41 (0.49)	0.42 (0.49)	0.021	0.42 (0.49)	0.42 (0.49)	<0.001
Age_Adolescent (mean (sd))	0.19 (0.39)	0.16 (0.37)	0.073	0.18 (0.38)	0.16 (0.37)	0.035
Age_Young (mean (sd))	0.35 (0.48)	0.37 (0.48)	0.05	0.36 (0.48)	0.37 (0.48)	0.016
Age_Old (mean (sd))	0.13 (0.33)	0.04 (0.20)	0.31	0.05 (0.22)	0.04 (0.20)	0.029
SpeedFlag (mean (sd))	0.23 (0.42)	0.34 (0.47)	0.237	0.34 (0.47)	0.34 (0.47)	0.004
HCurve (mean (sd))	0.15 (0.36)	0.25 (0.43)	0.26	0.25 (0.43)	0.25 (0.43)	0.001
TrainFlag (mean (sd))	0.00 (0.02)	0.00 (0.03)	0.018	0.00 (0.03)	0.00 (0.03)	0.011
Urban (mean (sd))	0.57 (0.50)	0.46 (0.50)	0.203	0.46 (0.50)	0.47 (0.50)	0.001
RoadClass_Street (mean (sd))	0.34 (0.47)	0.34 (0.47)	0.007	0.34 (0.47)	0.34 (0.47)	0.011
RoadClass_Interstate (mean (sd))	0.14 (0.35)	0.09 (0.28)	0.158	0.09 (0.29)	0.09 (0.28)	0.004

Female (mean (sd))	0.60 (0.49)	0.37 (0.48)	0.462	0.38 (0.48)	0.38 (0.48)	0.002
Safety_NoBelt (mean (sd))	0.02 (0.13)	0.10 (0.31)	0.375	0.10 (0.30)	0.10 (0.30)	0.015

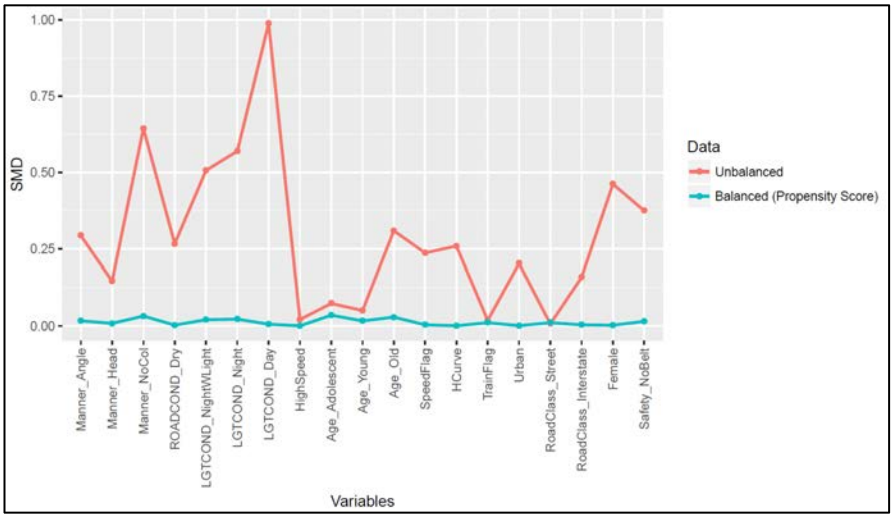
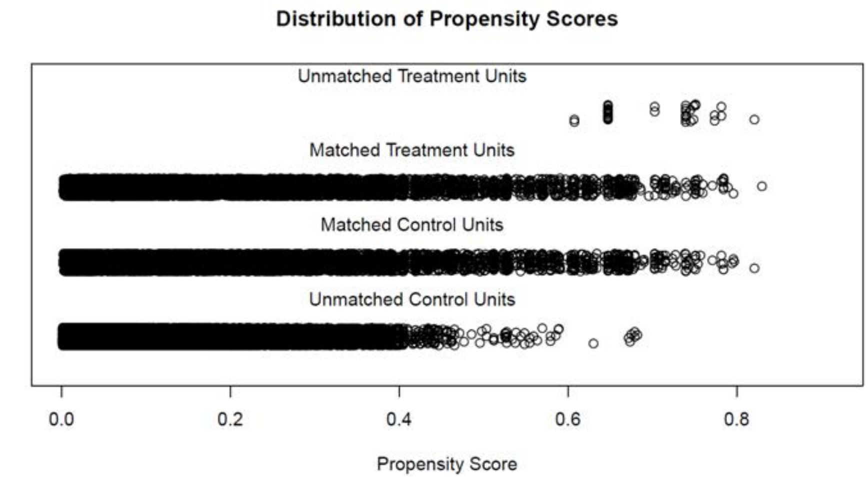


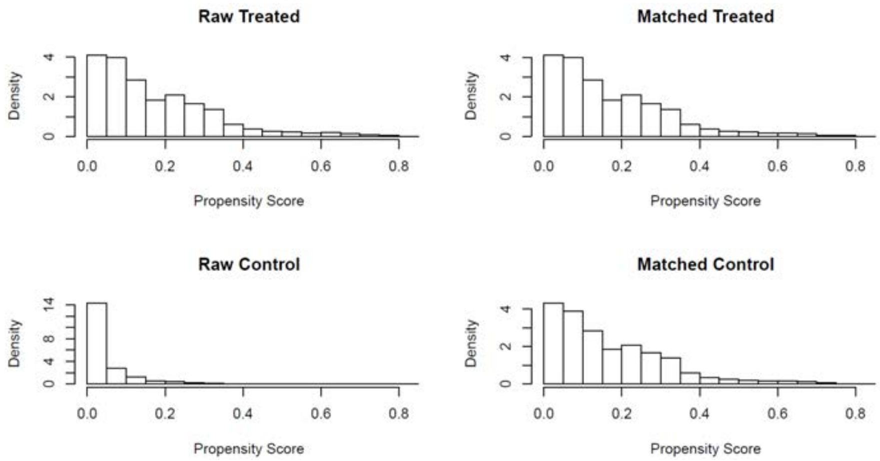
Figure 9-1 SMDs for balanced and unbalanced datasets

To assess the distribution of propensity scores after matching, two graphical methods including jitter plot and histogram plot were used. Jitter plot (upper plot in Figure 9-2) displays jittered estimated propensity scores of treated and control units. Jitter plot is broken up into four different sections: unmatched treatment units, matched treatment units, matched control units, and unmatched control units. Since the data was matched with a caliper, 42 observations from the treated group were not matched to observations from the control unit. Matched treatment units and matched control units have a similar distribution of propensity scores.

The histogram plot (lower plot in Figure 9-2) provides four histograms: the original treated and control groups and the matched treated and control groups. Results indicate that the propensity score distribution of treated and control units in unmatched (raw) data are different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-2 Distribution of propensity scores: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced, therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect ($E(Y_1) - E(Y_0)$). Since the sample size is relatively large, McNemar's test is also performed. The null hypothesis in McNemar's test is that there is no treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-4.

Table 9-4 Treatment Effect Analysis

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	16.54	11981	0.000	0.0474 0.0601	0.054	266.74	1	0.000

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is 0.0537 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, impaired driver involvement in a crash has a causal effect on fatality and incapacitating injuries. On average, the risk of being killed or severely injured in a crash increases by 5.4% when an impaired driver is involved in the crash.

Speed Flag

To evaluate the impact of speeding on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if a speeding citation was issued in crash } i \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if a speeding citation was not issued in crash } i \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-5) estimates the probability that a speeding driver was involved in a crash. All covariates except nighttime light condition (with and without street light), train flag, work zone, and urban indicator are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-5 Propensity Score Model for Speed Flag

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.81821	0.033486	-24.4346	0.000
Manner_Angle	-0.28383	0.025212	-11.2577	0.000
Manner_Head	0.463063	0.043083	10.74826	0.000
Manner_NoCol	0.793001	0.014266	55.58806	0.000
ROADCOND_Dry	-1.70158	0.012178	-139.723	0.000
LGTCOND_NightWLight	0.003391	0.032402	0.104643	0.916
LGTCOND_Night	0.028879	0.030246	0.954813	0.340
LGTCOND_Day	0.056701	0.028333	2.001284	0.045
HighSpeed	0.053488	0.015125	3.536447	0.000
Age_Adolescent	0.116479	0.016204	7.188374	0.000
Age_Young	0.272818	0.012527	21.77832	0.000

Age_Old	-0.46759	0.02235	-20.9212	0.000
ImpairedFlag	0.511606	0.023593	21.68482	0.000
HCurve	0.704755	0.014871	47.39022	0.000
VHill	0.096223	0.015308	6.285995	0.000
TrainFlag	-0.21874	0.346153	-0.63193	0.527
WorkZone	0.017488	0.036513	0.478937	0.632
Urban	-0.01156	0.015732	-0.73497	0.462
RoadClass_Street	-0.38914	0.017819	-21.8381	0.000
RoadClass_Interstate	0.519535	0.017706	29.34187	0.000
Female	-0.19131	0.012056	-15.8691	0.000
Safety_NoBelt	0.583472	0.035694	16.34634	0.000

Angle crashes, dry roadway conditions, old drivers, street class roadways and female drivers are associated with lower probabilities of speeding driver involvement while head-on crashes, no-collision crashes, daytime crashes, high-speed segments, young and adolescent drivers, impaired driving, horizontal curve, vertical hill, interstate roadways, and no-seatbelt are associated with higher probabilities of speeding driver involvement.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model cut out 5936 observations with speed flag. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-6. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-6 and shown in Figure 9-3. In the unmatched dataset, SMDs range between 0.011 and 0.93 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 and therefore the matched dataset is balanced.

Table 9-6 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for Speed Flag

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	157812	49745		43809	43809	
Manner_Angle	0.12 (0.33)	0.05 (0.22)	0.26	0.05 (0.22)	0.06 (0.23)	0.018
Manner_Head	0.02 (0.13)	0.02 (0.14)	0.026	0.02 (0.14)	0.02 (0.14)	0.005
Manner_NoCol	0.33 (0.47)	0.65 (0.48)	0.679	0.60 (0.49)	0.61 (0.49)	0.018
ROADCOND_Dry	0.74 (0.44)	0.32 (0.47)	0.93	0.35 (0.48)	0.36 (0.48)	0.013
LGTCOND_NightWLight	0.13 (0.34)	0.13 (0.33)	0.011	0.14 (0.34)	0.13 (0.34)	0.007
LGTCOND_Night	0.13 (0.34)	0.24 (0.43)	0.284	0.24 (0.43)	0.23 (0.42)	0.025
LGTCOND_Day	0.70 (0.46)	0.58 (0.49)	0.243	0.57 (0.49)	0.59 (0.49)	0.031
HighSpeed	0.37 (0.48)	0.55 (0.50)	0.375	0.53 (0.50)	0.54 (0.50)	0.023
Age_Adolescent	0.20 (0.40)	0.16 (0.37)	0.098	0.17 (0.38)	0.17 (0.37)	0.011
Age_Young	0.34 (0.47)	0.39 (0.49)	0.103	0.37 (0.48)	0.38 (0.48)	0.015
Age_Old	0.14 (0.35)	0.06 (0.24)	0.273	0.07 (0.26)	0.07 (0.25)	0.01
ImpairedFlag	0.05 (0.22)	0.08 (0.27)	0.129	0.09 (0.28)	0.08 (0.27)	0.022
HCurve	0.11 (0.31)	0.29 (0.46)	0.467	0.24 (0.43)	0.25 (0.43)	0.016

VHill	0.14 (0.34)	0.22 (0.41)	0.211	0.21 (0.41)	0.20 (0.40)	0.009
TrainFlag	0.00 (0.02)	0.00 (0.01)	0.013	0.00 (0.02)	0.00 (0.02)	<0.001
WorkZone	0.03 (0.18)	0.02 (0.15)	0.06	0.03 (0.16)	0.02 (0.15)	0.019
Urban	0.60 (0.49)	0.42 (0.49)	0.365	0.45 (0.50)	0.45 (0.50)	0.001
RoadClass_Street	0.38 (0.49)	0.21 (0.41)	0.394	0.23 (0.42)	0.22 (0.42)	0.023
RoadClass_Interstate	0.12 (0.32)	0.19 (0.39)	0.188	0.17 (0.38)	0.18 (0.38)	0.022
Female	0.61 (0.49)	0.51 (0.50)	0.213	0.52 (0.50)	0.52 (0.50)	0.008
Safety_NoBelt	0.02 (0.13)	0.03 (0.18)	0.101	0.03 (0.18)	0.03 (0.18)	0.005

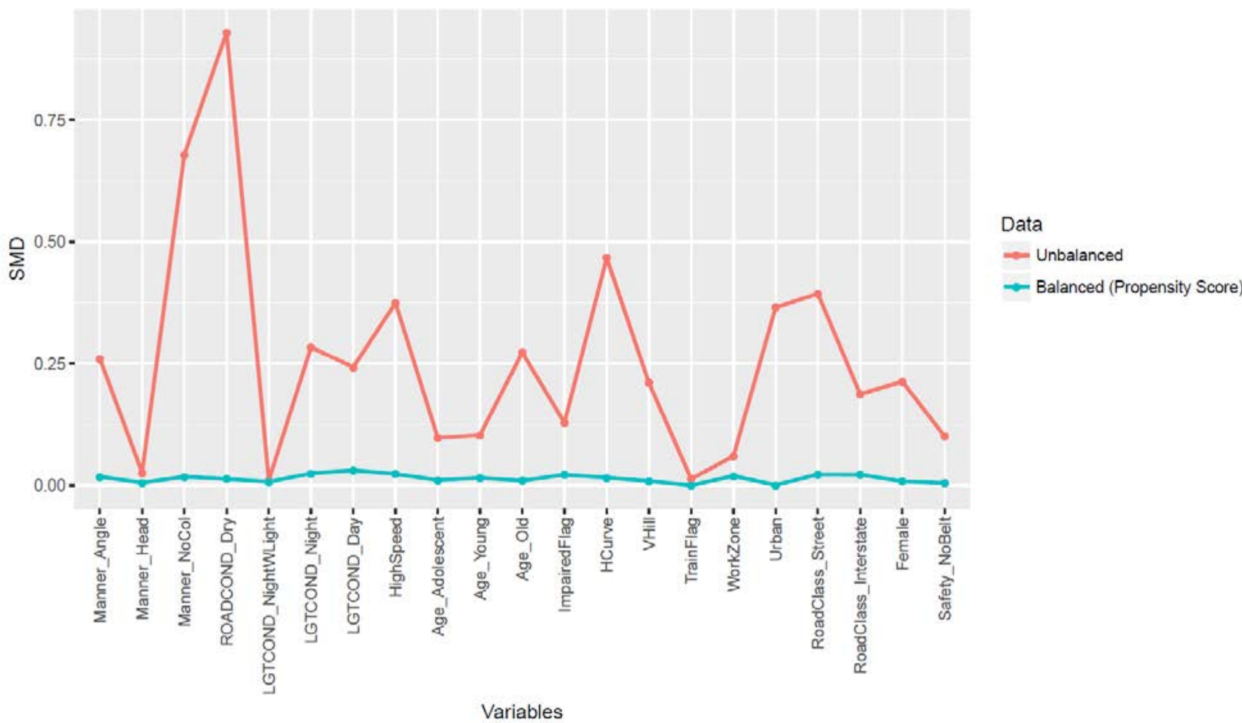
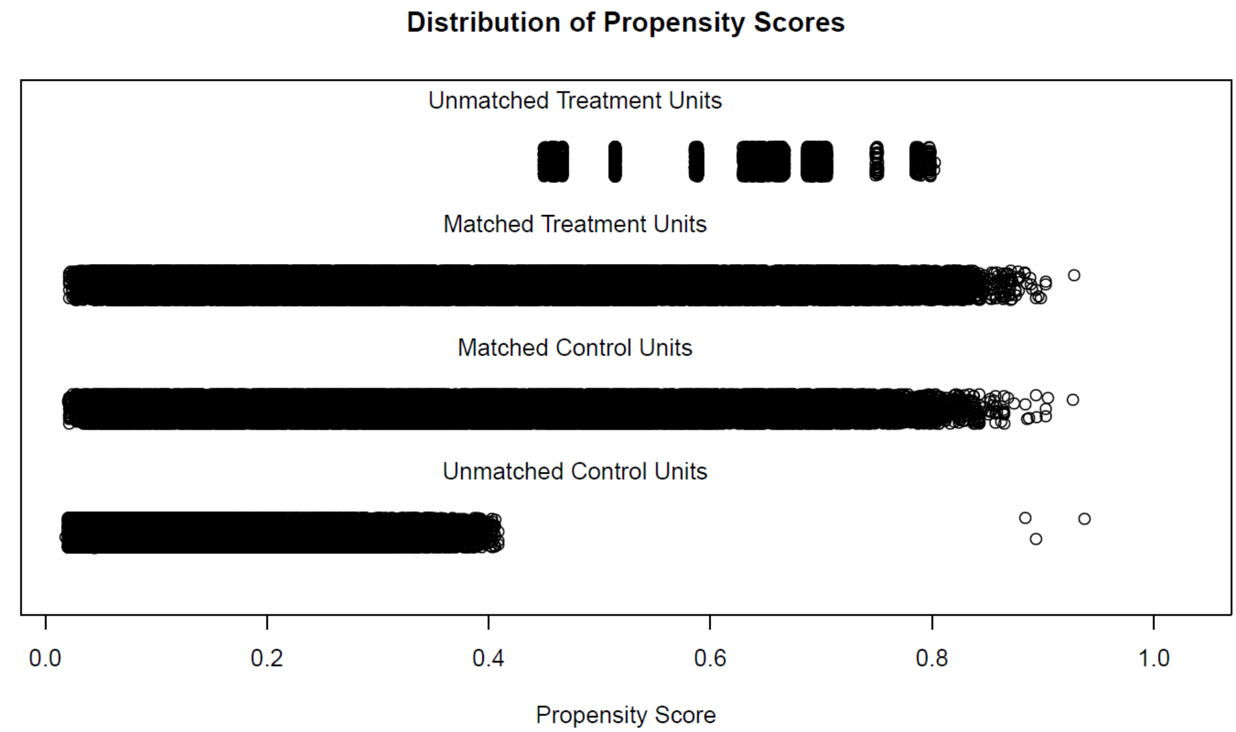
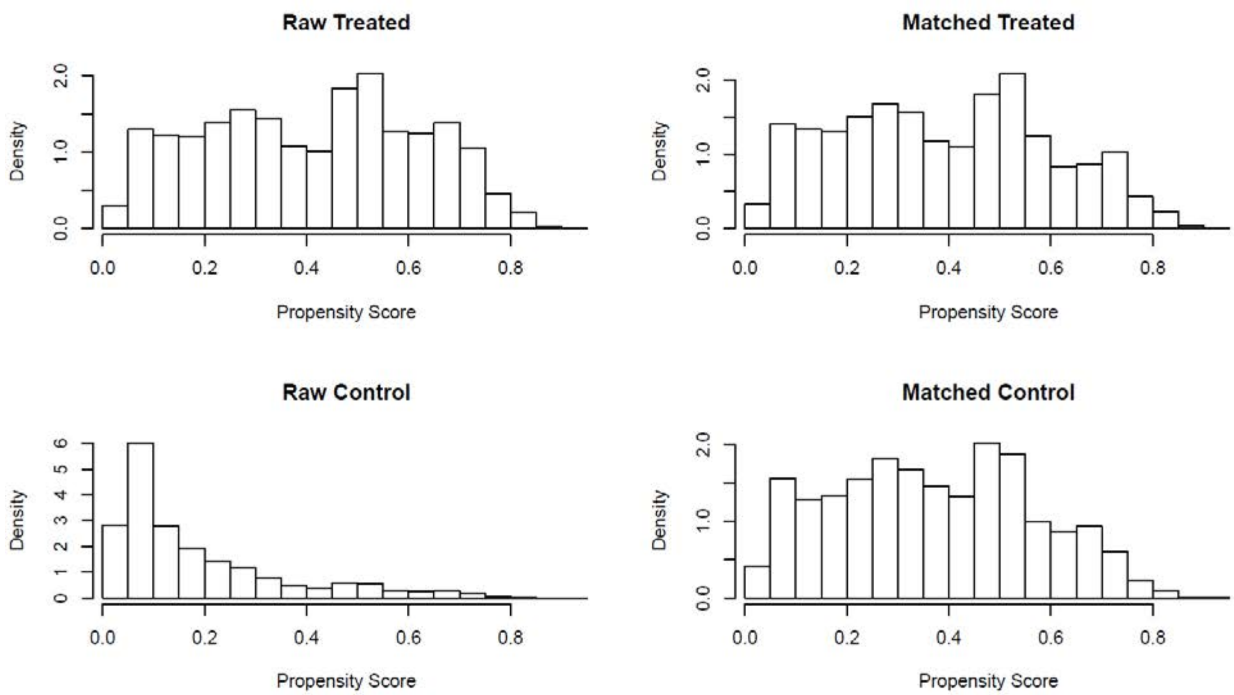


Figure 9-3 SMDs for balanced and unbalanced datasets for speed flag

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-4. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-4 Distribution of propensity scores for speed flag: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced,

therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-7.

Table 9-7 Treatment Effect Analysis for Speed Flag

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	8.88	43808	0.000	0.008 0.013	0.011	78.42	1	0.000

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is 0.011 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, speeding driver involvement in a crash has a causal effect on fatality and incapacitating injuries. On average, the risk of being killed or severely injured in a crash increases by 1.1% when a speeding driver is involved in the crash.

Seatbelt

To evaluate the impact of seatbelt usage on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if seatbelt was not used in crash } i \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if seatbelt was used in crash } i \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-8) estimates the probability that seatbelt is not used in a crash. All covariates except nighttime light condition (with and without street light), young and old drivers, horizontal curve, and street roadways are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-8 Propensity Score Model for Seatbelt

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.86802	0.095698	-50.8683	0.000
Manner_Angle	0.458998	0.067033	6.847376	0.000
Manner_Head	1.569659	0.079137	19.83471	0.000
Manner_NoCol	0.810688	0.043659	18.5688	0.000
ROADCOND_Dry	0.775546	0.037734	20.5532	0.000
LGTCOND_NightWLight	0.05315	0.085996	0.61806	0.537
LGTCOND_Night	0.023852	0.07832	0.304541	0.761
LGTCOND_Day	-0.15588	0.076008	-2.05085	0.040
HighSpeed	0.196994	0.040237	4.895803	0.000
Age_Adolescent	-0.35454	0.048481	-7.31291	0.000
Age_Young	0.064042	0.032957	1.943176	0.000
Age_Old	-0.07284	0.057402	-1.26901	0.052
ImpairedFlag	1.346964	0.039571	34.03958	0.000
SpeedFlag	0.532593	0.03601	14.79014	0.000
HCurve	0.003667	0.03876	0.094612	0.925
VHill	0.102726	0.040046	2.5652	0.010
TrainFlag	1.145905	0.438304	2.614404	0.009
WorkZone	-0.23786	0.115938	-2.05165	0.040
Urban	-0.29401	0.044852	-6.55526	0.000
RoadClass_Street	0.009793	0.049413	0.198195	0.843
RoadClass_Interstate	-0.74541	0.064324	-11.5883	0.000
Female	-0.24247	0.032364	-7.492	0.000

Daytime crashes, adolescent drivers, work zones, urban areas, interstate highways, and female drivers are associated with lower probabilities of no-seatbelt involvement while angle crashes, head-on crashes, no-collision crashes, dry roadway conditions, high-speed segments, impaired driving, speed flag, vertical hill, and train flag are associated with higher probabilities of no-seatbelt involvement.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model cut out two observations from the treatment units. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-9. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-9 and shown in Figure 9-5. In the unmatched dataset, SMDs range between 0.050 and 0.639 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 and therefore the matched dataset is balanced.

