PROBABILITY OF PEDESTRIAN DEATHS AND THE ROLE OF SOCIOECONOMIC FACTORS – A DISAGGREGATED APPROACH

FINAL REPORT

SOUTHEASTERN TRANSPORTATION CENTER

XUEHAO CHU, STEVE POLZIN, AND JODI GODFREY

CENTER FOR URBAN TRANSPORTATION RESEARCH, UNIVERSITY OF SOUTH FLORIDA

JULY 2018

US DEPARTMENT OF TRANSPORTATION GRANT DTRT13-G-UTC34

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Probability of Pedestrian Deaths and the Role of Socioeconomic Factors – A Disaggregated Approach		5. Report Date July 2018	
		6. Source Organization Code Budget	
7. Author(s) Chu, Dr. Xuehao; Polzin, Dr. Steven; Godfrey, Jodi Center for Urban Transportation Research, University of South Florida		8. Source Organization Report No.	
9. Performing Organization Name and Address Southeastern Transportation Center 309 Conference Center Building Knoxville, Tennessee 37996-4133 865.974.5255		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. DTRT12-G-UTC34	
12. Sponsoring Agency Name and Address US Department of Transportation Office of the Secretary of Transportation–Research 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered Final Report: May 2017 – July 2018	
		14. Sponsoring Agency Code USDOT/OST-R/STC	
15. Supplementary Notes: None			

16. Abstract

Purpose: The objective was to investigate the significance of socioeconomic factors, roadway characteristics, and other factors in explaining the probability of pedestrians dying once involved in vehicle crashes.

Methods: Using individual pedestrian crashes and related characteristics in urbanized areas of Florida during 2011-2015 as well as Census tract-level data on socioeconomic factors from the American Community Survey, this study statistically identified and determined the significance of socioeconomic factors, roadway characteristics, and other factors in explaining the probability of pedestrians dying once involved in vehicle crashes. The binary logistic regression was performed on the crash data and the odds ratio of statistically significant factors was estimated.

Results: While the socioeconomic environment as measured with ACS data at the tract level was not found to play any statistically significant role, many other attributes were found to play both a statistically and numerically significant role in increasing a pedestrian's chance of being killed once involved in a vehicle crash. These attributes include the vehicle's travel speed, substance abuse by pedestrians and drivers, senior pedestrians, heavy-duty trucks, curved roadways, dark or near dark conditions, etc. For example, the odds for a pedestrian being killed once being hit by a vehicle if the pedestrian's BAC level is at 0.08% or higher is about 58 times that if the pedestrian's BAC level is zero. Future research is suggested to assess the sensitivity of the significant impact of substance abuse by pedestrians on their chance of being killed on alternative specifications of substance abuse as independent variables, alternative specification of the dependent variables, and an explicit treatment of the possible endogenous nature of substance abuse testing results in crash data.

17. Key Words Florida; Pedestrians; Travel Speed; Logistic Regression; Probability of Death; Odds Ratio		18. Distribution Statement Unrestricted	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 58	22. Price N/A

Form DOT F 1700.7 (8-72)

TABLE OF CONTENTS

EXECUTIVE SUMMARY 1
DESCRIPTION OF PROBLEM 5
APPROACH AND METHODOLOGY 8
Introduction 8
Analytical Framework 8
Data 11
Socioeconomic Data 11
Crash Data 11
Combined Data 14
Model Specification 15
Statistical Model 15
Dependent Variable 16
Independent Variables 16
Regression Analysis 26
CONCLUSIONS 41
ACKNOWLEDGEMENTS 49
REFERENCES 50

TABLE OF FIGURES

Figure 1. A Framework on the Determinants of Pedestrian Injury Severity 10
Figure 2. Tampa-St. Petersburg Urbanized Area, 2010 Census 13
Figure 3. Pedestrian's Odds of Being Killed Relative to Odds for Speeds 01-14 MPH 38
Figure 4. Pedestrian's Probability of Being Killed by Vehicle's Travel Speed 39

TABLE OF TABLES

Table 1. Number of Pedestrian Crashes by Scenario 15
Table 2. Summary of Data Items by Category and Expected Effects 20
Table 3. Created Independent Variables and Expected Direction of Effects 23
Table 4. Binary Logistic Regression Results without Socioeconomic Attributes 28
Table 5. Odds Ratio for Select Dummy Variables 36

EXECUTIVE SUMMARY

The number of pedestrian fatalities increased significantly both in numbers and rates from 2009 to 2016 (NHTSA, 2011; undated). A total of 5,987 pedestrians died from vehicle crashes in 2016 versus 4,092 in 2009, a large 46.3% jump. As a share of total traffic fatalities, pedestrian deaths were 12.1% in 2009 but 16.0% in 2016.

The Southeastern (SE) region of the United States, including Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee, includes many of the states and metropolitan areas that are considered to be the most dangerous to pedestrians. Part of the Year 1 and Year 2 research of this MRI project observed the higher pedestrian fatalities per capita in the Southeastern region, than in other regions, primarily resulting from the probability of pedestrians dying once being involved in vehicle crashes being 85% higher in the Southeastern region than in other regions. In addition, the research explored the role of socioeconomic factors in the variation of the risk of pedestrian deaths across states and learned that aggregate analyses were not effective in explaining variations in the risk of pedestrian deaths across large geographies.

Using data on vehicle-pedestrian crashes that occurred in urbanized areas of Florida during the five-year period from 2011 to 2015, the objective of the current study was to use binary logistic regression to statistically identify and determine the significance of socioeconomic factors of Census tracts around crash locations, roadway characteristics, pedestrian and driver attributes, vehicle travel speeds and other vehicle attributes in explaining the probability of pedestrians dying once involved in vehicle crashes. The study adapted an analytical framework on the determinants of pedestrian injury severity from the literature and applied it to specify the empirical models for the regression analysis.

Two empirical models were specified and estimated. One includes the vehicle's travel speed prior to the crash as estimated by the investigating officer. The other does not include this variable but does include the posted speed limit as has been done in some previously published studies on pedestrian injury severity. In addition to the roles of pedestrian, driver, vehicle, roadway, and natural environmental attributes of a crash, this study also tested the role of the socioeconomic characteristics of the population in the environment surrounding a

crash site in the probability of pedestrians being killed once involved. The socioeconomic data came from the five-year American Community Survey data for 2011 to 2015 at the Census tract level.

Once other attributes were appropriately included, the socioeconomic variables were not found to be statistically significant. This negative outcome on the hypothesized role of the socioeconomic environment on pedestrian injury severity was somewhat a surprise but was not entirely unexpected. While most studies of pedestrian injury severity do not consider the socioeconomic environment, the few studies that do have shown some statistically significant results. As a result, the results presented for these two models do not include any socioeconomic variables. Other than the vehicle's travel speed, all other independent variables in the two estimated models were dummy variables (e.g., 1 for male and 0 for female). The following highlights results from the model with the vehicle's travel speed.

- Relative to the vehicle's travel speed under 15 mph, the pedestrian's odds of being killed becomes 1.615 times as large when the vehicle's travel speed reaches 15 to 24 mph, 4.200 times as large when it reaches 25 to 34 mph, 9.514 times as large when it reaches 35 to 44 mph, 16.182 times as large when it reaches 45 to 54 mph, and 26.055 times as large when it reaches 55 mph or higher.
- The top five variables in this ranked list are all related to alcohol or drug use by the driver or the pedestrian. Regardless of the vehicle's travel speed being included or not, these variables have dramatic effects on the pedestrian's chance of being killed.
    - The odds for a pedestrian being killed once being hit by a vehicle if the pedestrian's blood alcohol content (BAC) level is at 0.08% or higher is about 58 times that if the pedestrian's BAC level is zero. Even for a pedestrian with a BAC level below 0.08%, the odds of being killed once being hit is still 19 times that if their BAC level is 0.
    - Substance abuse by the pedestrian has a much greater effect on the pedestrian's chance of being killed than abuse by the driver. For example, the odds ratio for a positive drug test is about 40 for the pedestrian but is about 10 for the driver. The odds ratio for a BAC level below 0.08% is about 19 for the pedestrian but is about 10 for the driver.

- Being a senior pedestrian dramatically increases the odds of being killed. When the vehicle's travel speed is held constant, the odds for a pedestrian being killed if the pedestrian is 75 or older is about 8 times that if the pedestrian is 11 to 64 years of age. Further, the odds of being killed is significantly higher for older seniors than for younger seniors. In fact, the odds ratio for the pedestrian being 75 or older is 2.47 times that if the pedestrian is 65 to 74 years of age.
- Which part of the vehicle hit the pedestrian is much more important in determining the pedestrian's risk of death than what body type the vehicle was. The odds of a pedestrian being killed by the front of a vehicle is 5.350 times that by the sides or the back of the vehicle. In comparison, the risk in terms of odds ratio is lower at 2.950 when the pedestrian is hit by a heavy-duty truck relative to by a passenger car. The risk in terms of odds ratio is even lower at 1.754 when the pedestrian hit by a light-duty truck relative to by a passenger car.
- The pedestrian's odds of being killed on curved sections is 2.645 times that on straight sections.
- Relative to daylight conditions, the pedestrian's odds of being killed is 2.096 times as large under dark conditions and is 2.163 times as large under dusk/dawn conditions.
- Grade sections can raise the pedestrian's odds of being killed by 70.2% relative to flat sections.
- Being hit by a male driver raises the pedestrian's odds of being killed by 39.3% relative to by a female driver.

The current study may be extended in two ways. We still do not know whether and how some of the estimated effects of the various attributes of the transportation system and the natural environment for Florida may differ for other states and between the Southeastern and other regions. One direction of potential further research could extend the current study to a boarder geography. In addition, the estimated impact of substance abuse by pedestrians on their chance of being killed once being hit is greater than what has appeared in the literature. Another direction of research would be to assess the sensitivity of the estimated impact of substance abuse by pedestrians on their chance of being killed on alternative

specifications of substance abuse as independent variables, alternative specifications of the dependent variable, and an explicit treatment of the possible endogenous nature of substance abuse data.

DESCRIPTION OF PROBLEM

The number of pedestrian fatalities increased significantly both in numbers and rates from 2009 to 2016 (NHTSA, 2011; undated). A total of 5,987 pedestrians died from vehicle crashes in 2016 versus 4,092 in 2009, a large 46.3% jump. As a share of total traffic fatalities, it was12.1% in 2009 but was 16.0% in 2016.

The Southeastern (SE) region of the United States, including Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee, includes many of the states and metropolitan areas that are considered to be the most dangerous to pedestrians (McCann and DeLille, 2000; Smart Growth America, 2017). In fact, pedestrian fatalities per capita (i.e., the mortality rate) in 2009 were 44% higher in the SE region than in other regions (Chu, et al., 2017). Using an exposure measure that accounted for both vehicle and pedestrian activities, Chu et al. (2017) showed that this much higher pedestrian mortality rate for the SE region can be attributed to the region's 25% lower exposure per capita, 4% higher risk of pedestrians being involved in vehicle crashes per unit of exposure (i.e., the risk of involvement), and 85% higher probability of pedestrians dying once being involved in vehicle crashes (i.e., the probability of pedestrian deaths). That is, the primary reason for the region's high pedestrian mortality rate is the much higher probability of pedestrian deaths in the region. In addition, Chu at al. (2017) learned that aggregated analyses were not effective in explaining variations in the risk of pedestrian deaths across large geographies.

The objective of the current study was to use binary logistic regression to statistically identify and determine the significance of socioeconomic factors of Census tracts around crash locations, roadway characteristics, and other factors in explaining the probability of pedestrians dying once involved in vehicle crashes. Ideally, this study should have used pedestrian crashes of all severity from all states. This turned out not to be feasible due to limitations in data availability and definitions across states. A nationally representative sample of all pedestrian crashes, such as those in the Crash Report Sampling System (CRSS, formally the General Estimates System or GES before 2016) of the Federal Highway Traffic Safety Administration (NHTSA) would have been a reasonable alternative to the ideal situation. NHTSA codes the geographic coordinates of a crash in CRSS only if they are

available in the original police accident report from the state and does not include the coordinates in the publicly released database. As a result, the current study was limited to data from Florida but covered the five-year period from 2011 to 2015.

With this limitation to Florida, the study will not be able to directly inform why the probability of pedestrian deaths once involved in vehicle crashes is 85% higher in the SE region than in other regions. However, the hope was that it will indirectly inform through providing insights on what are important and what are not important in determining the probability of pedestrians dying from vehicle crashes.

There is a large body of work in the literature on identifying the determinants of pedestrian injury severities through regression analysis. Many of the studies focus on methodological issues in modeling pedestrian injury severities. Savolainen et al. (2011) is a recent review of this literature on both alternative statistical models and unique econometric issues with traffic crash injury severity data. Abay (2013) assesses the sensitivity of the estimation results from applying a given dataset on pedestrian injury severities to alternative statistical models. However, a common shortcoming of most studies in this large body of work is that they lack an implicitly spelled out theory or analytical framework to guide what and how independent variables should, or should not, enter the empirical models. Not having such an analytical framework to guide model specification can lead to significant variations in what and how independent variables are used, making it more difficult to compare the results across studies. More importantly, lacking such an analytical framework is more likely to lead to missing relevant variables and potentially biased estimation results.

