Final Report

Reducing Fatalities and Severe Injuries on Florida's High-Speed Multi-Lane Arterial Corridors

Part II
ANALYSIS OF THE CRASH LEVEL DATA

BD-548-22

Submitted to
The Florida Department of Transportation
Research Center
605 Suwannee Street, MS 30
Tallahassee FL 32399

Submitted By
Mohamed Abdel-Aty, PhD, PE
Anurag Pande, PhD
Chris Lee, PhD
Abhishek Das
Alexis Nevarez
Ali Darwiche
Premchand Devarasetty

College of Engineering and Computer Science
Center for Advanced Transportation Systems Simulation
University of Central Florida
PO Box 162450
Orlando, FL 32816-2450
Phone: (407) 823-5657
Fax: (407) 823-4676
E-mail: mabdel@mail.ucf.edu

April 2009

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
BD-548-22		

4. Title and Subtitle **Reducing Fatalities and Severe Injuries on Florida's High-Speed Multi-Lane Arterial Corridors;** **Part II: Analysis of the Crash Level Data**	5. Report Date April 28, 2009 6. Performing Organization Code 1650-7063
7. Author(s) **Mohamed Abdel-Aty, PhD, PE; Anurag Pande, PhD; Chris Lee, PhD; Abhishek Das, Alexis Nevarez, Ali Darwiche, Premchand Devarasetty**	8. Performing Organization Report No.
9. Performing Organization Name and Address **Center for Advanced Transportation Systems Simulation,** **University of Central Florida,** **P.O. Box 162450,** **Orlando, Fl 32816-2450**	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address **Florida Department of Transportation** **605 Suwannee St. MS 30** **Tallahassee, Florida 32399** **(850)414-4615**	13. Type of Report and Period Covered **Final Report** 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This report presents the analysis conducted to identify the factors that contribute to severe and fatal crash occurrence on multilane corridors. The authors' preliminary investigation using simultaneous ordered probit model provided enough evidence that a fixed influence area of intersections for all of the corridors is not justified. With the increase of an intersection's influence area, crash types that are more specific to segments get included and change the crash pattern for the overall intersection. Therefore, for investigation purposes, to treat the corridors in their entirety will result in much more insightful results than when treating the segments and intersections separately. The crash data were grouped into six major types as follows: 1) rear-end, 2) head-on, 3) angle/turning, 4) sideswipe, 5) crashes involving slow moving vehicles (e.g. cycles, mopeds, etc.), and 6) crashes involving single vehicles. Binary severity classification models were developed by using non-parametric conditional inference trees. Parameters like alcohol/drug use came out to be significant across all crash types and clusters. Lane changing on corridors with high truck traffic was found to be risky from a severity point of view. Poor pavement conditions and high permitted speed limits increased the likelihood of severe rear-end crashes. Non-use of safety equipment also increased the severity level provided the crash had occurred. Based on the results of the overall investigation certain recommendations were made taking the 4 Es (Engineering, Education, Enforcement and Emergency Management) into consideration.

17. Key Word	18. Distribution Statement
Severe crashes, multilane arterials, crash type	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	**Unclassified**	156	

EXECUTIVE SUMMARY

This report presents the analysis conducted to identify the factors that contribute to severe and fatal crash occurrence on multilane corridors. To meet the objective in question the investigators not only made certain innovative changes to the database but also used state of the art analytical methodologies. Before the actual methodologies and database changes were incorporated, certain preliminary investigations were carried out. Since crash data were being analyzed, it became necessary at a certain point this project's work that some of the actual crash reports be investigated individually.

The authors' preliminary investigation using simultaneous ordered probit model provided enough evidence that a fixed influence area of intersections for all of the corridors is not justified. With the increase of an intersection's influence area, crash types that are more specific to segments get included and change the crash pattern for the overall intersection; thus, the very purpose for which the influence area was used gets defeated. Therefore, for investigation purposes, to treat the corridors in their entirety will result in much more insightful results than when treating the segments and intersections separately. CHAPTER 3 looks into the details of the aforementioned investigation. Empirical evidence gathered while examining the individual crash reports suggested that the crashes' *site location* stored in the crash databases does not always reflect the actual ground situation. For the purpose of assigning crashes to the appropriate roadway elements, a set of heuristic rules were developed; CHAPTER 4 details these rules, which were developed by using *site location, traffic control* and *signalized node* information. The corridors were clustered again into four groups based on their length. The analysis that deals

with the identification of significant factors and that examines which categories of independent variables result in a higher proportion of severe crashes is reported in CHAPTER 7.

The crash data were grouped into six major types as follows: 1) rear-end, 2) head-on, 3) angle/turning, 4) sideswipe, 5) crashes involving slow moving vehicles (e.g. cycles, mopeds, etc.), and 6) crashes involving single vehicles. Binary severity classification models were developed by using non-parametric conditional inference trees. Parameters like alcohol/drug use came out to be significant across all crash types and clusters. Lane changing on corridors with high truck traffic was found to be risky from a severity point of view. Poor pavement conditions and high permitted speed limits increased the likelihood of severe rear-end crashes. Non-use of safety equipment also increased the severity level provided the crash had occurred. Presence of a driver/passenger within the vulnerable age group (< 3 years or >55 years) often resulted in an increased severity of injuries. Based on the results of the overall investigation certain recommendations were made taking the 4 Es (Engineering, Education, Enforcement and Emergency Management) into consideration; CHAPTER 8 contains the details on these recommendations.

Disclaimer

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AADT	Annual Average Daily Traffic
CAR	Crash Analysis Reporting System
DOT	Department of Transportation
FDOT	Florida Department of Transportation
FHWA	Federal Highway Administration
HSIS	Highway Safety Information System
NHTSA	National Highway Traffic Safety Administration
RCI	Roadway Characteristics Index

CHAPTER 1. INTRODUCTION

1.1 Background

Improving the safety of arterials, by reducing fatalities and injuries, has been the focus of transportation safety researchers and engineers across the world. Florida has one of the highest traffic fatality rates in the United States. In 2005, 3,543 fatalities occurred on Florida roadways, representing a 9% increase over the previous year. Traffic fatality rates are 26.49 per 100,000 drivers, 21.86 per 100,000 registered vehicles, and 19.92 per 100,000 of the population. The increase in fatalities in the state from 1975 through 2005 is 77% –the highest increase in fatalities among all states in the country–. In fact, the fatality rates within Florida's major cities –such as Jacksonville, Miami, Tampa, Fort Lauderdale and Orlando– are more than 15 fatalities per 100,000 of the population. Among the different road types, principal and minor arterials account for 57% of the total crashes in Florida (NHTSA, 2005). The proportion and sheer number of fatal crashes on Florida's principal arterials (excluding freeways and toll roads) are one of the highest in the nation as reported in 2005. In particular, speeding-related fatalities on arterials with speed limits $\geq$40 mph account for more than 72% of total speeding-related fatalities.

With the statistics just presented, there is a clear need to improve the safety of Florida's arterials –particularly of the ones that are high-speed and multilane–; this can be achieved through a reduction in the number of fatalities and severe injuries at such locations. Severe (i.e. incapacitating and fatal) crashes on arterials occur due to a combination of multiple factors. Therefore, in order to reduce fatalities and severe injuries on these arterials two approaches can be adopted: 1) to study the intersections and roadway segments –excluding the intersections– separately, and 2) to treat these two entity types together as a corridor. The first idea uses the concept of intersections' influence distance for separating the intersection-related crashes from the segment-related crashes. Das et al. (2008) showed by the method of simultaneous estimation that if the influence distance varied the crash characteristics associated with severe injuries also vary; this is due to the fact that the farther we move away from the center of an intersection, more crashes related to the connecting segment come into play. Wang et al. (2008) used frequency modeling for crashes with fixed as well as varying influence distance and found different sets of significant factors. These very recent studies show that the concept of using influence distance for assigning crashes to the roadway elements could be erroneous; however, it is believed that analyzing the crashes along a corridor will facilitate a more realistic identification of significant factors and to better understand the interaction between road design elements and traffic characteristics.

1.2 Tasks of the Project and Flow of the Paper

1.2.1 Tasks of the Project

The tasks carried out by the authors in this report are as follows:

1) Review of studies on crash occurrence at high-speed multilane arterials.
2) Identification of the databases.
3) Crash data analysis.
4) Problem identification and corresponding recommendations applicable to corridors.

1.2.2 Flow of the Report

CHAPTER 2 deals with the literature review of previous arterial road safety studies. CHAPTER 3 includes a preliminary study dealing with simultaneous modeling for both severity and crash location; the study showed that considering a fixed influence area may not be the best approach for studying all corridors. CHAPTER 4 includes the heuristic rules designed for assigning crashes to roadway elements. CHAPTER 5 deals with the data preparation. CHAPTER 6 and CHAPTER 7 deal with all the analysis performed in order to meet the project's objectives. Finally, the general recommendations are detailed in CHAPTER 8.

CHAPTER 2. LITERATURE REVIEW

2.1 Previous Studies

The current chapter summarizes some of the studies relevant to this investigation; particularly, past work relevant to high-speed multilane arterials in general is reviewed, focusing into severe (i.e. incapacitating and fatal) crashes. Under such premise, this literature review has been divided into three sections.

Section 1 presents past studies related to arterials. These studies were not always focused on safety issues but they always required some crash studies to be undertaken (e.g. studies dealing with median design guidelines, etc.). In addition, this section introduces selected and noteworthy work done from the late 1960s to mid 1990s; there were studies related to arterials but not exclusively dealing with the issue of corridor safety. At the end of the section, the overall results are summarized.

Section 2 deals with selected and important work performed on crash prediction models for roadway segments. The significant factors in the models will be discussed, as well as the research done to investigate the factors contributing to severe crashes. Overall, this discussion is essential since the main objective of this investigation is to reduce fatalities and severe injury-related crashes. This section also discusses the important issue of how the researchers' points of view vary when defining a roadway segment (i.e. the criteria to define a roadway segment differ); these discrepancies in the working definition of a segment can lead to confusing inferences. The section ends with a discussion on the results from the past work and on how vast is the problem of roadway segment definition.

Finally, Section 3 discusses selected papers dealing with the issue of corridor safety, especially from the access management point of view. Also, recent research work on signalized intersections, that demonstrates the spatial correlation among them and that they influence each other in several aspects, is also presented. To finalize the section, there is a discussion on the importance to address the safety aspect of the corridor as a whole (i.e. both roadway segments and intersections are considered).

2.1.1 Section 1: Arterial Safety Research

Mulinazzi and Michael (1967) developed crash prediction models for urban arterials; the number of high volume intersections per mile was found to be a significant factor. Walton et al. (1978) built up a regression equation to predict crashes at two-way left turn (TWLT) median lanes in Texas; their study specified that both the number of driveways per mile and area population were contributing factors. Average daily traffic (*adt*) and number of traffic signals per mile were identified as significant factors in both studies.

In his study on design guidelines for raised and transversable medians in Virginia, Parker (1983) found that the number of traffic signals per mile, number of driveways per mile, area population and *adt* had a significant effect on crashes for raised median sections. In an update of this work, Parker (1990) found the same results.

Squires and Parsonson (1989), based on their study of crash comparison of raised median sections and TWLT median lanes in Georgia, established that *adt* and number of traffic signals per mile were important factors.

Bowman et al. (1995) found land use, median width, number of driveways per mile, posted speed limit, as well as crash reporting threshold in dollars, all to be significant factors in crash prediction models for urban or suburban arterials' roadway sections with homogeneity in median type. Though the study included arterial sections with signalized intersections, they did not find the number of signalized intersections along the arterial section to be a significant factor.

Mountain et al. (1996) developed crash prediction models for road networks in seven counties of the U.K. Total two-way annual segment volume, length of the segment and number of minor intersections within the segment proved to be significant factors.

From the results of the studies just discussed, it can be concluded that certain design elements and traffic characteristics play a major role in crash occurrence along arterials. Among design elements, number of traffic signals per mile, number of driveways per mile (i.e. driveway density), median width, as well as length of the segment, all were found significant in most of the work. Regarding traffic characteristics, *adt*, speed limit and annual volume all were found to be the most significant. In addition to the factors just mentioned, other work also found land use and area population to be significant.

2.1.2 Section 2: Crash Prediction Models

Kim et al. (1995) investigated the predictors for crash and injury severity on roadways in Hawaii. Alcohol abuse and non-use of seat belts both were found to be important contributing factors to crash occurrence and to the increase of severe crashes.

O'Donnell and Connor (1996) tried to predict the injury severity of motor vehicle crashes. Their results showed non-use of seat belts, head-on collisions and alcohol abuse as significant factors; also, female drivers were found to be more involved in severe crashes than male drivers.

Bonneson and McCoy (1997) studied roadway segments for investigating the effect median-related treatments have on urban arterial safety. They defined the roadway segment as the section between two consecutive signalized intersections. Furthermore, they focused on segments with a minimum number of vehicles per day, speed limit, number of through lanes and length. Their crash model, which did not include crashes at intersections, had *adt*, segment length, driveway density, unsignalized public street approach density and land use as significant factors.

Milton and Mannering (1998) found section length, *aadt*, percentage of *aadt* occurring during the peak hour, percentage of trucks, speed limit, number of lanes, shoulder width, horizontal curves and tangent length as significant factors contributing to crash frequencies on highway sections, excluding signalized intersections. Delimiters for a section or segment were number of lanes, roadway width, shoulder width, state route number, road type, urban or rural location identifiers, speed, *aadt*, peak hour factors, as well as vertical and horizontal curve characteristics.

Chang and Mannering (1999) analyzed injury severity for non-truck- and truck-involved crashes. For non-truck-involved crashes, driver ejection, driver restrained systems, as well as alcohol impairment, all were found to contribute to fatalities and more severe injuries; on the other hand, truck-involved crashes appeared to be more serious.

Sawalha et al. (2000) examined safety of urban arterial roadway segments, the latter being defined as the part of the arterial between consecutive signalized intersections. A model was developed, which showed traffic volume, segment length, unsignalized intersection density, type of median, number of crosswalks, number of lanes and land use as significant factors.

Hanley et al. (2000) analyzed crash reduction factors on California state highways. The segments were chosen based on *aadt* and it is not very clear from the work whether intersections were included or not. Increments in shoulder width and curve correction with improved radius were found to be significant.

Zhang et al. (2000), through their study of the factors affecting the severity of motor vehicle crashes in Ontario, established that age, disobeying of traffic signs, non-use of seat belts, intersections without traffic control, speed, head-on and turning collisions, as well as overtaking maneuvers, increased the risk of a fatal or severe injury crash. Also, alcohol use and medical/physical condition of elderly drivers significantly increased the risk of fatalities.

Bedard et al. (2002), in their work on causes related to driver fatalities on roadways, found age, alcohol intake, point of impact, non-use of seat belts and speed as significant factors; for example, it was found that older male drivers were more prone to fatal crashes than older female drivers.

Kockelman and Kweon (2002) studied driver injury severity and found that increased driver age, vehicle age, head-on or rollover collision, number of vehicles involved and alcohol use were associated with more severe injuries. Also, female drivers and night time driving were related to increase the injury severity of two-vehicle crashes.