Table 9-9 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for Seatbelt

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	202433	4448		4446	4446	
Manner_Angle	0.10 (0.31)	0.07 (0.26)	0.121	0.07 (0.26)	0.07 (0.26)	0.002
Manner_Head	0.02 (0.13)	0.05 (0.22)	0.195	0.06 (0.24)	0.05 (0.22)	0.037
Manner_NoCol	0.40 (0.49)	0.66 (0.47)	0.551	0.66 (0.48)	0.66 (0.47)	0.016
ROADCOND_Dry	0.64 (0.48)	0.73 (0.44)	0.208	0.74 (0.44)	0.73 (0.44)	0.006
LGTCOND_NightWLight	0.13 (0.34)	0.15 (0.35)	0.05	0.15 (0.36)	0.15 (0.35)	0.003
LGTCOND_Night	0.16 (0.36)	0.31 (0.46)	0.369	0.32 (0.46)	0.31 (0.46)	0.013
LGTCOND_Day	0.67 (0.47)	0.50 (0.50)	0.365	0.49 (0.50)	0.50 (0.50)	0.022
HighSpeed	0.41 (0.49)	0.53 (0.50)	0.243	0.53 (0.50)	0.53 (0.50)	0.004
Age_Adolescent	0.19 (0.39)	0.13 (0.33)	0.181	0.13 (0.34)	0.13 (0.33)	0.019
Age_Young	0.35 (0.48)	0.38 (0.48)	0.057	0.38 (0.49)	0.38 (0.48)	0.012
Age_Old	0.12 (0.33)	0.08 (0.28)	0.13	0.09 (0.28)	0.08 (0.28)	0.008
ImpairedFlag	0.05 (0.22)	0.28 (0.45)	0.639	0.27 (0.44)	0.28 (0.45)	0.019
SpeedFlag	0.24 (0.43)	0.38 (0.48)	0.304	0.37 (0.48)	0.38 (0.48)	0.003
HCurve	0.15 (0.36)	0.25 (0.43)	0.239	0.24 (0.43)	0.25 (0.43)	0.008
VHill	0.16 (0.36)	0.19 (0.40)	0.103	0.20 (0.40)	0.19 (0.40)	0.015
TrainFlag	0.00 (0.02)	0.00 (0.04)	0.033	0.00 (0.04)	0.00 (0.04)	0.006
WorkZone	0.03 (0.17)	0.02 (0.13)	0.085	0.02 (0.13)	0.02 (0.13)	0.002
Urban	0.56 (0.50)	0.37 (0.48)	0.395	0.37 (0.48)	0.37 (0.48)	0.001
RoadClass_Street	0.34 (0.47)	0.26 (0.44)	0.186	0.26 (0.44)	0.26 (0.44)	0.006
RoadClass_Interstate	0.14 (0.34)	0.07 (0.25)	0.233	0.06 (0.24)	0.07 (0.25)	0.02
Female	0.59 (0.49)	0.44 (0.50)	0.309	0.44 (0.50)	0.44 (0.50)	0.01

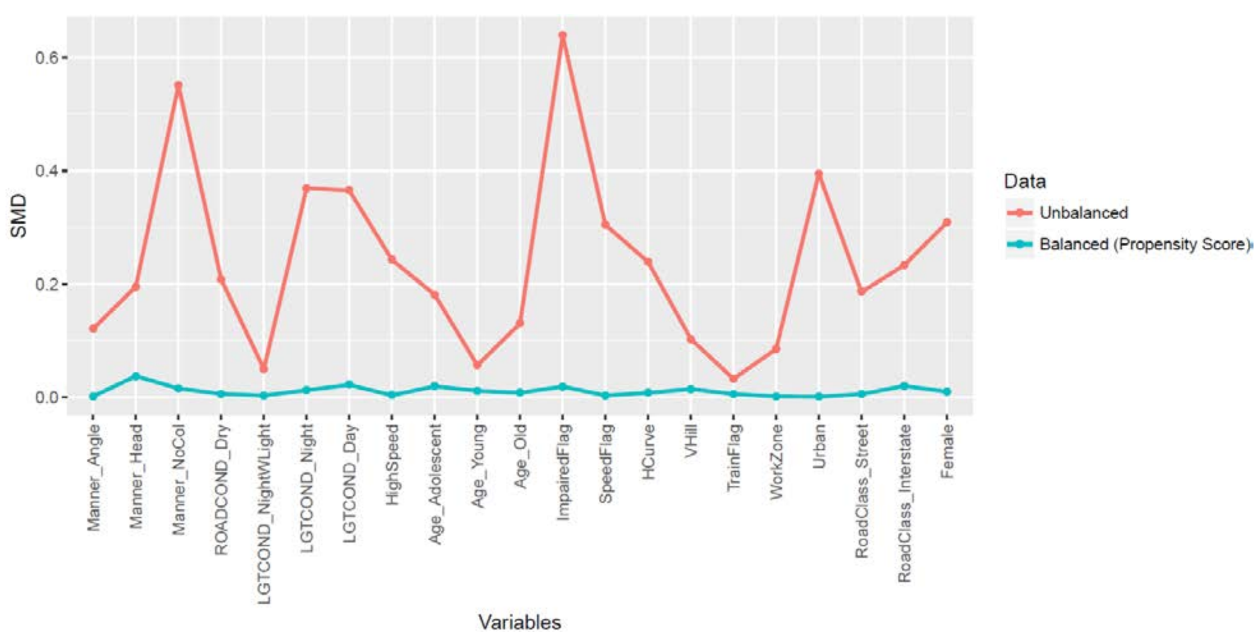
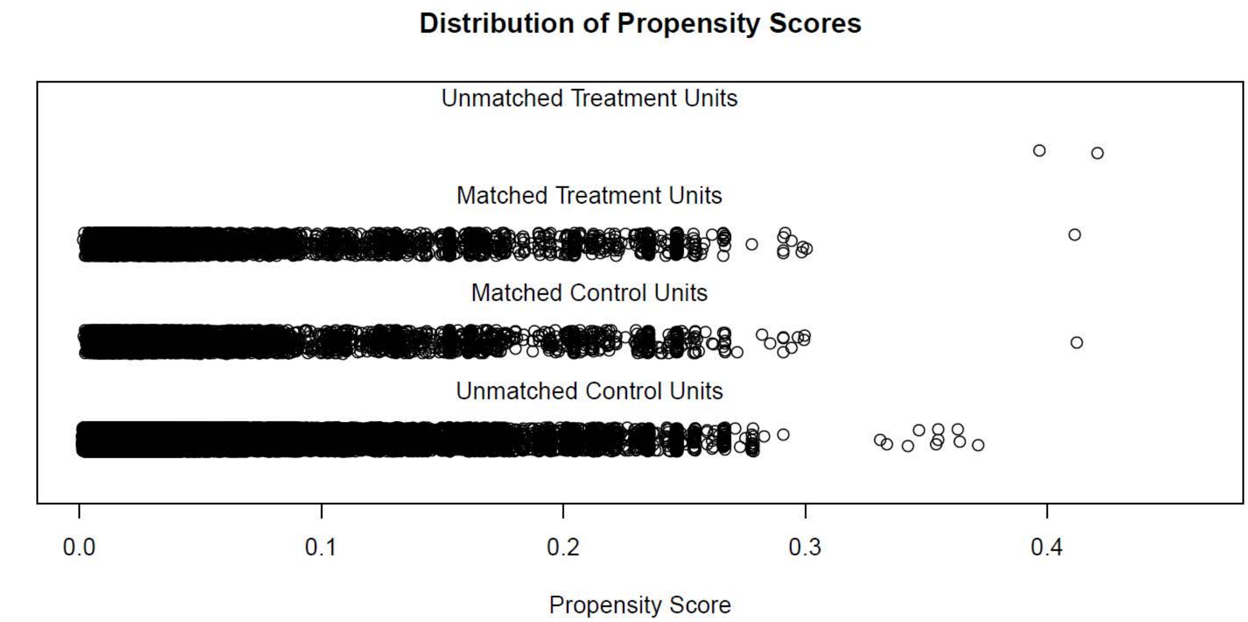
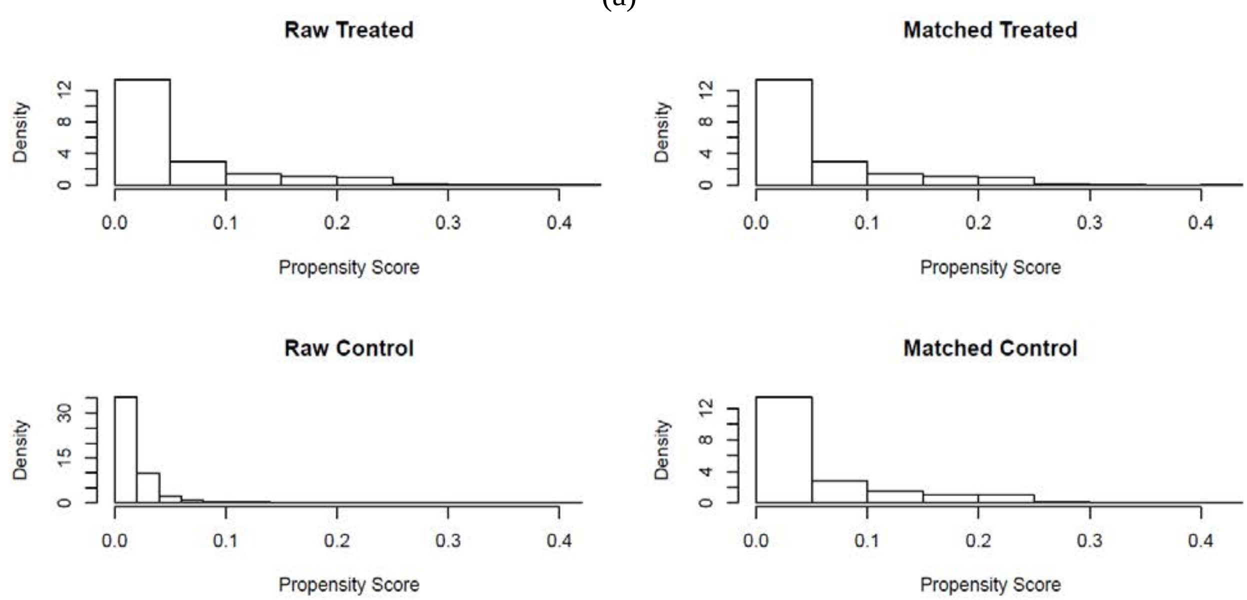


Figure 9-5 SMDs for balanced and unbalanced datasets for seatbelt

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-6. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-6 Distribution of propensity scores for seatbelt: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced,

therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-10.

Table 9-10 Treatment Effect Analysis for Seatbelt

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	33.04	4446	0.000	0.230 0.259	0.245	875.24	1	0.000

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is 0.245 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, no-seatbelt involvement in a crash has a causal effect on fatality and incapacitating injuries. On average, the risk of being killed or severely injured in a crash increases by 24.5% when seatbelt is not used in the crash.

Work Zone

To evaluate the impact of work zone on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if crash } i \text{ occured in a work zone} \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if crash } i \text{ did not occure in a work zone} \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-11) estimates the probability that a crash occurred in a work zone. All covariates except nighttime light condition (with and without street light), speed flag, impaired driving, and train flag are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-11 Propensity Score Model for Work Zone

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.18215	0.089397	-46.7819	0.000
Manner_Angle	-0.87461	0.06834	-12.798	0.000
Manner_Head	-0.7204	0.154105	-4.67477	0.000
Manner_NoCol	-0.61057	0.034779	-17.556	0.000
ROADCOND_Dry	0.889639	0.036638	24.28206	0.000
LGTCOND_NightWLight	-0.00116	0.084567	-0.01374	0.989
LGTCOND_Night	-0.01199	0.08448	-0.14193	0.887
LGTCOND_Day	0.205882	0.075507	2.726668	0.006
HighSpeed	0.26323	0.0312	8.436837	0.000
Age_Adolescent	-0.38324	0.042269	-9.0668	0.000
Age_Young	-0.08099	0.027797	-2.91356	0.004
Age_Old	0.079095	0.038615	2.048294	0.041
ImpairedFlag	-0.10092	0.069169	-1.45906	0.145
SpeedFlag	-0.00365	0.036799	-0.09911	0.921
HCurve	-0.2455	0.046649	-5.26282	0.000
VHill	0.107101	0.037589	2.849255	0.004
TrainFlag	0.041913	1.015429	0.041276	0.967
Urban	0.667735	0.032729	20.40219	0.000
RoadClass_Street	-1.43412	0.045156	-31.7596	0.000
RoadClass_Interstate	0.655237	0.031055	21.09933	0.000
Female	-0.10611	0.027026	-3.92619	0.000
Safety_NoBelt	-0.262	0.115825	-2.262	0.024

Angle crashes, head-on crashes, no-collision crashes, adolescent drivers, young drivers, horizontal curves, street roadways, female drivers and no-seatbelt are associated with lower probabilities of work zone crashes while dry roadways, daylight, high-speed segments, old drivers, vertical hills, urban areas, and interstate highways are associated with higher probabilities of work zone crashes.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model did not cut out any observations from the treatment units. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-12. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-12 Table 9-9and shown in Figure 9-7. In the unmatched dataset, SMDs range between 0.005 and 0.591 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 and therefore the matched dataset is balanced.

Table 9-12 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for Work Zone

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	200487	6394		6394	6394	
Manner_Angle	0.11 (0.31)	0.04 (0.19)	0.269	0.04 (0.20)	0.04 (0.19)	0.019
Manner_Head	0.02 (0.13)	0.01 (0.08)	0.095	0.01 (0.08)	0.01 (0.08)	0.014
Manner_NoCol	0.41 (0.49)	0.25 (0.43)	0.342	0.25 (0.43)	0.25 (0.43)	0.002
ROADCOND_Dry	0.63 (0.48)	0.83 (0.38)	0.452	0.83 (0.38)	0.83 (0.38)	0.004
LGTCOND_NightWLight	0.13 (0.34)	0.11 (0.31)	0.077	0.12 (0.32)	0.11 (0.31)	0.039
LGTCOND_Night	0.16 (0.37)	0.10 (0.30)	0.172	0.10 (0.30)	0.10 (0.30)	0.017
LGTCOND_Day	0.67 (0.47)	0.76 (0.43)	0.206	0.75 (0.43)	0.76 (0.43)	0.02
HighSpeed	0.41 (0.49)	0.54 (0.50)	0.265	0.54 (0.50)	0.54 (0.50)	0.009
Age_Adolescent	0.19 (0.39)	0.11 (0.31)	0.236	0.11 (0.31)	0.11 (0.31)	0.001
Age_Young	0.35 (0.48)	0.36 (0.48)	0.011	0.35 (0.48)	0.36 (0.48)	<0.001
Age_Old	0.12 (0.33)	0.14 (0.35)	0.05	0.14 (0.35)	0.14 (0.35)	0.004
ImpairedFlag	0.06 (0.23)	0.04 (0.19)	0.093	0.04 (0.19)	0.04 (0.19)	0.001
SpeedFlag	0.24 (0.43)	0.18 (0.38)	0.15	0.18 (0.38)	0.18 (0.38)	0.012
HCurve	0.16 (0.36)	0.09 (0.29)	0.199	0.10 (0.30)	0.09 (0.29)	0.022
VHill	0.16 (0.36)	0.14 (0.35)	0.036	0.15 (0.36)	0.14 (0.35)	0.029
TrainFlag	0.00 (0.02)	0.00 (0.01)	0.015	0.00 (0.00)	0.00 (0.01)	0.018
Urban	0.56 (0.50)	0.67 (0.47)	0.233	0.65 (0.48)	0.67 (0.47)	0.033
RoadClass_Street	0.35 (0.48)	0.11 (0.31)	0.591	0.10 (0.31)	0.11 (0.31)	0.02
RoadClass_Interstate	0.13 (0.33)	0.36 (0.48)	0.562	0.37 (0.48)	0.36 (0.48)	0.009
Female	0.59 (0.49)	0.59 (0.49)	0.005	0.58 (0.49)	0.59 (0.49)	0.015
Safety_NoBelt	0.02 (0.15)	0.01 (0.11)	0.071	0.01 (0.12)	0.01 (0.11)	0.019

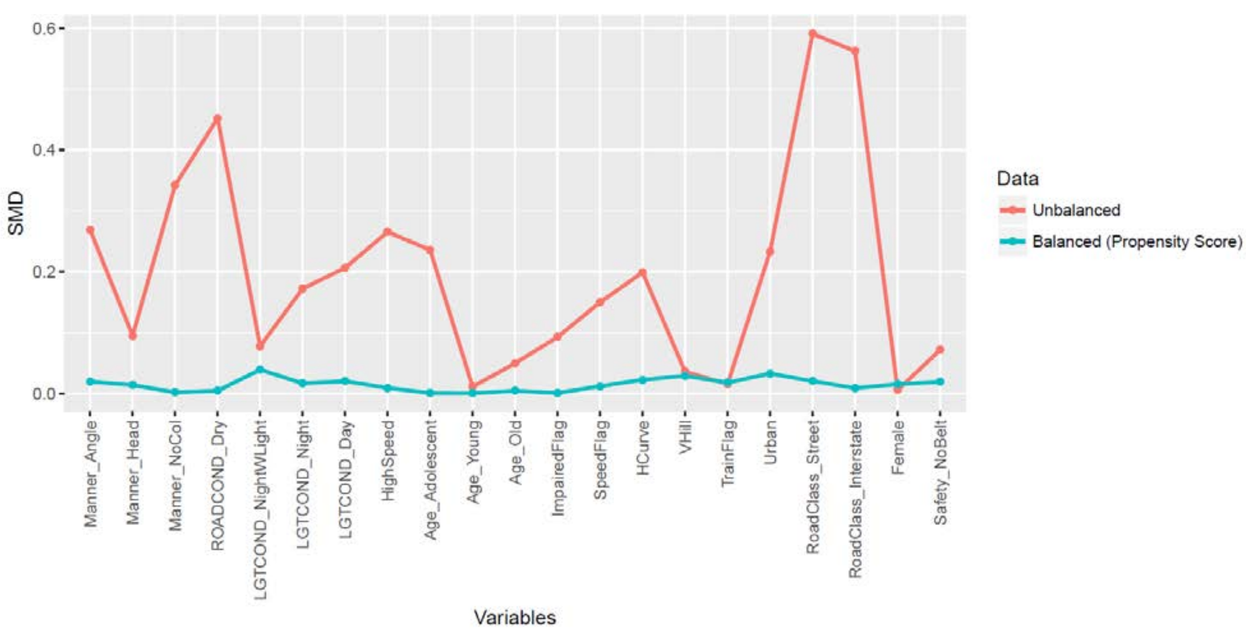
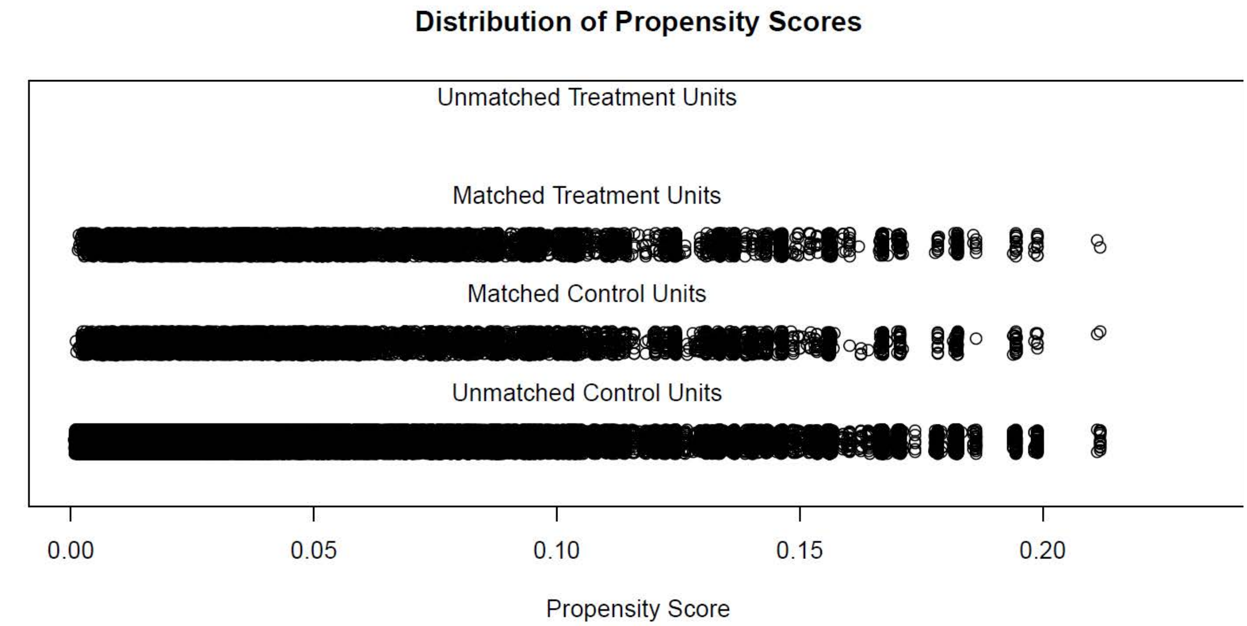
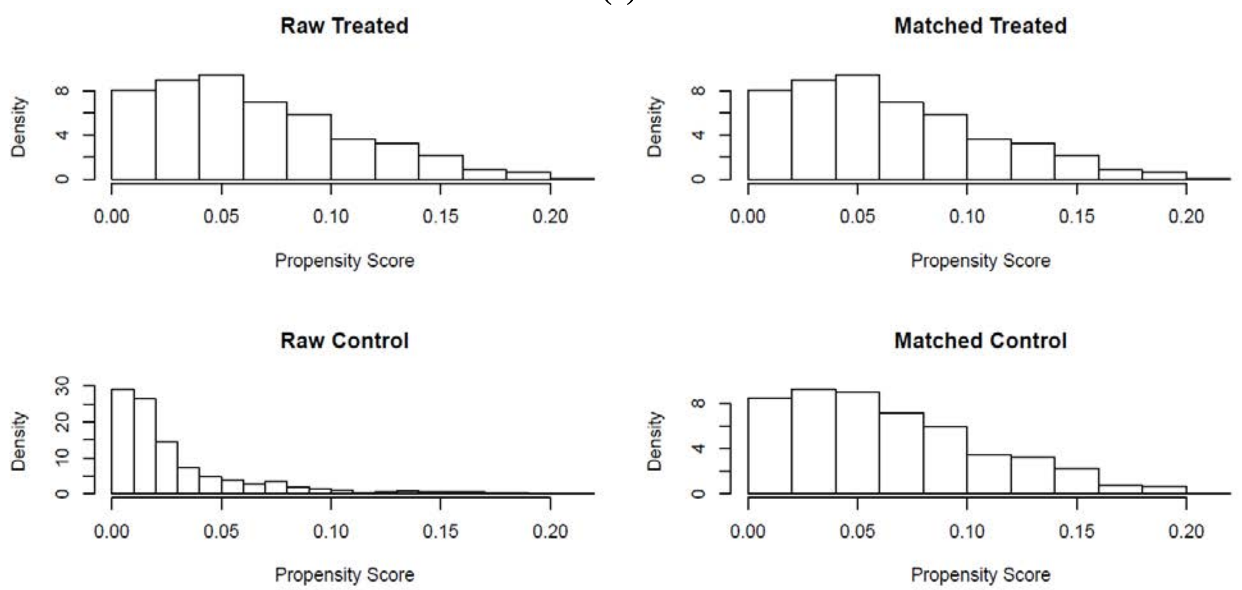


Figure 9-7 SMDs for balanced and unbalanced datasets for work zone

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-8. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-8 Distribution of propensity scores for work zone: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced, therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-13.

Table 9-13 Treatment Effect Analysis for Work Zone

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	-2.09	6393	0.036	-0.0097 -0.0003	-0.005	4.11	1	0.043

Paired t-test analysis on the matched dataset indicate that sample mean differences between matched treatment and control units is -0.005 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, work zone has a causal effect on fatality and incapacitating injuries. On average, the risk of being killed or severely injured in a crash decreases by 0.5% when the crash occurs in a work zone.

Roadway Condition

To evaluate the impact of roadway condition on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if crash } i \text{ occured on a not dry (icy/wet/snowy) surface} \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if crash } i \text{ occured on a dry surface} \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-14) estimates the probability that a crash occurred on not-dry roadways. All covariates except nighttime without streetlight, and train flag are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-14 Propensity Score Model for Roadway Condition

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.91438	0.029349	-31.1552	0.000
Manner_Angle	0.470071	0.017365	27.0699	0.000
Manner_Head	0.616133	0.038739	15.90455	0.000
Manner_NoCol	0.460076	0.012613	36.47536	0.000
LGTCOND_NightWLight	0.128868	0.028573	4.510057	0.000
LGTCOND_Night	-0.05166	0.027694	-1.86526	0.062
LGTCOND_Day	-0.431	0.025406	-16.9646	0.000
HighSpeed	0.116788	0.013425	8.699323	0.000
Age_Adolescent	-0.19909	0.013802	-14.4251	0.000
Age_Young	-0.0854	0.01101	-7.75654	0.000
Age_Old	-0.2214	0.016562	-13.3681	0.000
ImpairedFlag	-1.21335	0.02489	-48.7476	0.000
SpeedFlag	1.696154	0.012183	139.2183	0.000
HCurve	0.156815	0.014566	10.76557	0.000
VHill	0.353925	0.013855	25.54585	0.000
TrainFlag	0.258374	0.235155	1.098738	0.272
WorkZone	-0.89908	0.036183	-24.8482	0.000
Urban	-0.15096	0.013582	-11.1151	0.000
RoadClass_Street	0.033121	0.014522	2.280722	0.023
RoadClass_Interstate	-0.12792	0.01662	-7.6967	0.000
Female	0.115003	0.010652	10.79587	0.000
Safety_NoBelt	-0.8479	0.038261	-22.1612	0.000

Daytime crashes, adolescent drivers, young drivers, old drivers, impaired driving, work zone, urban areas, interstate highways and no-seatbelt are associated with lower probabilities of crashes on not dry roadways while angle crashes, head-on crashes, no-collision crashes, nighttime with streetlight, high-speed segments, speed flag, horizontal curve, vertical hill, street roadway class, and female drivers are associated with higher probabilities of crashes on not-dry roadways.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model cut out 19312 observations with speed flag. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-15. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-15 and shown in Figure 9-9. In the unmatched dataset, SMDs range between 0.060 and 0.792 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 and therefore the matched dataset is balanced.