Another emerging theme in the literature of pedestrian injury severities is the role of the socioeconomic environment of the local area around a crash location in pedestrian injury severities. Traditionally, the socioeconomic environment has been used to explain variations in pedestrian crashes across geographies (Aguero-Valverde and Jovanis, 2006; Maciag, 2014). It is argued that the socioeconomic environment affects the availability and quality of infrastructure through its influence on public policy and allocation of funding. More recently, the socioeconomic environment has been used to explain pedestrian injury severities (Kim et al., 2017; Xin et al., 2017). In most cases, however, the socioeconomic environment is used without explaining why and how it may affect pedestrian injury severity. This has occurred

partly because these studies lack a sound analytical framework to guide the specification of their empirical models.

Many studies focus on individual states (Ballasteros et al., 2004 on Maryland; Moudon et al., 2011 on Washington; Zajac and Ivan, 2003 on Connecticut; Islam and Jones, 2014 on Alabama; Kim et al., 2008 on North Carolina; Pour-Rouholamin and Zhou, 2016 on Illinois). Despite Florida occupying eight of the most dangerous metro areas in the country during the 10-year period from 2005 to 2014 (Smart Growth America, 2016), few studies of pedestrian safety focus on pedestrian injury severities in Florida. Both Lee and Abdel-Aty (2005) and Haleem et al. (2015) limit their studies to crashes at intersections rather than covering all locations. Siddiqui et al. (2006) and Xin et al. (2017) are exceptions to this limitation with the first focusing on the joint effects of crossing locations and light conditions and the second on presence of bus stops and the socioeconomic environment.

This study contributes to the literature on the determinants of the probability of pedestrians dying from being involved in vehicle crashes in several ways. Methodologically, this study contributes to the general literature by basing empirical model specifications on an analytical framework on the various levels of determinants of pedestrian injury severities. A related contribution of the current study is that it explains, through the analytical framework, why the socioeconomic environment may affect pedestrian injury severities and under what conditions it should be included as independent variables. Empirically, this study provides another piece of evidence to the emerging theme in the literature on the role of the socioeconomic environment in pedestrian injury severities. In addition, this study provides another analysis of the serious problem of pedestrian safety in the state of Florida. More importantly, the results from this study provide new insights into the role of the various direct and indirect determinants in the probability of pedestrians dying from being involved in vehicle crashes. The hope is that these insights can shed new light on potential reasons as to why Florida and other states in the SE region have had a dramatically higher probability of pedestrian deaths once being involved in vehicle crashes.

APPROACH AND METHODOLOGY

INTRODUCTION

The overall approach and methodology for this study was a binary regression analysis of the probability of pedestrians dying from being involved in vehicle crashes at the disaggregate level of individual crashes. This section is organized into the following sub-sections:

- **Analytical Framework**: describes a general structure on the direct and indirect determinants of pedestrian injury severity.
- **Data**: describes the data of pedestrian-vehicle crashes used in the regression analysis.
- **Model Specification**: specifies what dependent variable and independent variables to enter in the binary regression model.
- **Regression Analysis**: carries out the regression analysis through applying the specified regression model to the data and presents the results.

ANALYTICAL FRAMEWORK

The injury severity of pedestrians involved in vehicle crashes is determined largely by three sets of direct determinants: impact speed, impact configuration, and pedestrian attributes. The mass of an involved vehicle is an important determinant of injury severities to both its own occupants and the occupants of the other vehicles involved. Given an impact speed, on the other hand, the mass of the vehicle is unlikely to be a significant factor in determining the injury severity of a pedestrian as all vehicles are dramatically larger than pedestrians and able to inflict fatal injuries. As discussed later, the mass of the vehicle can still play a role in the injury severity of a pedestrian through its effect on the impact speed. Another direct determinant of the risk of pedestrian deaths once involved in vehicle crashes is the access and quality of emergency care. More generally about pedestrian injury severity, Siddiqui et al. (2006) summarize these direct determinants as follows:

- **Impact Speed**—The most important of these is the impact speed, which is the speed of the vehicle upon striking the pedestrian (Lee and Abdel-Aty, 2005; Sullivan and Flannagan, 2002; Jensen, 1999; Gårder, 2004; and Pitt et al.,1990). The chance of

survival by the pedestrian drops quickly between an impact speed of 20 mph and an impact speed of 40 mph (NHTSA, 1999).

- **Impact Configuration**—Besides impact speed, one set of determinants relates to impact configuration between the pedestrian and the vehicle (Yang, 2002). This impact configuration includes several aspects, including the angle at which the vehicle strikes the pedestrian (e.g., frontal versus side), the angle at which the pedestrian is struck (i.e., front, back, side), and the height of the impact on the pedestrian.
- **Pedestrian Attributes**—Another set of determinants relates to the characteristics of the pedestrian. The very young (Jensen, 1999; LaScala et al., 2001; Al_Ghamdi, 2002; and Fontaine and Gourlet, 1997) and the very old (Lee and Abdel-Aty, 2005; Jensen, 1999; LaScala et al., 2001; Al-Ghamdi, 2002; and Fontaine and Gourlet, 1997; and Zajac and Ivan, 2003)) are most vulnerable to suffering from severe injuries. Male pedestrians, being physically stronger and bigger on average than their female counterparts, may be less likely to sustain severe injuries.
- **Emergency Medical Service**—The availability and quality of emergency medical service are critical to the survivability of injured pedestrians.

Policy analysis of pedestrian safety, however, often requires an understanding of intermediate and indirect determinants of pedestrian injury severity that go beyond the direct determinants. Adapted from Figure 2 of Siddiqui et al. (2006), Figure 1 shows these intermediate and indirect determinants and how they play a role in pedestrian injury severity through their effects on the direct determinants.

- Vehicle attributes may affect both impact configuration and impact speed. High profile vehicles, such as SUVs and light trucks, are more likely to increase the height of the impact on a pedestrian. There is evidence that these high-profile vehicles represent a level of fatality risk to pedestrians that is several times as high as regular passenger cars (Lefle and Gabler, 2004). In addition, holding other factors constant, heavy vehicles are harder to stop, resulting in higher impact speeds for given vehicle travel speeds and hence more serious injuries to pedestrians.

- In addition to vehicle attributes, several sets of other factors affect the impact speed of a vehicle. These include the moving speed of the vehicle, driver attributes (e.g., attentiveness), road attributes (e.g., surface condition), and pedestrian visibility to the driver.

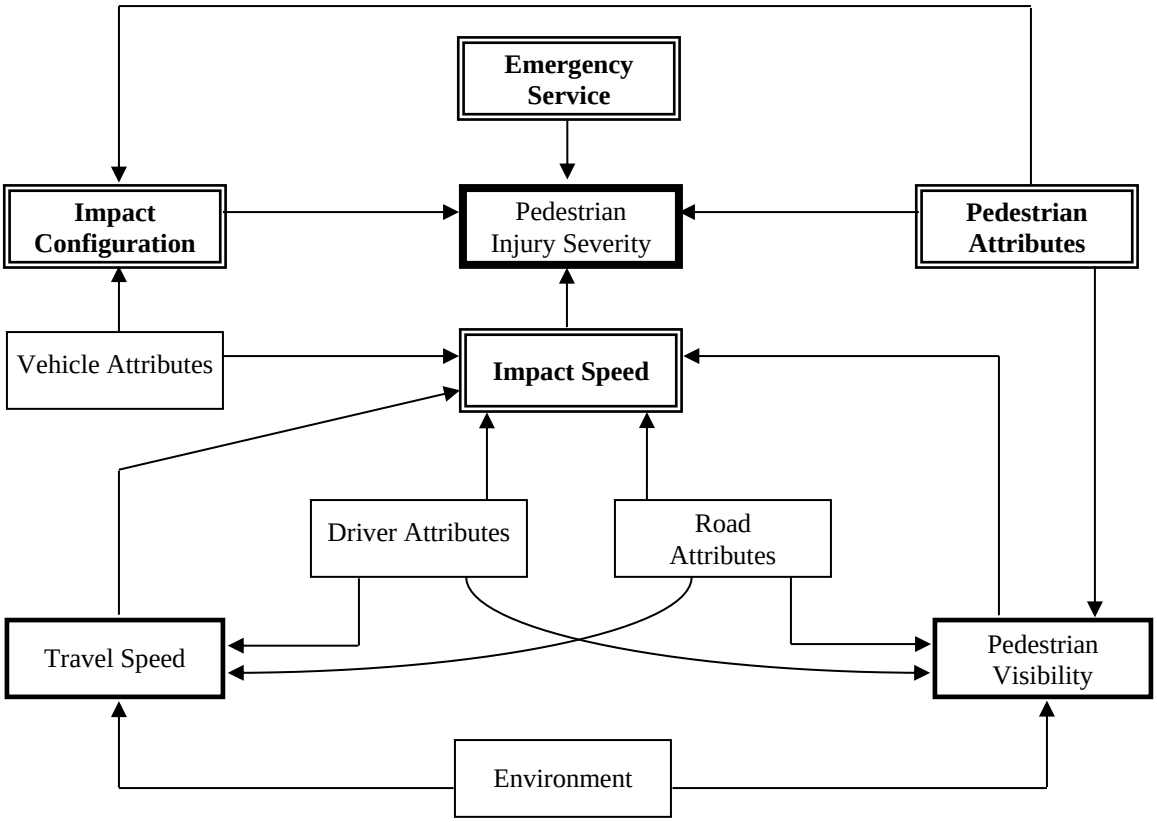

Figure 1. A Framework on the Determinants of Pedestrian Injury Severity

- Furthermore, both driver attributes (e.g., distracted drivers, drivers with poor vision, etc.) and road attributes (e.g., curve or steep grade sections) affect the moving speed and pedestrian visibility to the driver.
- Pedestrian attributes, such as whether they wear reflective clothing at night, affect pedestrian visibility to the driver. Besides vehicle attributes, some pedestrian attributes, such as person height, can also affect impact configuration.
- Finally, the environment can affect both the moving speed of the vehicle and pedestrian visibility to the driver. One part of the environment consists of natural

conditions, including light conditions, weather, etc. Both light and weather conditions can affect both the moving speed of the vehicle and pedestrian visibility. Another important part of the environment is the traffic stream within which the involved vehicle moves, including traffic density, composition of drivers, etc. Characteristics of the traffic stream affect the moving speed of the vehicle. The environment also includes the socioeconomic attributes of the area around each crash locations.

This general structure guided model specification for the regression analysis. However, model specification was influenced not only by this general structure but also the available data for the regression analysis.

DATA

The data used for this study includes individual pedestrian-vehicle crashes and socioeconomic characteristics at the Census tract level from the American Community Survey (ACS) for the five-year period of 2011 to 2015. This study period was chosen because 2015 was the latest year for which both crash data and socioeconomic data were available when the data were obtained.

Socioeconomic Data

The ACS data were downloaded at the tract level directly from American FactFinder. In addition to the socioeconomic characteristics of the population in individual tracts, the data also came with tract boundaries. A long list of ACS data items was available, but only a small number of them were selected and processed.

Crash Data

The crash data for this study included all pedestrian-vehicle crashes of all severity levels that involved a single vehicle and a single pedestrian on non-interstate non-toll public roadways in all urbanized areas of Florida.

The Safety Office of the Florida Department of Transportation provided the crash data from its Crash Analysis Reporting (CAR) database. The Safety Office receives long-form crash records from the Department of Highway Safety and Motor Vehicles (DHSMV) on an ongoing basis, uploads the long form reported crashes to its CAR database, and then adds location coordinates to each crash that has occurred either (a) on the State Highway System (SHS) or (b) on non-SHS local, public roadways. Crashes on the SHS receive location coordinates on the FDOT Transportation Data and Analytics Office's linear-reference system for the Roadway Characteristics Inventory (RCI). Crashes on non-SHS local, public roadways receive geospatial coordinates (latitude and longitude). Crashes in parking lots, on private property, on forest roads, or on private roads (i.e. those not involving public roadways) do not undergo location processing (FDOT, 2010).

The Safety Office provided the crash data for all crashes, including pedestrian-vehicle crashes, in three separate data files: a crash file, a non-motorist file, and a file on both vehicles and vehicle occupants. These three files can be linked to each other through the year of crash and a crash identification number that is unique to each crash within a given year. The crash file provided attributes of each crash, including the roadway system identifier (ownership), day of week and time of day, light conditions, weather conditions, the total number of vehicles and total number of pedestrians involved, etc. The non-motorist file included age, gender, use of alcohol, use of drugs, whether violation citation was issued, etc. The vehicle file included roadway features, vehicle body type, estimated vehicle travel speed, area of initial impact, drivers' age and gender, uses of alcohol and drugs by drivers, etc.

From these three raw data files a final crash file was created that included only pedestrian-vehicle crashes that involved a single vehicle and a single pedestrian on non-interstate non-toll public roadways in all urbanized areas of Florida.

- Focusing on urbanized areas has several reasons. Most pedestrian crashes occur in urbanized areas. Pedestrians with a lower socioeconomic status are more concentrated in urbanized areas. The size of census tracts is more uniform in urbanized areas. Figure 2 shows the boundaries of the Tampa-St. Petersburg urbanized area as defined by the 2010 Census (US Census Bureau, undated). This urbanized area includes not

only relatively high density urban areas in Clearwater, St. Petersburg, and Tampa but also low density suburban areas with many wide, high-speed arterials.