Martin (2002) focused on finding the relationship between crash rate and traffic flow on French interurban motorways. Hourly traffic, day of the week and number of lanes were contributing factors. Night time crashes and crashes occurring under light traffic conditions were found to result in more severe injuries. The roadway sections or segments were homogenous in terms of traffic between two motorway entry points; it is not apparent as to whether the entry points were signalized or unsignalized intersections.

Greibe (2003) developed crash prediction models for urban roads in Denmark where *adt*, land use and speed limit were important factors. Segments and intersections were treated independently; however, if intersections had a low flow rate these were included in the segments. Furthermore, it was not obvious whether the intersections were signalized or unsignalized.

Abdel-Aty (2003) analyzed driver injury safety levels at multiple locations and found driver's age, gender, seat belt use, point of impact, speed, vehicle type, weather condition and area type as major factors. His study also investigated segment and intersection crashes separately. It was found that non-use of seat belts, age, gender, speed, point of impact and alcohol consumption are important factors that contribute to severe injury-related crashes. In addition, crashes occurring on curved segments showed to have a higher probability of resulting in severe injuries. Abdel-Aty and Abdelwahab (2004) also found similar results for injury severity levels in traffic crashes. They also found that female drivers were more probable to be in a severe injury crash than male drivers, and that older people were more likely to be involved in a severe injury crash than younger drivers.

Hiselius (2004) conducted a study of Swedish rural roads that focused on roadway segments without intersections. His segment criteria were traffic flow, speed limit and road width.

In an Illinois county-level data study by Noland and Oh (2004), roadway sections were categorized based on location (urban or rural), cross section (divided or undivided), number of lanes, average median width, average shoulder width, as well as on horizontal and vertical curvature. The study did not indicate whether intersections were considered or not. Increase in number of lanes and in lane width both were found to be associated with higher occurrence of crashes and fatalities; on the other hand, increase in shoulder width resulted in fewer crashes.

Miaou and Song (2005) ranked sites for engineering safety improvements. They analyzed segments and intersections separately; the studied segments had low traffic volume.

Summarizing, the most important design elements from the roadway segment-related studies just mentioned are segment length, driveway density, number of lanes and shoulder width; other important design elements were road width, number of crosswalks, as well as horizontal and vertical curves. Among the most significant traffic characteristics are *adt* and speed limit; in addition to these, other work also showed that the standard deviation of traffic flow and the percentage of different vehicle types were also significant. In addition, land use was found to be significant in some studies.

With regards to factors contributing to fatal and severe injury crashes, it can be observed that more driver-related characteristics are responsible. Design- and traffic-related parameters are not ruled out, but their contribution to those crashes in specific is less. Non-use of seat belts, older driver age, alcohol use, and speeding are found to be significant in most studies dealing with crash severity. Head-on and angle collisions result in more fatalities and severe injuries than any other type of crashes. Some other studies show that crashes occurring at night and under light traffic conditions are more severe. Crash severity is also dependent on the point of impact of the crash, especially the ones hitting from the side. Intersections without any type of control are sites to more severe crashes.

Different researchers have their own view point as to how to define a road segment. A roadway is typically the section of the roadway between two consecutive signalized intersections. In some segment studies, unsignalized intersections have been included. Also, some work mentioned the inclusion of low volume intersections but do not clearly specify whether those are signalized or unsignalized. In most cases, the criteria to select segments to study are speed limit, number of lanes, *adt*, shoulder width and roadway width; other researchers add to the aforementioned selection criteria vertical and horizontal characteristics, road type, as well as urban or rural location for defining a segment. Thus, it can be observed that the vast literature has a confusing definition of segments in crash modeling.

2.1.3 Section 3: Corridor Safety

Jernigan (1999) compared the various corridor safety improvement efforts by Pennsylvania, California and Virginia. He also provided a model strategy for the development of these programs.

Levinson (1999) and Papayannoulis et al. (1999) developed a corridors safety model based on traffic volumes along corridors and access roads as well as access density. Increase in the number of crashes was found to be related to the increase in access density.

Brown and Tarko (1999) also found density of access points, proportion of signalized access points, outside shoulder, TWLT lanes, as well as presence of medians with no openings between signals as significant factors for the safety of urban arterials. They investigated the corridor as a whole.

Abdel-Aty and Radwan (2000) modeled traffic crash occurrence and involvement along Florida’s SR 50 and found *aadt*, degree of horizontal curvature, lane shoulder, median width, urban or rural location and section length to be significant factors. Their section definition included intersections.

Drummond et al. (2002) used a simulation approach for predicting the safety- and operational-related impacts of increased traffic signal density along entire corridors. The major contributing factors were found to be main-line delay, speed limit and stops.

Rees (2003) investigated full corridors in his corridor management studies; he also focused on applying access management treatments along corridors.

Recent signalized intersection modeling work done by Abdel-Aty and Wang (2006) has shown that there is a spatial correlation between crash patterns of successive signalized intersections.

Research on the spatial correlation of crash patterns corresponding to successive signalized intersections shows that there is a need to look at the sequence of signalized intersections along a corridor rather than treating each intersection as an isolated entity. It also has to be noted that intersections are also access points. The access management studies for corridor safety illustrate that both roadway segments and intersections are integral parts of a corridor. Therefore, the corridor should be improved as a whole, considering roadway segments and intersections, in order to significantly reduce fatal and severe crash occurrence.

2.2 Improvement Strategies Implemented by Different States and the Level of Success

Corridor Safety Improvement Programs (CSIPs) were created based on the fact that crashes are likely to occur along joined segments of highways. Some of these joined segments of highways or corridors are known to have a relatively high crash rate. In order to reduce the fatality and injury rate along these corridors it may not be sufficient that only spot improvements are done (Jernigan, 1997); therefore, multidisciplinary cooperation is necessary for achieving major safety and traffic changes on these corridors. This report summarizes the work done for improving safety along high-speed multilane arterials by different states in the U.S. The first example of such improvement task was carried out by the Pennsylvania Department of

Transportation (PennDOT) for U.S. Route 322 which had a series of fatal crashes; the success of this program led to similar work throughout the state, reason for which the Federal Highway Administration (FHWA) encouraged other states to work on similar projects. In 1991, the FHWA issued guidelines for developing a CSIP. The purpose of these guidelines was to establish a leadership-based program for overseeing the work done for improving safety along hazardous corridors. The guidelines had provision to have various agencies involved, to create a multidisciplinary team, to select corridors, to create an action plan, to implement the corresponding recommendations, and to evaluate the effectiveness of the actions taken. The states that carried out a CSIP for their roads, more or less followed these guidelines from the FHWA.

The following discussion focuses on the work done in ten different states across the nation: Pennsylvania, Washington, Virginia, California, Oregon, North Carolina, Kentucky, Arizona, Ohio and Florida. These states were selected for this report's purposes as relevant and substantial information from various sources could be gathered on their projects and/or work.

2.2.1 Pennsylvania

The pilot project in Pennsylvania's safety improvement program was targeted towards U.S. Route 322 in Delaware County, a high-volume and high-speed highway. This route was chosen on the behest of then Pennsylvania Governor Robert P. Casey after the occurrence of a crash on this corridor in 1988, resulting in multiple fatalities. The plan was successfully implemented in a period of six months, and the typical corridor safety problems could be identified for the selected corridor. Among the various countermeasures, highway design improvements, educational media programs and enforcement efforts for improving driver performance, as well as commercial truck safety inspections were the most important ones (Zogby et al., 1991). Emergency medical assistance along the designated corridor was also improved. The corridor had 40% less number of crashes three years after the improvements were implemented (Jernigan, 1999). Later, 55 corridors (totaling 880 miles of highway) were selected for the safety initiative; these sections not only accounted for 7% of the total fatalities but also for the maximum concentration of severe crashes per mile. The PennDOT, the state's Department of Health, as well as the state and local police, all worked in synergy. The improvements were applied over the entire length of the section and thus improved the overall safety along this path (Zogby et al., 1991). In 2002, Pennsylvania House Bill 2410 came into effect, allowing for fines to be doubled on the designated safety corridors; however, the safety effect of the bill has not yet been established.

2.2.2 Washington

Soon after the success of the Pennsylvania initiative encouraged by the FHWA, other states started to implement similar programs. Washington was one of the first states to start such a statewide program. This program, which started in 1992, still prevails to this day; several projects have been successfully completed and others are on the way. The Washington State Corridor Safety Program is a joint program between the Washington Traffic Safety Commission and the Washington Department of Transportation (WSDOT) and the goal is to reduce fatal and disabling crashes along the designated corridors. The corridors selected need to have statistical evidence of a crash occurrence problem and there must be local support for the project undertaken (Washington Traffic Safety Commission [WTSC], 2006). Some of the corridors like SR 14, which was one of the designated corridors, had safety issues like speeding, over the centerline-related crashes, driving under the influence (DUI) and operating defective equipment (National Highway Traffic Safety Administration [NHTSA], 2004). The action plan was primarily based on the 3 Es: education, enforcement and engineering (NHTSA, 1997). As of today, 21 projects have been completed and nine are still in progress. The number of crashes along 24 designated corridors has been reduced by 6%, reduction in traffic injuries is by 11%, alcohol-related crashes have gone down by 20% and, most importantly, fatality-disabling crashes have decreased by 34%. The fundamental elements of the program are education, enforcement and engineering solutions for improving safety along the designated corridors (WTSC, 2006).

2.2.3 Virginia

After the success of the Pennsylvania program, the state of Virginia also became active in the field of corridor safety in 1992. This program differed considerably from FHWA guidelines (Jernigan, 1997). The Virginia Department of Transportation (VDOT) and the Virginia Department of Motor Vehicles (DMV) co-sponsored two pilot projects: one urban and one rural. Apart from safety, the authorities wanted to identify the possible differences in the ability of the program to be effective (Jernigan, 1999). The selected urban corridor was a 5.5-mile segment on U.S. Route 144, while the selected rural corridor was a 19-mile stretch on U.S. Route 24. The significant safety problems along these corridors were driver's inattention, speeding, defective vehicles, DUI, rear-end crashes, angle crashes, fixed object crashes (run off the road), and sideswipe crashes. The suggested improvements for improving safety were: lowering the speed limit, enforcement, improving signage and sight distance, warning for DUI checkpoints along the corridor, installations of traffic signals, changes in the approach to intersections, guardrail installation, and addition of paved shoulders (Jernigan, 1997). After the improvements were implemented, there were 5% less number of injury crashes along the rural corridor and injuries decreased by more than 10%. For the urban corridor, the situation was something different; though the injury crashes decreased by 10%, the injuries increased by 5%. Virginia has also developed a methodology for determining, investigating and improving the safety of corridors (Fontaine and Read, 2006). The designated/selected corridors should definitely have above average crash rate and densities (Virginia's Surface Transportation Safety Executive Committee, 2006).

2.2.4 California

In 1992, California also started a corridor safety program lead by the California Highway Patrol; also, this was done in collaboration with Caltrans and California's Office of Traffic Safety. A 21-mile long corridor on SR 1 in Ventura County was selected. The recommendations for safety improvement included education and public information, enforcement, engineering solutions, as well as emergency response. The number of injury crashes and injuries on the corridor dropped by 25% (Jernigan, 1999); the crash rate decreased by 11% to 37% within a 3-year analysis period, and the injury crash rate decreased by 13% to 47% (Fontaine and Read, 2006). SR 41 and SR 46 were both designated as safety corridors after a severe collision resulted in multiple fatalities in 1995. The safety problems identified were unsafe turning, unsafe speed, right of way violations, DUI, and driver not at fault. The countermeasures implemented fell into the 4 Es: education, enforcement, engineering solutions and emergency response. In the end, these efforts paid off: fatalities were reduced by 10% and injury crashes decreased by 32% (Bichler-Robertson et al., 2001). In the recent past, State Highways 25, 49, 65 have been designated as safety corridors; for SR 25 in particular, the goal is to reduce the fatal and injury crashes (California Department of Transportation, 2006).

2.2.5 Oregon

In 1993, Oregon jumped into the scene of CSIPs. These were implemented along Oregon Routes 34 and 22. The common safety concerns on the corridor were speeding, speed variation and access-related crashes. Increased level of enforcements, dividing the highway and limiting the number of access points, provision of acceleration and deceleration lanes at major access points, limited use of traffic signal, as well as decreasing the speed limit, all were some of the recommendations for improving safety. The first phase of the program was a success (Hunter-Zaworski and Price, 1998), reflected by less fatalities and crashes along the designated safety corridors. In 2001, doubling fines were effective for safety improvements along Oregon's corridors. Through the implementation of this program, it was found that drivers have a higher perception of accident risks, traffic citations and fines when driving at work zones and school areas than when driving at safety corridors (Jones et al., 2002). For the new safety corridors' designation, the following three criteria must be met: 1) the three-year average of the fatality and injury crash rate must be greater than or equal to 110% of the three-year statewide average for similar types of roadways, 2) the state or the local law enforcement agencies must commit for making a certain corridor a patrol priority, and 3) the designated team should agree that the length is manageable from an enforcement and education point of view (Oregon Department of Transportation [ODOT], 2006). Oregon Routes 62, 22, 34, 11, 18, 99E, 140 and U.S. Routes 101, 199, 20, 26, 730 are the routes where Oregon's safety corridors are currently located (ODOT, 2007).

2.2.6 North Carolina

In 1998, the highway safety program of North Carolina was implemented in 21 counties across the state. Fatal truck-related crashes were the major safety problem, which increased both the number of roadside inspections and citations for Commercial Driving License (CDL) violations. Within a year of the program's implementation, there was a 4.6% reduction in crashes involving commercial motor vehicles (CMVs) at the designated counties and a 5.2% reduction in crashes involving CMVs at the non-designated counties. Also, there was a decrease in number of fatalities of 17.7% due to crashes involving CMVs in the designated counties, whereas fatalities increased by 7.6% due to crashes involving CMVs in the non-designated counties (Hughes, 1999). Overall, the crash rate did not change substantially.

2.2.7 Kentucky

In 1997, the Kentucky Transportation Cabinet started the Safety Corridor Program in the attempt to reduce the number of crashes and number of injuries and fatalities on state highways. A methodology for selecting high crash corridors was developed, and a crash analysis technique was proposed (Green and Agent, 2002). The designated corridor was U.S. Route 31W. The rural section of this corridor had a higher percentage of the fatal/injury crashes at intersections resulting from angle crashes; this section also had a high percentage of run off the road crashes. The urban section had a higher percentage of rear-end crashes, as well as more crashes on straight sections. In addition, the period from 12:00 p.m. (noon) to 6:00 p.m. reported a high number of crashes. Business and industrial districts were also related to a higher percentage of the crashes. Failure to yield, following too closely and driver's inattention were also major contributing factors to fatal/injury crashes (Green and Agent, 2002). Overall, this program focused on education and enforcement in order to alleviate the safety problems.