Table 9-15 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for roadway condition

Variable	Unmatched (Unbalanced) Data	Matched (Balanced) Data

	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	132213	74668		55356	55356	
Manner_Angle	0.11 (0.31)	0.10 (0.30)	0.027	0.11 (0.31)	0.11 (0.31)	0.014
Manner_Head	0.01 (0.12)	0.02 (0.14)	0.045	0.02 (0.14)	0.02 (0.14)	0.002
Manner_NoCol	0.33 (0.47)	0.54 (0.50)	0.424	0.47 (0.50)	0.46 (0.50)	0.015
LGTCOND_NightWLight	0.12 (0.33)	0.14 (0.35)	0.06	0.15 (0.36)	0.14 (0.35)	0.029
LGTCOND_Night	0.13 (0.34)	0.21 (0.41)	0.229	0.19 (0.40)	0.20 (0.40)	0.004
LGTCOND_Day	0.71 (0.45)	0.59 (0.49)	0.259	0.61 (0.49)	0.62 (0.49)	0.018
HighSpeed	0.37 (0.48)	0.49 (0.50)	0.237	0.44 (0.50)	0.45 (0.50)	0.019
Age_Adolescent	0.20 (0.40)	0.17 (0.37)	0.084	0.18 (0.38)	0.18 (0.38)	0.001
Age_Young	0.35 (0.48)	0.35 (0.48)	0.009	0.35 (0.48)	0.35 (0.48)	0.009
Age_Old	0.14 (0.35)	0.09 (0.29)	0.16	0.11 (0.31)	0.11 (0.31)	0.008
ImpairedFlag	0.07 (0.25)	0.04 (0.19)	0.128	0.06 (0.24)	0.05 (0.22)	0.054
SpeedFlag	0.12 (0.32)	0.45 (0.50)	0.792	0.27 (0.44)	0.26 (0.44)	0.02
HCurve	0.12 (0.32)	0.22 (0.41)	0.275	0.18 (0.38)	0.18 (0.38)	0.004
VHill	0.13 (0.33)	0.21 (0.41)	0.215	0.18 (0.38)	0.18 (0.39)	0.01
TrainFlag	0.00 (0.02)	0.00 (0.02)	0.003	0.00 (0.02)	0.00 (0.02)	0.002
WorkZone	0.04 (0.20)	0.01 (0.12)	0.156	0.02 (0.15)	0.02 (0.14)	0.026
Urban	0.61 (0.49)	0.47 (0.50)	0.271	0.52 (0.50)	0.51 (0.50)	0.012
RoadClass_Street	0.37 (0.48)	0.29 (0.45)	0.185	0.33 (0.47)	0.32 (0.47)	0.006
RoadClass_Interstate	0.13 (0.34)	0.14 (0.35)	0.018	0.13 (0.34)	0.13 (0.33)	0.011
Female	0.60 (0.49)	0.57 (0.50)	0.06	0.57 (0.50)	0.58 (0.49)	0.014
Safety_NoBelt	0.02 (0.16)	0.02 (0.13)	0.062	0.02 (0.15)	0.02 (0.14)	0.019

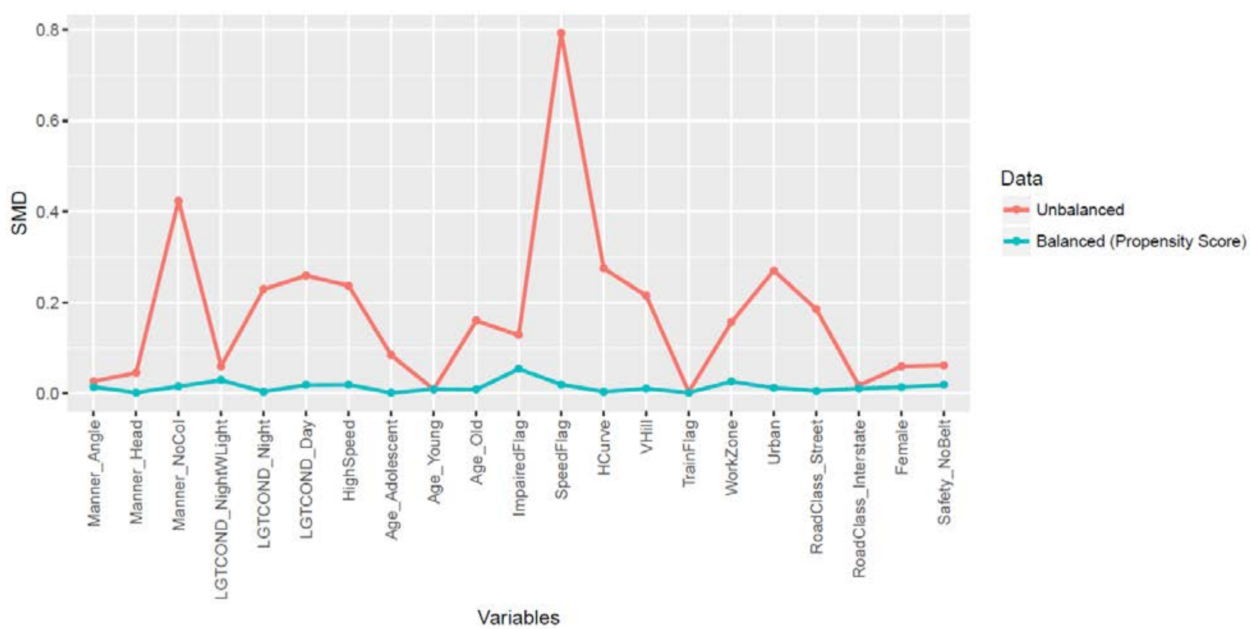
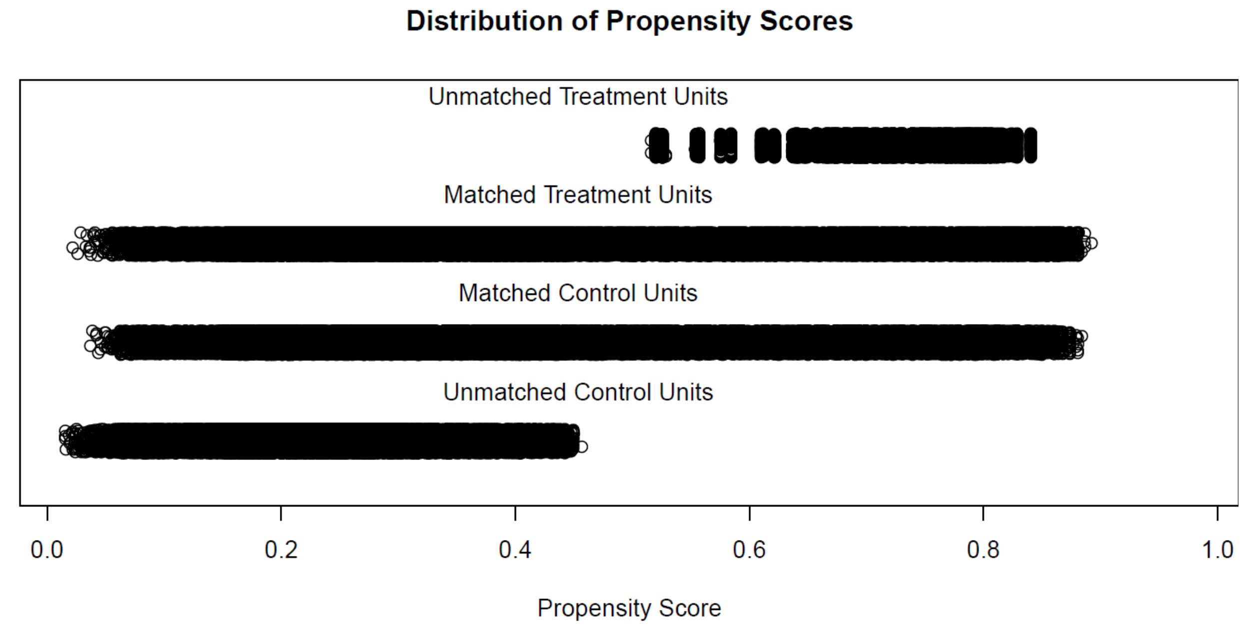
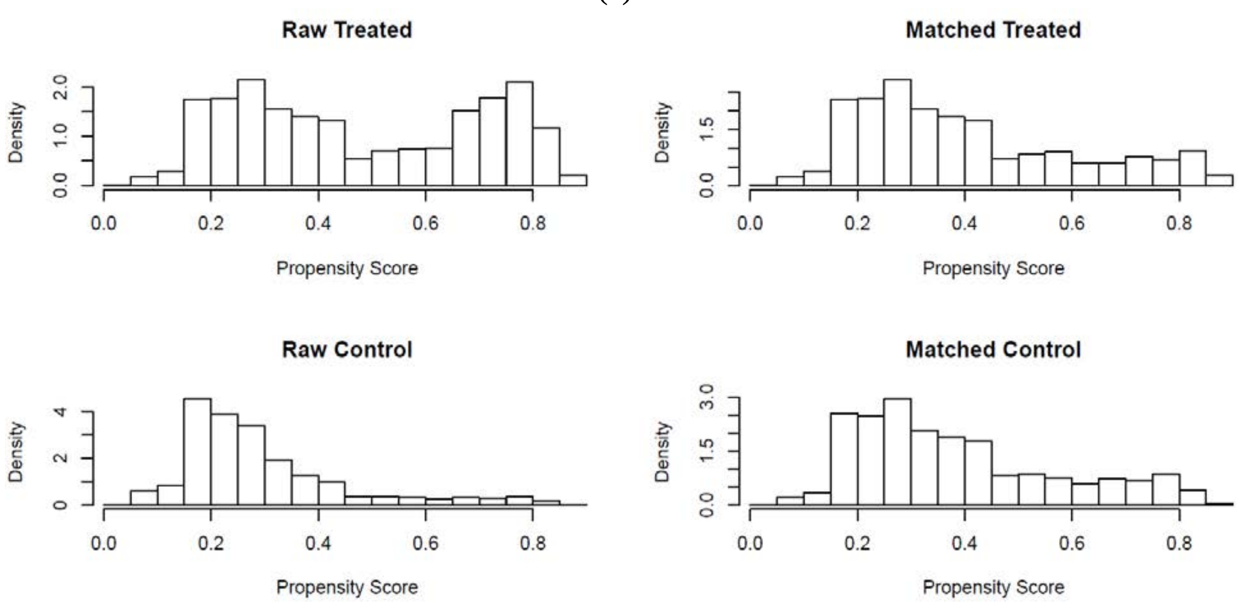


Figure 9-9 SMDs for balanced and unbalanced datasets for roadway condition

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-10. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-10 Distribution of propensity scores for roadway condition: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced,

therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-16.

Table 9-16 Treatment Effect Analysis for Roadway Condition

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	-15.54	55355	0.000	-0.017 -0.013	-0.015	239.98	1	0.000

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is -0.015 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, not-dry roadway surface has a causal effect on fatality and incapacitating injuries. On average, the risk of being killed or severely injured in a crash decreases y 1.5% when a crash occurs on a not-dry (snowy, wet, icy) roadway.

Street Light

To evaluate the impact of street lights on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if crash } i \text{ occured at nighttime on asegmet with street light} \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if crash } i \text{ occured at nighttime on asegmet without street light} \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-17) estimates the probability that a nighttime crash occurred on a segment with street light. All covariates except angle crashes, dry surface condition, adolescent drivers, train flag, street roadway class, female drivers, and no-seatbelt are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-17 Propensity Score Model for Street Light

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.24992	0.041803	-5.97844	0.000
Manner_Angle	0.023467	0.051969	0.451554	0.652
Manner_Head	-0.41322	0.081595	-5.06431	0.000
Manner_NoCol	-0.83286	0.028245	-29.4874	0.000
ROADCOND_Dry	0.01807	0.025912	0.697379	0.486
HighSpeed	-2.01307	0.030793	-65.3749	0.000
Age_Adolescent	0.056377	0.030939	1.822192	0.068
Age_Young	0.055431	0.025302	2.190736	0.028
Age_Old	-0.12881	0.050974	-2.52697	0.012
ImpairedFlag	0.146299	0.035198	4.156467	0.000
SpeedFlag	-0.11186	0.028586	-3.91296	0.000
HCurve	-0.26999	0.033005	-8.18021	0.000
VHill	-0.33508	0.033767	-9.92345	0.000
TrainFlag	-0.69132	0.402199	-1.71885	0.086
WorkZone	-0.15126	0.075563	-2.00174	0.045
Urban	2.458764	0.027886	88.17271	0.000
RoadClass_Street	0.038831	0.033561	1.157052	0.247
RoadClass_Interstate	0.759845	0.037814	20.09432	0.000
Female	0.001733	0.024731	0.070062	0.944
Safety_NoBelt	-0.08027	0.070854	-1.13285	0.257

Head-on crashes, no-collision crashes, high-speed segments, old drivers, speed flag, horizontal curves, vertical hills, and work zone are associated with lower probabilities of street light involvement while young drivers, impaired driving, urban areas, and interstate highways are associated with higher probabilities of street light involvement.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model cut out 17420 observations with street lights. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-18. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-18 Table 9-6and shown in Figure 9-11 In the unmatched dataset, SMDs range between 0.006 and 1.802 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 and therefore the matched dataset is balanced.

Table 9-18 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for Street Light

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	32804	27044		9624	9624	
Manner_Angle	0.03 (0.18)	0.09 (0.29)	0.246	0.07 (0.26)	0.07 (0.25)	0.02
Manner_Head	0.02 (0.14)	0.03 (0.16)	0.037	0.03 (0.16)	0.03 (0.16)	0.003

Manner_NoCol	0.75 (0.43)	0.38 (0.49)	0.785	0.51 (0.50)	0.51 (0.50)	0.002
ROADCOND_Dry	0.52 (0.50)	0.60 (0.49)	0.17	0.56 (0.50)	0.56 (0.50)	<0.001
HighSpeed	0.72 (0.45)	0.17 (0.38)	1.317	0.37 (0.48)	0.34 (0.48)	0.055
Age_Adolescent	0.15 (0.35)	0.30 (0.46)	0.37	0.23 (0.42)	0.22 (0.42)	0.021
Age_Young	0.37 (0.48)	0.40 (0.49)	0.062	0.39 (0.49)	0.38 (0.49)	0.016
Age_Old	0.05 (0.22)	0.07 (0.25)	0.059	0.07 (0.25)	0.07 (0.25)	0.006
ImpairedFlag	0.14 (0.35)	0.14 (0.35)	0.006	0.16 (0.36)	0.16 (0.37)	0.026
SpeedFlag	0.36 (0.48)	0.23 (0.42)	0.286	0.30 (0.46)	0.29 (0.45)	0.021
HCurve	0.26 (0.44)	0.12 (0.33)	0.348	0.17 (0.37)	0.18 (0.39)	0.042
VHill	0.21 (0.41)	0.11 (0.32)	0.261	0.16 (0.36)	0.16 (0.37)	0.013
TrainFlag	0.00 (0.02)	0.00 (0.03)	0.006	0.00 (0.03)	0.00 (0.04)	0.012
WorkZone	0.02 (0.14)	0.02 (0.15)	0.033	0.03 (0.18)	0.02 (0.15)	0.044
Urban	0.19 (0.39)	0.86 (0.35)	1.802	0.65 (0.48)	0.65 (0.48)	0.006
RoadClass_Street	0.11 (0.32)	0.54 (0.50)	1.021	0.36 (0.48)	0.42 (0.49)	0.11
RoadClass_Interstate	0.13 (0.33)	0.15 (0.35)	0.054	0.14 (0.35)	0.12 (0.32)	0.079
Female	0.45 (0.50)	0.52 (0.50)	0.141	0.50 (0.50)	0.50 (0.50)	<0.001
Safety_NoBelt	0.04 (0.20)	0.02 (0.15)	0.097	0.03 (0.17)	0.03 (0.16)	0.012

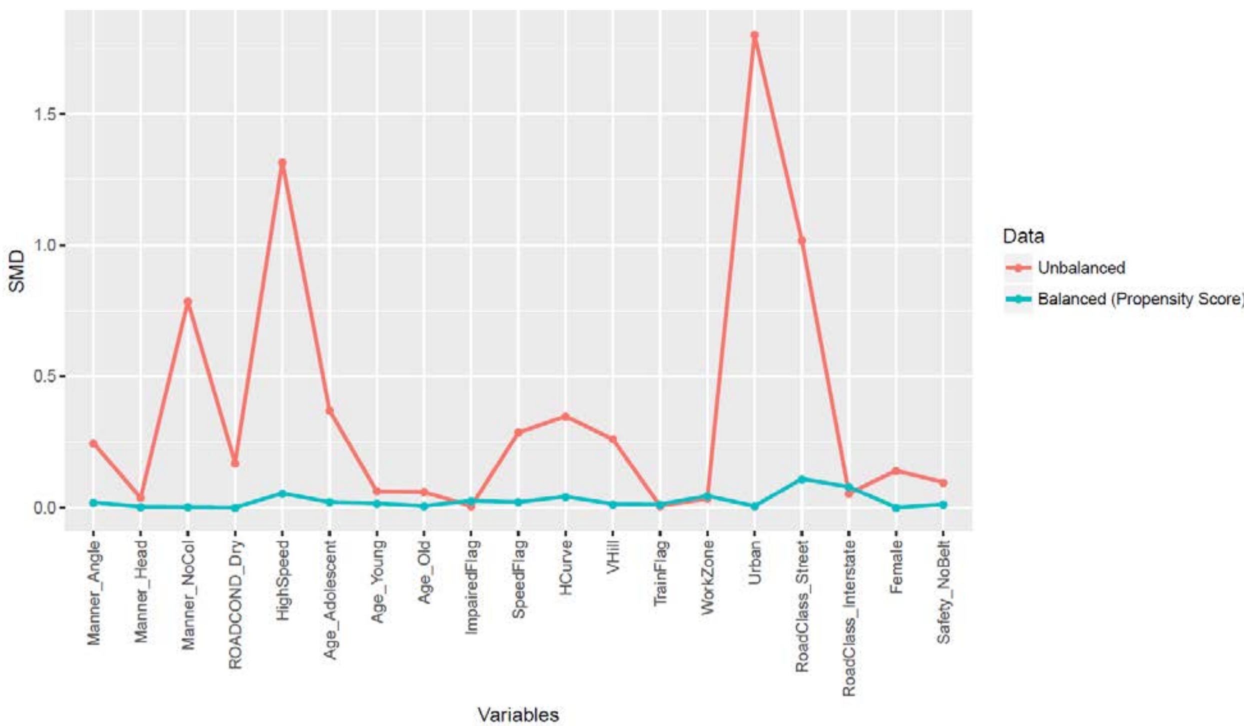
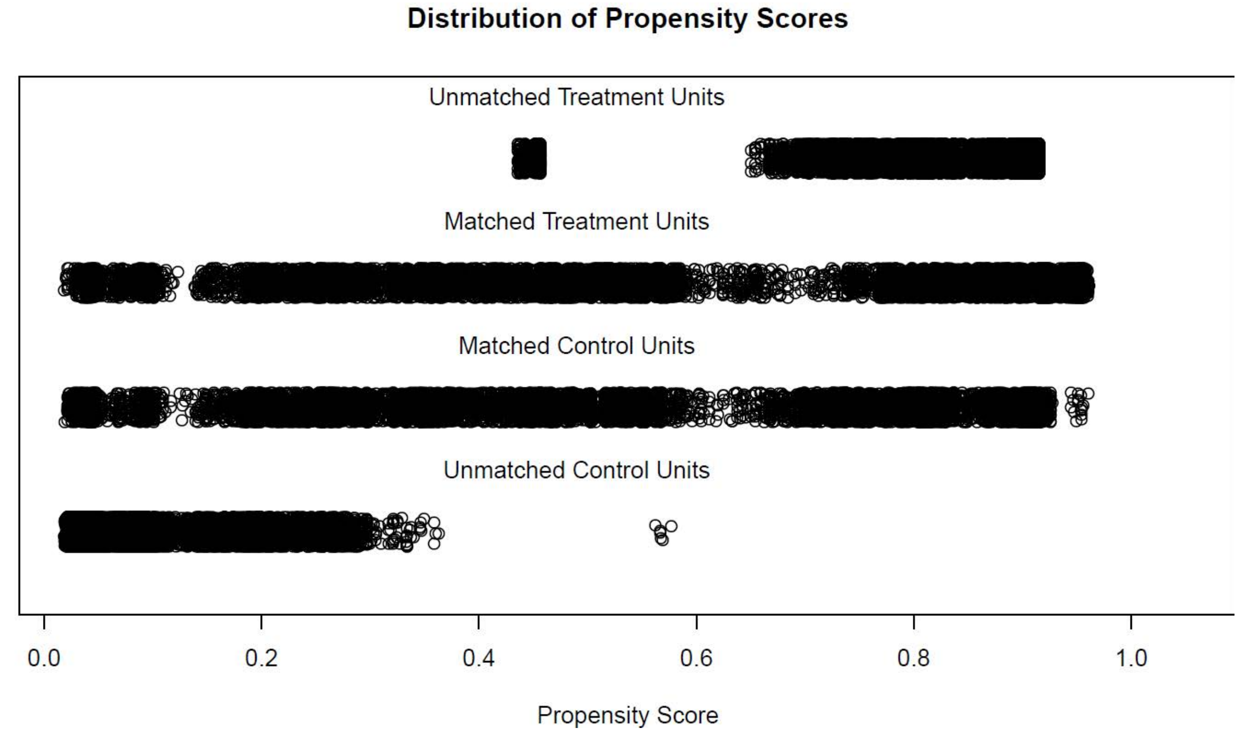


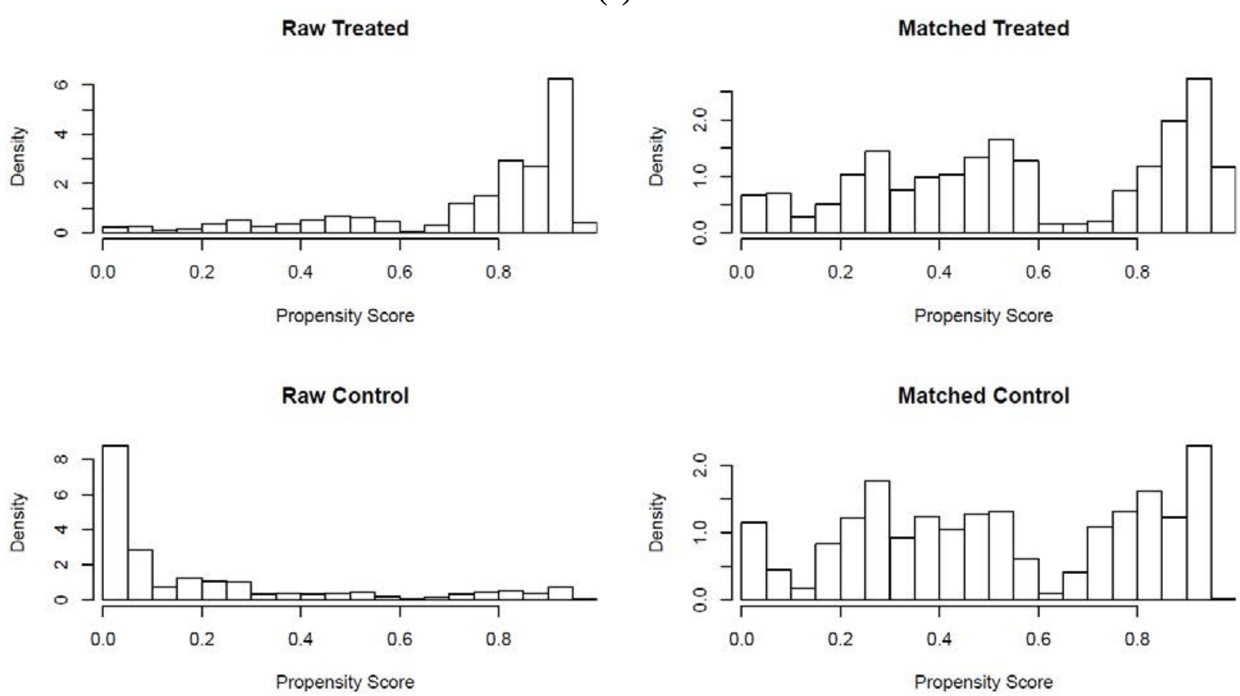
Figure 9-11 SMDs for balanced and unbalanced datasets for street light

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-12. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are

different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-12 Distribution of propensity scores for street light: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced, therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-19.

Table 9-19 Treatment Effect Analysis for Street Light

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	-5.07	9623	0.000	-0.016 -0.007	-0.011	25.22	1	0.000

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is -0.011 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, street light involvement in a crash has a causal effect on fatality and incapacitating injuries. On average, the risk of being killed or severely injured in a nighttime crash decreases by 1.1% when street lights are present.

Collision Manner

To evaluate the impact of collision manner on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if manner of collision for crash } i \text{ is head} - \text{on} \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if manner of collision for crash } i \text{ is not head} - \text{on} \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-20) estimates the probability that a crash manner was head-on. All covariates except nighttime without street light, high-speed segments, speed flag, vertical hills, and train flag are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-20 Propensity Score Model for Collision Manner

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.11857	0.098621	-41.7618	0.000
ROADCOND_Dry	-0.35452	0.03758	-9.43365	0.000
LGTCOND_NightWLight	0.305266	0.092977	3.283237	0.001
LGTCOND_Night	-0.11619	0.091849	-1.26498	0.206
LGTCOND_Day	-0.17129	0.085187	-2.01076	0.044
HighSpeed	0.067187	0.046518	1.444335	0.149
Age_Adolescent	0.183259	0.043597	4.203447	0.000
Age_Young	0.147737	0.036977	3.995398	0.000
Age_Old	0.494215	0.049175	10.05009	0.000
ImpairedFlag	0.768765	0.055806	13.77574	0.000
SpeedFlag	0.031474	0.042639	0.738155	0.460
HCurve	0.288032	0.044486	6.474722	0.000
VHill	0.081004	0.045669	1.773702	0.076
TrainFlag	-10.7809	93.90642	-0.1148	0.909
WorkZone	-0.44455	0.153591	-2.89439	0.004
Urban	-0.42956	0.047084	-9.12314	0.000
RoadClass_Street	0.466768	0.049258	9.475893	0.000
RoadClass_Interstate	-1.06537	0.091457	-11.6489	0.000
Female	0.198024	0.035983	5.503313	0.000
Safety_NoBelt	0.954626	0.073464	12.99442	0.000

Dry roadway conditions, daytime crashes, work zone, urban areas, and interstate highways are associated with lower probabilities of head-on collision involvement while nighttime with street light, adolescent drivers, young drivers, old drivers, impaired driving, horizontal curve, street roadways, female drivers and no-seatbelt are associated with higher probabilities of head-on collision involvement.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model did not cut out any observation. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-21. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-21 and shown in Figure 9-13. In the unmatched dataset, SMDs range between 0.032 and 0.353 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 and therefore the matched dataset is balanced.