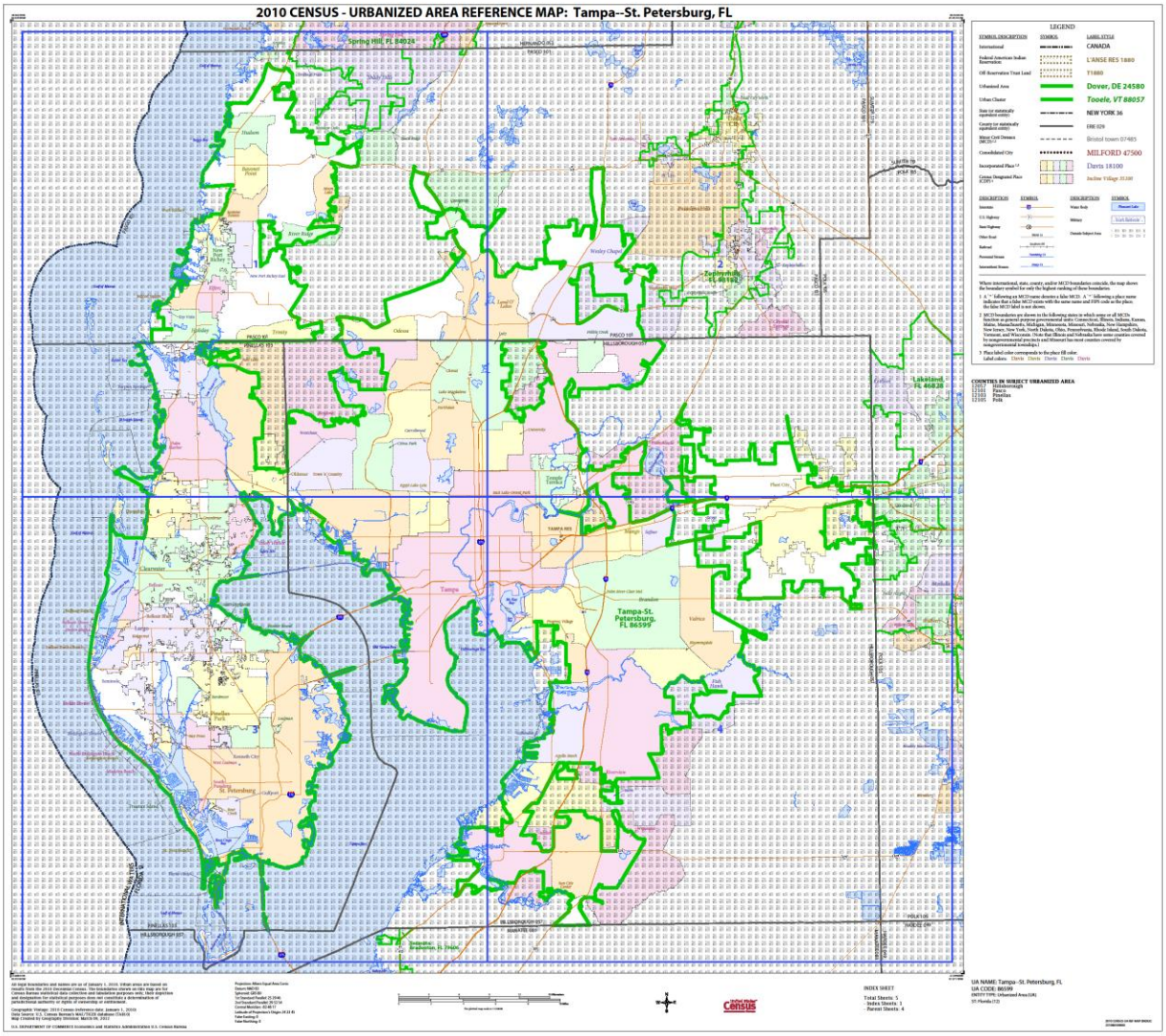

Figure 2. Tampa-St. Petersburg Urbanized Area, 2010 Census

- Limiting pedestrian-vehicle crashes to those involving a single-vehicle and a single-pedestrian was a short cut to linking each involved pedestrian to the vehicle that hit the pedestrian. Having this link was necessary so that the study can bring vehicle characteristics into the analysis. However, the non-motorist data file does not have a vehicle associated with each pedestrian involved. For those vehicle-pedestrian crashes

that involved more than one pedestrian or more than one vehicle, a unique one-to-one link between a pedestrian and a vehicle was not an easy task. Most pedestrian-vehicle crashes involve a single vehicle and a single pedestrian. Limiting single-vehicle pedestrian crashes also had the benefit of avoiding the complexities of possible secondary crashes between a pedestrian with other vehicles.

- Focusing on non-interstate crashes was associated with the study objective to examine the role of socioeconomic factors of the driver and pedestrian populations in the local area where a crash occurred. Interstate and toll-road users are much less likely to be coming from, or going to, the local area of the crash. As a result, the roadway system identifier was used to include crashes on U.S., state, county, or local roadways but to exclude all crashes on interstates and toll roads. Crashes on forest roads, private roadways, or parking lots were automatically excluded because they do not have location coordinates that were necessary to bring ACS data into the crash data. The exclusion of toll roads also meant that the final set of vehicle-pedestrian crashes did not include non-interstate limited-access expressways because most of them in Florida are toll roads.

Combined Data

Working through GIS software, the location coordinates of each vehicle-pedestrian crash in the crash data file was determined relative to the boundaries of each tract in Florida. The software automatically assigned crashes located on tract boundaries without any manual intervention. For this study, it was not critical for boundary crashes to fall into one tract or another because neighboring tracts are more likely to be similar than not.

Once this geographic relationship was established, the selected socioeconomic factors were merged into the crash data file. Finally, personal characteristics of pedestrians from the non-motorist data file and vehicle characteristics from the vehicle and occupant data file were merged into the crash data file that already had socioeconomic data. Table 1 summarizes the number of pedestrian crashes by scenario. Note that the count in the right column reflects the cumulative effects of the scenarios from the total number of pedestrian crashes in both rural and urban areas.

Table 1. Number of Pedestrian Crashes by Scenario

Scenario	Count
Total in rural and urban areas	43,939
In urbanized areas	28,170
On U.S., state, county, or local roads; excluding interstate, toll roads, forest or private roads	27,332
Involving only one vehicle and one pedestrian	23,768
Available vehicle and pedestrian data (e.g., vehicle data not available for hit-and-run crashes)	20,137
Vehicle in transport	19,816

The actual number of pedestrian crashes available for the regression analysis depended on the dependent variable used and on the set of independent variables used. The degree of valid data not available varied widely across data items directly from the crash data files and across independent variables defined from these data items. This variation affected the actual number of observations available for the regression analysis.

MODEL SPECIFICATION

The specification of an empirical model for the regression analysis later included three components: selecting a statistical model, specifying the dependent variable, and specifying independent variables in terms of what variables to use and how to use them.

Statistical Model

A range of statistical models have been used in the literature to study roadway crash injury severities. Savolainen et al. (2011) is a recent review of this literature. Most of these statistical models are applicable to pedestrian-vehicle crashes. Selecting a statistical model depends on several factors as reviewed by Savolainen et al. (2011). The following are some of these factors:

- The injury outcome can be treated as multiple responses (e.g., fatal, incapacitating, non-incapacitating, possible, or no injury) or as binary (e.g., fatal or non-fatal).
- Injury severity levels are ordinal by nature. In addition, some of the multiple response outcomes are in closely related categories such as no injury and possible injury.

- Underreporting of certain crashes is known to occur (Yamamoto et al., 2008; Ye and Lord, 2011; Mannering and Bhat, 2013). To the degree underreporting occurs, the crash data available become an outcome-based sample, i.e., the proportion of injury severities in the crash data is not the same as the proportion of injuries among all crashes, reported or not reported.
- The available data from crash reports and other sources may not have variables that should be included as independent variables.
- Injuries among the roadway users involved in the same crash are likely to be correlated due to unobserved elements of the crash.
- Crashes that occur closely in space or in time may be correlated due to them sharing unobserved spatial or temporal elements.

The current study chose the binary logit model for its simplicity and then focused on properly specifying the empirical model according to the analytical framework described earlier.

Dependent Variable

This study focused on the probability of the pedestrian dying from being hit by the vehicle in a single-pedestrian and single-vehicle crash. The dependent variable takes 1 if the pedestrian was fatally injured. It takes 0 if the pedestrian was involved but not fatally injured regardless of whether he was injured. Alternatively, it could have taken 0 if the pedestrian was injured but not fatally injured. This study chose the first option of defining the dependent variable for the 0 outcome. This is consistent with the 85% higher probability of pedestrians dying once involved in vehicle crashes in the SE region relative to other regions.

Independent Variables

Theoretical Specification. The analytical framework in Figure 1 calls for three possible alternative forms of specification for independent variables. One is based on the direct determinants only, one based on indirect determinants of impact speed without including impact speed or travel speed, and the other based on indirect determinants of impact speed without including impact speed but including travel speed.

- *Direct Determinants Only* – These include the impact speed and three groups of attributes relating to impact configuration, pedestrians involved, and medical emergency service. Each of these three groups may include more than one attribute. Impact configuration, for example, includes the height profile of the vehicle, the part of the vehicle that hit the pedestrian (i.e., initial area of impact in the crash file), the part of the pedestrian hit by the vehicle, etc. Pedestrian attributes include health, body mass, height, etc.
- *Indirect Determinants without Travel Speed* – One alternative is to replace impact speed by its intermediate determinants. As summarized in Figure 1, these indirect determinants include driver attributes, pedestrian visibility, vehicle attributes, and the travel environment. The travel environment includes light conditions, weather conditions, and socioeconomic characteristics of the population in the local area around the crash location. While this possibility is not shown in Figure 1, some have argued that the socioeconomic environment of an area affects pedestrian safety through the political system in terms of funding allocation across areas with different socioeconomic environments (Maciag, M., 2014). Although the vehicle's travel speed affects the impact speed, it is considered an intermediate determinant because travel speed is determined by other indirect determinants.
- *Indirect Determinants with Travel Speed* – Another alternative to the form of direct determinants only is to include the vehicle's travel speed as well, in addition to all indirect determinants of the impact speed. One argument for this would be that the indirect determinants available are not adequate in determining the vehicle's travel speed. More important, the vehicle's travel speed may be included as a policy variable.

Empirical Specification. Any empirical specification that is both theoretically sound and ready for regression analysis must consider both the theoretical specification and the available data. Before discussing model specification, it is necessary to discuss what the vehicle speed data available in the crash data represent.

- According to the official instructions to accident investigating officers (DHSMV, 2015), the speed data represent the speed at which the vehicle was traveling prior to the traffic crash as estimated by the investigating officer. Unfortunately, the instructions are not clear as to whether it represents the vehicle's speed prior or after any avoidance maneuvers. The name of the speed variable in the crash data is labeled as "vehicle speed," which is vague relative to being labeled as "travel speed."
- NHTSA (2017) takes the same speed data, interprets them as the vehicle's travel speed prior to any avoidance maneuvers for its FARS and CRSS databases, and labels the corresponding variable as "travel speed." With this NHTSA's interpretation, the available speed data probably represents the vehicle's travel speed rather than the impact speed in the analytical framework in Figure 1.
- Using pedestrian crash data from Florida, Lee and Abdel-Aty (2005) use the same data on vehicle speeds and treat them as the travel speed prior to any avoidance maneuvers.

Based on the above discussion, it is reasonable to assume that the vehicle speed available for each involved vehicle in the crash data represents the vehicle's travel speed prior to any avoidance maneuvers. As a result, the option of using impact speed directly without including any of its immediate or indirect determinants was unavailable for this study. This is not a limitation to the current study because one primary interest of the study was to examine the roles of socioeconomic factors and roadway attributes, etc. in the probability of pedestrians dying from being involved in vehicle crashes. If impact speed were directly used as an independent variable, these socioeconomic and roadway attributes would not be able to show up in the model. The other two available forms of the theoretical specification differ in the following way:

- One includes the vehicle's travel speed but not the speed limit of the roadway because the speed limit is just a determinant of the travel speed.
- The other form of the theoretical specification excludes the vehicle's travel speed and but include the speed limit.

Emergency medical service did not enter any empirical model for this study. The primary reason was that this study was limited to pedestrian-vehicle crashes in urbanized areas and the variation in the availability and quality of emergency medical service is relatively small across urbanized areas in Florida than across urbanized, urban cluster, and rural areas. Another reason was that data on the availability and quality of emergency medical service were not readily available at the level of Census tracts or similarly sized geographies. Table 2 lists specific data items from the crash files that this study considered for each of these categories:

- Impact configuration
- Pedestrian attributes
- Vehicle attributes
- Driver attributes
- Road attributes
- Pedestrian visibility
- Environment

Impact configuration was included for some of its attributes. The pedestrian's point of impact by the vehicle was not directly included because related data were not available in the crash files. Indirectly, both the age and gender of the pedestrian included as pedestrian attributes later, particularly children and seniors, can capture some effect of the height of the pedestrian's point of impact. In addition to these pedestrian attributes, vehicle size in terms of passenger cars, light-duty trucks (e.g., SUVs), and heavy-duty trucks and the initial area of impact on the vehicle with the pedestrian were also considered.

Pedestrian attributes include age, gender, pedestrian location at time of crash, pedestrian action prior to crash, action during crash, and use of alcohol or drugs.

Vehicle attributes include vehicle mass that affects braking distance and defects that may make braking less effective and affect both impact configuration and the vehicle's impact speed as shown in Figure 1. Vehicle body type would also capture differences in vehicle mass. The crash data had few cases of brake defects, but a much larger number of

missing data associated with the data item on vehicle defects. To avoid losing observations, brake defects were not considered as a vehicle attribute for this study.

While not explicitly shown in Figure 1, the height profile of the vehicle can also affect the driver's visibility of the pedestrian. In addition, the vehicle's location prior to the beginning of the sequence of crash events can affect its travel speed (e.g., turning lanes vs. through lanes).