2.2.8 Arizona and Ohio

A pilot study was conducted by the Arizona Department of Transportation (ADOT) in 1995 for observing how the CSIP takes shape. It was concurred that the tools considered for the pilot study could lead to progress in safety improvement identification and that they could be used by agencies other than the ADOT (Breyer and Joshua, 1999).

In 2005, Ohio's Highway Corridor Safety Program got started. Seven highways were identified: SR 37, 46, 49, 50, 60, 73 and 193 (Governor's Task Force on Highway Safety, 2005). The Governor's Task Force on Highway Safety has issued a handbook of guidelines and procedures including a process for safety corridor selection as well as a toolbox for safety studies and countermeasures.

2.2.9 Florida

The goal of the project set up in 1992 by Florida's Safety Management System was to establish a Corridor/Community Traffic Safety Program (C/CTSP) in the each of the 20 high crash counties across the state by 1996, with the aim to reduce the number of injuries and fatalities. The concept was pilot-tested in Lakeland, Florida, in collaboration with the Florida Department of Transportation (FDOT). The project became a success and a statewide C/CTSP coalition was formed (NHTSA, 1996). The designated corridor was Florida Avenue. Speeding, DUI, and non-use of seat belts were some of the safety concerns along this corridor and improvements were suggested accordingly. There was a reduction in number of crashes and injuries during the analysis period (Dummeldinger et al., 1994). In a recent research work on the safety of six-lane divided highways, it was suggested that reduction in horizontal curves, as well as the increase of median and shoulder widths, both can reduce the rate of severe and fatal crashes (Petritsch et al., 2007). As an additional note, human factors cause 94% of fatal crashes; therefore, to consider the 4 Es is recommended for reducing severe/fatal crashes (Spainhour et al., 2005).

2.2.10 Overview of Typical Safety Issues on Corridors

The safety issues affecting corridors can be broadly divided in two categories: 1) roadway design deficiencies, and 2) drivers' performance failures. Roadway design deficiencies include having several access points, a high number of traffic signals than what is actually required, inadequate shoulder, absence and/or inadequate length of acceleration/deceleration lanes, etc. Drivers' performance failures include speeding, DUI, CDL violations, over the centerline crashes, operating defective vehicles, right of way violations, non-use of seat belts, etc. The most common type of crashes observed were angle, rear-end, and fixed object (run off the road) crashes. Many safety corridors also had a high percentage of fatal/injury crashes involving trucks.

The improvements for the safety corridors being studied are based on the observed safety issues, and their implementation has been based on the 4 Es: education, enforcement, engineering solutions and emergency response. Education and media information has made the community aware of the hazardous corridors, urging people to proactively help in improving safety on the roads. Enforcement activities in many states include increased patrolling, doubling fines on the designated corridors, increased number of citations when violating traffic rules, booking drivers for DUI, as well as increased roadside inspection of commercial vehicles.

Changes in roadway design on sections of the safety corridors, reducing or increasing traffic signals, access management, adding paved shoulders, as well as modifying acceleration/deceleration lanes, all are among the recommended engineering changes for alleviating road safety conditions. In addition, having a good emergency response for improving the survival probability of crash victims has been of major concern for state agencies. Table 2-1 below presents a comparison of the work done in various states and the corresponding level of success as perceived by the authors of this report.

Table 2-1: Previous Corridor Improvement Work Done in Selected States

	Initial Initiatives (Yes or No)	**Success Measure of Initial Initiatives**	**New Initiatives / Projects**	**Success Measure of New Initiatives**
Pennsylvania	Yes	High	Doubling fines	No data
Washington	Yes	High	New projects	No data yet
Virginia	Yes	Relatively good	New projects	No data yet
California	Yes	High	Doubling fines	High
Oregon	Yes	Relatively good	Doubling fines	No data yet
North Carolina	Yes	Relatively good	-	-
Kentucky	Yes	-	-	-
Arizona	Yes	Relatively good	-	-
Ohio	Yes	-	-	-
Florida	Yes	Relatively good	New projects	-

CHAPTER 3. URBAN ARTERIAL CRASH CHARACTERISTICS RELATED WITH PROXIMITY TO INTERSECTIONS AND INJURY SEVERITY

3.1 Introduction

As mentioned in the opening paragraph of this document, in spite of the lower prevailing speeds –as compared to freeways/expressways–, arterials experience a significant proportion of severe/fatal crashes. For example, arterials are sites accounting for 57% of the fatal crashes in Florida (NHTSA, 2005). The safety on an arterial corridor may be affected by crash patterns on two seemingly distinct roadway elements: intersections and the segments between the intersections. A study by Abdel-Aty and Wang (2006) revealed the spatial correlation between crash patterns belonging to successive signalized intersections on an urban arterial; this indicates the need to look at the sequence of signalized intersections along a corridor rather than just analyzing each intersection as an isolated entity. For such approach, crashes occurring on and arterial segment(s) joining consecutive intersections would also be a critical part of the analysis. There is potential for achieving a better understanding of crash patterns on arterials if corridors are studied as a whole, instead of studying their parts independently (i.e. studying intersections and segments separately).

An important issue to be addressed for understanding safety of a corridor as a whole is the difference between crash patterns related to intersections and segments, especially as this relates to injury severity. There are significant variations among injury severity patterns that may be partially explained by the separation of the crash location from the intersection. For example, Abdel-Aty et al. (2006) found that the prevailing types of fatal or severe crashes at intersections are mostly angle and left turn crashes, while those on roadway segments farther from intersections are mostly fixed object collisions. Therefore, if one observes crashes only at the physical area of intersections crashes would involve a higher proportion of angle and/or left turn crashes, which tend to be more severe. However, as the definition of the intersection is changed for including some area around it (i.e. the influence area for an intersection is defined), rear-end and other crash types would be included in the sample and the severity patterns may be altered.

The influence area for an intersection is characterized by the distance from the center of the intersection along either of the two legs belonging to the corridor being considered. Crashes within this distance from any intersection (signalized or unsignalized) are categorized as intersection/intersection-related crashes, whereas the crashes beyond this distance are categorized as segment crashes. The current study aims to identify the factors associated with crashes and their severity on multilane arterials, while accounting for the variations that result from the location of crashes relative to intersections. This is accomplished by developing models for different distance thresholds that define the influence area for intersections. In addition, the methodology used in this study also accounts for the correlation between the factors explaining injury severity and the crash location (intersection vs. segment) at a particular threshold. The approach adopted herein provides a better understanding of the relationship between the crash location's relative proximity to intersections and the severity outcome; it may also contribute to the understanding of how changes made to an intersection affect the neighboring segments of the arterial.

Crash data from the SR 816 corridor in Broward County, Florida, are used in this study. The crashes belonging to intersections are separated from the crashes belonging to arterial segments; this has been done by using an ordinal variable defining injury severity, as well as a binary variable whose definition changes based on the specified intersection's influence distance. A detailed characterization of these two variables is provided in the next section along with the solution approach and modeling methodology. The section with details on the data used in the analysis is then followed by the results and conclusions of this investigation.

3.2. Solution Approach and Modeling Methodology

Relationships between the following variables are of interest in this study:

- A three-level ordinal variable representing the injury severity. This variable is created based on the injury severity information available in the Crash Analysis and Reporting (CAR) database, maintained by the FDOT.
- A binary variable representing crash location. It has a value of 1 if crashes occur within the intersection's threshold influence distance (intersection/intersection-related crashes), and has a value of 0 if crashes occur outside this threshold influence distance (segment crashes). In this study, the influence distance (taken from the center of the intersection) would vary in 50 ft increments on arterial corridors; therefore, there would be multiple binary variables that would be distinguished among crashes based on their location (i.e. intersection and non-intersection crashes).

An ordered probit modeling framework would be used for the first variable, since injury severity levels are naturally ordered. Ordered probit modeling has been applied to injury severity in several studies like those by O'Donnell and Connor (1996), Duncan et al. (1999) and Abdel-Aty (2003); however, none of these –with the exception of the one by Abdel-Aty (2003)– compared the factors that affect injury severity at different roadway locations. Abdel-Aty (2003) used the ordered probit model for studying crash severity at both roadway sections and signalized intersections; however, the analyses for these roadway elements (segments and intersections) were carried out independent of each other.

In the preliminary analysis, chi-square tests for association between injury severity and the binary variable(s) representing crash location suggested a possible association between them. Furthermore, the nature and strength of the association changes as the definition of the crash location variable varies. The results from these tests are later discussed. The straightforward way for assessing the impact of crash location (i.e. intersection) on injury severity would be by using the binary variable(s) representing crash location as an independent variable in the ordered probit model for injury severity; however, this binary variable would be related with the variables generally used in the model for injury severity. For example, crashes, under rainy conditions, are less likely to occur right at the intersection when compared to a roadway segment influenced by intersections. Similarly, left turn or angle crashes are more likely to occur within the physical area of the intersection, when compared to segments, and they also tend to be more severe. To avoid the confounding effects of other variables, it was decided that the models for injury severity (ordinal dependent variable) and for crash location (binary dependent variable) would be estimated simultaneously. Since the location variable may be associated with certain variables included in the severity model, its inclusion (i.e. recursive specification) would have also led to problems of correlated independent variables, as well as biased and inefficient estimates for the coefficients.

Simultaneous estimation of the two models would improve the coefficient estimates by accounting for the correlations between the unmeasured factors. The difference between independent estimation and the simultaneous (bivariate) modeling procedure is that the latter does not assume the errors for the two models to be uncorrelated; the latter procedure also provides the p-value for the statistical test on correlation with the null hypothesis being that the correlation coefficient is $\rho=0$.

3.3 Model Formulation

According to Long (1997), logit and probit models provide very similar results in terms of resulting classification and standardized effects for independent variables; however, convergence is more likely for bivariate probit models, even though it may require more computational time (Indiana University, 2007). The model specification for the simultaneously estimated probit model equations is as follows (Green, 2003):

$$Y_1^* = X_1'\beta_1 + \varepsilon_1 \tag{3.1}$$

$$Y_2^* = X_2'\beta_2 + \varepsilon_2 \tag{3.2}$$

where X_1 is the vector of independent variables explaining the roadway location of the crash and X_2 is the vector of independent variables explaining the crash injury severity. The disturbances ε_1 and ε_2 have the following specifications:

$$E[\varepsilon_1 \mid X_1, X_2] = E[\varepsilon_2 \mid X_1, X_2] = 0$$

$$Var[\varepsilon_1 \mid X_1, X_2] = Var[\varepsilon_2 \mid X_1, X_2] = 1$$

$$Cov[\varepsilon_1, \varepsilon_2 \mid X_1, X_2] = \rho$$

Note that the variables Y_1^* and Y_2^* are unobserved, latent and continuous. The binary and ordinal scale dependent variables, Y_1 (crash location) and Y_2 (injury severity), are observed when the respective latent variables Y_1^* and Y_2^* fall in certain ranges. The two independent variables observed as discrete categories (i.e. Y_1 and Y_2) are specified as follows:

$$Y_1 = \begin{cases} 0 & if \quad Y_1^* \leq 0 \\ 1 & if \quad Y_1^* > 0 \end{cases}$$

$$Y_2 = \begin{cases} 0 & if \quad Y_2^* \leq 0 \\ 1 & if \quad 0 < Y_2^* \leq \mu \\ 2 & if \quad Y_2^* > \mu \end{cases}$$

Previously shown Equation 3.1 (specified as a binary probit model) relates crash location with other crash characteristics, while Equation 3.2 (specified as an ordered probit model) relates injury severity with the independent variables. This formulation relates both crash location and injury severity, without confounding the effects of independent variables relating to both of them. Detailed descriptions of the variables constituting the vectors X_1 and X_2 are provided in the next section (see Table 3-1, page 29).

The estimates for model coefficients may be obtained using maximum likelihood estimation; the corresponding function incorporates the effect of correlation between the error terms. The coefficients for the models just specified (i.e. vectors β_1, β_2 along with $\rho(u_1, u_2)$) were estimated using SAS (2007). Details of the maximum likelihood estimation process may be found in the work done by Greene (2003).

Multiple sets of simultaneous models (corresponding to different thresholds on influence distance) based on the aforementioned specification would be estimated for the corridor being studied. The only difference between the sets of simultaneous models would be the definition of Y_1 (i.e. crash location variable); this definition would in turn depend on the threshold selected for separating intersection crashes from segment crashes. Details on these thresholds and the variables used in the analysis are also provided in the following section.

3.4 Data Preparation

The crash data used in this study are from a 9.72-mile corridor of arterial SR 816 in Broward County, Florida. Both signalized and unsignalized intersections are considered in this study. The intersection density (intersections per mile) for the corridor is 11.32. For this multilane arterial, the total number of crashes involving at least a possible injury over the four-year period (2002 through 2005) was found to be 1,575; from this number, 11.17% were either fatal or involved an incapacitating injury. The crash data for the aforementioned corridor were downloaded from the FDOT's CAR database.

Table 3-1: Variable Descriptions

Variable	Categories	Description
Independent Variables		
Traffic Condition (Based on time of day/day of week)	MPW	Morning peak traffic condition on weekday (7:00 a.m. – 9:30 a.m.)
	APW	Afternoon peak traffic condition on weekday (4:00 p.m. – 7:00 p.m.)
	FSN	Friday or Saturday night traffic condition (Friday 10:00 p.m. – Saturday 3:30 a.m.)
	OP	Off-peak traffic condition
Sectional *aadt*	1*	Section *aadt* <= 52,000
	2*	52,000 < Section *aadt* <= 58,000
	3*	58,000 < Section *aadt* <= 64,500
	4*	Section *aadt* > 64,500
Road Surface		Binary (1 = dry surface, 0 = all other cases)
Lighting		Binary (1 = day time, 0 = night time)
Weather		Binary (1 = clear, 0 = all other cases)
Road Curvature		Binary (1 = straight, 0 = curve)
Road Surface Type		Binary (1 = blacktop, 0 = all other cases)
Road Condition at time of Crash		Binary (1 = No defects, 0 = all other cases)
Vision Obstruction		Binary (1 = no obstruction, 0 = all other cases)
Alcohol/Drug Use		Binary (1 = No, 0 = Yes)
Pavement Surface Width		Width of the pavement (Continuous)
Shoulder Width1		Width of the shoulder closest to the travel lane (Continuous)
Shoulder Width2		Width of the shoulder farthest from the travel lane (Continuous)
Median Width		Width of the median (Continuous)
Speed Limit		Maximum posted speed limit (continuous)
Dependent Variables		
Crash Location (Y_1; location_D)	1	Crashes within the 'D' ft from the center of intersection
	0	Crashes beyond 'D' ft from the center of intersection
Injury Severity (Y_2)	2	Crashes resulting in incapacitating injuries or fatalities
	1	Crashes resulting in non-incapacitating injuries
	0	Crashes resulting in possible injuries

* The *aadt* values from various sections of the corridor have been split into four quartiles.