Table 9-21 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for Collision Manner

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	203383	3498		3498	3498	

ROADCOND_Dry	0.64 (0.48)	0.56 (0.50)	0.169	0.54 (0.50)	0.56 (0.50)	0.029
LGTCOND_NightWLight	0.13 (0.34)	0.19 (0.40)	0.177	0.19 (0.40)	0.19 (0.40)	0.001
LGTCOND_Night	0.16 (0.37)	0.18 (0.39)	0.067	0.19 (0.39)	0.18 (0.39)	0.012
LGTCOND_Day	0.67 (0.47)	0.58 (0.49)	0.196	0.57 (0.49)	0.58 (0.49)	0.003
HighSpeed	0.41 (0.49)	0.38 (0.49)	0.072	0.40 (0.49)	0.38 (0.49)	0.036
Age_Adolescent	0.19 (0.39)	0.22 (0.42)	0.088	0.22 (0.41)	0.22 (0.42)	0.014
Age_Young	0.35 (0.48)	0.37 (0.48)	0.033	0.36 (0.48)	0.37 (0.48)	0.001
Age_Old	0.12 (0.33)	0.16 (0.37)	0.107	0.16 (0.36)	0.16 (0.37)	0.013
ImpairedFlag	0.06 (0.23)	0.13 (0.34)	0.268	0.13 (0.34)	0.13 (0.34)	0.01
SpeedFlag	0.24 (0.43)	0.28 (0.45)	0.087	0.28 (0.45)	0.28 (0.45)	0.011
HCurve	0.15 (0.36)	0.22 (0.41)	0.165	0.22 (0.42)	0.22 (0.41)	0.011
VHill	0.16 (0.36)	0.18 (0.39)	0.071	0.18 (0.39)	0.18 (0.39)	0.001
TrainFlag	0.00 (0.02)	0.00 (0.00)	0.029	0.00 (0.00)	0.00 (0.00)	<0.001
WorkZone	0.03 (0.17)	0.01 (0.11)	0.128	0.01 (0.10)	0.01 (0.11)	0.022
Urban	0.56 (0.50)	0.50 (0.50)	0.125	0.48 (0.50)	0.50 (0.50)	0.041
RoadClass_Street	0.34 (0.47)	0.42 (0.49)	0.172	0.42 (0.49)	0.42 (0.49)	0.002
RoadClass_Interstate	0.14 (0.34)	0.04 (0.19)	0.353	0.04 (0.20)	0.04 (0.19)	0.004
Female	0.59 (0.49)	0.60 (0.49)	0.032	0.59 (0.49)	0.60 (0.49)	0.022
Safety_NoBelt	0.02 (0.14)	0.07 (0.25)	0.22	0.06 (0.23)	0.07 (0.25)	0.038

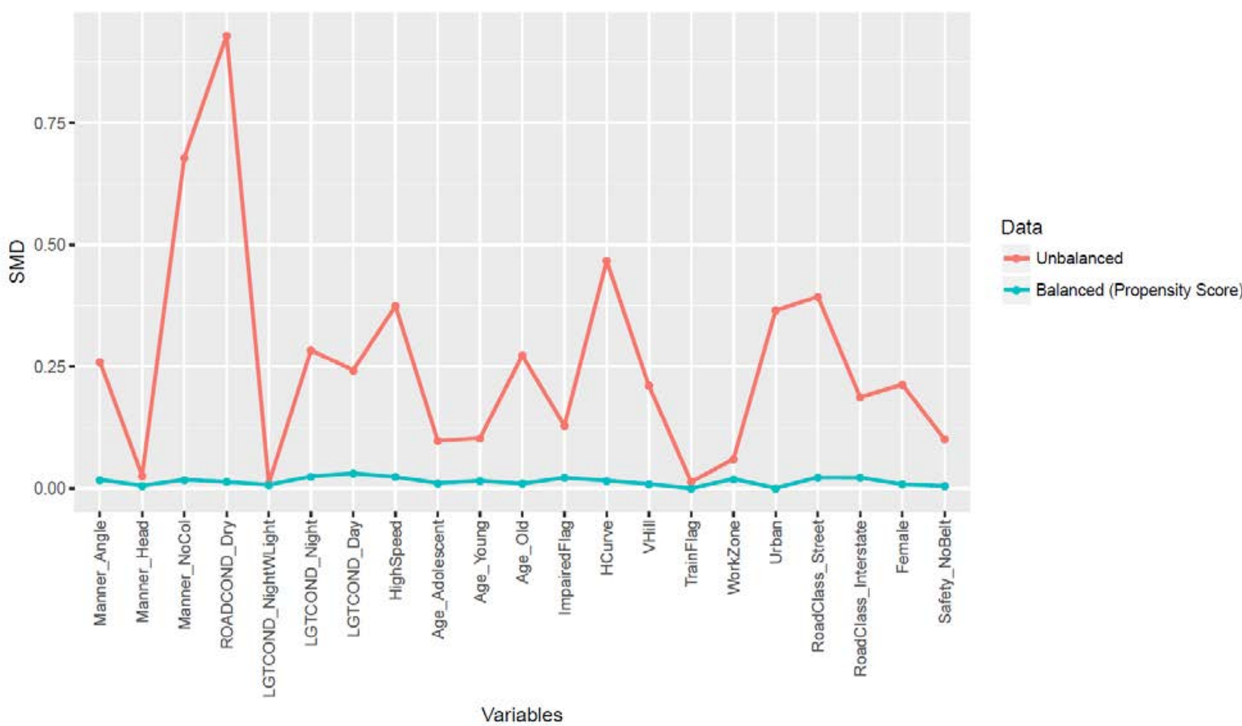
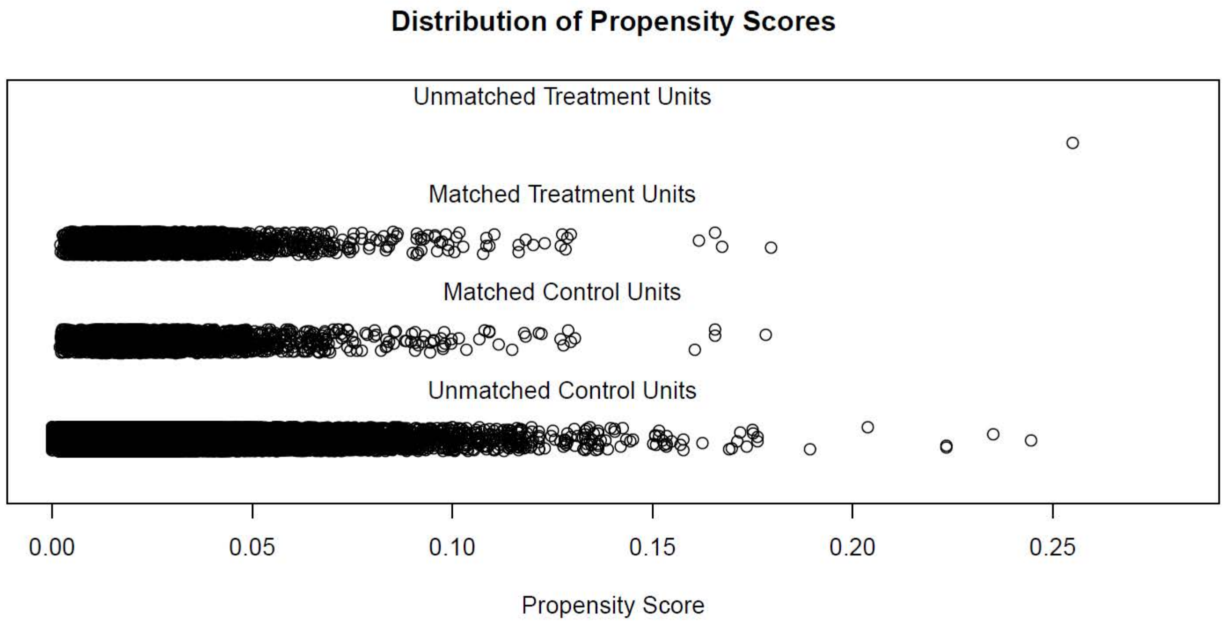


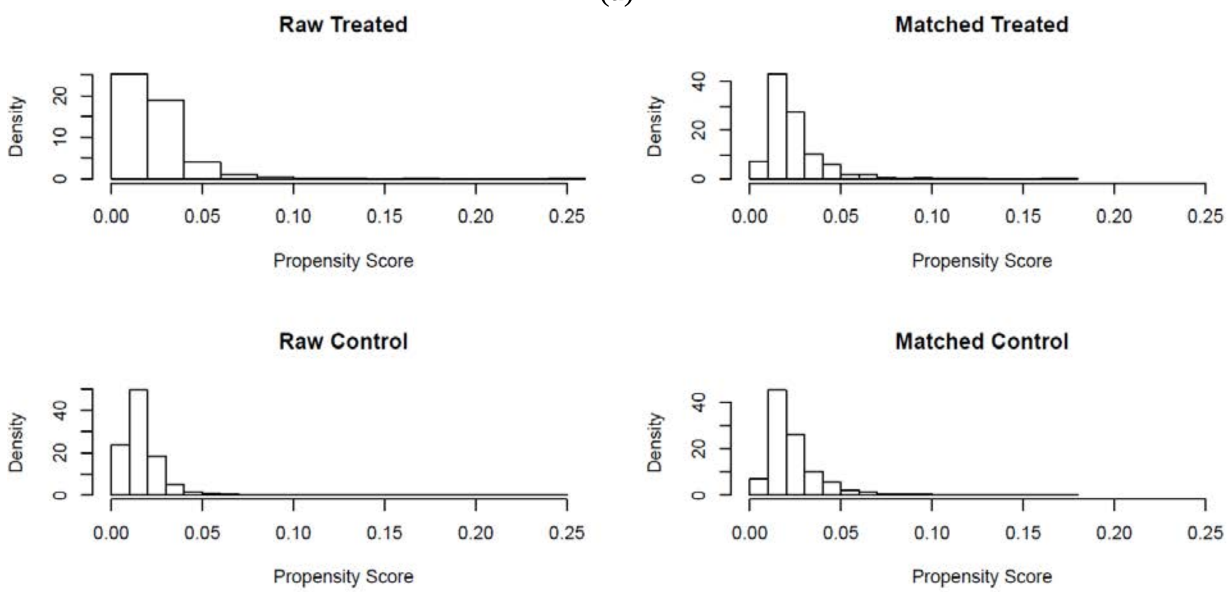
Figure 9-13 SMDs for balanced and unbalanced datasets for collision manner

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-14. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are

different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-14 Distribution of propensity scores for collision manner: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced, therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-22.

Table 9-22 Treatment Effect Analysis for Collision Manner

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	19.83	3497	0.000	0.125 0.153	0.139	352.01	1	0.000

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is 0.139 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, head-on collision involvement in a crash has a causal effect on fatality and incapacitating injuries. On average, the risk of being killed or severely injured in a crash increases by 13.9% when the manner of crash is head-on.

Horizontal Curve

To evaluate the impact of horizontal curves on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if crash } i \text{ occured on a horozontal curve} \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if crash } i \text{ did not occured on a horozontal curve} \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-23) estimates the probability that a crash occurred on a horizontal curve. All covariates except angle crashes, nighttime without street light, daytime crashes, old drivers, train flag, and no-seatbelt are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-23 Propensity Score Model for Horizontal Curve

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.32984	0.038585	-60.3826	0.000
Manner_Angle	0.008892	0.03127	0.284347	0.776

Manner_Head	1.088524	0.044814	24.28959	0.000
Manner_NoCol	1.289557	0.017112	75.35862	0.000
ROADCOND_Dry	-0.16873	0.014562	-11.587	0.000
LGTCOND_NightWLight	-0.08241	0.036802	-2.23934	0.025
LGTCOND_Night	0.042911	0.032958	1.301981	0.193
LGTCOND_Day	-0.0111	0.031289	-0.35472	0.723
HighSpeed	-0.21087	0.016755	-12.5856	0.000
Age_Adolescent	-0.07895	0.018721	-4.2174	0.000
Age_Young	0.060535	0.014254	4.24684	0.000
Age_Old	-0.04347	0.023681	-1.83576	0.066
ImpairedFlag	0.246387	0.024965	9.869175	0.000
SpeedFlag	0.699098	0.01485	47.07594	0.000
VHill	0.854195	0.015197	56.20936	0.000
TrainFlag	-0.75784	0.523033	-1.44894	0.147
WorkZone	-0.16562	0.046332	-3.57467	0.000
Urban	-0.37927	0.018077	-20.9811	0.000
RoadClass_Street	-0.23198	0.020623	-11.2486	0.000
RoadClass_Interstate	-0.30376	0.021986	-13.8162	0.000
Female	-0.05138	0.013503	-3.80474	0.000
Safety_NoBelt	0.07358	0.038327	1.919806	0.055

Dry roadway condition, nighttime with street light, high speed segments, adolescent drivers, work zone, urban areas, street roadways, interstate highways and female drivers are associated with lower probabilities of horizontal curve involvement while head-on crashes, no-collision crashes, young drivers, impaired driving, speed flag, and vertical hills are associated with higher probabilities of horizontal curve involvement.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model did not cut out any observations from the treatment units. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-24. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-24 and shown in Figure 9-15. In the unmatched dataset, SMDs range between 0.011 and 0.841 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 and therefore the matched dataset is balanced.

Table 9-24 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for Horizontal Curve

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	175524	32033		32033	32033	
Manner_Angle	0.11 (0.32)	0.04 (0.20)	0.265	0.04 (0.20)	0.04 (0.20)	0.006
Manner_Head	0.02 (0.12)	0.02 (0.15)	0.059	0.03 (0.16)	0.02 (0.15)	0.008
Manner_NoCol	0.34 (0.48)	0.73 (0.44)	0.841	0.73 (0.44)	0.73 (0.44)	0.011
ROADCOND_Dry	0.67 (0.47)	0.49 (0.50)	0.37	0.48 (0.50)	0.49 (0.50)	0.005

LGTCOND_NightWLight	0.14 (0.34)	0.11 (0.31)	0.091	0.10 (0.31)	0.11 (0.31)	0.005
LGTCOND_Night	0.14 (0.35)	0.27 (0.44)	0.328	0.28 (0.45)	0.27 (0.44)	0.015
LGTCOND_Day	0.69 (0.46)	0.57 (0.49)	0.232	0.57 (0.49)	0.58 (0.49)	0.009
HighSpeed	0.39 (0.49)	0.54 (0.50)	0.296	0.55 (0.50)	0.54 (0.50)	0.02
Age_Adolescent	0.20 (0.40)	0.15 (0.36)	0.114	0.15 (0.36)	0.15 (0.36)	0.002
Age_Young	0.35 (0.48)	0.35 (0.48)	0.011	0.35 (0.48)	0.35 (0.48)	0.011
Age_Old	0.13 (0.34)	0.08 (0.27)	0.159	0.09 (0.28)	0.08 (0.27)	0.018
ImpairedFlag	0.05 (0.22)	0.09 (0.29)	0.166	0.09 (0.29)	0.09 (0.29)	0.001
SpeedFlag	0.20 (0.40)	0.46 (0.50)	0.564	0.45 (0.50)	0.45 (0.50)	0.012
VHill	0.13 (0.33)	0.31 (0.46)	0.45	0.30 (0.46)	0.31 (0.46)	0.013
TrainFlag	0.00 (0.02)	0.00 (0.01)	0.02	0.00 (0.01)	0.00 (0.01)	0.003
WorkZone	0.03 (0.18)	0.02 (0.13)	0.095	0.02 (0.14)	0.02 (0.13)	0.004
Urban	0.60 (0.49)	0.34 (0.48)	0.528	0.34 (0.47)	0.35 (0.48)	0.012
RoadClass_Street	0.37 (0.48)	0.21 (0.41)	0.347	0.21 (0.41)	0.21 (0.41)	0.003
RoadClass_Interstate	0.14 (0.35)	0.11 (0.31)	0.092	0.11 (0.31)	0.11 (0.31)	0.005
Female	0.60 (0.49)	0.50 (0.50)	0.201	0.51 (0.50)	0.50 (0.50)	0.015
Safety_NoBelt	0.02 (0.14)	0.03 (0.18)	0.095	0.04 (0.18)	0.03 (0.18)	0.006

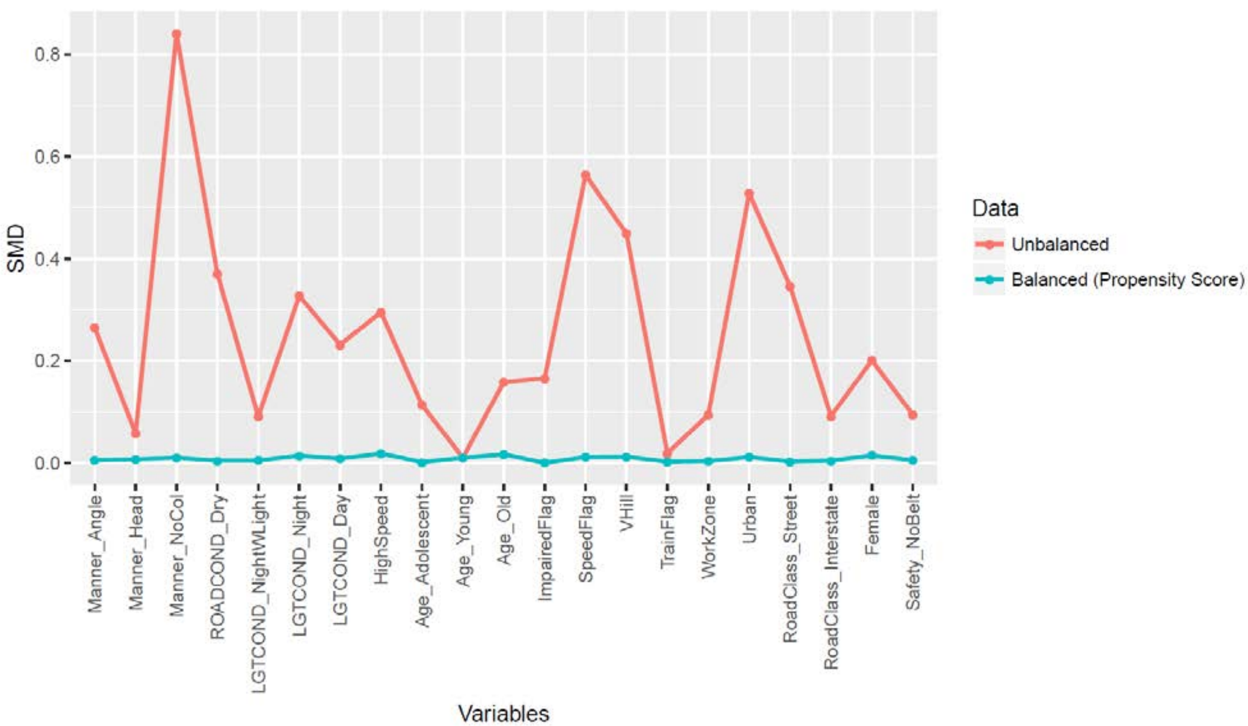
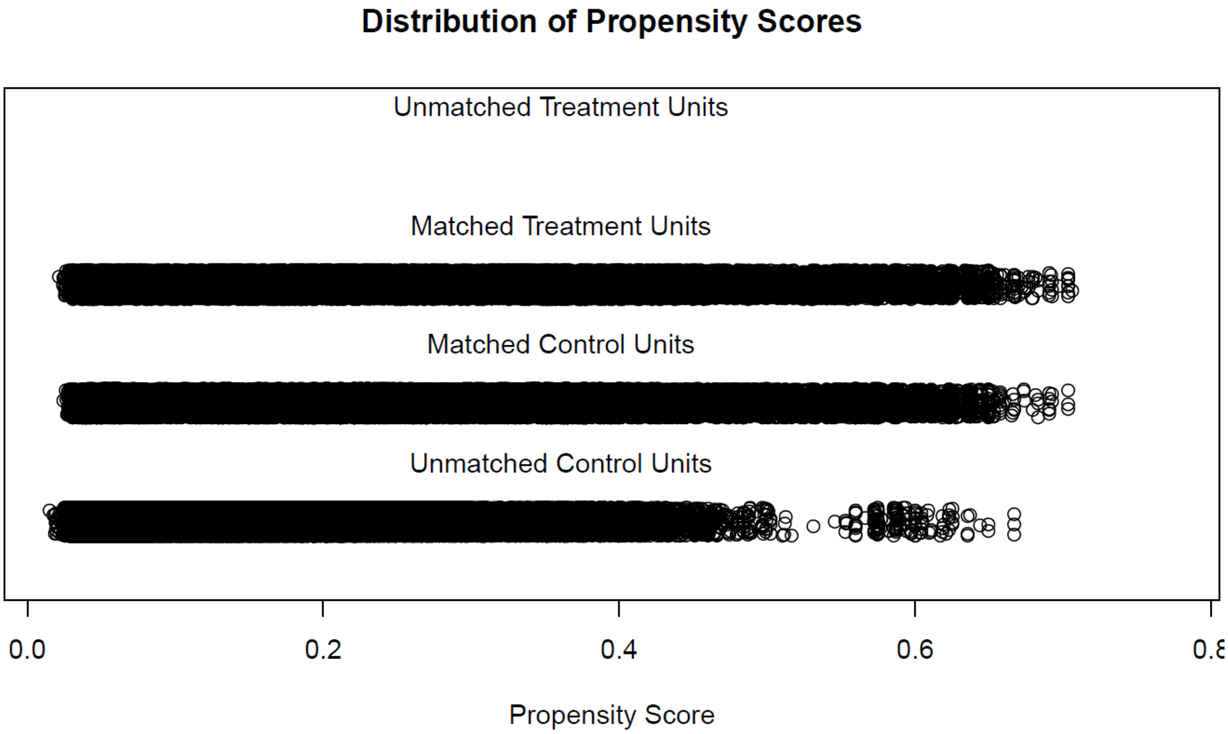
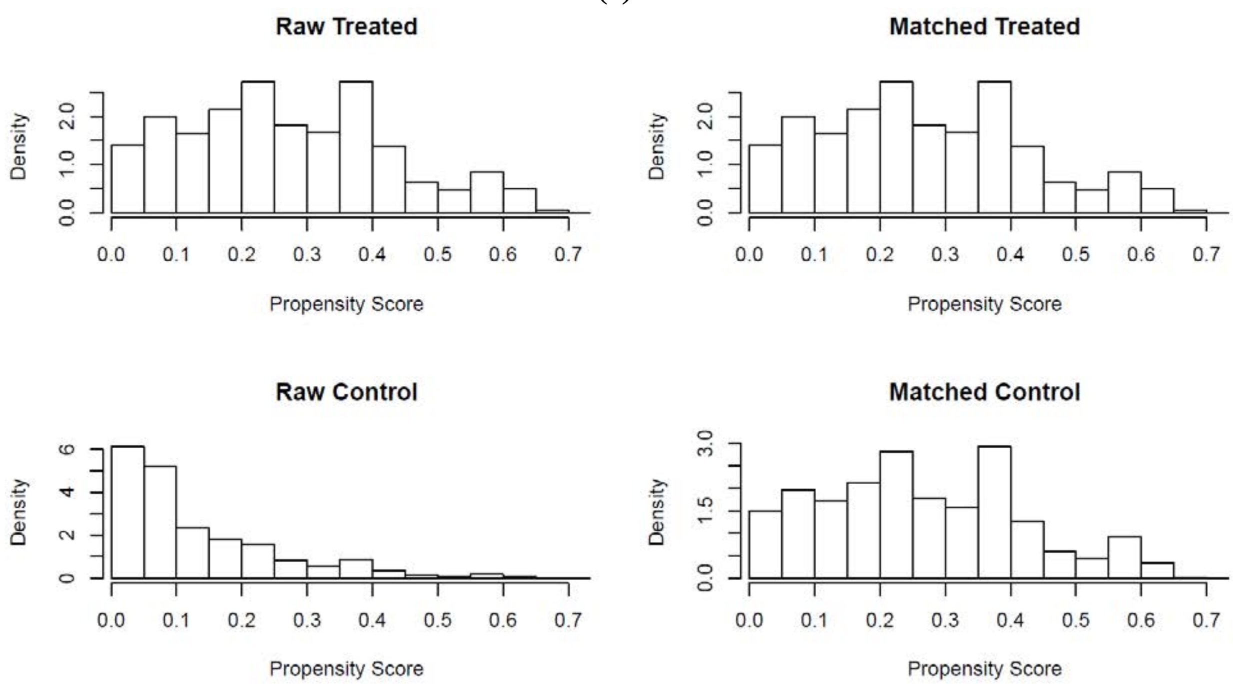


Figure 9-15 SMDs for balanced and unbalanced datasets for horizontal curve

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-16. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-16 Distribution of propensity scores for horizontal curve: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced,

therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-25.

Table 9-25 Treatment Effect Analysis for Horizontal Curve

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	2.13	32032	0.033	0.000 0.006	0.003	4.43	1	0.035

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is 0.003 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, horizontal curve involvement in a crash has a causal effect on fatality and incapacitating injuries. On average, the risk of being killed or severely injured in a crash increases by 0.3% when the crash occurs at a horizontal curve.

Vertical Hills

To evaluate the impact of vertical hills on fatality and incapacitating injuries, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if crash } i \text{ occured on a vertical hill} \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if crash } i \text{ did not occured on a vertical hill} \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for crash } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for crash } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-26) estimates the probability that a crash occurred on a vertical hill. All covariates except old drivers, train flag, and female drivers are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-26 Propensity Score Model for Vertical Hill

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.35809	0.03523	-38.5488	0.000

Manner_Angle	-0.13229	0.024666	-5.36324	0.000
Manner_Head	0.158189	0.046465	3.404509	0.001
Manner_NoCol	0.153392	0.015696	9.772929	0.000
ROADCOND_Dry	-0.36034	0.013804	-26.1042	0.000
LGTCOND_NightWLight	-0.32746	0.03476	-9.42067	0.000
LGTCOND_Night	-0.20101	0.03138	-6.40577	0.000
LGTCOND_Day	-0.16158	0.02908	-5.55655	0.000
HighSpeed	0.133774	0.016023	8.349065	0.000
Age_Adolescent	-0.0369	0.017286	-2.13435	0.033
Age_Young	-0.02962	0.013544	-2.18731	0.029
Age_Old	-0.02549	0.020589	-1.23823	0.216
ImpairedFlag	-0.09567	0.027646	-3.46047	0.001
SpeedFlag	0.095916	0.015191	6.314086	0.000
HCurve	0.851666	0.015155	56.19677	0.000
TrainFlag	-0.46072	0.396862	-1.16091	0.246
WorkZone	0.1333	0.037244	3.579073	0.000
Urban	-0.39997	0.016693	-23.9605	0.000
RoadClass_Street	-0.05478	0.018855	-2.90516	0.004
RoadClass_Interstate	-0.12026	0.020194	-5.95523	0.000
Female	-0.00966	0.012909	-0.74844	0.454
Safety_NoBelt	0.091861	0.040092	2.29129	0.022

Dry roadway condition, nighttime with street light, high speed segments, adolescent drivers, work zone, urban areas, street roadways, interstate highways and female drivers are associated with lower probabilities of horizontal curve involvement while head-on crashes, no-collision crashes, young drivers, impaired driving, speed flag, and vertical hills are associated with higher probabilities of horizontal curve involvement.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model did not cut out any observations from the treatment units. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-27. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-27 and shown in Figure 9-17. In the unmatched dataset, SMDs range between 0.010 and 0.447 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than 0.1 and therefore the matched dataset is balanced.