Table 2. Summary of Data Items by Category and Expected Effects

Category	Data Item from Crash Files and ACS	Impact Configuration	Pedestrian Visibility	Travel Speed	Impact Speed
Environment Attributes	Weather condition (e.g., rain, glare)		x	x	x
	Light condition		x	x	x
	Socioeconomic			x	x
Pedestrian Attributes	Age	x	x		
	Gender	x	x		
	Location on roadway at time of crash		x	x	
	Action prior to crash (e.g., crossing road)		x	x	
	Action contributed to crash (e.g., dart out)		x		x
	Use of alcohol		x		
	Use of drugs		x		
Driver Attributes	Age			x	x
	Gender			x	x
	Action contributed to crash (e.g., careless)			x	
	Distraction (e.g., inattentive)		x	x	x
	Vision obstruction (e.g., stopped vehicle)		x		x
	Use of alcohol		x	x	x
	Use of drugs		x	x	x
Vehicle Attributes	Body type	x	x		x
	Area of initial impact	x			
	Action prior to crash (e.g., turning left)			x	
Roadway Attributes	Speed limit			x	
	Width in number of lanes			x	
	Trafficway (e.g., divided, two-way)			x	
	Inclination (e.g., uphill)		x		
	Alignment (straight vs. curve)		x		
	Control device (e.g., no device)			x	
Pedestrian Visibility	No direct measurement				x
Travel Speed	Estimated by investigating officer				x

Driver attributes can affect pedestrian visibility, the vehicle's travel speed, and the impact speed through the driver's risk-taking behavior and choice of travel speeds, their attitudes towards pedestrians, their attentiveness to their surroundings, their health condition, etc. The crash data do not have items that directly measure all these driver attributes.

- The study used age, gender, and the driver's action at the time of crash that may have contributed to the crash to approximate the driver's risk-taking behavior and choice of travel speed.
- The study used the driver's distraction that may have influenced the driver's performance to measure driver attentiveness and its effect on pedestrian visibility.
- The study used the driver's external vision obstructions to measure the effects on pedestrian visibility.

Road attributes affect pedestrian visibility, the vehicle's travel speed, and its impact speed. Curved roads or roads at a steep grade can reduce pedestrian visibility to the driver. The road's design features, surface conditions, and the traffic attributes near the crash location at the time of crash affect the vehicle's travel speed. Surface conditions can also affect the vehicle's impact speed. Some of the roadway design features may be highly correlated and may not be used at once. The crash files did not provide data on volumes or any other traffic features. Indirectly, some of the above design features are highly correlated with traffic volumes and the vehicle's travel speed.

Pedestrian visibility can affect the time available for the driver to react and is not directly measured in the crash data. As discussed before, some of the driver and roadway attributes help capture pedestrian visibility to the driver. Also, some of the environment attributes discussed next affect the pedestrian's visibility to the driver. In addition to these attributes, this study used some of the pedestrian actions that may have contributed to the crash to capture pedestrian visibility. Specifically, these pedestrian actions included dart or dash into the road and not being visible with dark clothing at night.

Environment can affect both pedestrian visibility and the vehicle's travel speed. There is a natural component and a socioeconomic component. The natural component

includes weather conditions and lighting conditions. The socioeconomic component represents the collective attributes of the drivers and pedestrians in the area around a crash. The drivers and pedestrians present near the location of a crash do not necessarily live in the area around the crash. Many, particularly pedestrians, do. This study used the socioeconomic data aggregated at the Census tract level to approximate this socioeconomic environment.

It is hypothesized that the socioeconomic environment affects the average behavior of roadway users of all modes in the area. Drivers in areas with different socioeconomic environments may drive at different speeds and react differently when facing pedestrians under similar roadway and other conditions. Similarly, pedestrians in areas with different socioeconomic environments may behave differently in interacting with vehicle traffic. As a result, the socioeconomic environment can affect the vehicle's travel speed, the vehicle's impact speed, etc. As discussed in detail later, the socioeconomic environment would not play much of a role in pedestrian injury severity for an empirical model that includes the vehicle's impact speed. For models with no data on the vehicle's impact speed, however, the socioeconomic environment was expected to play a role in pedestrian injury severities.

For each of the data items in this study, researchers created one or more variables for consideration in the regression analysis. Table 3 summarizes these variables organized by category and by data item shown in Table 2. Most of these variables are dummies. One dummy for pedestrian gender, for example, is PED_MALE and it takes 1 if the pedestrian was male and 0 otherwise. For those data items that have more than one dummy variable defined, one serves as the base of comparison in the regression analysis. For those data items that have only one dummy variable defined, the base of comparison is what is not shown. Understanding the base of comparison for any defined dummy variable is critical because interpreting the result of a dummy variable must be relative to the corresponding base.

Table 3. Created Independent Variables and Expected Direction of Effects

Category	Data Item	Variables	Definition	Expected Direction of Effects
Natural Environment	Light condition	NE_DARK	1 if dark; 0 otherwise	±
		NE_DUSK_DAWN	1 if dusk or dawn; 0 otherwise	±
		NE_DAY	Base: daylight	N/A
	Contributing condition	NE_GLARE	1 if glare; 0 otherwise	±
	Weather	NE_WTHR_RAIN	1 if rail; 0 otherwise	±
		NE_WTHR_OT	1 if other weather condition; 0 otherwise	±
		NE_CLEAR	Base: clear weather	N/A
Driver Attributes	Gender	DR_MALE	1 if male; 0 otherwise	±
	Age	DR_AGE15_24	1 if age 15 to 24; 0 otherwise	±
		DR_AGE25_64	Base: 1 if age 25 to 64; 0 otherwise	N/A
		DR_AGE65P	1 if age 65 or over; 0 otherwise	±
	BAC level	DR_BAC_08_	1 if BAC 0.01 to 0.07; 0 otherwise	±
		DR_BAC_08P	1 if BAC 0.08 or over; 0 otherwise	±
		DR_BAC_0	Base: BAC=0	N/A
	Contributing action	DR_FAIL_YLD	1 if fail to yield right of way; 0 otherwise	+
		DR_CARELS	1 if careless negligent driving; 0 otherwise	+
		DR_OA	1 if other improper actions; 0 otherwise	+
		DR_PROPER	Base: no improper action	N/A
	Distraction	DR_DSTR	1 if distracted; 0 otherwise	+
	Vision obstruction	DR_OBST	1 if DR or PED vision obstructed; 0 otherwise	+
Pedestrian Attributes	Gender	PED_MALE	1 if male; 0 otherwise	±
	Age	PED_AGE0_10	1 if age 0 to 10; 0 otherwise	+
		PED_AGE11_64	Base: age 11 to 64	N/A
		PED_AGE65_74	1 if age 65 to 74; 0 otherwise	+
		PED_AGE75P	1 if age 75 or over; 0 otherwise	+
	BAC level	PED_BAC_08_	1 if BAC 0.01 to 0.07; 0 otherwise	+
		PED_BAC_08P	1 if BAC 0.08 or over; 0 otherwise	+
		PED_BAC_0	Base: BAC=0	N/A
	Action prior to crash	PED_CROSSING	1 if crossing roadway; 0 otherwise	+
		PED_W_ALONG	1 if walking along roadway; 0 otherwise	±
		PED_OT	Base: all other actions	N/A
	Contributing action	PED_DASH	1 if dart or dash out; 0 otherwise	+
		PED_OA	1 if any other improper action; 0 otherwise	+
		PED_NONE	Base: no improper action	N/A
	Location at crash	PED_INT_MRKD	1 if intersection marked crosswalk; 0 otherwise	-
		PED_INT_NMRKD	1 if intersection unmarked crosswalk; 0 otherwise	-
		PED_MID_MRKD	1 if midblock marked crosswalk; 0 otherwise	+
		PED_MID_NMRKD	1 if midblock without crosswalk; 0 otherwise	+
		PED_OT	Base: other locations	N/A
Roadway Attributes	Width in total number of lanes	RD_LANES3_5	1 if 3 to 5 lanes; 0 otherwise	+
		RD_LANES6P	1 if 6 or more lanes; 0 otherwise	+
		RD_LANES1_2	Base: 1 to 2 lanes	N/A
	Grade	RD_NOT_LVL	1 if not level; 0 otherwise	+
	Alignment	RD_CURVE	1 if curved; 0 otherwise	±
	Posted speed limit	RD_SPLMT25_	1 if 25 mph or under; 0 otherwise	-
		RD_SPLMT26_44	Base: 26 to 44 mph	N/A
		RD_SPLMT45P	1 if 45 mph or over; 0 otherwise	+
	Trafficway	RD_TWO_DVD	1 if two-way divided; 0 otherwise	+
		RD_TWO_NDVD	1 if two-way not divided; 0 otherwise	+
		RD_ONE_WAY	Base: one way	N/A
	Control device	RD_NO_CTRL	1 if no control device; 0 otherwise	+

Table 3. Created Independent Variables and Expected Direction of Effects (Continued)

Category	Data Item	Variable	Definition	Expected Direction of Effects
Vehicle Attributes	Travel speed	VEH_SP15_24	1 if 15 to 25 mph; 0 otherwise	+
		VEH_SP25_34	1 if 25 to 34 mph; 0 otherwise	+
		VEH_SP35_44	1 if 35 to 44 mph; 0 otherwise	+
		VEH_SP45_54	1 if 45 to 54 mph; 0 otherwise	+
		VEH_SP55P	1 if 55 or higher mph; 0 otherwise	+
		VEH_SP01_14	Base: 1 to 14 mph	N/A
	Area of initial impact in 12 clock points	VEH_FRONT	1 if 1, 11, 12; 0 otherwise	+
		VEH_SIDE_BACK	Base: other impact areas	N/A
	Body type	VEH_HDT	1 if medium-heavy duty truck; 0 otherwise	+
		VEH_LDT	1 if passenger van, pickup, SUV; 0 otherwise	+
		VEH_MTCYLC	1 if motorcycle or moped; 0 otherwise	-
		VEH_CAR	Base: passenger cars	N/A
	Action prior to crash	VEH_THRU_LN	1 if through lanes; 0 otherwise	+
Socio Economic Attributes	Population density	SE_DEN_POP	1 if thousand/square miles; 0 otherwise	-
	Race and ethnicity	SE_ShareHisp	Share Hispanic	±
		SE_ShareNHB	Share Non-Hispanic black	±
		SE_ShareNHW	Share Non-Hispanic white	±
		SE_ShareOT	Base: other race/ethnicity groups	N/A
	Household Income	SE_MED_INC	Median household income (in 2015 dollars)	-
		SE_PerCapINC	Per capita household income (in 2015 dollars)	-
		SE_SharePOVT	Share population at or below poverty; 0 otherwise	+
	Unemployment rate	SE_UnEmpRate	Share workers unemployed; 0 otherwise	+
	Education attainment	SE_ShareNoHS	Share population without HS diploma; 0 otherwise	+
	Vehicle ownership	SE_Share0Veh	Share workers with 0 vehicles; 0 otherwise	+
	Health insurance	SE_ShareNoHI	Share population without health insurance; 0	+

The last column on the right in Table 3 also shows the expected direction of effects for each of the created variables while holding other variables constant.

- In some cases, a specific direction of effects can be expected because the expected effects of these variables are expected in a single direction (e.g., either increasing or decreasing the probability of pedestrian fatality). A "+" sign for a variable indicates that increasing that variable would increase the probability.
- Similarly, a "-" sign indicates that increasing the variable can be expected to decrease the probability.
- In other cases, however, the effects of variables were expected to go both directions and the net of these opposite effects could not be expected in advance. A "±" sign is used for these cases. As Figure 1 shows, many attributes can affect more than one set of direct determinants of pedestrian injury severity, and this fact makes many of the variables not to have a specific expected direction of effects on pedestrian injury

severity. The net effect of increases in such a variable may be more severe pedestrian injuries (e.g., estimated coefficient is statistically significant and positive), may be less severe pedestrian injuries (e.g., estimated coefficient is statistically significant and negative), or even near zero on pedestrian injury severities (e.g., estimated coefficient is not statistically significant).

- Since variables used as the base of comparison are not directly estimated and would not have coefficients. As a result, they are shown with an "N/A" sign.

The reasons for the expected direction of effects are not discussed here for every variable, partially because many of these variables may not show up in the final regression equations. However, using several examples may help explain the expected direction of effects:

- **"+" example**: Relative to the vehicle's travel speed under 15 mph (VEH_SP01-14), VEH_SP35_44 was expected to increase the impact speed and the probability of the pedestrian dying from the crash. Another example is VEH_FRONT, the area of initial impact in the front of the vehicle. Relative to areas of initial impact on the sides or in the back, a front impact was expected to have a higher probability of the pedestrian dying from the crash.
- **"-" example**: Relative to roadway sections with a posted speed limit of 26 to 44 mph, crashes in roadway sections with a posted speed limit 25 mph or lower was expected to have a lower probability of the pedestrian dying from the crash.
- **"±" example**: Relative to clear weather conditions, rain can have multiple effects that can affect pedestrian injury severity differently. Rain may slow traffic down in general, lower the visibility between the driver and the pedestrian, make braking less effective, etc. The first effect may lower the vehicle's impact speed, but the other two effects may increase it.

Sometimes including interactions between two variables may bring additional insights that including individual variables only may not be able to. One example would be to include the product of CR_DARK and PED_DASH instead of just PED_DASH. It is possible that the pedestrian's dash and dart out actions may be deadly primarily at night rather than during day

light conditions. Because each of these two dummy variables has its own base of comparison and each may still be separately included as independent variables, interpreting the estimated result of this product variable can be challenging. As a result, this study did not use any interactions of variables in the represented models.