Before proceeding, some data issues required clarification. These issues were mainly related to the recorded crash location and the definition of influence distance. In the database used for this study, each crash is assigned to the intersection (node) nearest to its location; also, information on the distance of crash location from the node representing the center of the intersection is available in the database. Through a careful review of this information, it was noticed that a significant number of crashes are reported to have occurred at the milepost associated with the nodes (i.e. the distance between the crash location and the center of the intersection is reported as 0 ft); this does not necessarily mean that all of these crashes occurred at the midpoint of the intersection. Therefore, the significantly large number of such crashes implies that most crashes occurring inside the physical area of the intersection are reported to be 0 ft from the center of intersection. In addition, note that in the state of Florida an intersection's physical area is by default considered to be the area within 50 ft from the center of the intersection. Therefore, some of the crashes reported to be within 50 ft of the node (representing the intersection) in the database may be very close to the stop bar.

These crashes, while not strictly *at* the intersection, would most likely to be influenced by it. Therefore, the first two thresholds on influence distance were defined as 0 ft for intersection crashes and 50 ft for non-intersection crashes. The threshold of 0 ft indicates that the crashes occurring within 50 ft from the center of the intersection are classified as intersection crashes (i.e. only those crashes occurring within the physical area of the intersection). For the model corresponding to D=50, the crashes that have occurred within the physical area of the intersection and those that have occurred within 50 ft of the stop bar have been classified as intersection crashes. The successive thresholds were also defined by 50 ft increments (i.e. 100 ft, 150 ft, etc.). As mentioned earlier, this threshold defines one of the two simultaneously estimated dependent variables, Y_1 (refer to previous section).

It has to be noted that the selection of thresholds at 50 ft increments is somewhat arbitrary. Therefore, the results from the sets of simultaneous models estimated using different thresholds need to be interpreted in relative terms. For example, for the models with threshold at 100 ft, crashes closer to the intersection are treated as intersection/intersection-related crashes, as compared to the set of models with threshold at 150 ft. Table 3-1, page 29, lists the independent (regressors) and dependent (responses) variables used in this study; also, the last row of this table represents the crash location as binary variable *location_D*, which would take the value 1 for crashes within 'D' ft from the center of the intersection.

Crashes with fatalities and incapacitating injuries are combined into one category (of variable Y_2 representing injury severity) for two reasons. First, the relatively small frequency of fatal crashes, as compared to other injury severity levels, could create problems in the analysis. For example, the chi-square tests on contingency tables may not be valid due to low expected cell-frequency. Second, the crashes involving incapacitating injury could have easily been fatal, and vice versa, depending on the vulnerability of the subjects involved. Also, note that the variables shown in Table 3-1 are gathered from the long form (i.e. complete) crash reports filled out by law enforcement authorities in Florida; the information on crashes involving no injury is likely to be incomplete for this set of crashes (Abdel-Aty et al., 2004). Therefore, only crashes that involve at least a possible injury are included in this study, and the injury severity is categorized as a three-level ordinal variable.

It also has to be noted that some of the binary variables shown in Table 3-1 had more levels in the original database. Some of the categories belonging to these variables were quite infrequent and were therefore combined with each other. Note that the *aadt* of the road sections was divided into four quartiles, so that they have close to 25% cases in each of the categories;

this variable has been used as a nominal variable and not as an ordinal variable in the analysis, since the categorization may not follow the natural order in terms of the relationship between *aadt* with injury severity (Y_2). The rest of variables contained in the aforementioned table are self-explanatory.

3.5 Analysis and Results

As mentioned earlier, the association between the ordinal variable (crash injury severity) and the binary variable(s) (crash location) was first examined with chi-square tests. A contingency table is used in order to reliably assess the strength of this association through the chi-square test; each cell of the contingency table is required to have a minimum expected frequency. With the increase of the influence distance (starting from 0 ft) more crashes get assigned, as intersection crashes and the number of crashes assigned as non-intersection (segment crashes) is reduced. Beyond a certain influence distance, the frequency of non-intersection (segment) crashes becomes too low for the chi-square test statistic to be credible; therefore, a maximum allowable influence distance had to be chosen such that at least 10% of all crashes are assigned as non-intersection (segment) crashes. The maximum allowable threshold influence distance for SR 816 was found to be 200 ft using this criterion; for example, limiting the threshold distance to 200 ft helps in reducing the chances of having the influence area of one intersection overlap with other one. The chi-square test statistics and corresponding p-values for testing associations between Y_1 and Y_2 (with definition of Y_1 varying based on influence distance thresholds, D=0 ft through D=200 ft) are reported in Table 3-2, page 33.

Bivariate probit models, detailed earlier in this document, were developed for both the injury severity (Y_2; ordered probit) and crash location (Y_1; binary probit) variables. This bivariate formulation does not assume that the errors from the simultaneously estimated models are uncorrelated. The significance of correlation coefficient ρ is tested and reported along with the estimated coefficients (and their significance) for the independent variables included in the two models; this correlation accounts for the common factors associated with both dependent variables that are not explicitly included in the models. The last column of Table 3-2 below provides the estimates for ρ and their significance. Table 3-3, page 34, shows the detailed estimates of the variables' coefficients, and their significance along with error correlation coefficient estimates are shown in the last column of Table 3-2.

Table 3-2: Chi-Square Statistics and Error Correlation Coefficient Estimates

Influence Distance (ft)	**Chi-Square** (p-value) (from Contingency tables)	**Correlation coefficient** $\boldsymbol{\rho}$ (p-value) (from bivariate probit models)
0	**4.369** (0.113)	**0.053** (0.172)
50	**1.354** (0.508)	**-0.046** (0.266)
100	**1.285** (0.526)	**-0.055** (0.201)
150	**7.889** (0.019)	**-0.135** (0.005)
200	**5.950** (0.051)	**-0.120** (0.016)

It can be observed from Table 3-2 that the significance trend for ρ at various intersection influence distance values is similar to the corresponding significance trend of the chi-square statistic. In Table 3-2 and Table 3-3, page 34, cells with statistically significant parameters (at 90% confidence level) have been highlighted. It is worth mentioning that the values of μ, for converting the estimated latent continuous variable into the categorical injury severity variable, were also estimated for each of the five injury severity models and are in Table 3-3.

Table 3-3: Five Simultaneous Models for the Crash Location and Injury Severity Levels on SR 816 (D=threshold influence distances in ft)

		D = 0		D = 50		D = 100		D = 150		D = 200	
Parameter		Estimate	Approx p-value	Estimate	Approx p-value	Estimate	Approx p-value	Estimate	Approx p-value	Estimate	Approx p-value
Crash Location Model											
Traffic Condition	APW	-0.086	0.338	-0.166	0.074	-0.148	0.125	-0.157	0.152	-0.176	0.115
Traffic Condition	FSN	-0.090	0.504	-0.057	0.693	-0.076	0.618	-0.228	0.181	-0.222	0.211
Traffic Condition	MPW	-0.061	0.625	-0.055	0.671	-0.063	0.638	-0.062	0.685	-0.122	0.424
Traffic Condition	OP	0.000	.	0.000	.	0.000	.	0.000	.	0.000	.
Dry Road Surface		0.276	0.023	0.251	0.050	0.149	0.266	-0.010	0.948	-0.049	0.753
Daylight Condition		-0.150	0.039	-0.225	0.004	-0.246	0.003	-0.253	0.008	-0.246	0.013
Clear Weather		-0.110	0.298	-0.168	0.139	-0.167	0.157	-0.085	0.521	0.004	0.977
Straight Road Section		-0.483	0.159	-0.277	0.461	-0.402	0.339	0.016	0.970	0.071	0.870
Blacktop Road Surface		-0.112	0.245	0.007	0.944	0.064	0.537	0.201	0.080	0.223	0.056
No Vision Obstruction		-0.042	0.758	0.064	0.647	0.234	0.097	0.094	0.572	0.039	0.820
No Alcohol/Drug Use		-0.080	0.596	0.114	0.472	-0.099	0.570	-0.011	0.954	-0.189	0.394
Injury Severity Model											
Traffic Condition	APW	-0.216	0.017	-0.214	0.019	-0.215	0.018	-0.214	0.018	-0.214	0.019
Traffic Condition	FSN	0.154	0.229	0.154	0.228	0.154	0.229	0.154	0.229	0.154	0.231
Traffic Condition	MPW	0.045	0.710	0.042	0.729	0.040	0.737	0.039	0.745	0.040	0.741
Traffic Condition	OP	0.000	.	0.000	.	0.000	.	0.000	.	0.000	.
Dry Road Surface		0.095	0.425	0.094	0.433	0.093	0.435	0.093	0.433	0.094	0.430
Daylight Condition		0.044	0.535	0.044	0.527	0.045	0.520	0.047	0.507	0.047	0.508
Clear Weather		-0.021	0.837	-0.020	0.844	-0.020	0.845	-0.021	0.839	-0.021	0.836
Straight Road Section		-0.020	0.953	-0.019	0.956	-0.016	0.962	-0.018	0.958	-0.017	0.961
Blacktop Road Surface		-0.221	0.015	-0.223	0.014	-0.223	0.014	-0.224	0.014	-0.223	0.014
No Road Defects at time of Crash		0.029	0.881	0.046	0.812	0.047	0.806	0.072	0.708	0.064	0.742
No Vision Obstruction		-0.144	0.274	-0.150	0.256	-0.150	0.255	-0.157	0.233	-0.155	0.239
Pavement Surface Width		0.040	0.018	0.044	0.009	0.044	0.008	0.047	0.005	0.046	0.006
Closest Shoulder Width		-0.263	0.362	-0.258	0.371	-0.262	0.364	-0.259	0.368	-0.259	0.369
Farthest Shoulder Width		-0.181	0.370	-0.175	0.388	-0.171	0.399	-0.169	0.405	-0.171	0.399
Median Width		-0.013	0.016	-0.013	0.015	-0.013	0.015	-0.013	0.012	-0.013	0.012
Maximum Posted Speed Limit		0.023	0.023	0.021	0.033	0.021	0.033	0.020	0.040	0.020	0.037
AADT (1st Quartile)	1	0.329	0.003	0.333	0.002	0.335	0.002	0.326	0.003	0.323	0.003
AADT (2nd Quartile)	2	0.226	0.014	0.221	0.017	0.221	0.017	0.209	0.024	0.210	0.023
AADT (3rd Quartile)	3	0.044	0.638	0.045	0.631	0.043	0.648	0.028	0.762	0.028	0.762
AADT (4th Quartile)	4	0.000	.	0.000	.	0.000	.	0.000	.	0.000	.
No Alcohol/Drug Use		-0.322	0.021	-0.322	0.021	-0.322	0.021	-0.323	0.020	-0.324	0.020
µ (for classification)		0.916	<.0001	0.916	<.0001	0.916	<.0001	0.916	<.0001	0.916	<.0001
Correlation Coefficient											
Rho		0.053	0.172	-0.046	0.266	-0.055	0.201	-0.135	0.005	-0.120	0.016

The significance of ρ changes as the influence distance, used for defining crashes on intersections and segments, varies from 0 ft through 200 ft. For the models developed corresponding to intersection influence distances of 0, 50, and 100 ft, the ρ value is insignificant; this indicates that error terms in the two models are not significantly correlated with each other. On the other hand, the correlation coefficient becomes significant beyond the 100 ft influence distance. Table 3-2, page 33, also depicted a similar trend for the significance of the chi-square test statistic. This in effect means that on average, when intersection crashes are defined such that they include a smaller influence area (within about 100 ft of intersections for a corridor), severity on the arterial crashes may be modeled as independent from the crash location. Again, it is worth mentioning that the 100 ft is the distance from the center of the intersection; note that this distance may also vary from corridor to corridor depending on intersection density and traffic patterns. As mentioned earlier, models for D > 200 ft and beyond were not developed due to data constraints; it may be inferred that the correlation would have probably been significant.

From this point forward, the discussion deals with the factors found significant for the two simultaneously estimated models at various threshold distances. The crash location (Y_1) model(s) for various threshold influence distance values (D) show the factors that help discriminate intersection crashes from segments crashes. The crash injury severity (Y_2) model(s) for various threshold influence distance values (D) in Table 3-3, page 34, show the factors that significantly relate with the ordinal variable.

Figure 3-1, page 36, depicts the significant parameters for the ordinal crash injury severity model in the form of bubble plots, where the different bubble sizes reflect the relative significance of these parameters with respect to each other; in addition, note that the bubbles within a plot may be compared horizontally but not vertically: the plot on the left side shows the effect of the factors that decrease the crash severity (i.e. negative coefficients), whereas the plot on the right side depicts those that increase this severity (i.e. positive coefficients).

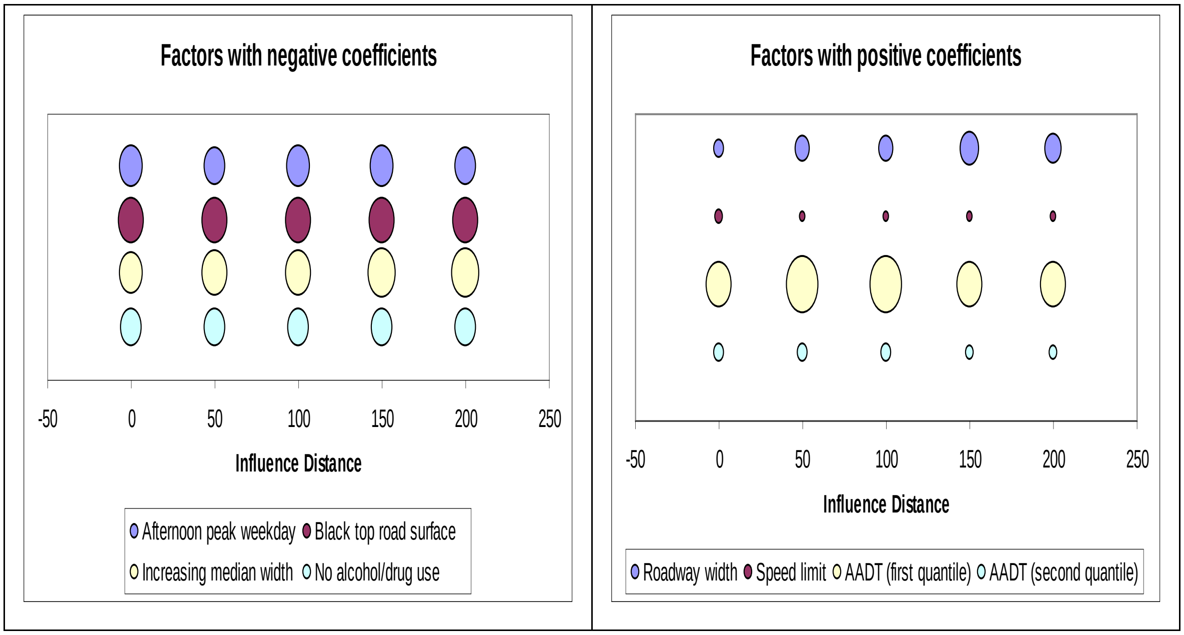


Figure 3-1: Significant Parameters for Crash Injury Severity Model

As it can be seen, Figure 3-1 above and Table 3-3, page 34, also illustrate that weekday afternoon peak period (APW) conditions (refer to Table 3-1, page 29), blacktop pavement surface and increase in median width decrease crash severity on SR 816 (see left side of Figure 3-1); non-use of alcohol/drugs also has the same effect, suggesting that alcohol/drug use increases the crash severity. Furthermore, during afternoon peak periods, speeds are generally lower due to congestion; therefore, crashes tend to be less severe. Likewise, higher median width may reduce the chances of severe crossover head-on collisions; this explains the significantly negative coefficient for median width. Similar results related to the severity of peak hour crashes at intersections were found by Abdel-Aty and Keller (2003); presence of median was also found to reduce crash severity in that study.