Table 9-27 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for Vertical Hill

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	175074	32483		32483	32483	
Manner_Angle	0.11 (0.31)	0.07 (0.26)	0.13	0.07 (0.25)	0.07 (0.26)	0.012
Manner_Head	0.02 (0.13)	0.02 (0.14)	0.025	0.02 (0.14)	0.02 (0.14)	0.007
Manner_NoCol	0.38 (0.48)	0.54 (0.50)	0.331	0.55 (0.50)	0.54 (0.50)	0.009

ROADCOND_Dry	0.66 (0.47)	0.52 (0.50)	0.286	0.52 (0.50)	0.52 (0.50)	<0.001
LGTCOND_NightWLight	0.14 (0.34)	0.10 (0.29)	0.132	0.09 (0.29)	0.10 (0.29)	0.016
LGTCOND_Night	0.15 (0.36)	0.21 (0.41)	0.167	0.22 (0.41)	0.21 (0.41)	0.013
LGTCOND_Day	0.67 (0.47)	0.64 (0.48)	0.076	0.64 (0.48)	0.64 (0.48)	0.004
HighSpeed	0.39 (0.49)	0.52 (0.50)	0.264	0.53 (0.50)	0.52 (0.50)	0.006
Age_Adolescent	0.19 (0.40)	0.17 (0.37)	0.071	0.16 (0.37)	0.17 (0.37)	0.013
Age_Young	0.35 (0.48)	0.34 (0.47)	0.019	0.34 (0.48)	0.34 (0.47)	0.006
Age_Old	0.13 (0.33)	0.11 (0.31)	0.063	0.11 (0.31)	0.11 (0.31)	0.003
ImpairedFlag	0.06 (0.23)	0.06 (0.24)	0.01	0.06 (0.24)	0.06 (0.24)	0.012
SpeedFlag	0.22 (0.42)	0.33 (0.47)	0.249	0.33 (0.47)	0.33 (0.47)	0.001
HCurve	0.13 (0.33)	0.31 (0.46)	0.447	0.29 (0.46)	0.31 (0.46)	0.023
TrainFlag	0.00 (0.02)	0.00 (0.01)	0.013	0.00 (0.01)	0.00 (0.01)	0.002
WorkZone	0.03 (0.17)	0.03 (0.17)	0.016	0.03 (0.17)	0.03 (0.17)	0.009
Urban	0.59 (0.49)	0.40 (0.49)	0.382	0.39 (0.49)	0.40 (0.49)	0.033
RoadClass_Street	0.36 (0.48)	0.24 (0.43)	0.252	0.24 (0.43)	0.24 (0.43)	0.016
RoadClass_Interstate	0.14 (0.34)	0.13 (0.33)	0.033	0.12 (0.33)	0.13 (0.33)	0.006
Female	0.59 (0.49)	0.55 (0.50)	0.075	0.55 (0.50)	0.55 (0.50)	0.004
Safety_NoBelt	0.02 (0.14)	0.03 (0.16)	0.041	0.03 (0.16)	0.03 (0.16)	0.003

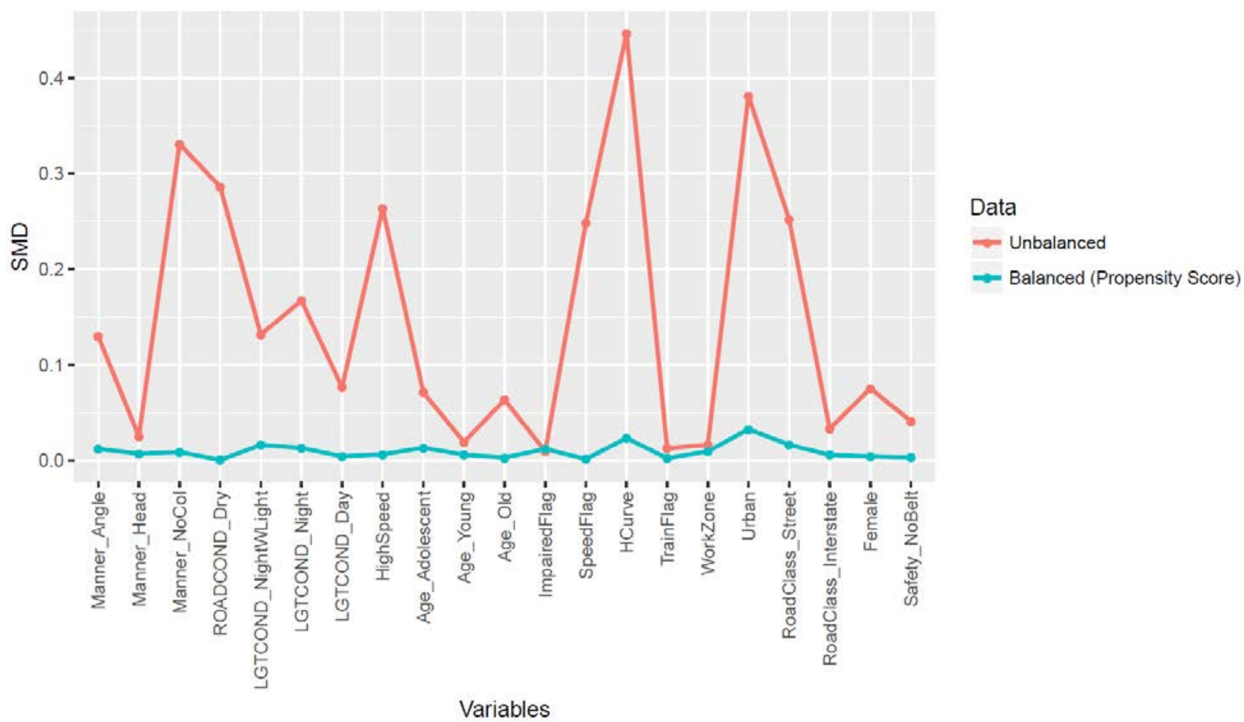
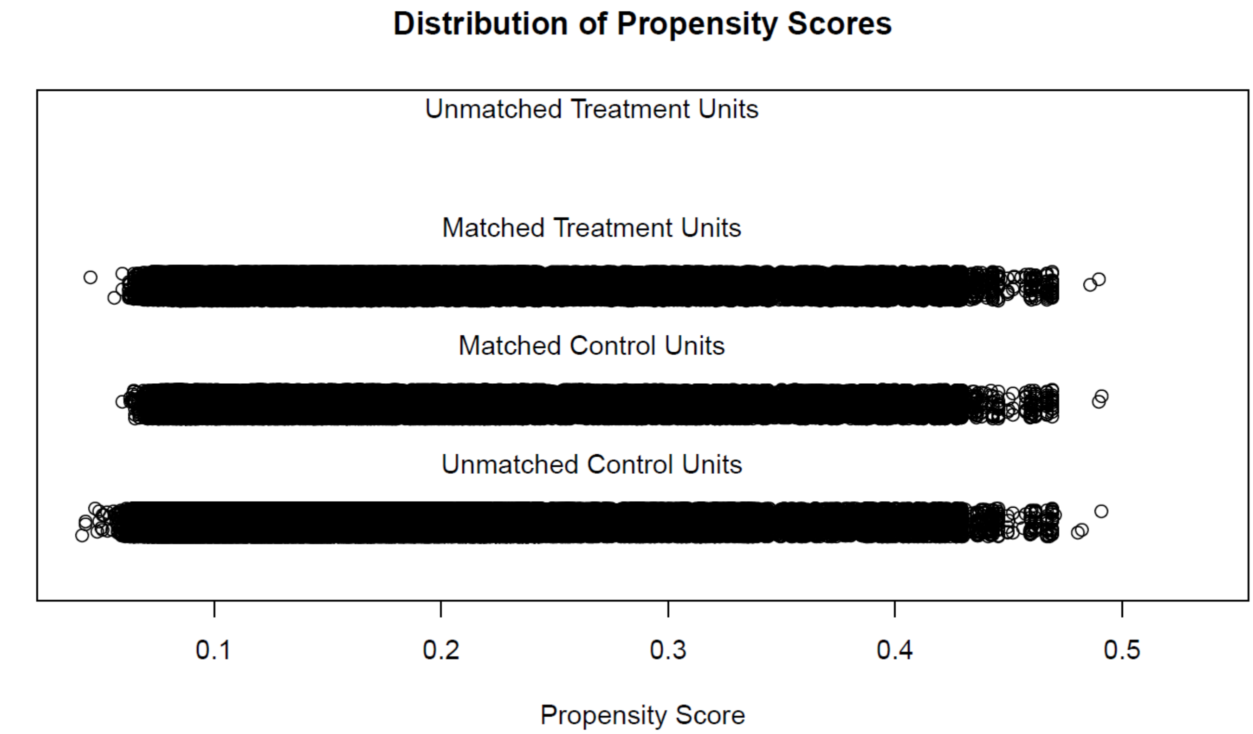


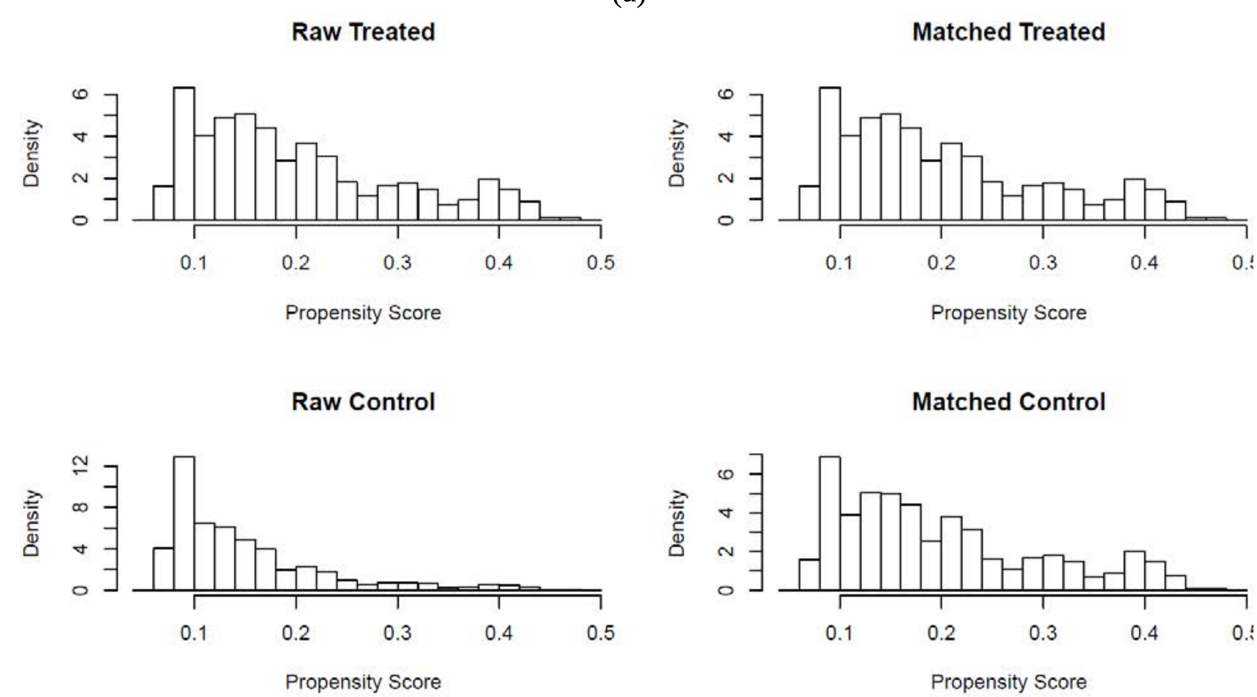
Figure 9-17 SMDs for balanced and unbalanced datasets for vertical hill

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-18. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are

different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-18 Distribution of propensity scores for vertical hill: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced, therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-28.

Table 9-28 Treatment Effect Analysis for Vertical Hill

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	1.85	32482	0.063	0.000 0.005	0.003	0.118	1	0.732

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is not statistically significant at 95% confidence level. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) cannot be rejected.

9.3 Site-specific

9.4 Data

The unit of operation in the site-specific analysis is a homogeneous road segment with uniform geometric characteristics including lane width, shoulder width, and number of lanes along a roadway segment. Roadway segment data including roadway geometry and mobility were obtained from Meta-Manager which is a data management system developed by the Wisconsin Department of Transportation (WisDOT). Meta-Manager stores roadway related data tables for Wisconsin State Trunk Network (STN) and consists the state highway system of Wisconsin, including the Interstate Highway System and the United States Numbered Highway System, and other state trunk highways. Different attributes for each STN roadway segment are stored and updated every year. In this study, two-way highways were analyzed and Meta-Manager data from February 2017 was used.

Segment-related crashes from 2012 to 2016 (5 years) were collected from WisTranPortal data hub (9) and filtered based on the following criteria: at least one car involved in the crash, no pedestrian involvement, no bicycle involvement, no motorcycle involvement, and no missing data points.

To control for event-oriented factors that may affect the severity of the crashes, target crashes were selected. Target crashes satisfy the following criteria: no drug involvement, no alcohol involvement, dry roadway condition, daylight, no work zone area, no train involvement, and no speeding involvement. Processed target crashes were mapped and spatially joined to the study segments using ArcGIS. From 9105 study segments, 600 segments had at least one target crash during the study period. Descriptive summary of the dataset is shown in Table 9-29.

Table 9-29 Descriptive Statistics of the Site-Specific Dataset

Continuous Variables					
Variable	Name	Minimum	Median	Mean	Maximum
Segment Length	SegmentLength	11.24	1602.97	1486.04	9254.03
Average Annual Daily Traffic	AADT	80	3114	3736	18990
Truck Percent	Truck_Percent	0	13.8	12.64	41.7
Number of Access Points	Num_AccessPoints	0	2	2.98	37
Traffic Signal Density	Signal_Density	0	0	0.02	2

Binary Variables			
Variable	Name	Definition	% value 1
Region	Region_NC	1 = If a segment is in the North Central region; 0 = Otherwise	22.55
	Region_NW	1 = If a segment is in the North West region; 0 = Otherwise	28.78
	Region_NE	1 = If a segment is in the North East region; 0 = Otherwise	12.89
	Region_SE	1 = If a segment is in the South East region; 0 = Otherwise	6.44
	Region_SW	1 = If posted speed l segment is in the South West region; 0 = Otherwise	29.34
Speed Limit	Speed_LTE_35	1 = If posted speed limit is less than equal 35 mph; 0 = Otherwise	1.03
	Speed_40	1 = If posted speed limit is 40 mph; 0 = Otherwise	0.3
	Speed_45	1 = If posted speed limit is 45 mph; 0 = Otherwise	5
	Speed_50	1 = If posted speed limit is 50 mph; 0 = Otherwise	0.83
	Base	1 = If posted speed limit is 55 mph or greater; 0 = Otherwise	92.84
Right Shoulder	NoRightShoulder	1 = If a segment does not have a right shoulder; 0 = Otherwise	0.2
No Passing	Poor_PerNoPassing	1 = If percent no passing zone is more than 49%; 0 = Otherwise	44.4
	Good_PerNoPassing	1 = If percent no passing zone is less than 26%; 0 = Otherwise	19.2
	Fair_PerNoPassing	1 = If percent no passing zone is between 26 and 49%; 0 = Otherwise	36.4
Horizontal Curve	Poor_HCURLE40	1 = If speed limit is less than equal 40 mph and the ratio of horizontal curve per mile is more than 0.99; 0 = Otherwise	4.36
	Fair_HCURLE40	1 = If speed limit is less than equal 40 mph and the ratio of horizontal curve per mile is between 0.001 and 0.99; 0 = Otherwise	7.88

	Good_HCURLE40	1 = If speed limit is less than equal 40 mph and the ratio of horizontal curve per mile is zero; 0 = Otherwise	87.58
	Poor_HCURGT40	1 = If speed limit is more than 40 mph and the ratio of horizontal curve per mile is more than 0.99; 0 = Otherwise	3.87
	Fair_HCURGT40	1 = If speed limit is more than 40 mph and the ratio of horizontal curve per mile is between 0.001 and 0.99; 0 = Otherwise	12.36
	Good_HCURGT40	1 = If speed limit is more than 40 mph and the ratio of horizontal curve per mile is zero; 0 = Otherwise	83.59
Observed Outcome	Outcome	1 = If a segment had at least one fatal and/or incapacitating (K and A crashes) crash during the study period; 0 = Otherwise	6.59

Analysis

No Passing Zone

To evaluate the impact of poor no passing zones on the risk of fatal and incapacitating injury crashes, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if segment } i \text{ has poor no passing condition} \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if segment } i \text{ has good or fair no passing condition} \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for segment } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for segment } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-30) estimates the probability that a segment has poor no passing condition. All covariates except AADT and 50 mph speed limit are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-30 Propensity Score Model for No Passing Zone

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.13896	0.13554	8.403	0.000
SegmentLength	-0.00008	0.00004	-2.230	0.026
AADT	-0.00002	0.00001	-1.858	0.063
Truck_Percent	0.01026	0.00488	2.101	0.036
Num_AccessPoints	0.12429	0.01278	9.725	0.000

Region_NC	-0.76038	0.06635	-11.460	0.000
Region_NW	-0.50973	0.06091	-8.368	0.000
Region_NE	-1.17135	0.08272	-14.160	0.000
Region_SE	0.57454	0.10706	5.367	0.000
Speed_LTE_35	0.69254	0.24920	2.779	0.005
Speed_40	1.16754	0.52175	2.238	0.025
Speed_45	0.52418	0.11724	4.471	0.000
Speed_50	0.32967	0.25331	1.301	0.193
NoRightShoulder	2.06343	0.64691	3.190	0.001
Good_HCURLE40	-0.50561	0.09630	-5.251	0.000
Poor_HCURLE40	2.12355	0.26316	8.069	0.000
Good_HCURGT40	-1.10678	0.07678	-14.415	0.000
Poor_HCURGT40	2.00144	0.29506	6.783	0.000

Long segments, and good horizontal curve conditions are associated with lower probabilities of poor no passing conditions. NC, NE, and SE regions have lower chances of poor no passing conditions while SE region has higher odds in comparison to SW region. Segments with speed limits less than 50 mph have higher chances of having poor no passing condition in comparison to the segments with speed limits of more than 50 mph. Greater truck percent, greater number of access points, no right shoulder, and poor horizontal curve conditions are associated with higher probabilities of poor no passing conditions.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model cut out 1121 observations with speed flag. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-31. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-31 and shown in Figure 9-19. In the unmatched dataset, SMDs range between 0.030 and 0.567 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than equal 0.1 and therefore the matched dataset is balanced.

Table 9-31 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for No Passing Zone

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	5062	4043		2922	2922	
SegmentLength	1508.96 (655.46)	1457.34 (693.76)	0.076	1441.53 (657.45)	1444.00 (714.44)	0.004
AADT	3864.49 (2508.06)	3574.31 (2985.10)	0.105	3917.53 (2607.86)	3894.14 (2868.71)	0.009
Truck_Percent	12.97 (4.81)	12.23 (5.16)	0.15	13.01 (5.05)	12.96 (5.05)	0.011
Num_AccessPoints	2.79 (1.99)	3.22 (2.47)	0.192	3.11 (2.32)	3.20 (2.37)	0.039
Region_NC	0.01 (0.11)	0.02 (0.15)	0.059	0.02 (0.13)	0.02 (0.13)	0.016
Region_NW	0.25 (0.43)	0.19 (0.40)	0.136	0.20 (0.40)	0.24 (0.43)	0.10
Region_NE	0.31 (0.46)	0.26 (0.44)	0.104	0.30 (0.46)	0.32 (0.47)	0.036

Region_SE	0.17 (0.38)	0.08 (0.26)	0.294	0.09 (0.29)	0.10 (0.30)	0.041
Speed_LTE_35	0.04 (0.19)	0.10 (0.30)	0.26	0.06 (0.24)	0.07 (0.26)	0.045
Speed_40	0.01 (0.08)	0.02 (0.12)	0.096	0.01 (0.10)	0.02 (0.12)	0.046
Speed_45	0.00 (0.03)	0.01 (0.07)	0.079	0.00 (0.04)	0.00 (0.05)	0.015
Speed_50	0.03 (0.18)	0.07 (0.25)	0.161	0.05 (0.22)	0.06 (0.24)	0.039
NoRightShoulder	0.01 (0.08)	0.01 (0.10)	0.03	0.01 (0.09)	0.01 (0.09)	0.004
Good_HCURLE40	0.00 (0.02)	0.00 (0.06)	0.067	0.00 (0.03)	0.00 (0.06)	0.056
Poor_HCURLE40	0.95 (0.22)	0.78 (0.41)	0.497	0.92 (0.27)	0.92 (0.28)	0.02
Good_HCURGT40	0.00 (0.06)	0.09 (0.29)	0.429	0.01 (0.08)	0.01 (0.11)	0.07
Poor_HCURGT40	0.93 (0.26)	0.72 (0.45)	0.567	0.88 (0.33)	0.86 (0.34)	0.041

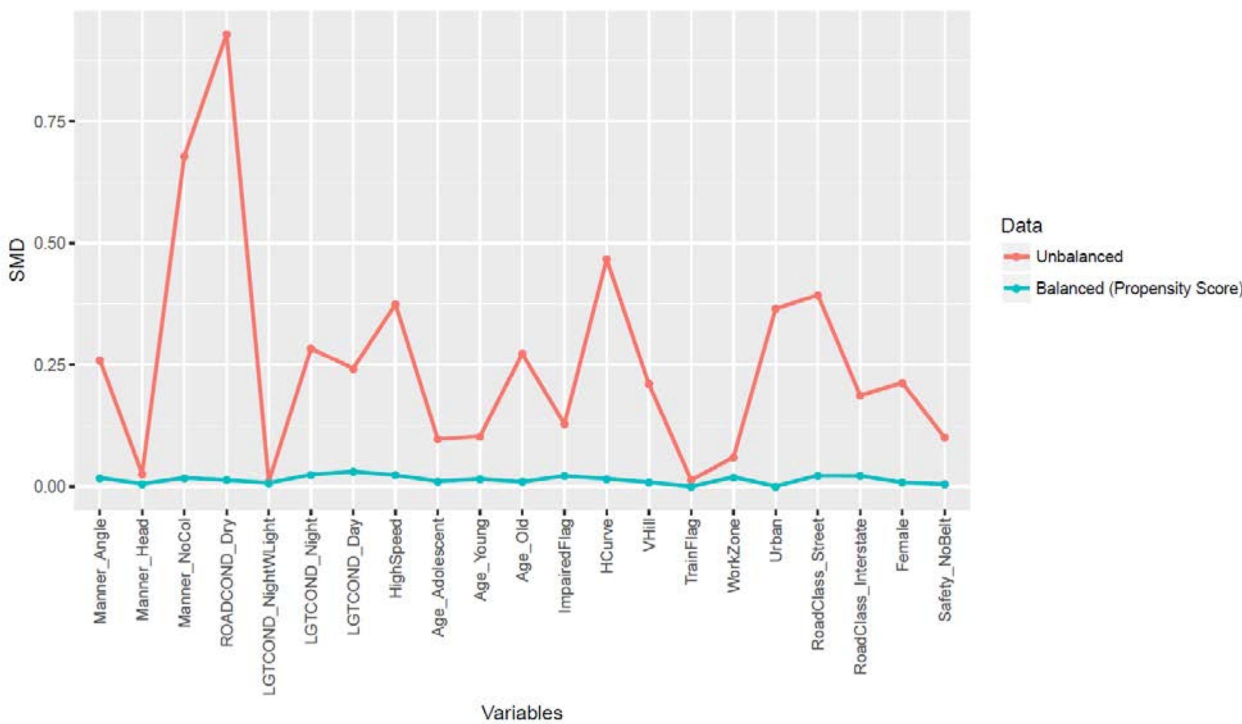
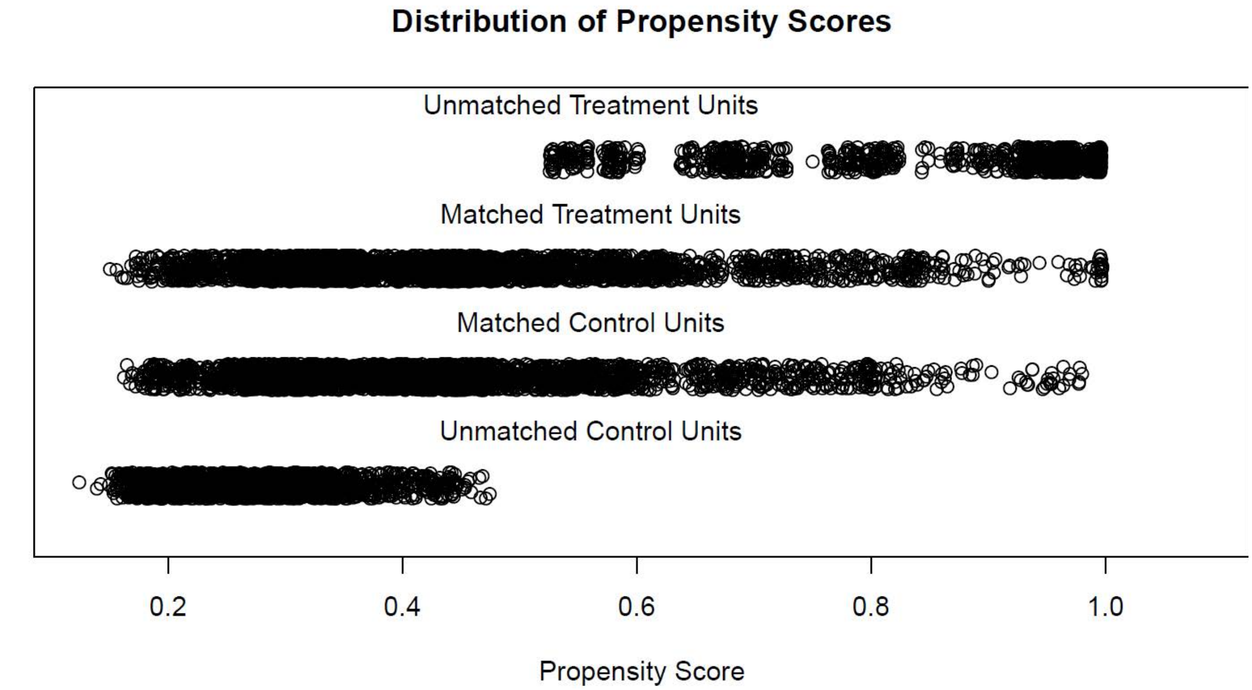
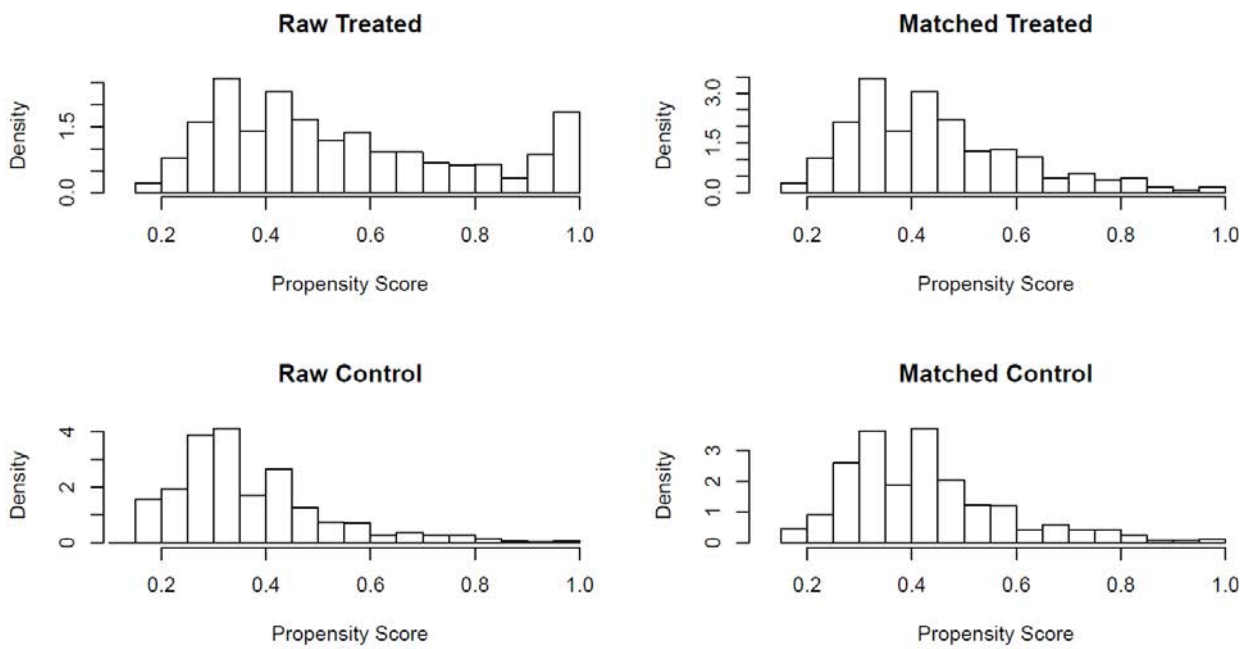


Figure 9-19 SMDs for balanced and unbalanced datasets for no passing zone

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-20. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-20 Distribution of propensity scores for no passing zone: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced, therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-32.