REGRESSION ANALYSIS

The injury severity of an involved pedestrian was assessed and determined by the investigating officer in the field in most cases. In a sequentially more severe order, the possible levels of injury were:

- None
- Possible
- Non-incapacitating
- Incapacitating
- Fatal

The pedestrian's severity level would be modified later if the pedestrian died within 30 days of the crash due to injuries suffered from the crash. Some pedestrians died due to causes not related to the crash (e.g., heart attack, natural causes, suicide, homicide), and the related crashes were excluded from this study. Also excluded were crashes without a valid severity level coded for the pedestrian.

This study defined the dependent variable as a binary outcome: 1 if the pedestrian died from injuries suffered from the crash; and 0 if the pedestrian was involved but was either not injured or not fatally injured. One option would have been to exclude crashes where the pedestrian was not injured. These no-injury crashes were included in this study partly because including them is consistent with the results from Years 1 and 2 of this overall project as discussed in the introduction of this report. From a policy point of view, a non-injury outcome is better than any injury outcome. As a result, it does not make much policy sense to exclude the no-injury crashes from a study of pedestrian injury severity.

To keep the analysis simple, this study chose logistic regression as the statistical technique for analysis. With logistic regression, the probability of the pedestrian dying from

a crash has a closed form; model estimation is fast; and its results typically are similar to those from a binary probit model. In addition, this study chose not to explicitly deal with potential spatial autocorrelation among spatially related crashes to avoid additional complexities.

Table 4 shows the results from the binary logistic regression analysis for two models. One model includes the vehicle's travel speed (shown in three columns in the middle), while the other does not (shown in the three columns on the right). When the vehicle's travel speed is included, roadway and vehicle related variables believed to be associated solely with the vehicle's travel speed are excluded. These roadway related variables include the number of lanes, speed limit, and whether the roadway was one-way and was divided. The vehicle related variables indicate the vehicle's action prior to the crash (e.g., turning). When the vehicle's travel speed is not included, the posted speed limit is included along with the other roadway attributes. Typically, regression models for pedestrian injury severity use the specification without including the vehicle's travel speed.

Those variables defined from the same data item are listed together. Further, those variables for the same attribute group (e.g., vehicle attributes) are indicated with a common two- or three-letter combination in the beginning of the variable names (e.g., VEH for vehicle attributes) and are listed together. Lacking space for additional columns, these group names and data item names are not shown in this table.

For each variable in a given model, the table shows its coefficient (Coeff.), the significance level (Sig.), and the odds ratio. The statistical significance level is highlighted in bold if it is at least 5% or better (i.e., Sig.≤0.050) and is underlined if it is worse than 5% (i.e., Sig.>0.050) but at least 10% (i.e., Sig.≤0.100). Each model also includes a constant, and its estimates are listed after all other independent variables. Although listed in the table, the base variable for each data item with more than one dummy variable included in the model was not part of the model estimation and hence has no coefficient. For example, CR_DAY is a dummy variable for daylight conditions and serves as the base for the two other dummy variables for light conditions: CR_DARK and CR_DUSK_DAWN.

Table 4. Binary Logistic Regression Results without Socioeconomic Attributes

Variable	Brief Description	Including Travel Speed			Excluding Travel Speed		
		Coeff.	Sig.	Odds Ratio	Coeff.	Sig.	Odds Ratio
NE_DARK	1 if dark	0.740	**0.000**	2.096	1.030	**0.000**	2.800
NE_DUSK_DAWN	1 if dusk or dawn	0.771	**0.000**	2.163	0.943	**0.000**	2.568
NE_DAY	Base: daylight						
NE_GLARE	1 if glare	0.442	0.503	1.555	0.481	0.459	1.617
NE_WTHR_RAIN	1 if rain	-0.141	0.563	0.869	-0.269	0.260	0.764
NE_WTHR_OT	1 if other weather conditions	0.102	0.450	1.107	0.091	0.491	1.096
NE_CLEAR	Base: clear weather						
DR_MALE	1 if male	0.331	**0.002**	1.393	0.336	**0.001**	1.400
DR_AGE15_24	1 if age 15 to 24	-0.276	**0.043**	0.759	-0.168	0.210	0.846
DR_AGE65P	1 if age 65 or over	-0.168	0.257	0.845	-0.195	0.182	0.823
DR_AGE25_64	Base: 1 if age 25 to 64						
DR_BAC_08_	1 if BAC 0.01 to 0.07	2.312	**0.008**	10.096	2.761	**0.001**	15.810
DR_BAC_08P	1 if BAC 0.08 or over	0.521	0.243	1.685	0.912	**0.040**	2.490
DR_BAC_0	Base: BAC=0						
DR_DRUG	1 if drug test positive	2.347	**0.000**	10.440	2.859	**0.000**	17.443
DR_FAIL_YLD	1 if fail to yield right of way	-0.048	0.888	0.953	-0.035	0.920	0.966
DR_CARELS	1 if careless negligent driving	-0.076	0.793	0.927	-0.020	0.943	0.980
DR_OA	1 if other improper actions	0.257	0.226	1.293	0.426	**0.041**	1.531
DR_PROPER	Base: no improper action						
DR_DSTR	1 if distracted	0.100	0.700	1.105	0.149	0.564	1.161
DR_OBST	1 if DR or PED vision obstructed	-0.165	0.470	0.848	-0.172	0.449	0.842
PED_MALE	1 if male	0.056	0.604	1.057	0.098	0.353	1.103
PED_AGE1_10	1 if age 0 to 10	0.454	0.109	1.575	0.391	0.167	1.478
PED_AGE65_74	1 if age 65 to 74	1.211	**0.000**	3.358	1.237	**0.000**	3.444
PED_AGE75P	1 if age 75 or over	2.117	**0.000**	8.304	2.101	**0.000**	8.178
PED_AGE11_64	Base: age 11 to 64						
PED_BAC_08_	1 if BAC 0.01 to 0.07	2.965	**0.000**	19.392	3.366	**0.000**	28.952
PED_BAC_08P	1 if BAC 0.08 or over	4.057	**0.000**	57.807	4.312	**0.000**	74.618
PED_BAC_0	Base: BAC=0						
PED_DRUG	1 if drug test positive	3.694	**0.000**	40.188	4.023	**0.000**	55.882
PED_CROSSING	1 if crossing roadway	0.260	0.071	1.297	0.155	0.275	1.168
PED_W_ALONG	1 if walking along roadway	-0.014	0.983	0.986	0.168	0.496	1.183
PED_ACT_OT	Base: all other actions						
PED_DASH	1 if dart or dash out	0.273	0.287	1.314	0.234	0.350	1.264
PED_OA	1 if any other improper action	0.715	**0.002**	2.045	0.677	**0.003**	1.968
PED_NONE	Base: no improper action						
PED_INT_MRKD	1 if intersection marked crosswalk	-0.004	0.981	0.996	0.073	0.690	1.076
PED_INT_NMRKD	1 if intersection unmarked crosswalk	0.067	0.751	1.069	0.096	0.649	1.101
PED_MID_MRKD	1 if midblock marked crosswalk	-0.066	0.865	0.936	-0.056	0.883	0.945
PED_MID_NMRKD	1 if midblock without crosswalk	0.218	0.072	1.243	0.217	0.069	1.243
PED_LOC_OT	Base: other locations						

Table 4. Binary Logistic Regression Results without Socioeconomic Attributes (Continued)

Variable	Brief Description	Coeff.	Sig.	Odds Ratio	Coeff.	Sig.	Odds Ratio
RD_LANES3_5	1 if 3 to 5 lanes				0.246	0.093	1.279
RD_LANES6P	1 if 6 or more lanes				0.450	**0.011**	1.568
RD_LANES1_2	Base: 1 to 2 lanes						
RD_NOT_LVL	1 if not level	0.532	**0.027**	1.702	0.752	**0.001**	2.122
RD_CURVE	1 if curved	0.973	**0.000**	2.645	1.057	**0.000**	2.877
RD_SPLMT25_	1 if 25 mph or under				-0.608	**0.013**	0.544
RD_SPLMT45P	1 if 45 mph or over				0.569	**0.000**	1.767
RD_SPLMT26_44	Base: 26 to 44 mph						
RD_TWO_DVD	1 if two-way divided				0.324	0.318	1.383
RD_TWO_NDVD	1 if two-way not divided				0.029	0.927	1.030
RD_ONE_WAY	Base: one way						
RD_NO_CTRL	1 if no control device				0.189	0.147	1.208
VEH_SP15_24	1 if 15 to 25 mph	3.260	**0.000**	26.055			
VEH_SP25_34	1 if 25 to 34 mph	2.784	**0.000**	16.182			
VEH_SP35_44	1 if 35 to 44 mph	2.253	**0.000**	9.514			
VEH_SP45_54	1 if 45 to 54 mph	1.435	**0.000**	4.200			
VEH_SP55P	1 if 55 or higher mph	0.479	**0.073**	1.615			
VEH_SP01_14	Base: 1 to 14 mph						
VEH_FRONT	1 if 1, 11, 12 vehicle clock points	1.677	**0.000**	5.350	1.681	**0.000**	5.371
VEH_HDT	1 if medium-heavy duty truck	1.082	**0.000**	2.645	0.946	**0.000**	2.576
VEH_LDT	1 if passenger van, pickup, SUV	0.562	**0.000**	1.754	0.537	**0.000**	1.711
VEH_MTCYLC	1 if motorcycle or moped	-0.893	0.055	0.409	-0.768	0.096	0.464
VEH_CAR	Base: passenger cars						
VEH_THRU_LN	1 if through lanes				0.081	0.863	1.085
VEH_TURNING	1 if turning				-1.317	**0.010**	0.268
VEH_BACKING	1 if backing				-0.237	0.787	0.789
VEH_ACT_OT	Base: other vehicle actions						
Constant		-7.890	**0.000**	0.000	-7.022	**0.000**	0.001
Observations		9,065					
Odds Ratio for Fatal in Data		0.103					
Chi-Square		2,598			2,477		
Degree of freedom		42			47		
Significance level		<0.001			<0.001		
Nagelkerke R^2		0.539			0.517		

Notes: 1. The dependent variable, PED_FATAL, =1 if fatal injury and 0 for no injury or non-fatal injuries. 2. Sig. in bold indicates significance at the 5% level. 3. Sig. underlined indicates significance at the 10% level.

The bottom of the table shows several summary statistics. These models were estimated with 9,065 crashes, which are about 45 percent of the number of crashes that met the selection criteria of this study described earlier in relation to Table 1. This significant drop in the number of crashes reflects the fact that the crash files used had serious problems

in terms of missing data or data items without valid codes. Most data items varied in the degree of missing or bad codes. Because of the many data items included in the analysis, their joint degree of missing or bad codes is significant. Because the two models differ slightly in the variables used, the available observations normally would differ as well. To keep comparisons easier, the same observations were used for both models.

As indicated by the Chi-Square, degree of freedom, and significance level for each model, each model is significantly better than the constant only model because the independent variables jointly as a whole are statistically significant at a better than 0.001 significance level. In terms of one measure (Nagelkerke R^2) of the variability of the dependent variable in the data, the included set of independent variables explains 53.9% of the variation when the vehicle's travel speed is included and 51.7% without travel speed. There are many ways to measure this variability for logistic regression, and there is no consensus on which one is best. Nagelkerke R^2 is reported here because it is scaled between 0 and 1, just like the traditional R^2 for linear regression, and it happens to be provided by SPSS used for model estimation. The model with the vehicle's travel speed fits the data better both in terms of the Chi-Square value and the degree of variation explained.

The statistical significance of individual independent variables differs little between the two models for their common variables. Only two variables differ in that they are statistically significant at the 5% level in one model but not in the other model.

- DR_BAC_08P, the driver's BAC level at 0.08% or higher. While having a positive coefficient in both models, it is statistically significant at the 5% level when the vehicle's travel speed is excluded but is insignificant in the other model. The reasons for this differential statistical performance is not fully clear. One possibility is that the driver's higher level of BAC may play a big role in predicting his travel speed and it may become largely irrelevant once his travel speed is included as well.
- DR_OA, the driver's contributing action to the crash other than failing to yield right of way and careless negligent driving. The other contributing actions include, running a stop sign or a red light, speeding, improper turn or backing, disregarded other traffic signs or markings, etc. Similar to the case for the driver's BAC level at 0.08% or

higher, this variable is significant in the model without the vehicle's travel speed but is insignificant in the other model. The same argument made in the case of the driver's BAC level being 0.08% or higher may be made here to explain this difference between the two models.

The statistical significance of independent variables, however, varies dramatically within each model. It is common for a regression model in a publicly released study to contain only statistically significant independent variables. One reason for this common practice is that empirical models are typically specified somewhat arbitrarily without a sound analytical framework. This study took a different approach in this regard. Both models include all variables that should be included based on the analytical framework in Figure 1. The results from a significant variable and an insignificant variable can both be informative.

- **Natural Environment**: Relative to daylight conditions, both dark and near dark conditions have highly statistically significant increasing effects at the 5% level on pedestrian's likelihood of dying from a crash. Glare conditions to the driver, however, are not statistically significant. Neither are non-clear weather conditions relative to clear conditions.

- **Driver Attributes**: Being a male driver has a highly statistically significant increasing effect at the 5% level on the pedestrian's chance of being killed relative to female drivers.