Regarding the blacktop surface variable, it was found to negatively affect severity in all five models (D=0 ft through D=200 ft). Also, note that this variable is significant for separating intersection vs. segment crashes, particularly when intersection crashes include those occurring within 150 ft and 200 ft of the intersection (refer to models for D=150 ft and D=200 ft in Table 3-3, page 34; also, see Figure 3-2, page 38). For intersection crashes occurring within 0, 50 and 100 ft of the intersection, this variable was not significant in the crash location model. These crashes are not only less severe (Abdel-Aty and Keller, 2003; Ma and Kockelman, 2004) but tend to be more frequent in segments within 150-200 ft from the intersection. The findings also seem to corroborate with a study that found that asphalt pavements may lead to a higher frequency of peak period crashes (Abdel-Aty et al., 2006). Since crashes on blacktop surfaces with asphalt base seem to have higher frequencies during peak periods and within 150-200 ft of the intersection, this suggests that these pavement surfaces *might* increase the odds of rear-end crashes; also, this may explain the negative coefficient of the variable representing blacktop surfaces with asphalt base in the injury severity model –since rear-end crashes tend to be less severe–.

It was also found that an increase in the roadway width and the speed limit contribute to the increase of crash severity. In addition, *aadt* values below the median value (both 1st and 2nd quartiles) are also positively associated with crash severity (refer to right side of Figure 3-1, page 36). Among the factors positively influencing injury severity, lower *aadt* is the most significant factor. Also, it is interesting to see that roadway width has a higher effect when the influence distance is greater than 0 ft.

Figure 3-2 below depicts significant parameters for five binary crash location models, each estimated simultaneously with the corresponding injury severity model. The model coefficients were provided in Table 3-3, page 34. Once again, the different bubble sizes reflect the relative significance of the parameters. Note that some parameters do not have a corresponding bubble at certain values of D (i.e. intersection influence distance); this indicates that if crashes at intersections are defined by these particular values of influence distance, the corresponding parameters do not contribute in discriminating the crash location. As it can be noted from the figure, the plot on the left side shows the effect of the factors that decrease the likelihood of a crash being within a particular distance from the intersection (i.e. negative coefficients), whereas the plot on the right side depicts those that increase this likelihood (i.e. positive coefficients).

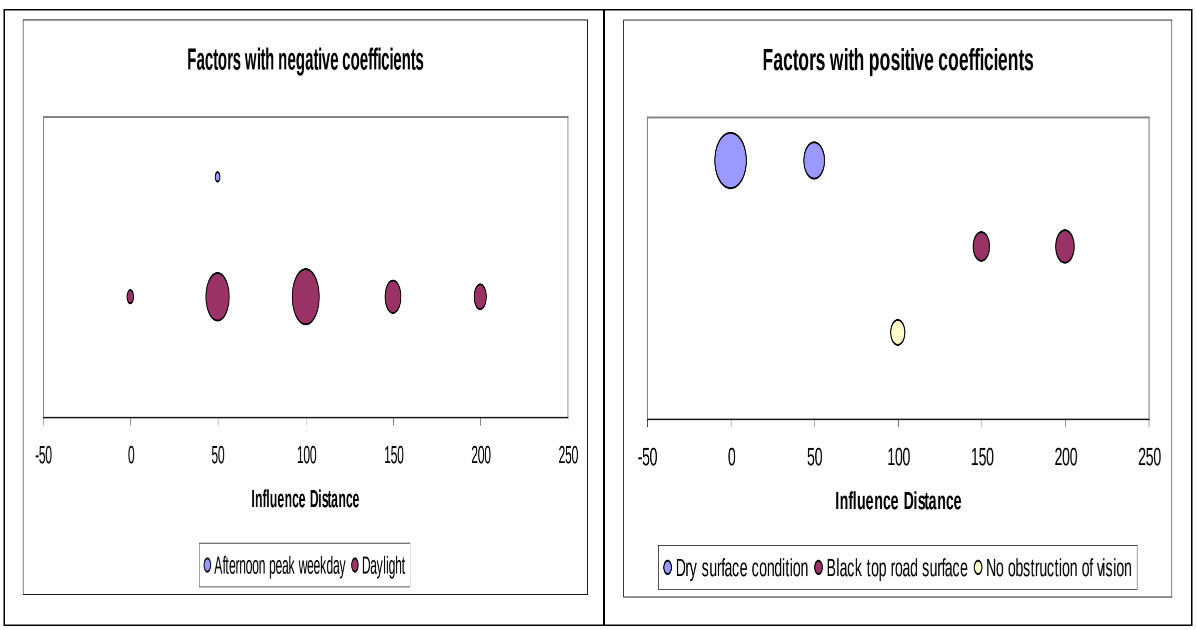


Figure 3-2: Significant Parameters for the Crash Location Model

From Figure 3-2, page 38, and Table 3-3, page 34, it can also be observed that during weekday afternoon peak period (APW) conditions the likelihood of a crash occurring within 50 ft of the intersection is less when compared to off-peak (OP) traffic conditions. Note that while this difference is insignificant when considering influence distances of 100, 150 and 200 ft (no corresponding bubble in left side of Figure 3-2, page 38), the p-value is much closer to 0.10 (refer to Table 3-3). Note that this difference between afternoon peak weekdays (APW) and off-peak (OP) conditions (refer to Table 3-1, page 29) is insignificant if one examines the relative likelihood of a crash occurring within the physical area of an intersection (influence distance=0 ft); this is probably because during the afternoon peak hours, drivers expect congestion and expect to slow down and/or to stop as they approach an intersection, which reduces the likelihood of crashes prevailing at the vicinity of intersections. Also, the modeling results show that the variable representing normal daylight is significant in separating crashes at intersection and segments regardless of the specified influence distance; this significance is higher when considering influence distance values of 50 and 100 ft and, while it is hard to conclude, the smaller coefficient of this variable at influence distance values of 150 and 200 ft might be definitely caused by the dilemma zone phenomenon.

Among the variables having positive coefficients (refer to right hand side of Figure 3-2, page 38) blacktop road surface is significant for separating intersection crashes from segment crashes if the influence distance is 150 ft or 200 ft; the implications of this result were discussed earlier. Dry surface conditions also increase the likelihood of a crash to occur at the physical area of an intersection or within 50 ft of it; from these two cases, it can be observed that the significance is higher if considering the physical area of the intersection. The latter statement can be interpreted as follows: if the influence area of the intersection is increased, then the weather conditions' ability to discriminate between intersection and segment crashes is diminished; this might be due to crashes occurring during wet weather conditions that are more prevalent on approaches to intersections. In addition, a result that was not clearly understood was the vision obstruction variable, which was found to be significant for identifying intersection crashes from segment crashes when the influence distance is set at 100 ft; on the other hand, this variable was not significant at any other influence distance value and the respective p-values were not even on the margin, which can be due to peculiar issues of the corridor being studied (e.g. a few intersections with vision obstruction problems along the corridor, or the demographics of Broward County having a considerable proportion of older drivers).

3.6 Concluding Remarks

To understand safety on urban arterials is a complex problem, especially since it is affected by the different traffic pattern interactions at the intersections and segments connecting them. The implementation of certain safety improvements at intersections may lead to unanticipated changes in the safety/operation performance of nearby segments, or vice versa; therefore, an improved understanding of safety may be achieved if consecutive intersections on arterial corridors and the segments connecting them are all examined as a whole, instead of examining them as isolated entities. The analysis presented in this document is an effort towards that direction, which focuses on the injury outcomes of crashes occurring at such locations.

A simultaneous analysis of crash characteristics, which can explain considering both crash location (intersection vs. segment) and severity, was performed. More specifically, this was carried out through a simultaneous estimation of models for crash location and injury severity at five different values of intersection influence distance D (i.e. the distance from the center of an intersection along the corridor, up to which crashes are categorized as intersection-related), which varied from 0 ft through 200 ft in 50 ft increments. This simultaneous estimation of crash location and injury severity models made it possible to account for the correlation between the errors in the two models; this correlation is more likely to be the result of unknown common factors affecting both variables but that are not explicitly included in either model.

Regarding the model for the crash location variable, it suggested that during the peak hour crashes are less likely to occur at or in the vicinity of intersections. It was also found that an increase in pavement surface width and speed limits increase crash severity; similarly, lower *aadt* values have been found to be positively associated with crash severity. From the latter statement, it may be inferred that certain conditions that make the driving task easier, such as a wide roadway and a low *aadt*, can lead to an increased severity of crashes.

Finally, it should be noted that the results obtained in this study may be specific to the corridor being considered; however, it may be expected that similar results (e.g. the influence distance threshold beyond which the error correlation coefficient becomes significant) could be obtained for corridors with comparable intersection density. Along the same note, the results also suggest that for corridors with higher intersection density (i.e. more closely spaced intersections) the errors may not be correlated; therefore, the crash location and injury severity may be modeled independent of each other. This inference is based on an insignificant correlation between errors for the simultaneous models developed D values of 0, 50 and 100 ft. On the other hand, for arterials on which intersections are fewer and farther from each other, the injury severity models need to account for crash location (i.e. intersection vs. segment crashes).

CHAPTER 4. RULES FOR CRASH ASSIGNMENT

4.1 Background

Signalized and unsignalized intersections, as well as the segments connecting them, constitute the three basic elements of any given arterial. In common practice, crashes are assigned to these elements based on the crash location. For this study's purposes, signalized intersections will be referred as intersections, while unsignalized intersections will be considered as a type of access point (e.g. any street intersecting the arterial and that has a control type other than a signal; it could be a county road or a private driveway). In the U.S., most states have a defined intersection influence area for their jurisdictions. For example, in the state of Florida crashes that occurred within 250 ft of any intersection are referred as intersection-related crashes (Abdel-Aty and Wang, 2006; Wang et. al, 2006). However, having a defined intersection influence area for assigning crashes may arise some problems and/or discrepancies; this can be explained by the following:

- To have an influence distance could cause misclassification; for example, segment crashes could be misclassified as intersection-related.
- Recalling the method of simultaneous estimation conducted by Das et al. (2008), it was observed that if the influence distance varied the crash characteristics associated with severe injuries also vary; this is due to the fact that the farther we move away from the center of an intersection, more crashes related to the connecting segment come into play. Furthermore, Wang et al. (2008) conducted crash frequency modeling with fixed as well as varying influence distance; different set of significant factors

were found. Overall, these two very recent studies show that the concept of using influence distance for assigning crashes to roadway elements could be erroneous.

Apart from the aforementioned problems that relate to the influence distance, there are other problems related to the ways crashes are reported. In most cases, police officers do not take an actual measurement of the crash distance. The crash distance, which also determines the crash location, is the distance from the center of an intersection to the exact location of the crash; however, this distance is sometimes also taken from the intersection's stop bar on the arterial. In addition to this, the state of Florida has a 50 ft default intersection size; since not all intersections are of the same size, no matter how good the police officer is at guessing the location indicated in the crash report, which could turn out to be a very rough approximation. Therefore, to use the influence distance for classifying intersection-related crashes is not recommended.

Currently, there is no standard guideline for unsignalized intersections, such as the one of influence distance for signalized intersections. Therefore, if the *site location* is used to determine the location of a crash, the only access-related crashes that could be identified are those having a site location value of *driveway access*.

For the corridor level analysis presented here, it has been critical to know how to assign a crash to its appropriate roadway element; the goal is to assign crashes to segments, intersections or access points. In the case of crashes that have occurred at an unsignalized intersection, police officers often report these as intersection-related; this example suggests that the site location parameter could be a weak indicator for assigning crashes, and to use it alone for this purpose could lead to erroneous results. This problem lead to an investigation in order to find other crash record parameters that could be used for assigning crashes correctly. A meticulous study of crash reports revealed that traffic control in combination with the site location did a superior job in identifying the most appropriate roadway element to be assigned/selected; hence the method of assigning a crash based on crash characteristics. However, in certain cases the two aforementioned crash parameters may not facilitate the identification of crashes that are related to intersections or access points; in those cases it is necessary to check whether the particular node (i.e. intersection) is signalized or unsignalized.

Based on the aforementioned study, 377 crash reports were examined. Certain rules, in the form of *if-then-else* statements, were developed for the purpose of assigning crashes correctly; these rules had an overall accuracy of 93.63% as compared to 57.82% of accuracy obtained when only the site location parameter is used.

In the following sections, details for each rule will be provided; these will enable the reader not only to understand these rules but also to know how they were developed. The authors have used numeric representation for the parameters *site location* and *traffic control*; Table 4-1 below and Table 4-2, page 45, provide the meaning of all the numeric values assigned to these two parameters.

Table 4-1: Legend for *Site Location*

Site Location	Numeric Representation
Not at Intersection / RR / Bridge	1
At Intersection	2
Influenced by Intersection	3
Driveway Access	4
Railroad	5
Bridge	6
Entrance Ramp	7
Exit Ramp	8
Parking Lot – Public	9
Parking Lot – Private	10
Private Property	11
Toll Booth	12
Public Bus Stop Zone	13
All Other	77

Table 4-2: Legend for *Traffic Control*

Traffic Control	Numeric Representation
No Control	1
Special Speed Zone	2
Speed Control Sign	3
School Zone	4
Traffic Signal	5
Stop Sign	6
Yield Sign	7
Flashing Light	8
Railroad Signal	9
Officer / Guard / Flag person	10
Posted No U-Turn	11
No Passing Zone	12
All Other	77

It is important to note several observations from Table 4-1, page 44. The authors recommend the reader to focus on site location values 1, 5 and 6. As can be observed, the site location value of 1 relates to crashes not at intersection or at railroad or bridges; however, common practice identifies crashes occurring at railroads and bridges with the corresponding and exclusive site location values of 5 and 6, respectively. In addition, crashes that have occurred near a railroad, which have a site location value of 1, will have traffic control value of 9. For this study's purposes, the authors recommend to focus on site location values 4, 9, 10 and 11 since all of these represent access-related crashes; however, the data will almost never have any crashes with site location values of 9, 10 or 11 since these denote driveway-related crashes and are already included under the site location value of 4. Site location value of 12, which represents toll booth, has not been considered in this study.