Table 9-32 Treatment Effect Analysis for no passing zone

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	2.46	2921	0.014	0.003 0.029	0.016	5.81	1	0.016

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is 0.016 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, poor no passing zone condition of a segment has a causal effect on the risk of fatal and incapacitating injury crashes. On average, the risk of at least one fatal or incapacitating injury crash occurrence in a segment during the study period increase by 1.6% when the condition of no passing zone is poor.

Divided Segments

To evaluate the impact of divided segments on the risk of fatal and incapacitating injury crashes, the following potential outcome framework is used:

- Treatment:

$$T_i = \begin{cases} 1 & \text{if segment } i \text{ is divided} \\ 0 & \text{otherwise} \end{cases}$$

- Control:

$$C_i = \begin{cases} 1 & \text{if segment } i \text{ is undivided} \\ 0 & \text{otherwise} \end{cases}$$

- Potential outcome:

$$Y_{di} = \begin{cases} Y_{1i} & \text{potential outcome for segment } i \text{ with treatment} \\ Y_{0i} & \text{potential outcome for segment } i \text{ without treatment} \end{cases}$$

Propensity scores were estimated using a binary logistic regression. The developed propensity score model (Table 9-33) estimates the probability that a segment has poor no passing condition. All covariates except good no passing zone condition are statistically significant (at 95% confidence level) in the estimation of the propensity scores.

Table 9-33 Propensity Score Model for Divided Segments

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-10.113	1.096	-9.232	0.000
VMT	-74.016	3.113	-23.779	0.000

Truck_Percent	-0.045	0.010	-4.686	0.000
PostedSpeed	0.031	0.013	2.418	0.016
TravelWay_Width	0.206	0.032	6.547	0.000
NoRightShoulder	4.129	0.785	5.258	0.000
Num_AccessPoints	0.102	0.018	5.696	0.000
Signal_Density	2.220	0.294	7.540	0.000
Good_PerNoPassing	0.198	0.112	1.756	0.079
Good_HCURLE40	2.204	0.259	8.493	0.000
Good_HCURGT40	1.408	0.191	7.361	0.000

Greater vehicle miles traveled (VMTs) and truck percent are associated with lower probabilities of a segment being divided, while greater speed limits, travel way width, number of access points, signal densities as well as no right shoulder, and good horizontal curves are associated with higher probabilities of a segment being divided.

Using the developed propensity model, treated ad control observations were matched. The caliper assigned in the model cut out 44 observations with speed flag. Summary statistics of the unmatched and matched datasets for treatment and control groups are listed in Table 9-34. Calculated SMDs for both unbalanced and balanced (with PSM) dataset are listed in Table 9-34 and shown in Figure 9-21. In the unmatched dataset, SMDs range between 0.054 and 1.135 which show that the treatment and control groups are unbalanced. SMDs for all covariates in the matched dataset are less than equal 0.1 and therefore the matched dataset is balanced.

Table 9-34 Descriptive Statistics of Unmatched and Matched Datasets and SMDs for Divided Segments

Variable	Unmatched (Unbalanced) Data			Matched (Balanced) Data		
	Control	Treatment	SMD	Control	Treatment	SMD
# of Observation	8413	709		665	665	
VMT	0.06 (0.05)	0.02 (0.02)	1.135	0.02 (0.02)	0.02 (0.02)	0.052
Truck_Percent	12.73 (5.05)	11.57 (3.91)	0.256	11.42 (4.67)	11.67 (3.90)	0.06
PostedSpeed	54.28 (3.19)	53.38 (3.85)	0.254	52.98 (5.46)	53.44 (3.81)	0.096
TravelWay_Width	23.76 (1.46)	23.99 (2.94)	0.098	23.93 (2.05)	23.95 (2.96)	0.011
NoRightShoulder	0.00 (0.04)	0.01 (0.09)	0.1	0.01 (0.09)	0.01 (0.09)	0.017
Num_AccessPoints	2.89 (2.17)	4.08 (2.59)	0.5	4.19 (3.87)	4.00 (2.54)	0.061
Signal_Density	0.01 (0.12)	0.06 (0.25)	0.232	0.03 (0.17)	0.03 (0.17)	0.009
Good_PerNoPassing	0.19 (0.39)	0.21 (0.41)	0.054	0.19 (0.39)	0.20 (0.40)	0.034
Good_HCURLE40	0.87 (0.34)	0.97 (0.16)	0.405	0.98 (0.14)	0.97 (0.16)	0.04
Good_HCURGT40	0.83 (0.38)	0.95 (0.22)	0.403	0.96 (0.20)	0.95 (0.22)	0.035

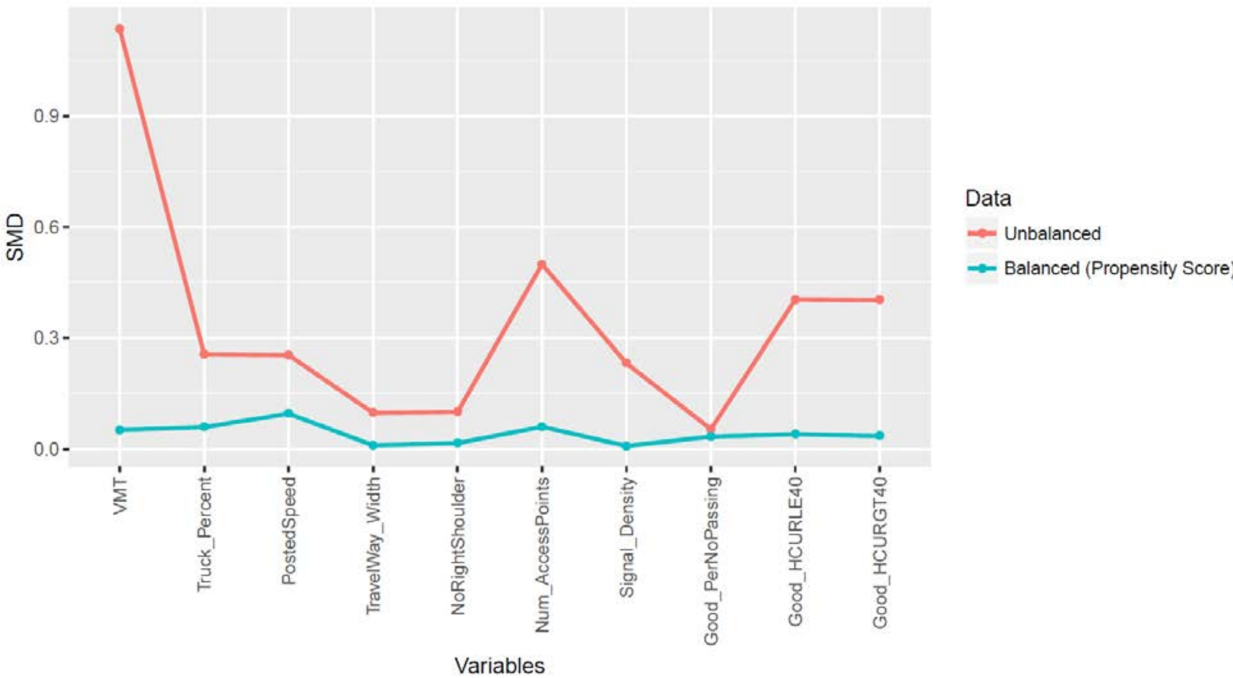
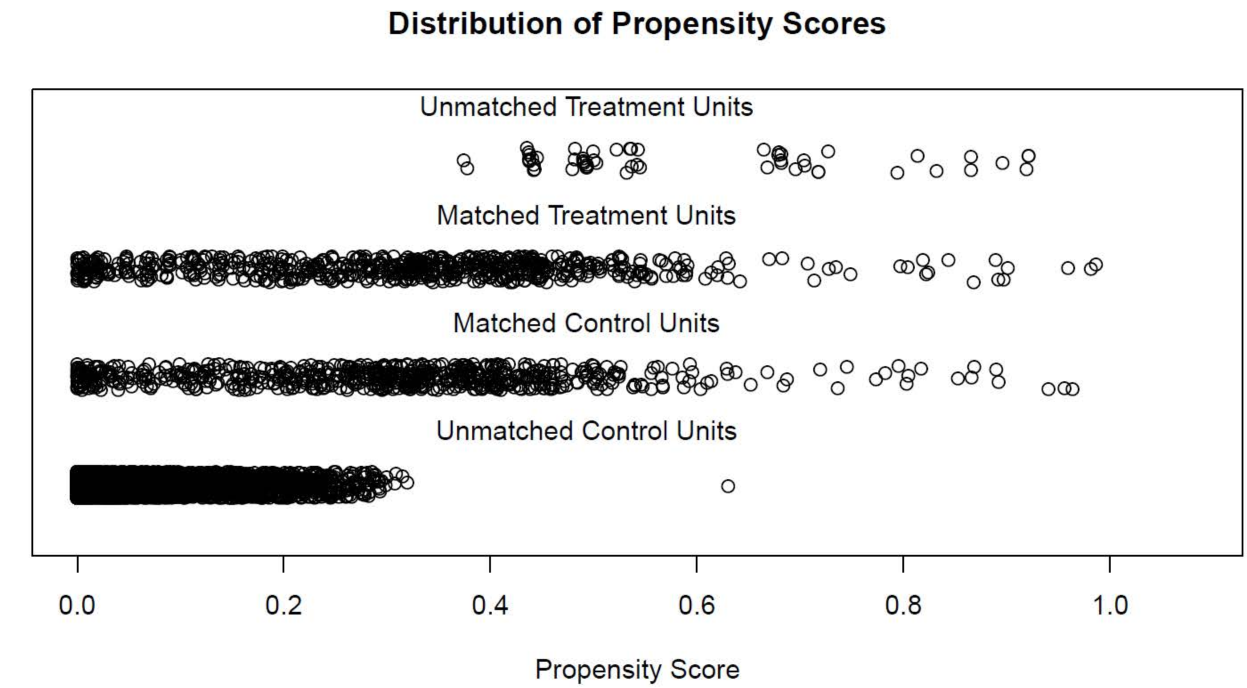
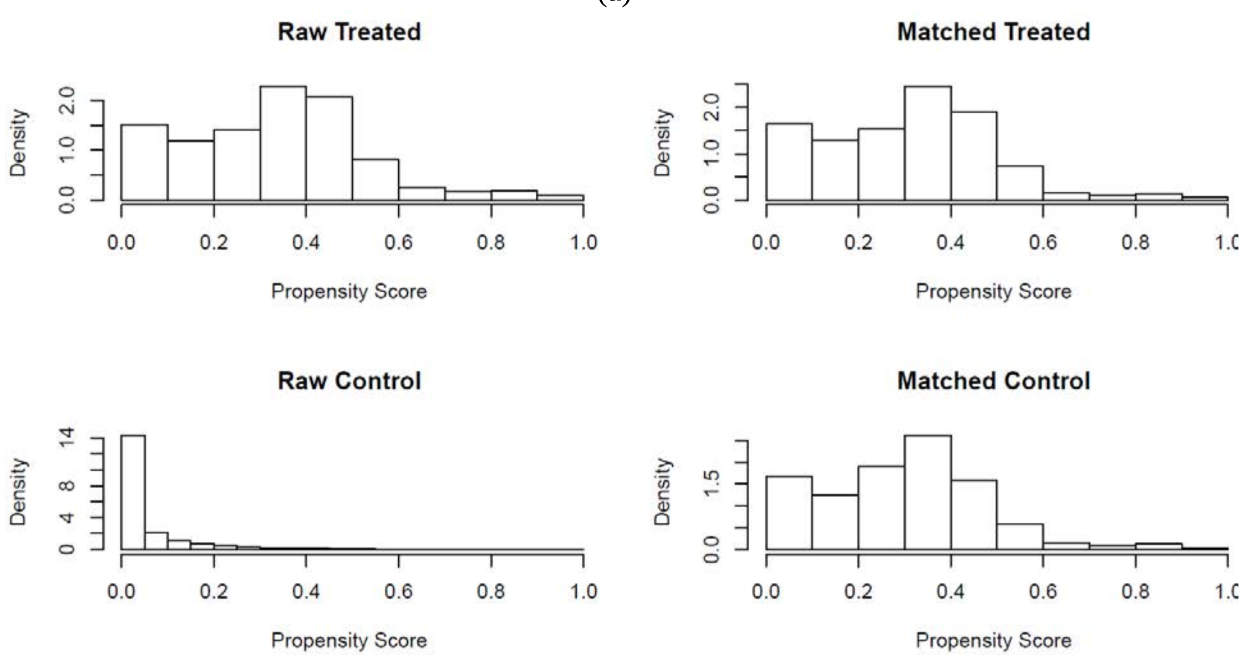


Figure 9-21 SMDs for balanced and unbalanced datasets for divided segments

Jitter plot and histogram plot were used to assess the distribution of propensity scores after matching which are shown in Figure 9-22. Results of jitter plot and histogram plot indicate that the propensity score distribution of treated and control units in unmatched (raw) data are different while the distributions are similar after matching. Therefore, the matching method was successful in improving the covariate balance.



(a)



(b)

Figure 9-22 Distribution of propensity scores for divided segments: a) jitter plot, b) histogram

After obtaining adequate balance and successful matching, outcome analysis was conducted to estimate the treatment effect. All covariates in the matched dataset are balanced, therefore, a mean difference between treated and control units in the matched data is sufficient to estimate the average treatment effect. Results of the paired t-test and McNemar's test are listed in Table 9-35.

Table 9-35 Treatment Effect Analysis for Divided Segments

Matching Method	Paired t-test					McNemar's Test		
	t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	chi-squared	df	p-value
Propensity Score	-3.072	664	0.002	-0.042 -0.009	-0.026	8.26	1	0.004

Paired t-test analysis on the matched dataset indicates that sample mean differences between matched treatment and control units is -0.026 which is statistically significant. McNemar's chi-squared test also shows that the null hypothesis (no treatment effect) can be rejected. Therefore, dividing a segment has a causal effect on the risk of fatal and incapacitating injury crashes. On average, the risk of at least one fatal or incapacitating injury crash occurrence in a segment during the study period reduces by 2.6% when the segment is divided.

9.5 Summary

This paper performed a causal inference analysis in the context of an important transportation safety issue. In the event-oriented analysis, the outcome variable was modeled as a binary variable indicating whether a crash resulted in severe injury (fatal or incapacitating) or not. In the site-specific analysis, the outcome variable was modeled as a binary variable indicating whether a severe injury (fatal or incapacitating) crash occurred in a segment during the study period or not.

Various treatment (exposure) variable including impaired driving, speed flag, seatbelt, work zone, roadway condition, street light, collision manner, horizontal manner, vertical hills, no passing zone, and divided segments were analyzed. For each factor, potential confounders including driver characteristics, driver behavior, site-specific, and environmental conditions were considered. Potential outcome framework was used, and the covariates were balanced using the Propensity Score Matching method.

A five-year period segment-related crash data was analyzed. It was shown that the Propensity Score method was successful in matching and balancing the covariates. The results of the outcome analysis are summarized in Table 9-36.

Table 9-36 Summary of the Outcome Analysis

Factor	Treatment	Control	Outcome	Paired T-Test					% Change of Risk
				t	df	p-value	95 Percent Confidence Interval	Sample Estimates Mean	
Event-Oriented									

Impaired Driving	Impaired Driver Involvement in a Crash	All Other Crashes	K and A Crash Severity	16.54	11981	0	0.0474 0.0601	0.054	5.4
Speed Flag	Speeding Involvement in a Crash	All Other Crashes	K and A Crash Severity	8.88	43808	0	0.008 0.013	0.011	1.1
Seatbelt	No Seatbelt Involvement in a Crash	All Other Crashes	K and A Crash Severity	33.04	4446	0	0.23 0.259	0.245	24.5
Work Zone	Work Zone Involvement in a Crash	All Other Crashes	K and A Crash Severity	-2.09	6393	0.036	-0.0097 -0.0003	-0.005	-0.5
Roadway Condition	Non-Dry Pavement Involvement in a Crash	All Other Crashes	K and A Crash Severity	-15.54	55355	0	-0.017 -0.013	-0.015	-1.5
Street Light	Street Light Involvement in Nighttime Crashes	Nighttime Crashes Without Street Light	K and A Crash Severity	-5.07	9623	0	-0.016 -0.007	-0.011	-1.1
Collision Manner	Head-On Manner Involvement in a Crash	All Other Crashes	K and A Crash Severity	19.83	3497	0	0.125 0.153	0.139	13.9
Horizontal Curve	Horizontal Curve Involvement in a Crash	All Other Crashes	K and A Crash Severity	2.13	32032	0.033	0 0.006	0.003	0.3
Vertical Hill	Vertical Hill Involvement in a Crash	All Other Crashes	K and A Crash Severity	1.85	32482	0.063	0 0.005	0.003	0.3
Site-Specific									
No Passing Zone	Poor No Passing Condition of a Two-Way Highway Segment	All Other Two-Way Highway Segments	At Least One K and A Crashes During Study Period	2.46	2921	0.014	0.003 0.029	0.016	1.6
Divided	Divided Two-Way Highway Segments	All Other Two-Way Highway Segments	At Least One K and A Crashes During Study Period	-3.072	664	0.002	-0.042 -0.009	-0.026	-2.6

Results indicate that no seatbelt involvement has the greatest impact on the fatality and incapacitating injuries. No seatbelt involvement increases the risk of K and A severities by 24.5%. Head-on collision manner and impaired driving increase the risk by 13.9% and 5.4%, respectively. Non-dray pavement conditions, street lights, and work zones reduce the risk of fatal and incapacitating injury crashes by 1.5%, 1.1%, and 0.5%, respectively. Results of the site-specific analysis showed that during the study period, poor no passing zone condition increases the risk of K and A crash occurrence by 1.6% while dividing a segment decreases the risk by 2.6%.

Results of this study can be used by transportation organizations, traffic safety workgroups, and traffic safety commissions to enhance safety programs and support policy decisions. In addition, the results can be used to quantify the safety impacts of countermeasure related to impaired driving.

In this study, intersection-related crashes were not analyzed. Future studies can explore the impact of impaired driving on the severity of intersection-related crashes. In addition, crashes involving pedestrian, bicycle, and motorcycle require further investigation.

10.REFERENCES

[1] Wisconsin Department of Motor Vehicle. *Final Year-End Crash Statistics*. https://wisconsindot.gov/Pages/about-wisdot/newsroom/statistics/final.aspx

[2] NHTSA. National motor vehicle crash causation survey: Report to congress.In *National Highway Traffic Safety Administration Technical Report DOT HS, No. 811*, National Highway Traffic Safety Administration, 2008. p. 059.

[3] WisDOT. Safety Data Resource Guide.In, Wisconsin, 2008.

[4] Zou, Y., L. Wu, and D. Lord. Modeling over-dispersed crash data with a long tail: examining the accuracy of the dispersion parameter in negative binomial models. *Analytic methods in accident research*, Vol. 5, 2015, pp. 1-16.

[5] Shirazi, M., D. Lord, S. S. Dhavala, and S. R. Geedipally. A semiparametric negative binomial generalized linear model for modeling over-dispersed count data with a heavy tail: Characteristics and applications to crash data. *Accident Analysis & Prevention*, Vol. 91, 2016, pp. 10-18.

[6] Geedipally, S. R., D. Lord, and S. S. Dhavala. The negative binomial-Lindley generalized linear model: Characteristics and application using crash data. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 258-265.

[7] Rahman Shaon, M. R., and X. Qin. Use of mixed distribution generalized linear models to quantify safety effects of rural roadway features. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2583, 2016, pp. 134-141.

[8] Anastasopoulos, P. C., and F. L. Mannering. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis & Prevention,* Vol. 41, No. 1, 2009, pp. 153-159.

[9] El-Basyouny, K., and T. Sayed. Accident prediction models with random corridor parameters. *Accident Analysis & Prevention*, Vol. 41, No. 5, 2009, pp. 1118-1123.

[10] Yamamoto, T., J. Hashiji, and V. N. Shankar. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. *Accident Analysis and Prevention,* Vol. 40, No. 4, 2008, pp. 1320-1329.

[11] Greene, W. H. *Econometric Analysis*. Prentice Hall, Upper Saddle River, N.J., 2000.

[12] Shaon, M. R. R., X. Qin, Z. Chen, and J. Zhang. Exploration of Contributing Factors Related to Driver Errors on Highway Segments. *Transportation Research Record*, 2018, p. 0361198118790617.

[13] Wisconsin Department of Motor Vehicle. *Citations and Convictions*. https://wisconsindot.gov/Documents/about-wisdot/newsroom/statistics/factsfig/citation.pdf.

[14] Tracey Wood & Associates. *Prohibited Alcohol Concentration (PAC) in Wisconsin*. http://www.wisconsinowilaws.net/prohibitedalcoholconcentrationpacinwisconsin.html.

[15] Grieve Law. https://www.grievelaw.com/MilwaukeeCriminalDefenseAttorneyBlog/ProhibitedAlcoholConcentrationPAC.

[16] Lee, J. D., K. L. Young, and M. A. Regan. Defining driver distraction. *Driver distraction: Theory, effects, and mitigation,* Vol. 13, No. 4, 2008, pp. 31-40.

[17] NHTSA. *Distracted Driving*. https://www.nhtsa.gov/risky-driving/distracted-driving.

[18] Wikipedia. *Restrictions on cell phone use while driving in the United States*. https://en.wikipedia.org/wiki/Restrictions_on_cell_phone_use_while_driving_in_the_United_States.