  Relative to drivers aged 25 through 64, being a younger driver is statistically significant at the 5% level when the vehicle's travel speed is held constant. At the same time, being a young driver has a decreasing effect on the pedestrian's chance of being killed when the vehicle's travel speed is held constant. It is unclear what may explain this result. However, being an older driver does not have a statistically significant effect on the pedestrian's chance of being killed in either model.

  Relative to zero BAC, the driver's BAC level being below 0.08% is statistically significant at the 5% level. So is the driver's positive drug test result.

Relative to no improper actions by the driver, those improper actions by the driver considered to be contributing to the crash by the investigating officer turned out not to be statistically significant in most cases. Neither are driver distractions or obstructions to their vision. These results may be counter-intuitive. Assuming these improper actions of the driver reflecting the truth, one possibility is that these improper actions of the driver may be significant in explaining the pedestrian being hit by the driver but not significant in explaining the pedestrian's chance of being killed. After all, they were thought by the investigating officer to have contributed to the crash but not necessarily to the specific injury severity of the pedestrian.

The variable capturing the driver's improper actions other than failing to yield to the right of way or careless negligent driving, however, is an exception when the vehicle's travel speed is not held constant. Without the vehicle's travel speed being held constant, the driver's other improper actions tend to increase the vehicle's travel speed, which in turn increases the vehicle's impact speed and the pedestrian's chance of being killed. When the vehicle's travel speed is given, on the other hand, these other improper actions do not directly affect the vehicle's impact speed and hence have no significant effect on the pedestrian's chance of being killed.

- **Pedestrian Attributes**: Different from being a male driver, being a male pedestrian is not statistically significant to the pedestrian's chance of being killed.

  While being a child pedestrian aged 10 or under is not significant relative to pedestrians aged 11 to 64, being an older pedestrian aged 65 or over is highly statistically significant to the pedestrian's chance of being killed. It is not clear why being child pedestrians is not statistically significant in this case. However, for older pedestrians, this result most likely comes from them being more physically fragile.

  Like drivers, having a positive test result for either alcohol or drug is statistically significant for pedestrians. Unlike drivers, the level of statistical significance for pedestrians is much higher.

  What the pedestrian was doing prior to the crash (e.g., crossing vs. walking along a road) is not statistically significant at the 5% level relative to all other actions,

probably due to them having similar effects as all other actions on the injury outcome. At the 10% level (actually at 5.4%), however, crossing a roadway is statistically significant relative to all actions other than crossing or walking along a roadway when the vehicle's travel speed is held constant.

Relative to no improper action by the pedestrian, it is a surprise that dash or dart actions (PED_DASH) are not significantly significant, but other improper contributing actions as a group (PED_OA) are significant. While not separately reported due to difficulties in interpreting the estimated results, the product of PED_DASH and NE_DARK (e.g., dark relative to daylight conditions) when replacing PED_DASH is statistically significant at the 10% level in both models. The pedestrian's dash and dart out actions can significantly affect the pedestrian's injury outcome under dark conditions but not necessarily under other light conditions.

Relative to other locations, being in a marked crosswalk (PED_INT_MRKD or PED_MID_MRKD) tends to lower the pedestrian's chance of being killed in both models. The effects, however, are not statistically significant. Being in an unmarked intersection crosswalk (PED_INT_NMRKD) or a midblock location without a crosswalk (PED_MID_NMRKD) tends to raise the pedestrian's chance of being killed. In the case of being in an unmarked intersection crosswalk, the effect is not statistically significant. For being in a midblock location without a crosswalk, however, the effect is statistically significant at the 10% level in both models.

- **Roadway Attributes**: Both roadway grade and alignment are statistically significant at the 5% level in the pedestrian's chance of being killed. These apply to both models.

  The other attributes apply to the model without the vehicle's travel speed. Relative to roadway width with 1 to 2 lanes, having 3 to 5 lanes is significant at the 10% level but having 6 or more lanes is statistically significant at the 5% level. Relative to a posted speed limit from 26 to 44 mph, having a lower speed limit has a statistically significant decreasing effect at the 5% level on the pedestrian's chance of being killed but having a higher speed limit has a statistically significant increasing effect. Relative to a one-way roadway, being a two-way roadway is not statistically

significant, nor is whether control device was present. Once roadway width and the posted speed limit are held constant, traffic way features (i.e., one-way vs. two-way) and presence or absence of traffic control devices do not have additional significant correlational effects on the vehicle's travel speed.

Vehicle Attributes: Relative to the vehicle's travel speed under 15 mph, all five related dummy variables have a statistically significant increasing effect on the pedestrian's chance of being killed with the significance level at 10% if it is from 15 mph to 24 mph and at 5% level for all other four dummy variables.

Being hit by the front of a vehicle (relative to the sides or the back) or by either a light-duty truck (e.g., passenger van) or a heavy-duty truck (relative to passenger cars) has a highly statistically significant increasing effect on the pedestrian's chance of being killed. Being hit by a motorcycle or moped has a statistically significant decreasing effect at the 10% level on the pedestrian's chance of being killed in both models. However, the decreasing effect of being hit by a motorcycle or moped is greater when the vehicle's travel speed is included. For a given vehicle's travel speed, speed can be lowered faster for a motorcycle or moped due to its smaller mass than a passenger car; and the impact configuration on the hit pedestrian is much smaller for a motorcycle or moped than for a passenger car. When the vehicle's travel speed is not included in a model and hence is not necessarily the same between a motorcycle and a passenger car, these potential advantages of a motorcycle over a passenger car may be lower, potentially because the vehicle's travel speed typically is greater for motorcycles or mopeds than for passenger cars.

Relative to the vehicle's other maneuver actions prior to the crash, moving in through lanes or backing is not statistically significant in the pedestrian's chance of being killed. Making turns, including left turns, right turns, and U-turns, however, has a statistically significant decreasing effect at the 5% level. These variables only show up in the model without the vehicle's travel speed.

Partially due to how the dummy variables were defined for each given data item, almost all variables that are statistically significant have a positive coefficient. As already mentioned, there are three exceptions: 1) when the vehicle was a motorcycle or moped; 2) when the vehicle was turning; and 3) the posted speed limit is 25 mph or lower. With the above discussions, no further discussion is needed on how these significant variables may impact the pedestrian's chance of being killed. What is important to emphasize is the numerical significance of the effects of these statistically significant variables.

With logistic regression, it is difficult to interpret the estimated coefficients directly. It is common to look at the estimated odds ratio, i.e., the ratio of the pedestrian's odds of being killed over its odds of not being killed. As mentioned earlier, the value of this odds ratio is 10.3% in the data used, i.e., once hit by a vehicle, the odds of a pedestrian being killed among the 9,056 crashes in the analysis data set was about 10.3% of the odds of not being killed. If one labels the pedestrian's gender by g (= m for male and f for female) and the pedestrian's probability of being killed once being hit by P_g, the pedestrian's probability of not being killed would be $1\text{-}P_g$, and the pedestrian's odds of being killed then would be the ratio of these two probabilities: $O_g = P_g/(1\text{-}P_g)$. To compare a male pedestrian's odds of being killed with a female pedestrian's odds of being killed, it is convenient to use the ratio of their odds: Odds Ratio = O_m / O_f. If this odds ratio happens to be 2, for example, then the pedestrian's odds of being killed for males is twice that for females.

The following discussion focuses on dummy variables that are not related to the vehicle's travel speed and have a statistically significant positive effect on the pedestrian's chance of being killed once being involved in a collision. Instead of trying to interpret the coefficient of every one of these variables, the discussion focuses on those with the largest numerical effect in terms of the estimated value of their odds ratio and they are discussed in a declining order of their odds ratio as summarized in Table 5. They are ranked according to their odds ratio in the model without the vehicle's travel speed.

Table 5. Odds Ratio for Select Dummy Variables

Variable		Including Travel Speed	Excluding Travel Speed
PED_BAC_08P	1 if BAC 0.08 or over	57.807	74.618
PED_DRUG	1 if drug test positive	40.188	55.882
PED_BAC_08_	1 if BAC 0.01 to 0.07	19.392	28.952
DR_DRUG	1 if drug test positive	10.454	17.443
DR_BAC_08_	1 if BAC 0.01 to 0.07	10.096	15.810
PED_AGE75P	1 if age 75 or over	8.304	8.178
VEH_FRONT	1 if 1, 11, 12 clock points on vehicle	5.350	5.371
PED_AGE65_74	1 if age 65 to 74	3.358	3.444
RD_CURVE	1 if curved	2.645	2.877
CR_DARK	1 if dark	2.096	2.800
VEH_HDT	1 if medium-heavy duty truck	2.950	2.576
CR_DUSK_DAWN	1 if dusk or dawn	2.163	2.568
DR_BAC_08P	1 if BAC 0.08 or over		2.490
RD_NOT_LVL	1 if not level	1.702	2.122
PED_OA	1 if any other improper action	2.045	1.968
RD_SPLMT45P	1 if 45 mph or over		1.767
VEH_LDT	1 if passenger van, pickup, SUV	1.754	1.711
RD_LANES6P	1 if 6 or more lanes		1.568
DR_MALE	1 if male	1.393	1.400
RD_LANE3_5	1 if 3 to 5 lanes		1.279
PED_MID_NMRKD	1 if midblock without crosswalk	1.243	1.243

Note: taken from Table 4.

The top five variables in this ranked list are all related to alcohol or drug use by the driver or the pedestrian. Regardless of the inclusion of the travel speed of the vehicle, these variables have dramatic effects on the pedestrian's chance of being killed. The following highlights some of the observations for the model with the vehicle's travel speed:

- The odds for a pedestrian being killed once being hit by a vehicle if the pedestrian's BAC level is at 0.08% or higher is about 58 times that if their BAC level is zero. Even for a pedestrian with a BAC level below 0.08%, the odds of being killed once being hit is still 19 times that if his BAC level is 0.
- Substance abuse by the pedestrian has a much greater effect on the pedestrian's chance of being killed than abuse by the driver. The odds ratio for a positive drug test

is about 40 for the pedestrian but is about 10 for the driver. The odds ratio for a BAC level below 0.08% is about 19 for the pedestrian but is about 10 for the driver.

- Being a senior pedestrian dramatically increases the odds of being killed. When the vehicle's travel speed is held constant, the odds for a pedestrian being killed if they are 75 or older is about 8 times that of a pedestrian's odds at 11 to 64 years of age. Further, the odds of being killed is significantly higher for older seniors than for younger seniors. In fact, the odds ratio for the pedestrian being 75 or older is 2.47 (=8.304/3.358) times that if they are 65 to 74 years of age.
- Which part of the vehicle hit the pedestrian is much more important in determining the pedestrian's risk of death than what body type the vehicle was. The odds of a pedestrian being killed by the front of a vehicle is 5.350 times that by the sides or the back of the vehicle. In comparison, the risk in terms of odds ratio is lower at 2.950 when the pedestrian is hit by a heavy-duty truck relative to by a passenger car. The risk is in terms of odds ratio is even lower at 1.754 when the pedestrian hit by a light-duty truck relative to by a passenger vehicle.

Besides the above highlights, Table 5 also provides several important insights on the determinants of the pedestrian's chance of being killed once being hit by a vehicle. When the vehicle's speed is held constant:

- The pedestrian's odds of being killed on curved sections is 2.645 times that on straight sections.
- Relative to daylight conditions, the pedestrian's odds of being killed is 2.096 times as large under dark conditions and is 2.163 times as large under dusk/dawn conditions.
- Grade sections can raise the pedestrian's odds of being killed by 70.2% relative to flat sections.
- Being hit by a male driver raises the pedestrian's odds of being killed by 39.3% relative to by a female driver.

When the vehicle's speed is not held constant:

- Having a speed limit at 45 mph or higher can raise the pedestrian's odds of being killed by 76.7% relative to having a speed limit at 25 mph or lower.

- Relative to roadway sections with 1 to 2 lanes in both directions, the pedestrian's odds of being killed is 27.9% higher on sections with 3 to 5 lanes and 56.8% higher on sections with at least 6 lanes in both directions.

One important consideration was the numerical significance of increases in the vehicle's travel speed on the pedestrian's chance of being killed once being hit. Figure 3 shows this numerical significance in terms of how fast the pedestrian's odds of being killed increases relative to speeds under 15 mph. Relative to speeds under 15 mph, the pedestrian's odds of being killed becomes 1.615 times as large when the vehicle's travel speed reaches 15 to 24 mph, 4.200 times as large when it reaches 25 to 34 mph, 9.514 times as large when it reaches 35 to 44 mph, 16.182 times as large when it reaches 45 to 54 mph, and 26.055 times as large when it reaches 55 mph or higher.