4.2 Site Location 1: Not at Intersection / RR Xing / Bridge

By observing the site location value of 1 alone, one would assign all the crashes to segments; this is true for crashes where traffic control has a value of 1, 2, 3, 4, 10 or 12. When traffic control has a value of 5, 6, 7, 8, 9 or 11, it suggests that crashes do not always occur due to segment characteristics. For example, given that the site location is 1 and traffic control is 5 for a crash, then a closer look at the crash reports reveal that this particular crash occurred due to signalized intersection-related causes. Similarly, when the traffic control is 6 for a crash, then an investigation into the crash reports shows that this particular crash is related to an unsignalized intersection with a stop sign, which is considered to be an access point in this study. The examples just mentioned are described in Figure 4-1, page 47, through Figure 4-4, page 48. Figure 4-1 and Figure 4-2, both in page 47, are from crash report #769122280, where the site location is 1 and the traffic control is 5. In this particular instance, the at fault driver rear-ended the stationary vehicle in front view (i.e. the stationary vehicle was stopped at an intersection and was waiting for the red light to turn green); even though it has been classified as a *not at intersection* crash, this crash is definitely related to the signalized intersection and needs to be classified/assigned as such.

FLORIDA TRAFFIC CRASH REPORT BE USED ONLY FOR PURPOSES OF THE FDOT. SEE TITLE 23, USC, SECTION 409.

NARRATIVE / DIAGRAM

MAIL TO: DEPT. OF HIGHWAY SAFETY & MOTOR VEHICLES
TRAFFIC CRASH RECORDS
TALLAHASEE, FLORIDA 32399-0500

DO NOT WRITE IN THIS SPACE

TIME EMS NOTIFIED (FATALITIES ONLY)	TIME EMS ARRIVED (FATALITIES ONLY)	DATE OF CRASH	COUNTY / CITY CODE	INVEST. AGENCY REPORT NUMBER	HSMV CRASH REPORT NUMBER
AM PM	AM PM	05/27/2006	07 / 00	FHPD06OFF056456	76912228

(NARRATIVE)

Vehicle One (V-1) was traveling west on State Road 50 (SR50/Colonial Dr) in the outside lane. Vehicle Two (V-2) was stopped in the outside westbound lane for a red traffic light at the intersection of Dean Rd. For unknown reasons, V-1 failed to stop as it approached V-2 from behind. As a result, the front of V-1 collided with the rear of V-2. V-1 then left the scene without exchanging information required by law. V-2 was moved from final rest prior to my arrival.

V-1 was described as a blue Oldsmobile Cutlass. The driver of V-1 was described as a thinly built Hispanic male. No other information was available. This case is closed pending further leads.

Figure 4-1: Crash Narrative by the Police Officer

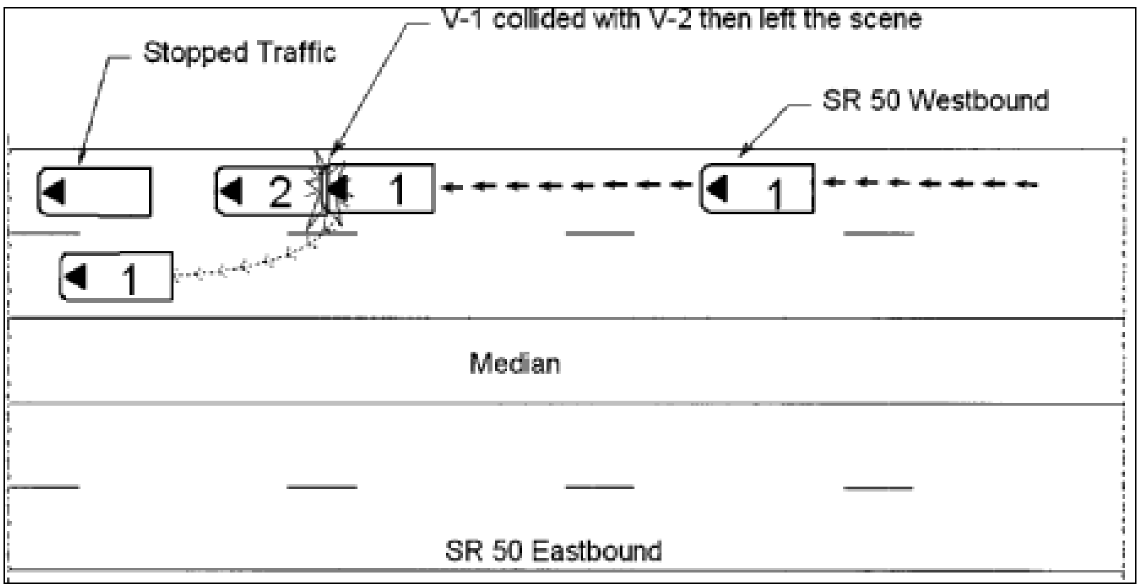
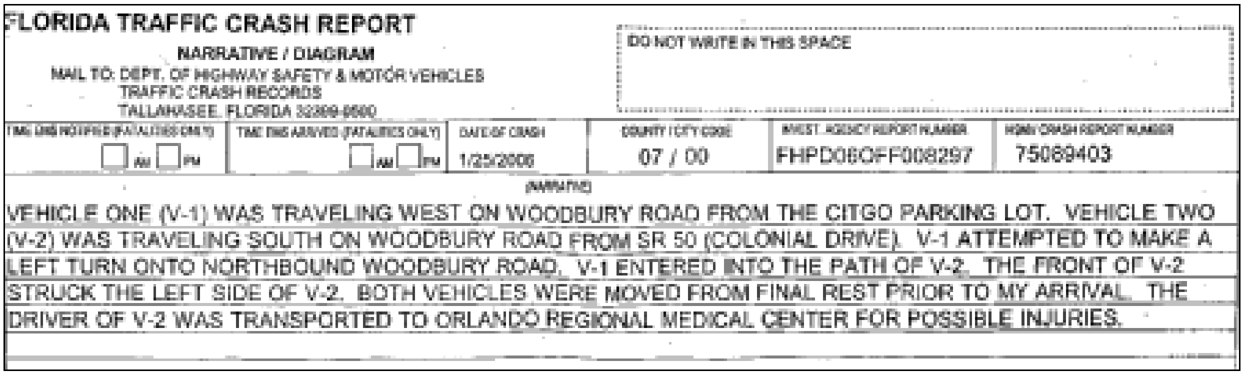


Figure 4-2: Police Officer's Graphical Representation of How the Crash Has or May Have Occurred

Likewise, Figure 4-3 and Figure 4-4, both in page 48, are from crash report #750894030, where the site location is 1 and the traffic control is 6. The description in Figure 4-3 and the illustration of Figure 4-4 clearly indicate that the crash is related to the unsignalized intersection rather than the segment. The at fault driver was getting out of a driveway, and while attempting to make a left turn came in collision course of the other vehicle resulting in an angle crash.



FLORIDA TRAFFIC CRASH REPORT

NARRATIVE / DIAGRAM

MAIL TO: DEPT. OF HIGHWAY SAFETY & MOTOR VEHICLES
TRAFFIC CRASH RECORDS
TALLAHASEE, FLORIDA 32399-0500

DO NOT WRITE IN THIS SPACE

TIME EMS NOTIFIED (FATALITIES ONLY)	TIME EMS ARRIVED (FATALITIES ONLY)	DATE OF CRASH	COUNTY / CITY CODE	INVEST. AGENCY REPORT NUMBER	HSMV CRASH REPORT NUMBER
☐ AM ☐ PM	☐ AM ☐ PM	1/25/2006	07 / 00	FHPD06OFF008297	75089403

(NARRATIVE)

VEHICLE ONE (V-1) WAS TRAVELING WEST ON WOODBURY ROAD FROM THE CITGO PARKING LOT. VEHICLE TWO (V-2) WAS TRAVELING SOUTH ON WOODBURY ROAD FROM SR 50 (COLONIAL DRIVE). V-1 ATTEMPTED TO MAKE A LEFT TURN ONTO NORTHBOUND WOODBURY ROAD. V-1 ENTERED INTO THE PATH OF V-2. THE FRONT OF V-2 STRUCK THE LEFT SIDE OF V-2. BOTH VEHICLES WERE MOVED FROM FINAL REST PRIOR TO MY ARRIVAL. THE DRIVER OF V-2 WAS TRANSPORTED TO ORLANDO REGIONAL MEDICAL CENTER FOR POSSIBLE INJURIES.

Figure 4-3: Crash Narrative by the Police Officer

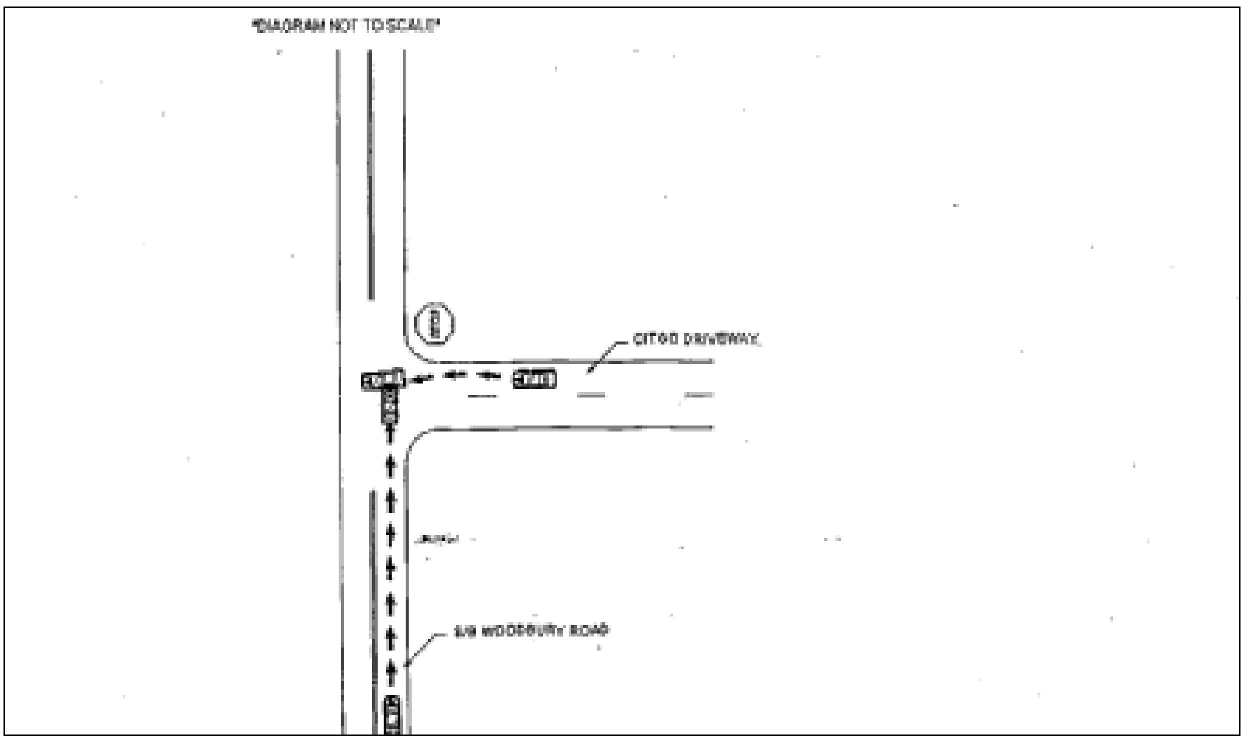


Figure 4-4: Police Officer's Graphical Representation of How the Crash Has or May Have Occurred

Therefore, it is now clear that the site location alone should not be used for assigning crashes to the different roadway locations. At least a combination of both site location and traffic control is required in order to correctly assign crashes where the site location is 1.

Figure 4-5 below is the flow chart of how a crash can be appropriately assigned to one of the three roadway locations, when considering a site location value of 1. The flow chart is essentially a set of *if-then-else* statements which can be conveniently understood. After all the checks for the traffic control parameter are made, the crash can be assigned to the correct roadway component.

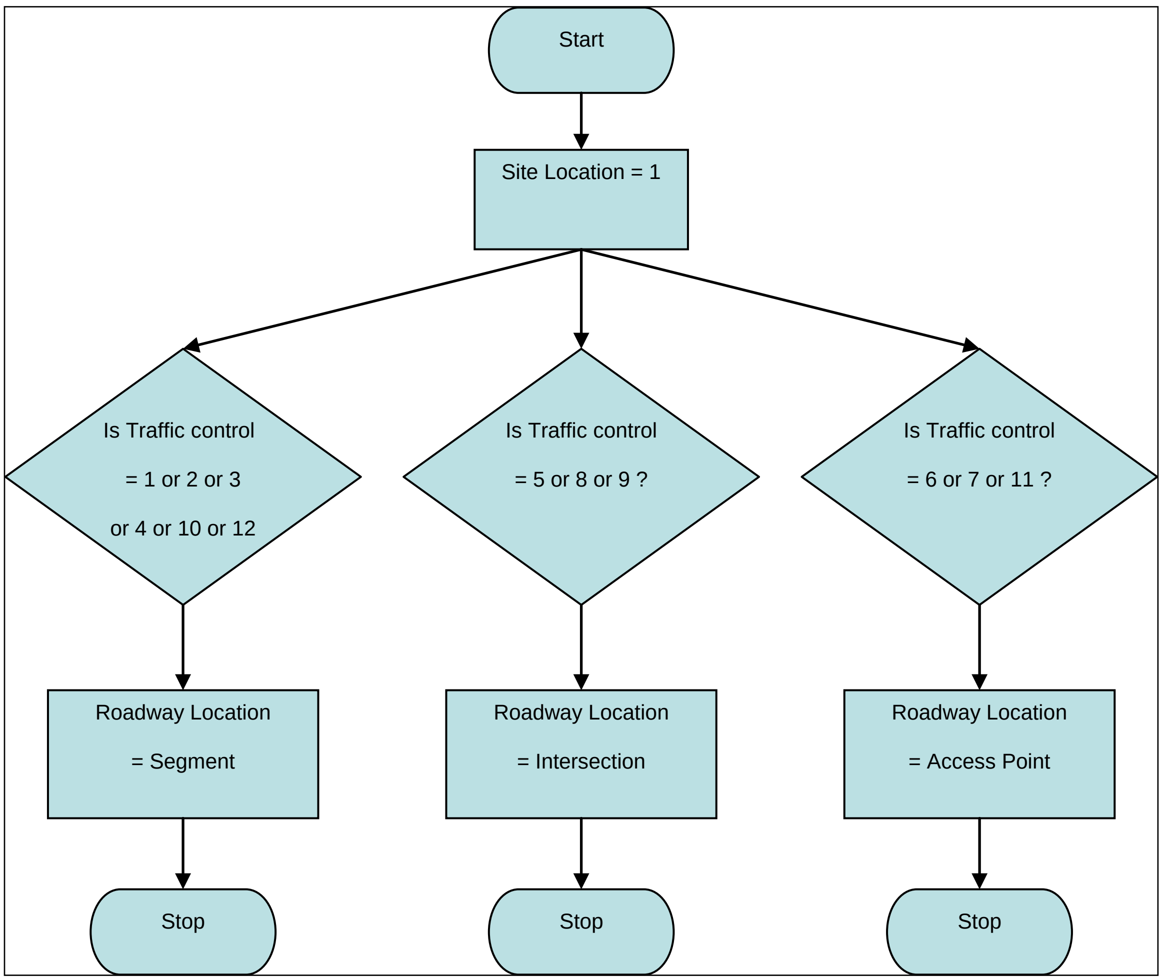
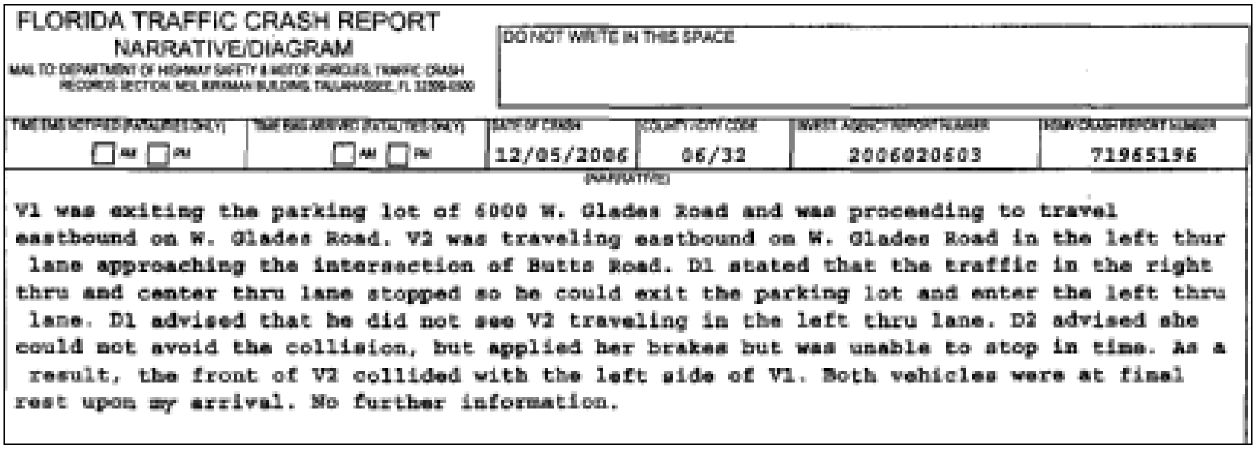