[19] Wisconsin Department of Transportation (WisDOT). *Cell Phones, Driving and the Law*. https://wisconsindot.gov/Documents/dmv/shared/cell.pdf.
[20] Barraclough, P., A. af Wåhlberg, J. Freeman, B. Watson, and A. Watson. Predicting crashes using traffic offences. A meta-analysis that examines potential bias between self-report and archival data. *PLoS one,* Vol. 11, No. 4, 2016, p. e0153390.
[21] Shawky, M., Y. Al-Badi, and A. Al-Ghafli. Relationship between Socio-Demographic of Drivers and Traffic Violations and Crashes Involvements.In *2nd World Congress on Civil, Structural, and Environmental Engineering*, Barccelona, Spain, 2017.
[22] Federal Bureau of Investigation (FBI). *The Number of Police Officers for Every Wisconsin City*. https://patch.com/wisconsin/waukesha/number-police-officers-every-wisconsin-city
[23] Washington, S. *Incorporating safety into long-range transportation planning*. Transportation Research Board, 2006.
[24] Washington, S., J. Metarko, I. Fomunung, R. Ross, F. Julian, and E. Moran. An inter-regional comparison: fatal crashes in the southeastern and non-southeastern United States: preliminary findings. *Accident Analysis & Prevention,* Vol. 31, No. 1, 1999, pp. 135-146.
[25] Miaou, S.-P., J. J. Song, and B. K. Mallick. Roadway traffic crash mapping: a space-time modeling approach. *Journal of Transportation and Statistics*, Vol. 6, 2003, pp. 33-58.
[26] Girasek, D. C., and B. Taylor. An exploratory study of the relationship between socioeconomic status and motor vehicle safety features. *Traffic injury prevention,* Vol. 11, No. 2, 2010, pp. 151-155.
[27] Wang, Y., and K. M. Kockelman. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention,* Vol. 60, 2013, pp. 71-84.
[28] Levine, N., K. E. Kim, and L. H. Nitz. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis & Prevention*, Vol. 27, No. 5, 1995, pp. 675-685.
[29] Xu, P., H. Huang, N. Dong, and M. Abdel-Aty. Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. *Accident Analysis & Prevention,* Vol. 70, 2014, pp. 110-120.
[30] NHTSA, N. H. T. S. A. National motor vehicle crash causation survey: Report to congress. *National Highway Traffic Safety Administration Technical Report DOT HS,* Vol. 811, 2008, p. 059.
[31] Administration, N. H. T. S. 2014 motor vehicle crashes: overview. *Traffic safety facts research note,* Vol. 2015, 2015, pp. 1-9.
[32] Wisconsin Department of Transportation (WisDOT). *Preliminary review of 2015 crash fatality trends.* Wisconsin Department of Transportation (WisDOT), Wisconsin. http://wisconsindot.gov/Documents/about-wisdot/newsroom/newsletters/safety/tsr-vol19no2.pdf. Accessed April 21, 2017.
[33] Mitra, S., and S. Washington. On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis & Prevention,* Vol. 39, No. 3, 2007, pp. 459-468.
[34] Lord, D., S. P. Washington, and J. N. Ivan. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention,* Vol. 37, No. 1, 2005, pp. 35-46.
[35] Lord, D., and F. Mannering. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part a-Policy and Practice,* Vol. 44, No. 5, 2010, pp. 291-305.

[36] Mannering, F. L., V. Shankar, and C. R. Bhat. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, Vol. 11, 2016, pp. 1-16.
[37] Mannering, F. L., and C. R. Bhat. Analytic methods in accident research: Methodological frontier and future directions. *Analytic methods in accident research*, Vol. 1, 2014, pp. 1-22.
[38] Vangala, P., D. Lord, and S. R. Geedipally. Exploring the application of the Negative Binomial–Generalized Exponential model for analyzing traffic crash data with excess zeros. *Analytic Methods in Accident Research*, Vol. 7, 2015, pp. 29-36.
[39] Geedipally, S. R., D. Lord, and S. S. Dhavala. A caution about using deviance information criterion while modeling traffic crashes. *Safety science*, Vol. 62, 2014, pp. 495-498.
[40] Parker, S. T., and Y. Tao. WisTransPortal: a Wisconsin traffic operations data hub.In *Applications of Advanced Technology in Transportation*, 2006. pp. 611-616.
[41] WisDOT. *Wiscosnin Crash Data User Guide*. http://transportal.cee.wisc.edu/documents/database/crash/crash-data-user-guide.pdf.
[42] Devlin, A., N. Candappa, B. Corben, and D. Logan. Designing safer roads to accommodate driver error. *Psychopharmacology*, Vol. 10, 2011, pp. 193-212.
[43] Vehicles, D. o. M. *Law Enforcement Officer's Instruction Manual for Completing the Wisconsin Motor Vehicle Accident Report Form (MV 4000).* Wisconsin Department of Transportation, Madison, WI. https://wisconsindot.gov/Documents/dmv/shared/manual-mv4000.pdf.
[44] *Revised Uniform State Traffic Deposit Schedule*, Wisconsin Judicial Conference. http://www.wicourts.gov/publications/fees/docs/bondsched12.pdf.
[45] Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 64, No. 4, 2002, pp. 583-639.
[46] WisDOT. *Wisconsin Crash Data User Guide*. http://trans-portal.cee.wisc.edu/documents/database/crash/crash-data-user-guide.pdf. Accessed March, 2018.
[47] Shaon, M. R. R., X. Qin, M. Shirazi, D. Lord, and S. R. Geedipally. Developing a Random Parameters Negative Binomial-Lindley Model to analyze highly over-dispersed crash count data. *Analytic methods in accident research*, Vol. 18, 2018, pp. 33-44.
[48] Yang, C. D., and W. G. Najm. Examining driver behavior using data gathered from red light photo enforcement cameras. *Journal of safety research*, Vol. 38, No. 3, 2007, pp. 311-321.
[49] Choi, E.-H. Crash factors in intersection-related crashes: An on-scene perspective.In, 2010.
[50] Keay, L., S. Jasti, B. Munoz, K. A. Turano, C. A. Munro, D. D. Duncan, K. Baldwin, K. J. Bandeen-Roche, E. W. Gower, and S. K. West. Urban and rural differences in older drivers' failure to stop at stop signs. *Accident Analysis & Prevention*, Vol. 41, No. 5, 2009, pp. 995-1000.
[51] Preusser, D. F., A. F. Williams, S. A. Ferguson, R. G. Ulmer, and H. B. Weinstein. Fatal crash risk for older drivers at intersections. *Accident Analysis & Prevention*, Vol. 30, No. 2, 1998, pp. 151-159.
[52] Braitman, K. A., B. B. Kirley, S. Ferguson, and N. K. Chaudhary. Factors leading to older drivers' intersection crashes. *Traffic injury prevention*, Vol. 8, No. 3, 2007, pp. 267-274.
[53] Bonneson, J. A., K. Zimmerman, and M. Brewer. *Guidelines to Reduce Red-light-running*. Texas Transportation Institute, Texas A & M University System, 2003.
[54] Qin, X., K. Wang, and C. Cutler. Logistic regression models of the safety of large trucks. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2392, 2013, pp. 1-10.

[55] Staplin, L., K. Lococo, S. Byington, and D. Harkey. Highway design handbook for older drivers and pedestrians.In, 2001.
[56] Campbell, J. L. *Human factors guidelines for road systems*. Transportation Research Board, 2012.
[57] Tay, R. The effectiveness of enforcement and publicity campaigns on serious crashes involving young male drivers: Are drink driving and speeding similar? *Accident Analysis & Prevention,* Vol. 37, No. 5, 2005, pp. 922-929.
[58] Golembiewski, G. A., and B. Chandler. Intersection safety: A manual for local rural road owners.In, 2011.
[59] Wentz, B., D. Warzala, and K. Harder. The effect of rumble strips on drivers approaching rural, stop-controlled intersections. *Minnesota: Minnesota Department of Transportation,* Vol. 8, 2006, p. 9.
[60] Cummings, P., T. D. Koepsell, J. M. Moffat, and F. P. Rivara. Drowsiness, counter-measures to drowsiness, and the risk of a motor vehicle crash. *Injury Prevention,* Vol. 7, No. 3, 2001, pp. 194-199.
[61] Wisconsin Department of Transportation (WisDOT). *Wisconsin Strategic Highway Safety Plan, 2017-2020.* Available Online, http://wisconsindot.gov/Documents/safety/education/frms-pubs/strategichwy-17-20.pdf, 2017.
[62] Zegeer, C. V., D. Nabors, and P. Lagerwey. *PEDSAFE: Pedestrian Safety Guide and Countermeasure Selection System.* Federal Highway Administration. Available online: http://www.pedbikesafe.org/PEDSAFE/index.cfm, August 2013.
[63] Sundstrom, C., D. Nabors, M. Hintze, B. Schultheiss, P. Lagerwey, and K. Langford. *BIKESAFE: Bicycle Safety Guide and Countermeasure Selection System.* Federal Highway Administration. Available online:http://www.pedbikesafe.org/bikesafe/authors.cfm, September 2014.
[64] Federal Highway Administration. *Proven Safety Countermeasures.* Office of Safety. Available online, https://safety.fhwa.dot.gov/provencountermeasures/.
[65] Harwood, D. W., K. M. Bauer, K. R. Richard, D. K. Gilmore, J. L. Graham, I. B. Potts, D. J. Torbic, and E. Hauer. Pedestrian safety prediction methodology.In, Final Report for National Highway Cooperative Research Program Project 17-28, Available online: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w129p3.pdf, March 2008.
[66] Schneider, R., M. Diogenes, L. Arnold, V. Attaset, J. Griswold, and D. Ragland. Association between roadway intersection characteristics and pedestrian crash risk in Alameda County, California. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2198, 2010, pp. 41-51.
[67] Turner, S., G. Wood, T. Hughes, and R. Singh. Safety performance functions for bicycle crashes in New Zealand and Australia. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2236, 2011, pp. 66-73.
[68] Nordback, K., W. E. Marshall, and B. N. Janson. Bicyclist safety performance functions for a US city. *Accident Analysis & Prevention,* Vol. 65, 2014, pp. 114-122.
[69] Alluri, P., M. A. Raihan, D. Saha, W. Wu, A. Huq, S. Nafis, and A. Gan. Statewide analysis of bicycle crashes. *Prepared for Florida Department of Transportation*, No. Available online, http://www.fdot.gov/research/Completed_Proj/Summary_SF/FDOT-BDV29-977-23-rpt.pdf, 2017.
[70] Zegeer, C., C. Lyon, R. Srinivasan, B. Persaud, B. Lan, S. Smith, D. Carter, N. J. Thirsk, J. Zegeer, and E. Ferguson. Development of crash modification factors for uncontrolled pedestrian

crossing treatments. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2636, 2017, pp. 1-8.
[71] Clifton, K. J., and K. Kreamer-Fults. An examination of the environmental attributes associated with pedestrian–vehicular crashes near public schools. *Accident Analysis & Prevention*, Vol. 39, No. 4, 2007, pp. 708-715.
[72] Dumbaugh, E., and W. Li. Designing for the safety of pedestrians, cyclists, and motorists in urban environments. *Journal of the American Planning Association*, Vol. 77, No. 1, 2010, pp. 69-88.
[73] El-Basyouny, K., and T. Sayed. Safety performance functions using traffic conflicts. *Safety science*, Vol. 51, No. 1, 2013, pp. 160-164.
[74] Environmental Protection Agency. Smart location database. *Available online, https://www.epa.gov/smartgrowth/smart-location-mapping#SLD*, 2014.
[75] Liggett, R., H. Huff, R. Taylor-Gratzer, N. Wong, D. Benitez, T. Douglas, J. Howe, J. Cooper, J. Griswold, D. Amos, and F. Proulx. Bicycle Crash Risk: How Does It Vary, and Why?In, Prepared for California Department of Transportation, Available online, https://www.its.ucla.edu/wp-content/uploads/sites/6/2016/08/Final-Report-to-Caltrans-Bicycle-Crash-v3.pdf, 2016.

76 Pearl, J., Causal Inference in Statistics: An Overview, Statistics Surveys, Vol. 3, 2009, pp. 96-146
77 Holland, P.: Statistics and Causal Inference, Journal of the American Statistical Association, Vol. 81, 1986, pp. 945–970
78 King, G., R. Nielsen, C. Coberley, J. E. Pope, A. Wells, Comparative Effectiveness of Matching Methods for Causal Inference, 2011
79 Stuart, E. A., Rubin, D. B., Best Practices in Quasi-Experimental Designs: Matching Methods for Causal Inference, *In Best Practices in Quantitative Methods*, SAGE Publications, London, 2008, pp. 155–176
80 Austin, P., Optimal Caliper Widths for Propensity-Score Matching When Estimating Differences in Means and Differences in Proportions in Observational Studies, *Pharmaceutical Statistics*, Vol. 10, 2011, pp. 150–161
81 Austin, P., Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research, *Communications in Statistics—Simulation and Computation*, Vol. 8, 2009, pp.1228–1234
82 Austin, P., An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies, *Multivariate Behavioral Research*, Vol. 46, 2011, pp. 399–424
83 Lin, M., H. C. Lucas, G. Shmueli, Too Big to Fail: Large Samples and the p-Value Problem, *Information Systems Research*, Vol. 24, No. 4, 2013, pp. 906-917
84 The WisTransPortal System. Wisconsin Traffic Operations and Safety Laboratory Website. (URL: https://transportal.cee.wisc.edu/)

11.APPENDIX A

Table 11-1 Distribution of Violations by Charge Type

Charge Code	Charge Description	2016 Count	2016 % Male	2016 Conviction Rate	2017 Count	2017 % Male	2017 Conviction Rate
SI	Speeding intermediate (11-19 over)	94934	57.42%	99.67%	86722	57.31%	99.65%
OWS	Operating while suspended	77878	63.15%	99.22%	78456	62.19%	99.12%
FFS	Failure to fasten seat belt	70498	70.19%	99.76%	58632	70.47%	99.77%
CNI	Compulsory insurance - no insurance	63878	61.39%	98.59%	62307	61.21%	98.68%
CNP	Compulsory insurance - no proof	55291	59.95%	99.33%	50838	60.50%	99.31%
S	Speeding (1 - 10 over)	48878	56.94%	99.84%	43500	56.78%	99.84%
UV	Unregistered vehicle	35131	63.38%	98.35%	37323	62.99%	98.18%
OWL	Operating without driver license	30079	67.54%	99.24%	28598	67.32%	99.14%
OWI	Operating while intoxicated	29673	73.74%	83.30%	28906	73.55%	82.60%
SE	Speeding excess (20 or more over)	25966	65.04%	99.31%	26254	64.53%	99.20%
FOS	Failure to obey traffic sign or signal	22926	62.11%	99.21%	23004	62.62%	99.29%
PAC	Prohibited alcohol concentration	19090	74.15%	18.87%	18868	74.02%	18.07%
ORS	Operating while registration suspended	18851	58.63%	98.48%	19138	57.72%	98.73%
DS	Defective speedometer	15714	58.12%	99.94%	14613	58.46%	99.90%
OAR	Operating after revocation	14727	75.98%	99.46%	15184	74.72%	99.33%
LNP	License not on person	13618	65.02%	99.53%	13034	65.09%	99.46%
FPF	Failure to pay forfeiture	11311	66.91%	99.50%	10389	67.90%	99.36%
ID	Inattentive driving	10122	58.18%	99.30%	8897	58.94%	99.35%
FYR	Failure to yield right of way	10060	52.04%	99.23%	9738	51.61%	99.26%
UAL	Underage alcohol	8958	65.90%	98.04%	8279	65.56%	96.76%
FPJ	Failure to pay forfeiture - juvenile	8538	64.29%	99.77%	7133	62.72%	99.72%
OT	Obstructing traffic	8311	56.61%	99.87%	7184	57.53%	99.92%
SVL	Signal violation	8038	58.68%	99.93%	8385	58.52%	99.92%
FTC	Following too closely	6954	54.50%	99.28%	6637	56.83%	99.41%
IP	Improper plates	6599	70.00%	97.98%	6496	70.75%	97.98%
IL	No or improper lights	5114	64.47%	96.11%	4811	63.00%	95.86%
FVC	Failure to keep vehicle under control	4568	67.16%	97.46%	4970	66.92%	97.97%
IS	Imprudent speed	4559	74.82%	98.75%	4631	75.53%	98.79%
DLT	Deviating from lane of traffic	4084	60.01%	97.85%	4164	61.48%	97.93%
IC	Implied consent	3072	77.99%	90.01%	3108	77.54%	92.95%

CSR	Child safety restraint	3020	38.18%	99.27%	2680	38.47%	99.40%
IVO	Intoxicant in vehicle - operator	2766	74.80%	96.35%	2787	75.31%	96.66%
IT	Illegal turn	2651	59.22%	99.40%	2770	59.57%	99.21%
PI	Passing illegally	2589	67.36%	98.65%	2689	65.75%	98.96%
FRA	Failure to report accident	2449	72.27%	97.43%	2491	72.54%	97.23%
FPS	Failure to pay support	2244	84.54%	100.00%	2304	85.76%	99.96%
TFC	Too fast for conditions	2103	64.81%	98.72%	2311	63.44%	98.36%
RPS	Restrictions on parking and stopping	2040	68.04%	99.80%	2947	67.53%	99.97%
DWS	Driving on wrong side of highway	1935	69.66%	97.26%	1854	69.69%	97.09%
DOF	Interfere w/ traffic sign/signal	1928	59.54%	100.00%	1927	58.43%	99.90%
OV	Obstructed view or control	1901	70.54%	97.84%	1958	73.44%	97.50%
GPV	GDL passenger violation	1845	59.46%	98.97%	1632	62.56%	99.33%
RD	Reckless driving	1733	78.48%	97.23%	1760	79.15%	96.42%
VOR	Violation of restriction	1629	68.82%	98.53%	1749	67.52%	97.94%
IM	Improper muffler	1483	84.56%	98.25%	1321	82.89%	97.73%
T	Truancy	1432	57.89%	98.88%	1522	59.86%	98.69%
BI	Backing illegally	1427	59.50%	99.58%	1342	57.00%	99.33%
DSP	Duty upon striking property	1414	74.26%	98.23%	1394	74.68%	96.92%
FSU	Failure to stop after accident - unattended vehicle	1264	62.26%	98.18%	1272	63.76%	98.27%
JA	Juvenile alcohol	1199	56.38%	98.42%	1184	57.35%	97.47%
FSB	Failure to stop for school bus	1028	50.19%	98.35%	1015	47.98%	98.52%
AEO	Attempt to elude officer	1026	79.34%	99.03%	1183	75.32%	99.15%
POH	Parking on highway	800	61.50%	99.75%	742	68.60%	99.73%
PUP	Permit unauthorized person to operate	753	30.54%	99.34%	617	33.71%	98.38%
VUF	Vehicle used in commission of felony	703	87.06%	99.72%	683	87.99%	99.56%
IVP	Intoxicant in vehicle - passenger	652	71.47%	98.77%	596	68.79%	98.99%
FSA	Failure to stop after accident	639	76.06%	99.84%	643	73.87%	99.22%
FNC	Failure to notify of address or name change	630	60.00%	97.30%	597	61.47%	97.82%
IE	Improper equipment	585	73.68%	98.63%	535	75.51%	98.69%
DOW	Driving over walk	578	68.51%	99.13%	495	63.23%	99.60%
JNK	Non-trackable	561	68.81%	98.75%	482	66.80%	98.34%
D	Drugs	548	75.73%	99.09%	477	73.38%	98.74%
UAO	Underage alcohol operation	509	70.92%	78.00%	438	72.37%	81.05%
OII	Operating while intoxicated causing injury	485	80.82%	40.21%	668	73.80%	39.82%
TWD	Texting while driving	460	53.26%	98.91%	350	49.71%	99.14%
IIV	Intoxicants in vehicle carrying underage person	309	77.35%	97.09%	300	71.00%	98.00%

FGS	Failure to give signal	301	72.76%	97.34%	273	68.86%	97.44%
FTT	Failure to transfer title	299	77.93%	96.99%	349	72.78%	97.71%
IDT	Ignition/immobilization device tampering	299	82.61%	99.33%	268	77.24%	99.25%
IB	Improper brakes	216	62.04%	99.07%	205	70.24%	99.02%
UID	Underage id	210	67.62%	96.19%	200	71.50%	95.50%
SLL	Special limitations on load	189	93.65%	99.47%	210	92.86%	98.57%
UA	Unnecessary acceleration	187	91.98%	97.86%	154	91.56%	95.45%
FDL	Failure to dim lights	176	63.07%	97.73%	186	68.28%	98.39%
CSI	Commercial speeding intermediate (15-19 over)	151	98.68%	99.34%	169	94.08%	99.41%
CUL	Commercial unlawful operation	140	98.57%	98.57%	148	99.32%	98.65%
GCV	Gdl curfew violation	138	68.12%	97.10%	97	69.07%	97.94%
JCS	Juvenile controlled substance	138	66.67%	98.55%	108	72.22%	97.22%
DAT	Driving against traffic	136	69.85%	100.00%	139	64.75%	99.28%
CDL	Commercial deviating from lane	132	91.67%	100.00%	175	96.00%	97.71%
CFC	Commercial following too closely	128	95.31%	99.22%	132	95.45%	97.73%
R	Racing	115	93.91%	97.39%	134	94.78%	95.52%
UN	Unnecessary noise	109	83.49%	100.00%	56	75.00%	98.21%
GBH	Great bodily harm	104	84.62%	46.15%	121	78.51%	42.15%
TPV	Transport person or vehicle illegally	71	76.06%	100.00%	53	64.15%	100.00%
FYL	Flashing yellow signal violation	60	68.33%	98.33%	57	52.63%	96.49%
LH	Littering highway	53	92.45%	100.00%	53	94.34%	100.00%
DDH	Driving on divided highway improperly	52	78.85%	100.00%	41	70.73%	97.56%
NHI	Negligent homicide intoxicated	50	68.00%	50.00%	90	80.00%	56.67%
MDO	Gdl miscellaneous driving offense	47	70.21%	95.74%	55	60.00%	96.36%
NH	Negligent homicide	42	78.57%	100.00%	49	81.63%	100.00%
FA	Falsified application	39	66.67%	100.00%	51	66.67%	100.00%
CWI	Commercial operating while intoxicated	37	94.59%	78.38%	48	100%	81.25%
CPI	Commercial passing illegally	36	100.00%	100.00%	40	97.50%	100.00%
UTD	Using telephone while driving w/prob IP	28	42.86%	100.00%	17	58.82%	100.00%
CTF	Commercial too fast for conditions	21	100.00%	90.48%	30	96.67%	96.67%
CFR	Commercial failure to report accident	20	95.00%	95.00%	15	93.33%	93.33%
CDS	Commercial duty upon striking property	18	100.00%	77.78%	19	94.74%	63.16%
COO	Commercial absolute sobriety	17	94.12%	29.41%	10	100%	50.00%

MSC	Miscellaneous	16	43.75%	100.00%	8	50.00%	100.00%
CSE	Commercial speeding excess (20 or more over)	13	100.00%	100.00%	25	96.00%	96.00%
OSO	Operating while out of service	13	100.00%	84.62%	8	100%	87.50%
FAR	Falsified accident report	12	50.00%	100.00%	14	42.86%	92.86%
CPB	Commercial possession of intoxicating beverage	10	100.00%	80.00%	12	100%	50.00%
CIS	Commercial imprudent speed	9	77.78%	100.00%	8	87.50%	100.00%
OML	Operating with multiple licenses	9	88.89%	100.00%	0	0.00%	0.00%
CFH	Crossing fire hose	8	87.50%	87.50%	5	60.00%	100.00%
PLS	Projecting loads on side of vehicle	7	85.71%	100.00%	7	71.43%	100.00%
TCC	Transporting children in cargo areas of motor vehicle	7	85.71%	100.00%	9	66.67%	100.00%
IUL	Illegal use of operator's license	6	100.00%	100.00%	5	80.00%	100.00%
CFI	Compulsory insurance-fraud	5	80.00%	100.00%	5	100%	100.00%
CRD	Commercial reckless driving	5	100.00%	80.00%	6	100%	100.00%
RVL	Roadway violation	5	100.00%	100.00%	8	62.50%	100.00%
OWD	Operating while disqualified	5	100.00%	80.00%		0.00%	0.00%
FEM	Following emergency vehicle	4	100.00%	100.00%	11	72.73%	100.00%
ICU	Implied consent underage	4	50.00%	100.00%	2	100%	100.00%
IPW	Using cell phone while driving in work zone	4	50.00%	100.00%	260	50.38%	99.23%
RRF	Railroad failure to stop	4	100.00%	100.00%	1	100%	100.00%
CA	Commercial alcohol	3	100.00%	100.00%	2	100%	50.00%
CFU	Commercial failure to stop after accident-unattended	3	100.00%	66.67%	3	100%	33.33%
CIC	Commercial implied consent	3	100.00%	100.00%	2	100%	50.00%
CTU	Commercial telephone use while driving	3	100.00%	100.00%	52	100%	100.00%
FAV	Fraudulent application	2	100.00%	100.00%	2	100%	100.00%
HDS	Haz commercial duty upon striking property	1	100.00%	0.00%	1	100%	0.00%
HWI	Haz commercial operating while intoxicated	1	100.00%	100.00%	2	100%	0.00%
JID	Juvenile ID	1	100.00%	100.00%	2	50.00%	100.00%
SLR	Surrender of licenses and registration upon rev or sus	1	0.00%	100.00%	4	50.00%	75.00%
CII	Commercial OWI causing injury	1	100.00%	0.00%	0	0.00%	0.00%
ADL	Altering driver license	0	0.00%	0.00%	2	0.00%	100.00%
CFS	Commercial failure to stop after accident	0	0.00%	0.00%	1	100%	100.00%
FD	Found delinquent	0	0.00%	0.00%	2	50.00%	100.00%
HCA	Haz commercial alcohol	0	0.00%	0.00%	1	100%	100.00%
IR	Illegal riding	0	0.00%	0.00%	1	0.00%	100.00%

OCS	OWI - controlled substance	0	0.00%	0.00%	1	0.00%	0.00%
SOL	Surrender of licenses upon cancel	0	0.00%	0.00%	2	100%	0.00%
Grand Total		790662	63.45%	96.57%	754140	63.38%	96.37%

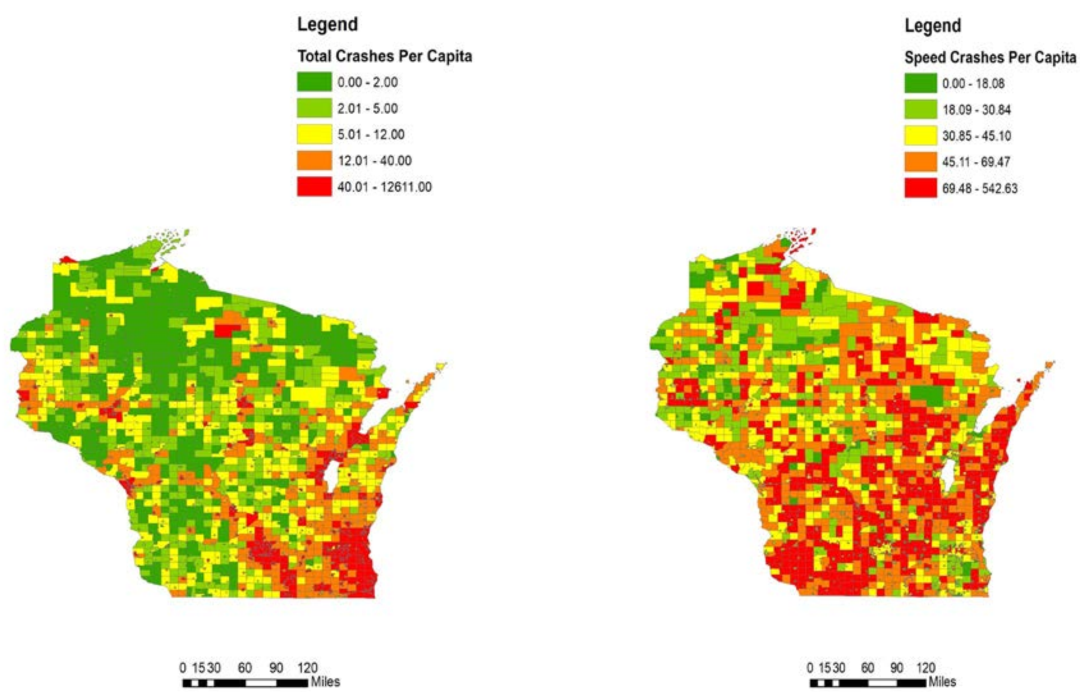
12.APPENDIX B

The exploratory analysis of traffic violation rate and crash rate indicates that there is no significant correlation between these variables. The correlation matrix shows that the population count in each municipality has a higher correlation with crashes compared with different types of traffic violations. Thus, the population at each municipality was used as an explanatory variable to predict total number of crashes using linear regression model. The model formulation is provided below:

$$Crash\ Count_i = \beta_0 + \beta_1 \times Population_i \quad \text{(B-1)}$$

Where,
$Crash\ Count_i$ is the count of crashes at i-th municipality, $Population_i$ is the population estimate of i-th municipality and, $\beta_0\ and\ \beta_1$ are estimated model coefficients.