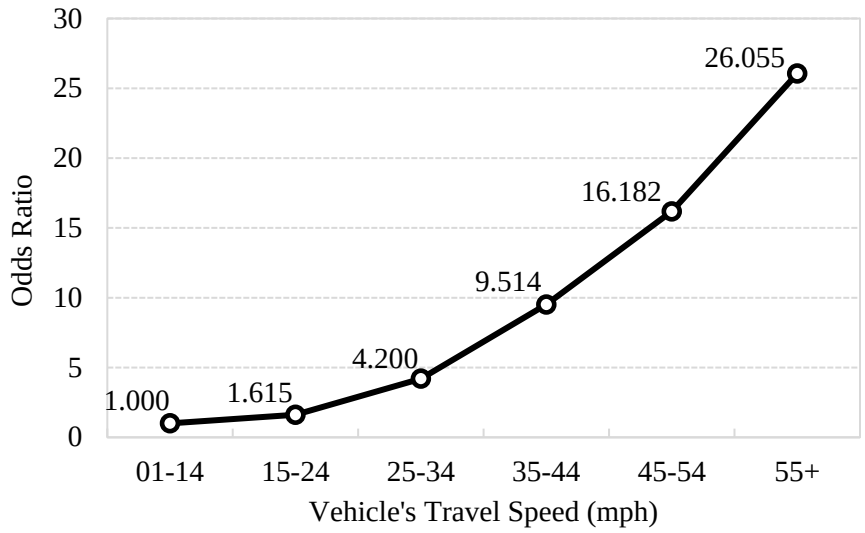

Figure 3. Pedestrian's Odds of Being Killed Relative to Odds for Speeds 01-14 MPH

While not being reported in Table 4, this study also estimated an alternative specification of the model with the vehicle's travel speed included. Instead of the vehicle's travel speed being entered as a range of dummy variables, the alternative specification included it as a continuous variable. This alternative specification did not fit the data as well as the specification reported in Table 4. With this alternative specification, the vehicle's

travel speed had an odds ratio of 1.067, indicating that increasing the vehicle's travel speed by 1 mph would increase the odds of the pedestrian being killed by 6.7%. With some additional computation, one can find out that the odds of the pedestrian being killed would double from increasing the vehicle's travel speed by 10.7 mph.

With this alternative specification, one can look at the effects of the vehicle's travel speed directly on the pedestrian's probability of being killed once being hit. Unlike the odds of the pedestrian being killed once being hit, the pedestrian's probability of being killed also depends on all other variables in the model. Assuming each of all other independent variables taking its mean value across all its available observations, Figure 4 shows the derived relationship from the model on how the vehicle's travel speed affects the pedestrian's probability of being killed once being involved.

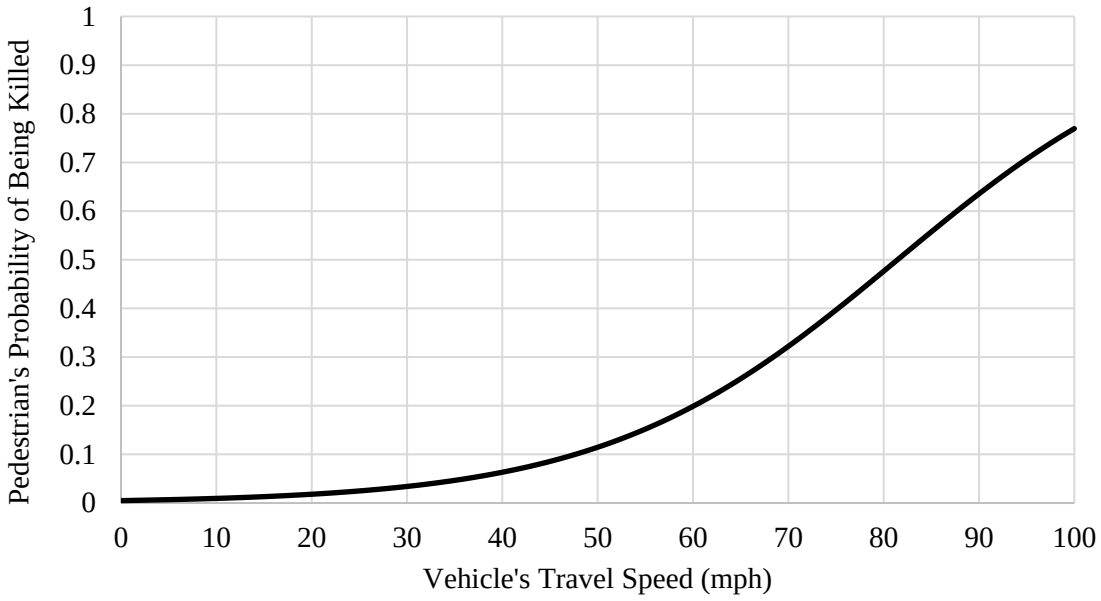

Figure 4. Pedestrian's Probability of Being Killed by Vehicle's Travel Speed

One objective of this study was to test any role of the socioeconomic environment surrounding a crash site in the probability of pedestrians being killed once involved. The models in Table 4 do not include any socioeconomic variables. After these two models were estimated, these socioeconomic variables were tested individually and collectively in each of

these two models. No socioeconomic variables turned out to be statistically significant. As a result, no estimation results are separately reported here. This negative outcome on the hypothesized role of the socioeconomic environment on pedestrian injury severity is somewhat surprising. While most studies of pedestrian injury severity do not consider the socioeconomic environment, the few studies that do have shown some statistically significant results. There are several possible reasons for the lack of statistical significance.

- The presumption of significance is premised on the expectation that the pedestrian and/or the driver share the socioeconomic characteristics of the geography in which the crash occurred. While such a hypothesis might be logical for many walk trips, particularly those in residential areas, it is far less likely that the geographic socioeconomic characteristics around the crash location are shared by the driver.
- In addition, some pedestrian crashes occur not at the residential/home end of the trip but rather might be in a place of employment or destination where the pedestrian's characteristics may be dissimilar to those of the local geography. It should be noted here that the tract-level socioeconomic attributes from the Census do characterize the residential areas of the population.
- Additionally, social demographic characteristics may be more influential in determining the probability of the pedestrian being involved in the crash versus determining the probability of being killed once being involved.

CONCLUSIONS

Using data on pedestrian-vehicle crashes that occurred in urbanized areas of Florida during 2011-2015 and spatially matched socioeconomic data at the Census tract level from the 2011-2015 American Community Surveys, this study attempted to statistically identify and determine the role of attributes of the roadway system, of the vehicles and users of this system, of the natural environment (e.g., light and weather conditions), and of the socioeconomic environment in a pedestrian's chance of being killed once being hit by a vehicle.

Before discussing what this study has learned, and has not learned, about the determinants of the probability of pedestrian deaths once being involved in vehicle crashes, it is important to point out several issues for properly understanding and interpreting the results reported in this study.

- The results are conditional on the fact that the pedestrian in question was involved and hit by the vehicle in question. This study did not try to qualify what was critical in determining the pedestrian's involvement in the crash. Nor did it address the factors that were critical in determining both the pedestrian's initial involvement in the crash and their injury severity from that initial involvement. Rather, the question being addressed here was, once being hit by a vehicle, what were critical factors in determining the pedestrian's odds of being killed from that collision.
- For a given estimated model, the coefficient of an independent variable, and the derived odds ratio, reflect the isolated effect of this independent variable on the pedestrian's chance of being killed while all other independent variables also included in the same model are being held constant.

Understanding and interpreting the reported results should also keep in mind this study's limitations. Three study limitations include:

- The final analysis relied on a relatively small portion of available crashes. There were about 24,000 pedestrian crashes that occurred on non-interstate or toll roads that involved only a single vehicle and a single pedestrian. Limiting to single-vehicle and

single-pedestrian crashes made linking vehicle and pedestrian data to crashes cleaner. However, the linking process reduced the number of crashes by about 3,500 due to hit-and-run crashes and crashes with no pedestrian information. More seriously, high degrees of missing data and invalidly coded data further reduced the number of crashes to about 9,000 that were used in the final analysis. If the approximately 15,000 crashes thrown out differed in a systematic way from the 9,000 crashes used, unintentional bias could have been introduced into the data and the results.

- Without taking complex econometric techniques to deal with potential econometric issues in the crash data, this study essentially assumed that the data used necessarily satisfy the conditions of the binary logistic regression model. To the degree this assumption is not entirely valid, some of the reported results could be altered if the related econometric issues were explicitly addressed.
- Using data of the socioeconomic attributes of the residents living in tracts of crash location from the American Community Survey (ACS), this study assumed that these socioeconomic attributes can reasonably represent those attributes of the pedestrians and drivers passing through these tracts. The degree to which this assumption may be valid depends highly on the mix of land uses in these tracts. For a given mix of land uses, this assumption is more likely to be true for pedestrians than for drivers.

From this study, we gained several important insights about what factors may not be statistically significant determinants of the probability of pedestrians dying once being involved in vehicle crashes. Some of these may well be important to the probability of pedestrians being involved in vehicle crashes, but they are not so significant to the probability of pedestrian deaths in this study:

- The socioeconomic environment as measured with ACS data at the tract level was not found to play any statistically significant role. There are several possible reasons for the lack of significance. The presumption of significance is premised on the expectation that the pedestrian and/or the driver share the socioeconomic characteristics of the geography in which the crash occurred. While such a hypothesis might be logical for many walk trips, particularly those in residential areas, it is far less likely that the geographic socioeconomic characteristics around the crash location

are shared by the driver. In addition, some pedestrian crashes occur not at the residential home end of the trip but rather might be in a place of employment or destination where the pedestrian's characteristics may be dissimilar to those of the local geography. It should be noted here that the tract-level socioeconomic attributes from ACS characterize the residential areas of the population. Additionally, social demographic characteristics may be more influential in determining the probability of the pedestrian being involved in the crash versus determining the probability of the chance of being killed.

- Attributes of the natural environment, including glare to the driver and non-clear weather conditions, are not found to have any statistically significant effect on the pedestrian's chance of being killed.
- Relative to drivers aged 25 through 64, drivers 65 years of age or older are not found to be statistically significantly correlated to the pedestrian's chance of being killed once involved in a collision.
- Relative to no improper actions, those driver actions that were thought to have contributed to the crash by the investigating officer do not play any statistically significant role. These improper actions include failing to yield right of way, careless negligent driving, etc.
- Male pedestrians are not found to be more likely to be killed than female pedestrians.
- Relative to pedestrians aged 11 to 64, younger pedestrians were not found to be more likely to be killed once being hit by a vehicle.
- Relative to no improper actions, the pedestrian's dash or dart out actions, regardless of light conditions, were not found to be statistically significant in the chance of being killed. However, the pedestrian's dash or dart out actions under dark conditions were likely to be statistically significant.
- Besides the dash or dart out pedestrian actions, other pedestrian actions that were thought to have contributed to the crash by the investigating officer, were found to be statistically significant in the pedestrian's chance of being killed.
- Relative to other locations of the pedestrian, being in an intersection crosswalk (marked or unmarked) or in a midblock crosswalk was not found to play a statistically

significant role in the pedestrian's chance of being killed. Being in a midblock location without a crosswalk, however, was found to have a statistically significant increasing effect on the pedestrian's chance of being killed.

This study also gained insights on what factors are likely to be important in the pedestrian's chance of being killed once being involved in a collision. As expected, one of these factors was the vehicle's travel speed prior to any maneuver actions. Relative to speeds under 15 mph, the pedestrian's odds of being killed becomes 1.615 times as large when the vehicle's travel speed reaches 15 to 24 mph, 4.200 times as large when it reaches 25 to 34 mph, 9.514 times as large when it reaches 35 to 44 mph, 16.182 times as large when it reaches 45 to 54 mph, and 26.055 times as large when it reaches 55 mph or higher.

When entered as a continuous variable, the vehicle's travel speed has an odds ratio of 1.067, indicating that increasing the vehicle's travel speed by 1 mph would increase the odds of the pedestrian being killed by 6.7%. With some additional computation, one can find out that the odds of the pedestrian being killed would double from increasing the vehicle's travel speed by 10.7 mph.

Other than the vehicle's travel speed, all other independent variables in the estimated models were dummy variables (e.g., taking 1 for male and 0 for female). Among those dummy variables found to have a statistically significant increasing effect on the pedestrian's chance of being killed, the top five variables with the highest numerical impacts are all related to substance abuse by the driver or the pedestrian. The following are some of the results when the vehicle's travel speed is being held constant:

- The odds for a pedestrian being killed once being hit by a vehicle if the pedestrian's BAC level is at 0.08% or higher is about 58 times that if their BAC level is zero. Even for a pedestrian with a BAC level below 0.08%, the odds of being killed once being hit is still 19 times that if their BAC level is 0.
- Substance abuse by the pedestrian has a much greater effect on the pedestrian's chance of being killed than abuse by the driver. For example, the odds ratio for a positive drug test is about 40 for the pedestrian but is about 10 for the driver. The odds ratio for a BAC level below 0.08% is about 19 for the pedestrian but is about 10 for the driver.

- Being an elder pedestrian dramatically increases the odds of being killed. When the vehicle's travel speed is held constant, the odds for a pedestrian being killed if the pedestrian is 75 or older is about 8 times that if he is 11 to 64 years of age. Further, the odds of being killed is significantly higher for older elderlies than for younger elderlies. In fact, the odds ratio for the pedestrian being 75 or older is 2.47 times that if the pedestrian is 65 to 74 years of age.
- Which part of the vehicle hit the pedestrian is much more important in determining the pedestrian's risk of death than what body type the vehicle was. The odds of a pedestrian being killed by the front of a vehicle is 5.350 times that by the sides or the back of the vehicle. In comparison, the risk in terms of odds ratio is lower at 2.950 when the pedestrian is hit by a heavy-duty truck relative to by a passenger car. The risk is in terms of odds ratio is even lower at 1.754 when the pedestrian is hit by a light-duty truck relative to by a passenger vehicle.
- The pedestrian's odds of being killed on curved sections is 2.645 that on straight sections.
- Relative to daylight conditions, the pedestrian's odds of being killed is 2.096 times as large under dark conditions and is 2.163 times as large under dusk/dawn conditions.
- Grade sections can raise the pedestrian's odds of being killed by 70.2% relative to flat sections.
- Being hit by a male driver raises the pedestrian's odds of being killed by 39.3% relative to by a female driver.