Figure 4-5: Rules for Assigning Crashes to Roadway Locations, Based on Site Location = 1

4.3 Site Location 2: At Intersection

The site location value of 2 indicates that the crash has taken place inside a signalized intersection. However, as mentioned earlier, several of the crashes that have occurred at unsignalized intersections are also sometimes reported as intersection crashes. Therefore, for these crashes some new parameters, in addition to site location and traffic control, have to be considered in order to distinguishing signalized from unsignalized intersection-related crashes. Here, the node information (i.e. whether the crash is signalized or unsignalized) is used for correctly assigning crashes to intersections or access points; that particular variable is not necessary for traffic control values of 5, 6, 7, 8, 9 or 12, where no conflict can be identified. Figure 4-6, page 51, through Figure 4-9, page 52, will illustrate how a conflict may arise and thus support the use of this new binary variable. For example, Figure 4-6 and Figure 4-7, both in page 51, of crash report #719651960 denote an access-related crash. The sole combination of site location value of 2 and traffic control value of 1 cannot help resolve the problem of misclassification; therefore, it is necessary to know the signal information of that particular node.



FLORIDA TRAFFIC CRASH REPORT
NARRATIVE/DIAGRAM

DO NOT WRITE IN THIS SPACE

		DATE OF CRASH	COUNTY / CITY CODE		
☐ AM ☐ PM	☐ AM ☐ PM	12/05/2006	06/32	2006020603	71965196

(NARRATIVE)

V1 was exiting the parking lot of 6000 W. Glades Road and was proceeding to travel eastbound on W. Glades Road. V2 was traveling eastbound on W. Glades Road in the left thur lane approaching the intersection of Butts Road. D1 stated that the traffic in the right thru and center thru lane stopped so he could exit the parking lot and enter the left thru lane. D1 advised that he did not see V2 traveling in the left thru lane. D2 advised she could not avoid the collision, but applied her brakes but was unable to stop in time. As a result, the front of V2 collided with the left side of V1. Both vehicles were at final rest upon my arrival. No further information.

Figure 4-6: Crash Narrative by the Police Officer

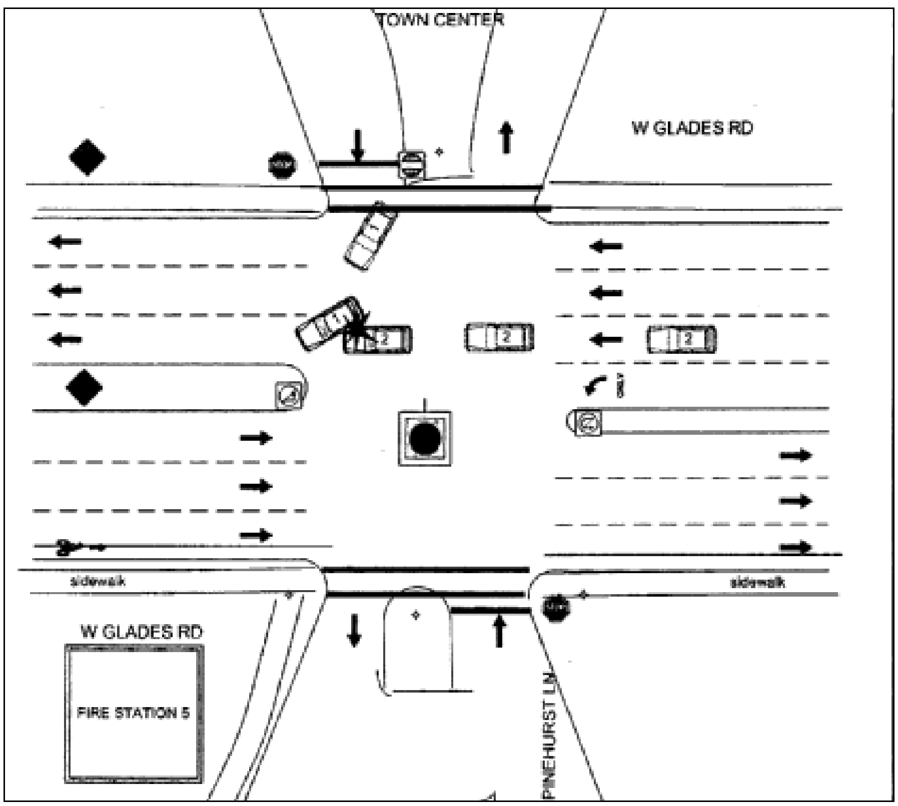


Figure 4-7: Police Officer's Graphical Representation of How the Crash Has or May Have Occurred

Figure 4-8 and Figure 4-9, both in page 52, of crash report #719651790 clearly indicate that the crash is a signalized intersection crash. The site location is 2 and the traffic control is 1.

By just observing the site location or the simple combination of the site location and traffic control variables some of the crashes, but not most of them, can be correctly assigned; therefore, the node check variable is needed in this situation.

FLORIDA TRAFFIC CRASH REPORT
NARRATIVE/DIAGRAM
MAIL TO: DEPARTMENT OF HIGHWAY SAFETY & MOTOR VEHICLES, TRAFFIC CRASH RECORDS SECTION, NEIL KIRKMAN BUILDING, TALLAHASSEE, FL 32399-0500

DO NOT WRITE IN THIS SPACE

TIME EMS NOTIFIED (FATALITIES ONLY)	TIME EMS ARRIVED (FATALITIES ONLY)	DATE OF CRASH	COUNTY / CITY CODE	INVEST. AGENCY REPORT NUMBER	HSMV CRASH REPORT NUMBER
☐ AM ☐ PM	☐ AM ☐ PM	12/07/2006	06/32	2006020754	71965179

(NARRATIVE)

V1 was traveling Northbound on NW 15th Ave in the inside left turn lane to head westbound on W. Glades Rd. V2 was traveling Northbound on NW 15th Ave in the inside left turn lane to head westbound on W. Glades Rd. and stopped for traffic in the intersection with Glades Rd directly infront of V1. V1 failed to stop intime causing the front of her vehicle to collide into the rear of V2.

It should be noted, D1 stated a third vehicle was involved. The third vehicle was directly infront of V2. The license plate given to me of this vehicle (CW tag:131SVA) came back no record found in NCIC/FCIC.

Figure 4-8: Crash Narrative by the Police Officer

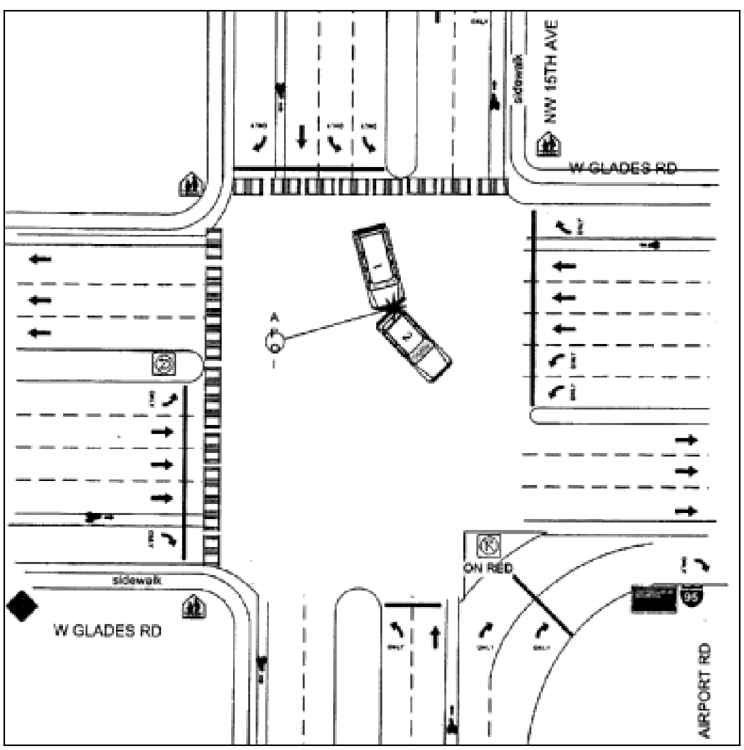
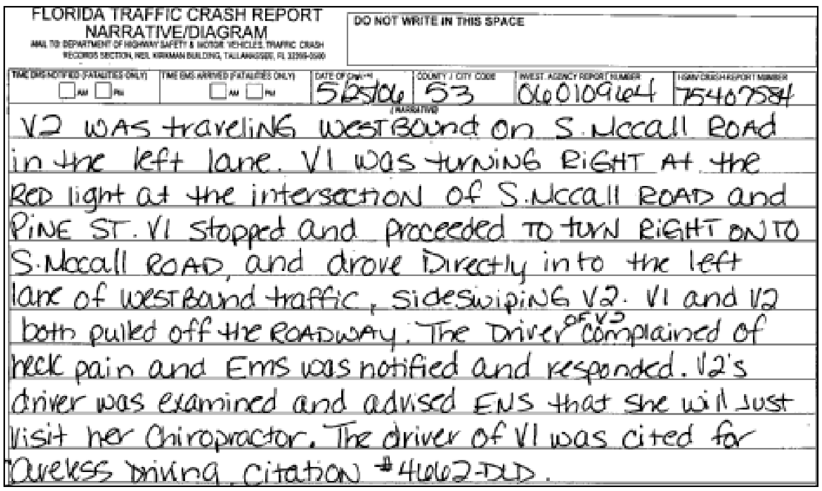


Figure 4-9: Police Officer's Graphical Representation of How the Crash Has or May Have Occurred

Also, Figure 4-10 and Figure 4-11 below provide an example when the site location has a value of 2 and where the node information is not necessary, so that the simple rules may be applied. Crash report #754075840 has a traffic control value of 5 (i.e. traffic signal), which is a clear example of a signalized intersection-related crash.



FLORIDA TRAFFIC CRASH REPORT
NARRATIVE/DIAGRAM
MAIL TO: DEPARTMENT OF HIGHWAY SAFETY & MOTOR VEHICLES, TRAFFIC CRASH RECORDS SECTION, NEIL KIRKMAN BUILDING, TALLAHASSEE, FL 32399-0500

DO NOT WRITE IN THIS SPACE

TIME EMS NOTIFIED (FATALITIES ONLY)	TIME EMS ARRIVED (FATALITIES ONLY)	DATE OF CRASH	COUNTY / CITY CODE	INVEST. AGENCY REPORT NUMBER	HSMV CRASH REPORT NUMBER
☐ AM ☐ PM	☐ AM ☐ PM	5/25/06	53	060109464	75407584

NARRATIVE

V2 WAS traveling westbound on S. Mccall Road in the left lane. V1 was turning RIGHT At the Red light at the intersection of S. Mccall ROAD and PINE ST. V1 stopped and proceeded TO turn RIGHT ONTO S. Mccall ROAD, and drove Directly into the left lane of westbound traffic, sideswiping V2. V1 and V2 both pulled off the ROADWAY. The Driver of V2 complained of neck pain and EMS was notified and responded. V2's driver was examined and advised EMS that she will just visit her Chiropractor. The driver of V1 was cited for Careless Driving. Citation #4662-DLD.

Figure 4-10: Crash Narrative by the Police Officer

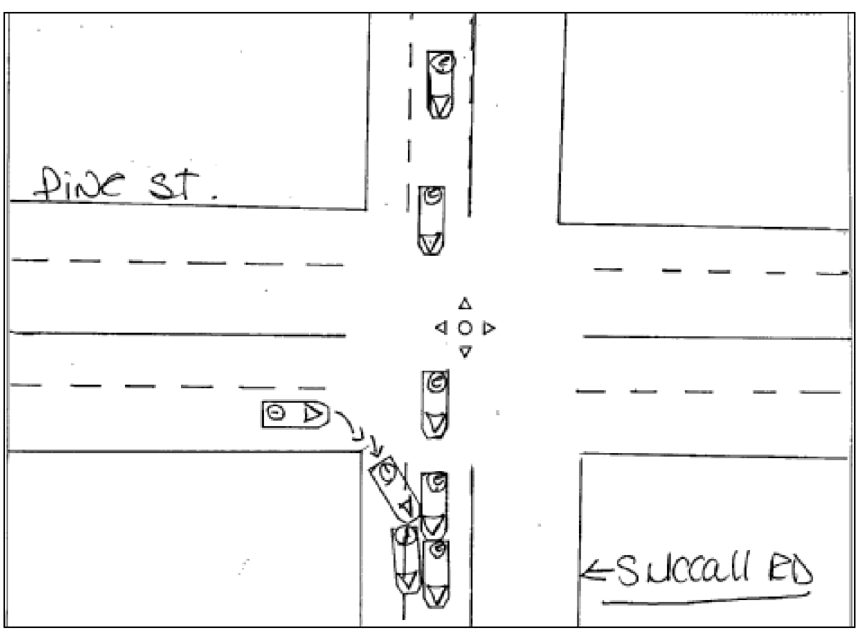


Figure 4-11: Police Officer's Graphical Representation of How the Crash Has or May Have Occurred

Similarly, when the traffic control is 6 (i.e. stop sign) crashes are usually access-related. Therefore, the traffic control value is sufficient in some cases for distinguishing signalized intersection crashes from access-related crashes; however, some traffic control values are not enough for this purpose. Figure 4-12 below is the flow chart of how a crash has to be appropriately assigned to one of the three roadway elements when considering a site location value of 2. After all the checks related to traffic control and node signalization are made, a crash can be assigned to the correct roadway component.