A spatial distribution of total crashes, speed-related crashes, impairment-related crashes and inattentive driving-related crashes along with population distribution between municipalities is provided in Figure 12-1. The parameter estimates of crash-population regression models are presented in Table 12-1. To further explore the relationship between crash counts and population, a regression model was developed for each quartile of population. The scatterplot with the fitted regression line is provided in Figure 12-2.



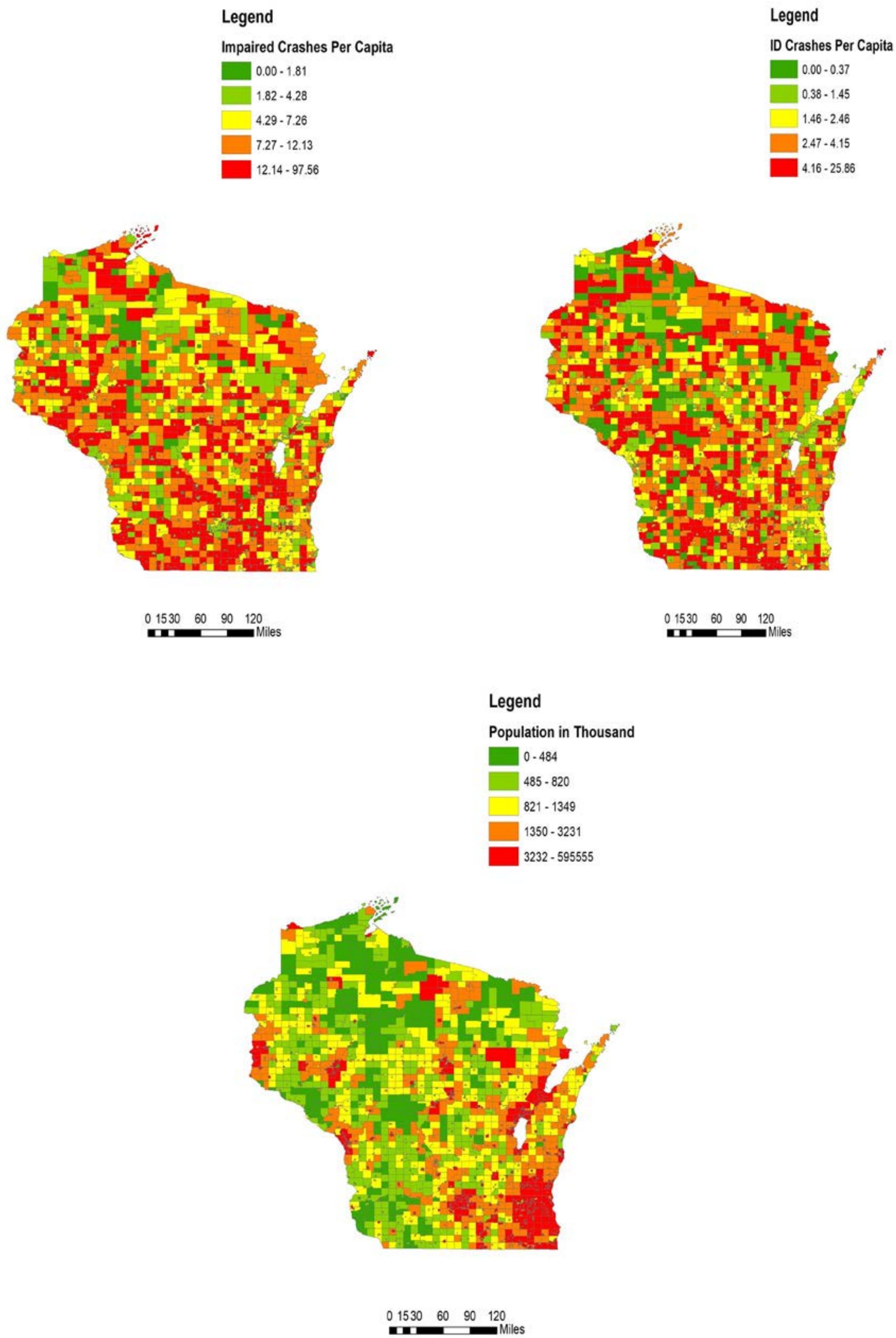
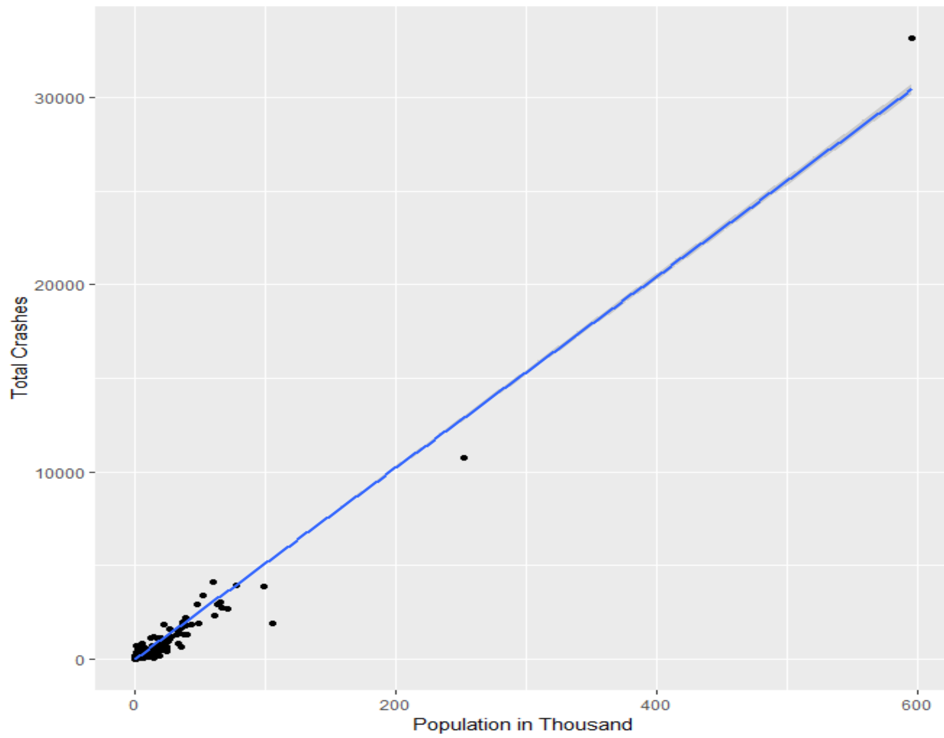


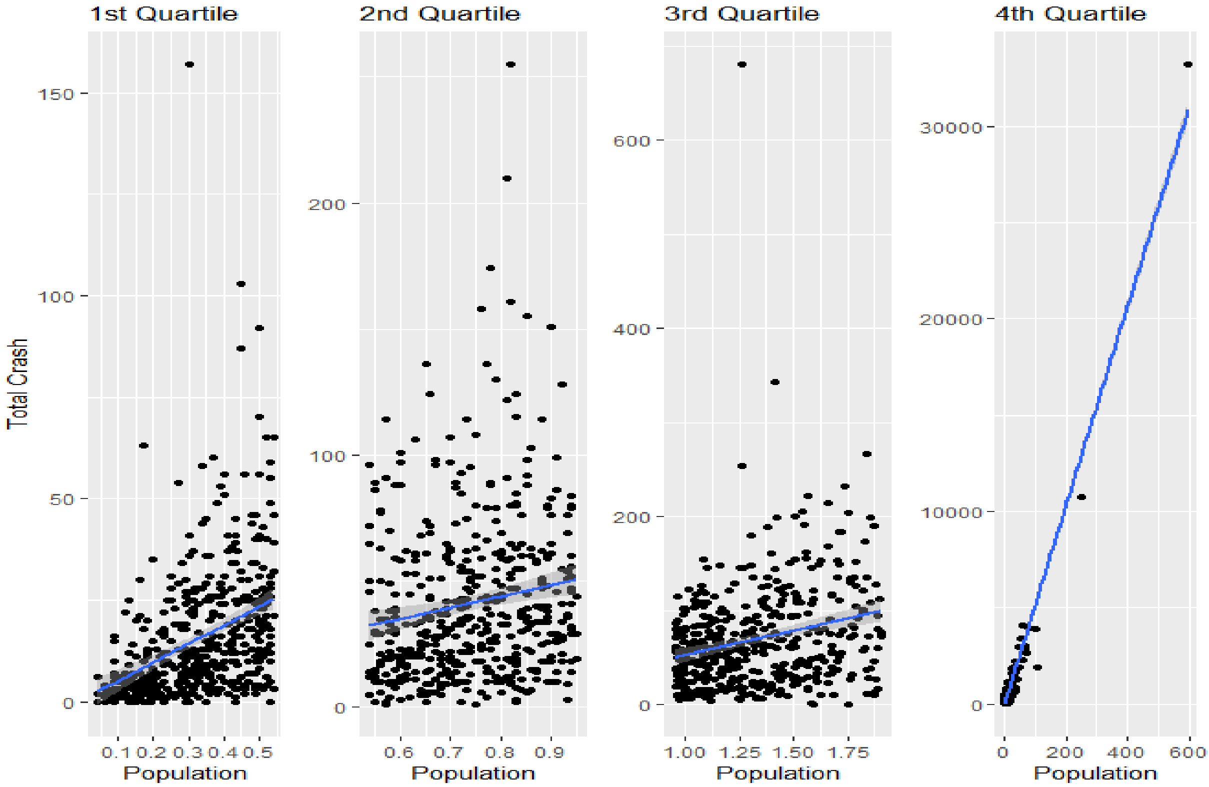
Figure 12-1 Spatial Distribution of Crashes and Population.

Table 12-1 Model Description and Parameter Estimates for Modeling Total Crashes.

Models	Min. Population	Mean Population	Max. Population	β_0	β_1	R-Square
Full Model	39	3085	595555	-25.33	51.22	0.963
1st Quantile	39	330	538	0.5973	45.55	0.145
2nd Quantile	539	739.2	947	8.17	44.66	0.025
3rd Quantile	948	1316	1907	1.60	51.43	0.061
4th Quantile	1913	9980	595555	-112.78	51.95	0.967



(a)



(b)

Figure 12-2 Scatterplot for (a) Full Model, (b) Models with Population Quartiles.

As noted in Table 12-1, the full model that includes all municipalities provides a very good model fit with an R-square value of 0.963. However, for models with population quantiles, model performance was very low for 1st, 2nd, and 3rd quantile. The model performance is again very high with 4th quantile indicating population estimate and crash counts are highly correlated for higher population municipalities.

Risky driving behavior has been identified as a major contributor of crash occurrence. The behavior related crashes such as speed-related, alcohol-impaired and inattentive driving-related crashes were also modeled using population estimate in each municipality. Similar to total crash count modeling, the behavior related crashes were modeled for complete dataset and quartiles of population. The parameter estimates of crash count by behavioral type and population regression models are presented in Table 12-2. The parameter estimates and model performance indicate that the behavior-based regression models also follows a similar pattern as total crash modeling.

Table 12-2 Model Description and Parameter Estimates for Modeling Crash Counts by Behavior Type.

Model Description	Speed Crash			Impaired Crash			Distracted Crash		
	β_0	β_1	R-Square	β_0	β_1	R-Square	β_0	β_1	R-Square
Full Model	-0.89	6.55	0.921	0.69	1.80	0.966	-1.4	6.62	0.940
1st Quantile	0.34	7.08	0.107	-0.01	3.11	0.123	0.19	4.36	0.044
2nd Quantile	1.48	7.76	0.016	-0.46	3.60	0.040	-0.43	5.67	0.031
3rd Quantile	-0.12	9.00	0.042	1.00	2.11	0.030	-0.88	6.17	0.053
4th Quantile	-10.35	6.63	0.923	-0.004	1.80	0.969	-2.53	6.63	0.938

It can be concluded from the regression modeling of crash counts and population are:

- More than 96% variability in total crash counts for municipality can be explained by population.
- The regression model parameter estimates with different behavior related crashes also follow similar trend as total crashes.
- 93% to 97% variability in different behavior related crashes can be explained by population estimates in each municipality.
- Similar to total crashes, the model performance was very low in 1st, 2nd and 3rd quantiles of population for all behavior related crashes.

13.APPENDIX C

Pedestrian and Bike Crash: Corridor Data Collection Dictionary

1. **Total number of signalized intersections:** Record the total number of signalized intersections in the selected 1-mile corridor. An intersection can be identified as signalized intersections if mounted traffic signal is provided at the intersection to control traffic flow.
2. **Total number of unsignalized intersections:** Record the total number of unsignalized intersections in the selected 1-mile corridor. An intersection can be identified as unsignalized intersections if no mounted traffic signal is provided and/or the intersection is controlled by traffic signs.
3. **Total number of marked crosswalks:** Record the total number of marked crossing within 1-mile corridor of roadway. Please include all marked crossings that are either at intersections or at mid-block of any segments within selected corridor.

Figure 13-1 Marked Crosswalk.

4. **Total number of marked midblock crosswalks:** Record the total number of marked midblock crosswalks within selected 1-mile corridor.

Figure 13-2 Marked Midblock Crosswalk.

5. **Average pavement width:** Estimate the curb-to-curb distance in urban areas and edge of pavement to edge of pavement distance for average pavement width in rural areas. If the selected 1-mile corridor is homogeneous (no change in number of lanes with selected corridor), estimate pavement width after each 0.50 miles and provide average estimate of pavement width. If the number of lanes changes within selected corridor, record the pavement widths for different number of lanes and provide average estimate of pavement width. Record the data to the closest foot. For example, if first half-mile section of the corridor is 40 feet wide and the second half-mile section is 60 feet wide, record 50 feet for the corridor.
6. **Total number of residential driveways:** Record total number of residential driveways within selected 1-mile corridor. A residential driveway can be defined as is a type of private road for local access to one or a small group of residence.

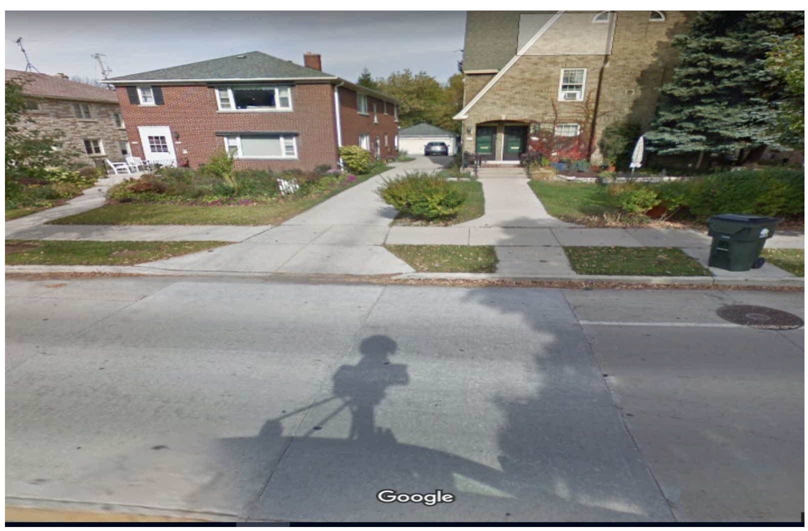


Figure 13-3 Residential Driveway.

7. **Total number of non-residential driveways:** Record total number of non-residential driveways within selected 1-mile corridor. A non-residential driveway can be defined as

is a type of private road for local access to one or a small group of business locations such as gas station, grocery store etc.

Figure 13-4 Non-Residential Driveway.

8. **Average number of through lanes:** Record average number of through lanes excluding exclusive left-turn/ right-turn lanes. If the number of through lanes changes within selected 1-mile corridor, estimate the average number of through lanes.
9. **Total number of left-turn lanes approaching intersections:** Record the total number of left-turn lanes at all intersections on the selected 1-mile corridor only. The left-turn lanes on the intersecting roadways should not be included in this attribute. For two-way traffic, record a left turn lane if there is a presence of LT-lane in any direction of travel of selected corridor at any intersection. In other words, if there is a continuous, two-way left-turn lane for the entire length of the corridor, you should count both approaches for every intersection along the corridor.
10. **Total number of right-turn lanes approaching intersections:** Record the total number of right-turn lanes on the approaching road at all intersections on the selected 1-mile corridor only. The right-turn lanes on the intersecting roadways should not be included in this attribute. For two-way traffic, record a right turn lane if there is a presence of RT-lane in any direction of travel of selected corridor at any intersection.
11. **Percent of corridor with a median:** Record the percent of corridor with a raised median (more common in urban areas) or a grass median (more common in rural areas). As the corridor length is restricted to 1-mile, an approximate length of roadway with median can be converted to record this attribute using (length of corridor with raised median in miles x 100) %. Record this attribute to the closest 10%.
12. **Total number of crosswalks that have a curb extension (bumpout) on at least one end:** Record total number of crosswalks that have a curb extension (bumpout) on at least one-end of the crosswalk.

Figure 13-5 Curb Extension.

13. **Percent of corridor with Sidewalk present:** Record the percent of corridor with a sidewalk on the selected 1-mile corridor in any direction. As the corridor length is restricted to 1-mile, an approximate length of roadway with sidewalk can be converted to record this attribute using [(length of corridor with sidewalk in miles / 2) x 100] %. Record this attribute to the closest 10%. Note: the value of 50% can be reached in two ways: 1) a sidewalk is present on one side only for the entire one-mile corridor, or 2) sidewalks are present on both sides of the roadway for one half (0.5 miles) of the one-mile corridor.
14. **Percent of corridor with Paved shoulder present**: Record the percent of corridor with a paved shoulder on the selected 1-mile corridor in any direction. To count as a paved shoulder, the paved area outside of the white line must be at least 4 feet wide. As the corridor length is restricted to 1-mile, an approximate length of roadway with paved shoulder can be converted to record this attribute using [length of corridor with paved shoulder in miles / 2) x 100] %. Record this attribute to the closest 10%. Note: the value of 50% can be reached in two ways: 1) a paved shoulder is present on one side only for the entire one-mile corridor, or 2) paved shoulders are present on both sides of the roadway for one half (0.5 miles) of the one-mile corridor.

Figure 13-6 Unpaved Shoulder.

Figure 13-7 Paved Shoulder.

15. **Percent of corridor with designated Bike lane present:** Record the percent of corridor with a designated bike lane on the selected 1-mile corridor in any direction. As the corridor length is restricted to 1-mile, an approximate length of roadway with a designated bike lane can be converted to record this attribute using [(length of corridor with designated bike lane in miles / 2) x 100] %. Record this attribute to the closest 10%. Note: the value of 50% can be reached in two ways: 1) a bike lane is present on one side only for the entire one-mile corridor, or 2) bike lanes are present on both sides of the roadway for one half (0.5 miles) of the one-mile corridor.

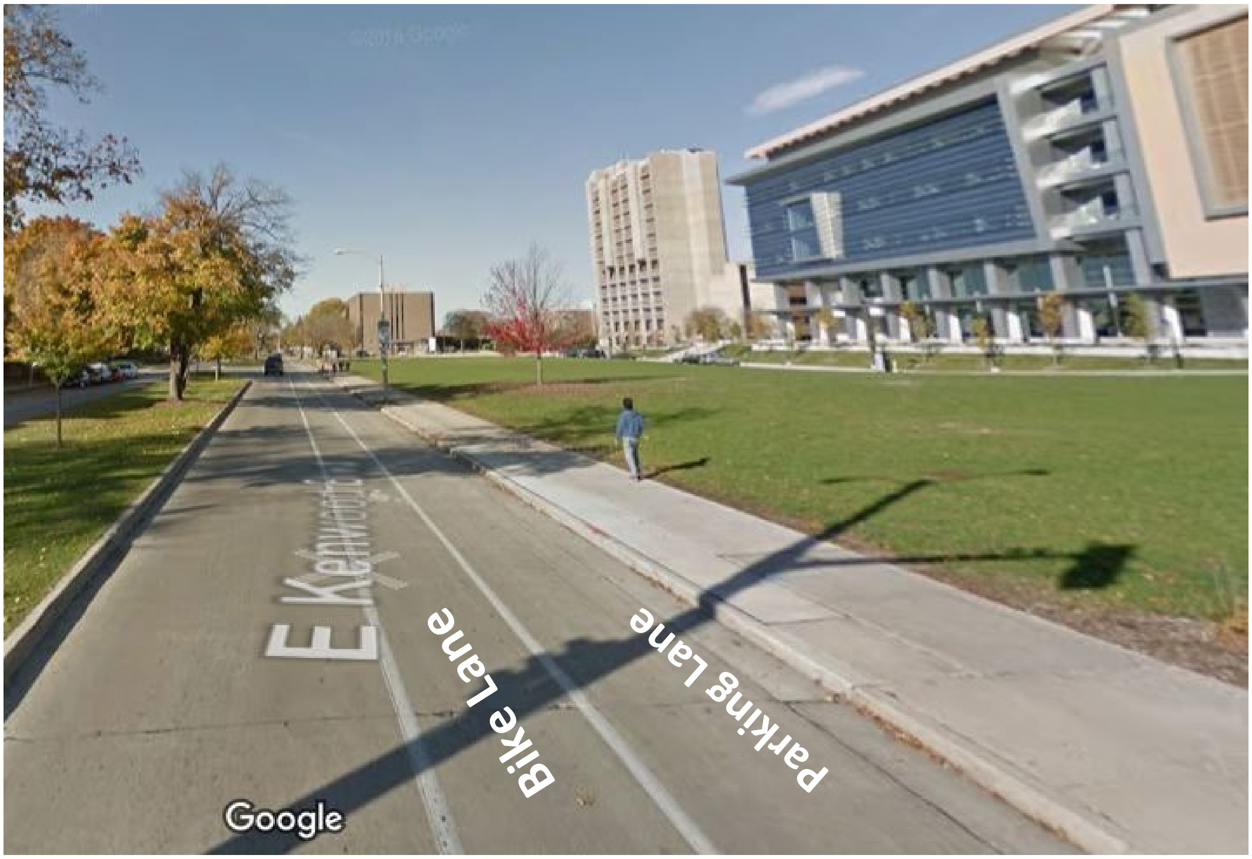


Figure 13-8 Designated Bike Lane.

16. **Percent of corridor with Sidepath:** Record the percent of corridor with a sidepath on the selected 1-mile corridor in any direction. A sidepath can be identified as any parallel roadway along the selected corridor with a width >= 8 feet. As the corridor length is restricted to 1-mile, an approximate length of roadway with a sidepath can be converted to record this attribute using [length of corridor with a sidepath in miles / 2) x 100] %. Record this attribute to the closest 10%. Note: the value of 50% can be reached in two ways: 1) a sidepath is present on one side only for the entire one-mile corridor, or 2) sidepaths are present on both sides of the roadway for one half (0.5 miles) of the one-mile corridor. For corridors with sidepaths, it will be rare to find a sidepath on both sides of the roadway, so 0% or 50% will be common values.

Figure 13-9 Side Path.

17. **Posted speed limit for the corridor:** Record average posted speed limit on the selected 1-mile corridor from speed limit signs visible from Google Street View. If Google Street View does not have this information, use MetaManager to obtain posted speed limit for the selected corridor. If the speed limit changes within the one-mile segment, estimate the distance-weighted average speed limit.
18. **Percent of corridor with Two-way Left-turn lane:** Record the percent of corridor with a two-way left-turn lane on the selected 1-mile corridor. Record this attribute to the closest 10%.

Figure 13-10 Two-way Left-turn Lane.

19. Number of crashes: Record the total number of crashes manually from each selected 1-mile corridor.
    Alternative: Record the total number of crashes on the selected 1-mile corridor by spatially joining crashes with each corridor (Issue: the selected 1-mile corridor may not always resemble with STN network shapefile prepared by WisDOT)
20. Average AADT: Estimate the average of AADT to the nearest 1000 on 1-mile corridor using https://trust.dot.state.wi.us/roadrunner/. There might be multiple count locations within 1-mile corridor. For multiple count locations, estimate the average AADT to the nearest 1000 value. If AADT value is not available on the online map, use MetaManager dataset to estimate average AADT for the 1-mile corridor.
21. Number of multi-use trail crossings: While most of the 1-mile corridors will not have multi-use trail crossings, this variable is of particular interest to WisDOT and may have a positive correlation with the number of observed crashes in the corridor. If possible, we will identify trail crossings using available multi-use trail GIS layers rather than finding them in Google Maps.