This study, while providing insights into the factors that contribute to fatalities given a pedestrian-vehicle accident, does not appear to fully account for the serious problem of pedestrians in Florida and other Southeastern states having a significantly higher probability of dying once being involved in vehicle crashes. For example, we still do not know whether and how some of the estimated effects of the various attributes of the transportation system and the natural environment for Florida may differ for other states and between the Southeastern and other regions. To help fill this knowledge gap, potential further research could extend the current study to a boarder geography. To the degree that the socioeconomic

environment in fact does not play a significant role beyond the probability of having an accident, one possibility would be to use the national representative sample of crashes of all severity in the CRSS (formally GES) of the National Highway Traffic Safety Administration. The current study did not choose CRSS data primarily because of the need to match Census tract level socioeconomic data with the crash data through the location coordinates of each crash. Without having location coordinates, it was not feasible to use CRSS data for the current study. Using CRSS data can have several advantages. The data are likely to be much cleaner in terms of a lower degree of missing data and invalid data. It is possible to test how some of the results reported here may differ across the four Census regions (Northeast, Midwest, South, and West). Being a sample across the nation rather than a census in a state, issues like spatial correlation among crashes are much less likely to occur.

Another potentially beneficial direction of further research relates to estimating the impact of substance abuse by pedestrians on their chance of being killed once being involved. When included as independent variables in modeling pedestrian injury severity, substance abuse typically has been found to be both statistically and numerically significant. The current study included both alcohol and drug abuses for both drivers and pedestrians in explaining the probability of a pedestrian being killed versus not being killed once being involved through a binary regression analysis. The most significant result of the current study is the substantially greater impact of substance abuse by pedestrians than by drivers. If true, the single most profound outcome of the study, in the opinion of the authors, is the need to recognize and address the importance of alcohol and substance presence in the pedestrian victims of pedestrian-vehicle accidents and direct attention to education and enforcement to mitigate this situation. However, this result from the current study differs from most of the previous studies in the literature in several ways and its sensitivity should be tested.

1. **Specification of Independent Variables** – Some of the difference likely may have resulted from differences in whether and how substance abuse is included for analysis. Previous studies vary widely on the specification of substance abuse as independent variables. Some do not consider substance abuse at all for either drivers or pedestrians (Aziz, et al., 2013; Islam and Jones, 2014; Haleem, et al., 2015; Olszewski, et al., 2015). While most studies do consider substance abuse, they still

vary in many ways. Some consider substance abuse by drivers only (Kim, et al., 2010; Kwigizile, et al., 2011; Kim, et al., 2017) or by pedestrians only (Miles-Doan, 1996; Lee and Abdel-Aty, 2005). Some consider alcohol but not drug abuse (Kim, et al., 2010; Tay et al., 2011; Abay, 2013; Pour-Rouholamin and Zhou, 2015). When considering both alcohol and drug abuses, some treat them jointly in a single variable (Lee and Abdel-Aty, 2005; Xin et al., 2017). Few consider substance abuse by both drivers and pedestrians (Kim et al., 2008; Moudon et al., 2011; Abay, 2013). Part of the future research on the impact of substance abuse would be to test how alternative specifications of substance abuse as independent variables may influence the estimated results.

2. **Specification of the Dependent Variable** – Another part of the difference between the current study and previous studies in the literature may have resulted from the nature of the dependent variable defined. The current study defined the dependent variable as a binary one with 1 for the pedestrian being killed and 0 for not being killed. However, most previous studies that consider substance abuse by pedestrians define the dependent variable as an ordered multi-outcome variable based on the injury levels available in the crash data. A few of the previous studies that consider substance abuse by pedestrians also use a binary dependent variable but define it differently from the current study. Instead of defining it as fatal vs. non-fatal as in the current study, Moudon et al. (2011) define it as high-injury (fatal and incapacitating) vs. low-no injury (all other severity levels). Part of the future research on the impact of substance abuse would be to test how alternative specifications of the dependent variable may influence the estimated impact of substance abuse by pedestrians on their chance of being killed.
3. **Nature of Substance Abuse Data** – A more significant reason for the dramatic impact of pedestrian substance abuse found in the current study on their chance of being killed may have something to do with the nature of crash data on substance abuse. The current study used testing results for both alcohol and drug uses to measure substance abuse. To have a test result for a given substance and a given person involved in a vehicle-pedestrian crash involves at least two steps: 1) someone

must request a test; and 2) the test is successfully conducted and recorded. It appears that a test is almost always requested for anyone who was fatally injured but is much less likely to be requested for non-fatally injured persons. In addition, fatally injured persons cannot reject a test, but non-fatally injured persons frequently reject a request for such a test. As a result, substance abuse by the pedestrian may affect his chance of being killed and the pedestrian being killed also affects the likelihood of him having a test result.

This is a form of endogeneity problem where independent variables are potentially influenced by in-injury severity outcomes. As Savolainen et al. (2011) point out, the crash-injury severity literature has only begun to scratch the surface on this important issue. While the differential treatment of testing substance abuse probably is true for both pedestrians and drivers, but the effect on whether a person has a test result in the crash data differs substantially between pedestrians and drivers, especially for crashes with a single vehicle and a single pedestrian. The driver in a crash with a single vehicle and a single pedestrian rarely dies from the crash. Consequently, the nature of the data on substance abuse as used in this study does not have much effect for drivers. The specification of the dependent variable, however, can have significantly affect how the nature of substance abuse data influences the estimated impacts of substance abuse on the pedestrian's chance of being killed once being involved.

One aspect of future research in relation to this endogeneity problem would be to use crash data from entities that follow more objective procedures for requesting and testing for substance abuse across different severity levels. Another aspect would be to test the sensitivity of the results by defining binary dependent variables differently: fatal vs. non-fatal, fatal or incapacitating vs. all other levels, fatal vs. incapacitating, etc. One may also deal with the endogeneity problem directly through econometric techniques by simultaneously modeling injury severity and availability of test results for substance abuse.

ACKNOWLEDGEMENTS

The authors are thankful for support received from the Southeastern Transportation Center (United States Department of Transportation grant number DTRT13-G-UTC34). The authors greatly appreciate the sponsorship of the Office of Secretary of Transportation. Benjamin Jacobs, Crash Records and Research Administrator of Florida Department of Transportation's Safety Office, provided the crash data for this study. Aditya Sharma of the USF Department of Computer Science and Engineering matched the American Community Survey data to the pedestrian crash data used in the study.

REFERENCES

Abay, K.A., "Examining pedestrian-injury severity using alternative disaggregate models", *Research in Transportation Economics* 43 (2013): 123-136.

Aguero-Valverde, J. and P.P. Jovanis (2006). Spatial Analysis of Fatal and Injury Crashes in Pennsylvania. *Accident Analysis and Prevention* 38: 618-625.

Al-Ghamdi, A.S., "Pedestrian–vehicle crashes and analytical techniques for stratified contingency table," *Accident Analysis and Prevention* 34 (2002): 205–214.

Aziz, H.M.A., S.V. Ukkusuri, and S. Hasan, "Exploring the determinants of pedestrian-vehicle crash severity in New York City," *Accident Analysis and Prevention* 50 (2013): 1298-1309.

Ballasteros, M., P. Dischinger, P. Langenberg, Pedestrian injuries and vehicle type in Maryland, 1995–1999. *Accident Analysis and Prevention* 36 (2004): 73–81.

Chu, X., S. Polzin, and J. Godfrey, *Exploration of Socio-Demographic Characteristics and Culture as Factors in Differential Safety Performance Across Geography*, Southeastern Transportation Center, Knoxville, Tennessee. 2017.

Department of Highway Safety and Motor Vehicles, *Instructions for Completing the Florida Uniform Traffic Crash Report Forms*, HSMV 90010S, 2015.

Florida Department of Transportation, *Frequently Asked Questions*, 2010. https://fdotewp1.dot.state.fl.us/TrafficSafetyWebPortal/publications.aspx.

Fontaine, H. and Y. Gourlet, "Fatal pedestrian accidents in France: A typological analysis," *Accident Analysis and Prevention* 29 (1997): 303-312.

Gårder, P.E., "The Impact of speed and other variables on pedestrian safety in Maine," *Accident Analysis and Prevention* 36 (2004): 522-542.

Haleem, K., P. Alluri, and A. Gan, "Analyzing pedestrian crash injury severity at signalized and non-signalized locations," *Accident Analysis and Prevention* 81 (2015): 14-23.

Islam, S., and S.L. Jones, "Pedestrian at-fault crashes on rural and urban roadways in Alabama," *Accident Analysis and Prevention* 72 (2014): 267:276.

Jensen, S.U., "Pedestrian safety in Denmark," *Transportation Research Record* 1674 1999): 61-69.

Kim, J.K., G. Ulfarsson, V. Shankar, S. Kim, "Age and pedestrian injury severity in motor-vehicle crashes: a heteroskedastic logit analysis. *Accident Analysis and Prevention* 40 (2008), 1695–1702.

Kim, M., S.Y. Kho, and D.K. Kim, "Hierarchical ordered model for injury severity of pedestrian crashes in South Korea," *Journal of Safety Research* 61 (2017): 33-40.

Kwigizile, V., T. Sando, and D. Chimba, "Inconsistencies of ordered and unordered probability models for pedestrian injury severity," *Transportation Research Record* 2264 (2011): 110-118.

LaScala, E.A., F.W. Johnson, and P.J. Gruenewald, "Neighborhood characteristics of alcohol-related pedestrian injury collisions: A geostatistical analysis," *Prevention Science* 2 (2001): 123-134.

Lee, C., and M. Abdel-Aty, "Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida," *Accident Analysis and Prevention* 37 (2005): 775–786.

Lefler, D.E., and H.C. Gabler, "The fatality and injury risk of light truck impacts with pedestrians in the United States," *Accident Analysis and Prevention* 36 (2004): 295-304.

Maciag, M., *America's Poor Neighborhoods Plagued by Pedestrian Deaths*, a Governing Magazine Research Report, Washington, D.C., 2014.

Mannering, F.L., C.R., Bhat, "Analytic methods in accident research: Methodological frontier and future directions," Analytical Methods in Accident Research 1 (2014): 1-22.

McCann, B. and B. DeLille. *Mean Streets 2000: Pedestrian Safety, Health and Federal Transportation Spending*. Surface Transportation Policy Project. 2000.

Moudon, A.V., L. Lin, J. Jiao, P. Hurvitz, and P. Reeves, "The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington," *Accident Analysis and Prevention* 43 (2011): 11–24.

National Highway Traffic Safety Administration, *Literature Review on Vehicle Travel Speeds and Pedestrian Injuries*. U.S. Department of Transportation, Washington, D.C. 1999.

National Highway Traffic Safety Administration, *Traffic Safety Facts 2009*. DOT HS 811 402, U.S. Department of Transportation, Washington, D.C. 2011.

National Highway Traffic Safety Administration, *2016 FARS/CRSS Coding and Validation Manual*. DOT HS 812 449, U.S. Department of Transportation, Washington, D.C. 2017.

National Highway Traffic Safety Administration, *Traffic Safety Facts Annual Report Tables*. https://cdan.nhtsa.gov/tsftables/tsfar.htm, accessed on May 15, 2018. U.S. Department of Transportation, Washington, D.C. undated.

Olszewski, P., P. Szagała, M. Wolański, and A. Zielińska, "Pedestrian fatality risk in accidents at unsignalized crosswalks in Poland," *Accident Analysis and Prevention* 84 (2015): 83-91.

Pitt, R., B. Guyer, C. Hsieh, and M. Malek, "The severity of pedestrian injuries in children: An analysis of the pedestrian injury causation study," *Accident Analysis and Prevention* 22 (1990): 549–559.

Pour-Rouholamin, M., and H. Zhou, "Investigating the risk factors associated with pedestrian injury severity in Illinois," *Journal of Safety Research* 57 (2016): 9-17.

Savolainen, P.T., F.L. Mannering, D. Lord, and M.A. Quddus, "The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives," *Accident Analysis and Prevention* 43 (2011): 1666-1676.

Siddiqui, N., Chu, X., Guttenplan, M., "Crossing locations, light conditions, and pedestrian injury severity," *Transportation Research Record* 1982 (2006): 141–149.

Smart Growth America. *Dangerous by Design*. Washington, D.D. 2017.

Sullivan, J.M., and M.J. Flannagan, "The Role of ambient light level in fatal crashes: inferences from daylight saving time transition," *Accident Analysis and Prevention* 34 (2002): 487-498.

Tay, R., J. Choi, L. Kattan, and A. Khan, "A multinomial logit model of pedestrian–vehicle crash severity," *International Journal of Sustainable Transportation* 5 (2011): 233-249.

United States Census Bureau, *2010 Census Urban Area Reference Maps*, https://www2.census.gov/geo/maps/dc10map/UAUC_RefMap/ua/ua86599_tampa--st_petersburg_fl/DC10UA86599_000.pdf, accessed on May 15, 2018, undated.

Xin, C., R. Guo, Z. Wang, Q. Lu, and P.S. Lin, "The effects of neighborhood characteristics and the built environment on pedestrian injury severity: a random-parameters generalized ordered probability model with heterogeneity in means and variances," *Analytical Methods in Accident Research* 16 (2017): 117-132.

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* 40 (2008): 1320–1329.

Yang, J., *Review of Injury Biomechanics in Car-Pedestrian Collisions*. Crash Safety Divisions, Chalmers University of Technology, Sweden. 2002.

Ye, F., and D. Lord, "Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit, and mixed logit," *Transportation Research Record* 2241 (2011): 51–58.

Zajac S.S., and J.N. Ivan, "Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut," *Accident Analysis and Prevention* 35 (2003): 369-379.