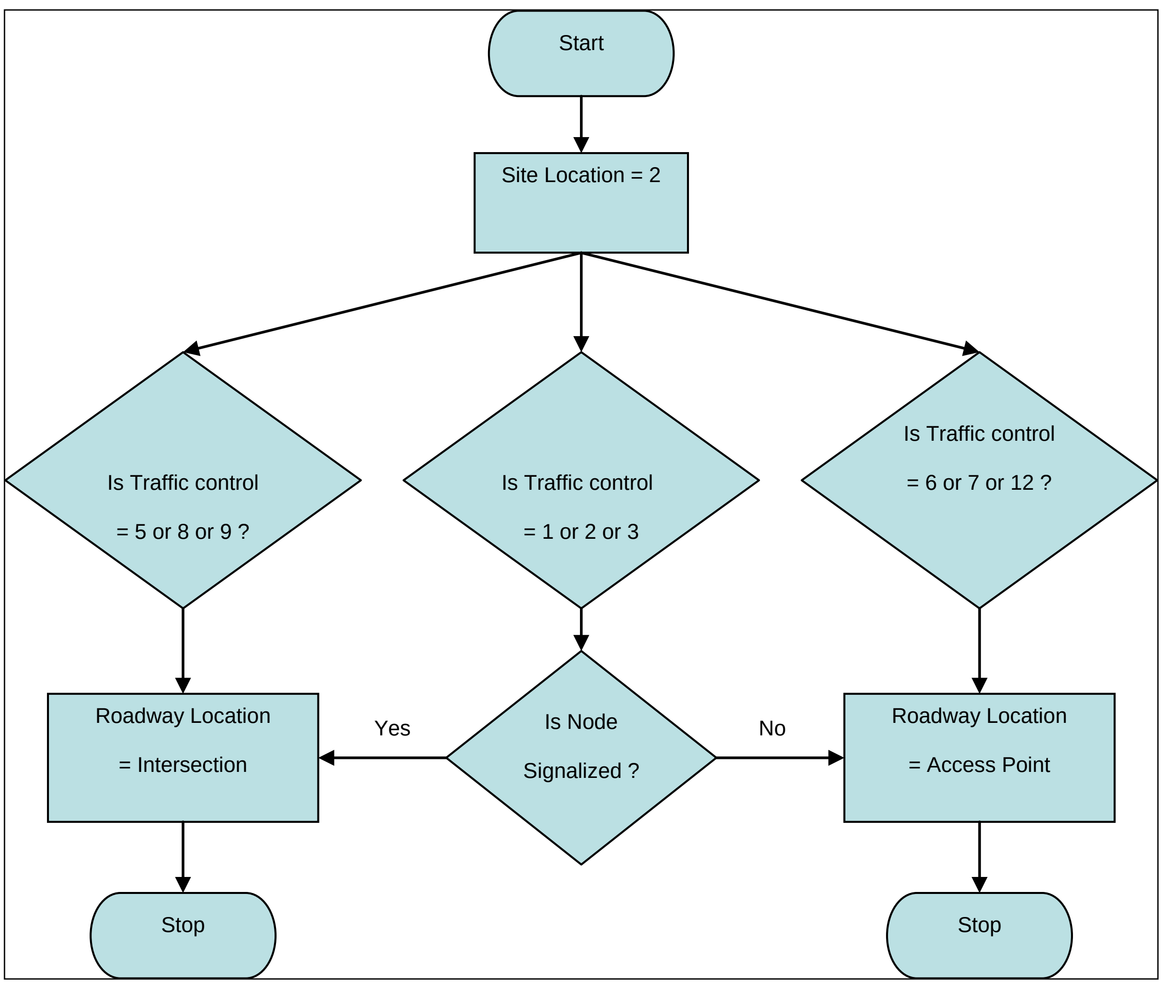


Figure 4-12: Rules for Assigning Crashes to Roadway Locations, Based on Site Location = 2

4.4 Site Location 3: Influenced by Intersection

The site location value of 3, which identifies crashes influenced by intersection site location, has the exact same issues as the site location value of 2. Certain values of traffic control are capable of correctly assigning crashes while some others are not; therefore, these rules are almost similar to those developed for the site location value of 2. Figure 4-13 below provides the rules for the present site location.

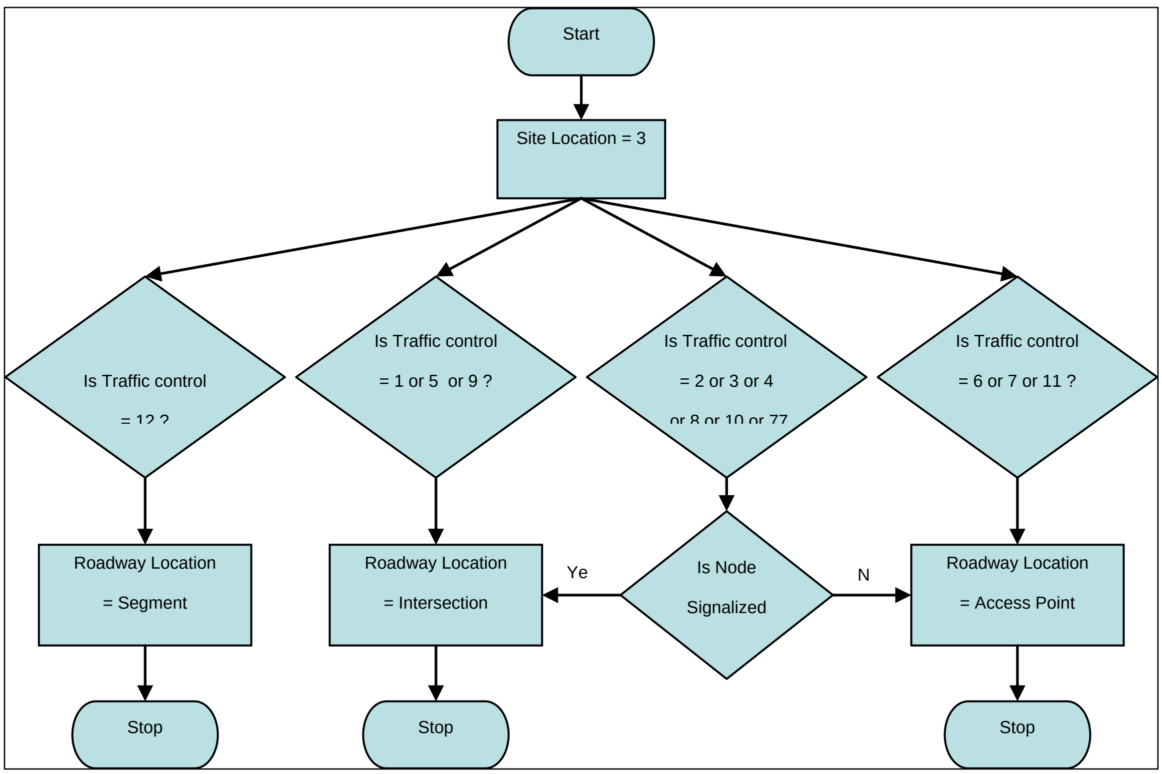


Figure 4-13: Rules for Assigning Crashes to Roadway Locations, Based on Site Location = 3

4.5 Site Location 4: Driveway Access

Driveway access-related crashes have been assigned as access-related crashes. It was mentioned earlier that for the present analysis driveways and some other unsignalized intersections are considered to be access-related. For this particular site location, most crashes are related to access points, except for the cases when the traffic control is 5 (i.e. traffic signal) or 8 (i.e. flashing light). The corresponding rules are provided in Figure 4-14 below.

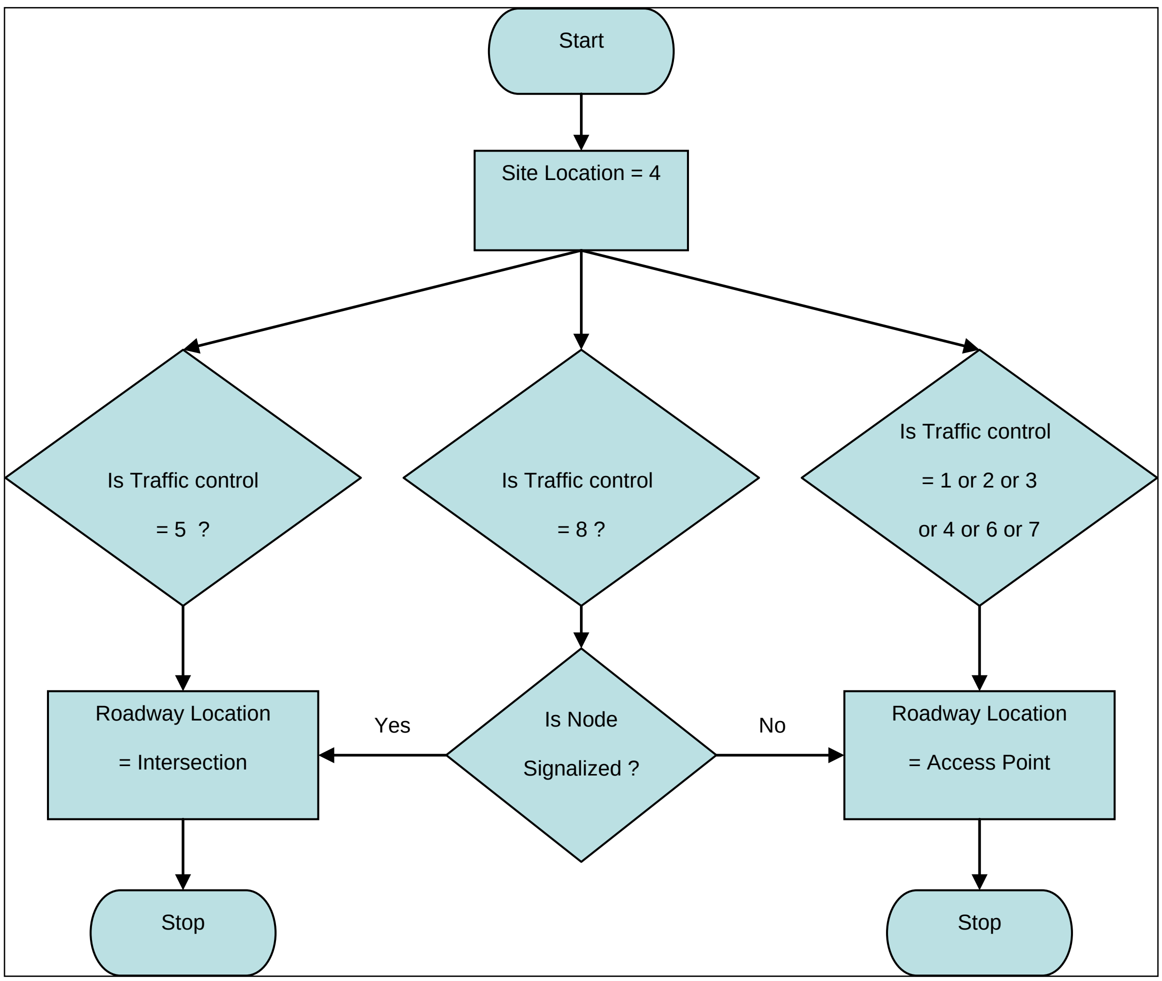


Figure 4-14: Rules for Assigning Crashes to Roadway Locations, Based on Site Location = 4

4.6 Site Location 5: Railroad

The site location value of 5 (i.e. railroad) helps in identifying those crashes which have occurred at or near a railroad intersection. The rules for node checking are required for certain values of traffic control. The corresponding rules for correctly assigning the site location are provided in Figure 4-15 below.

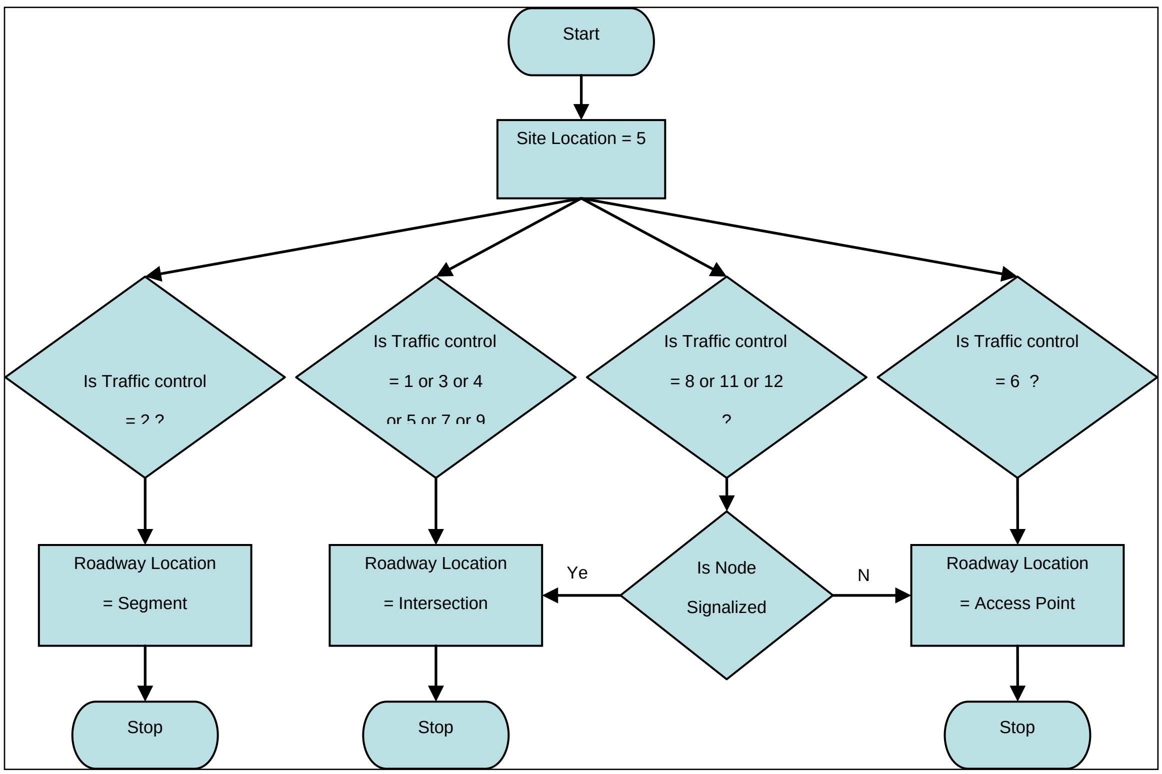


Figure 4-15: Rules for Assigning Crashes to Roadway Locations, Based on Site Location = 5

4.7 Site Location 6: Bridge

The site location value of 6 (i.e. bridge) helps in identifying those crashes which have occurred at or near a bridge. Most of these crashes are segment-related; still, the assigned rules have to be followed in order to correctly assign the crashes that could not have been related to the respective segment. The corresponding rules' flow chart is provided in Figure 4-16 below.

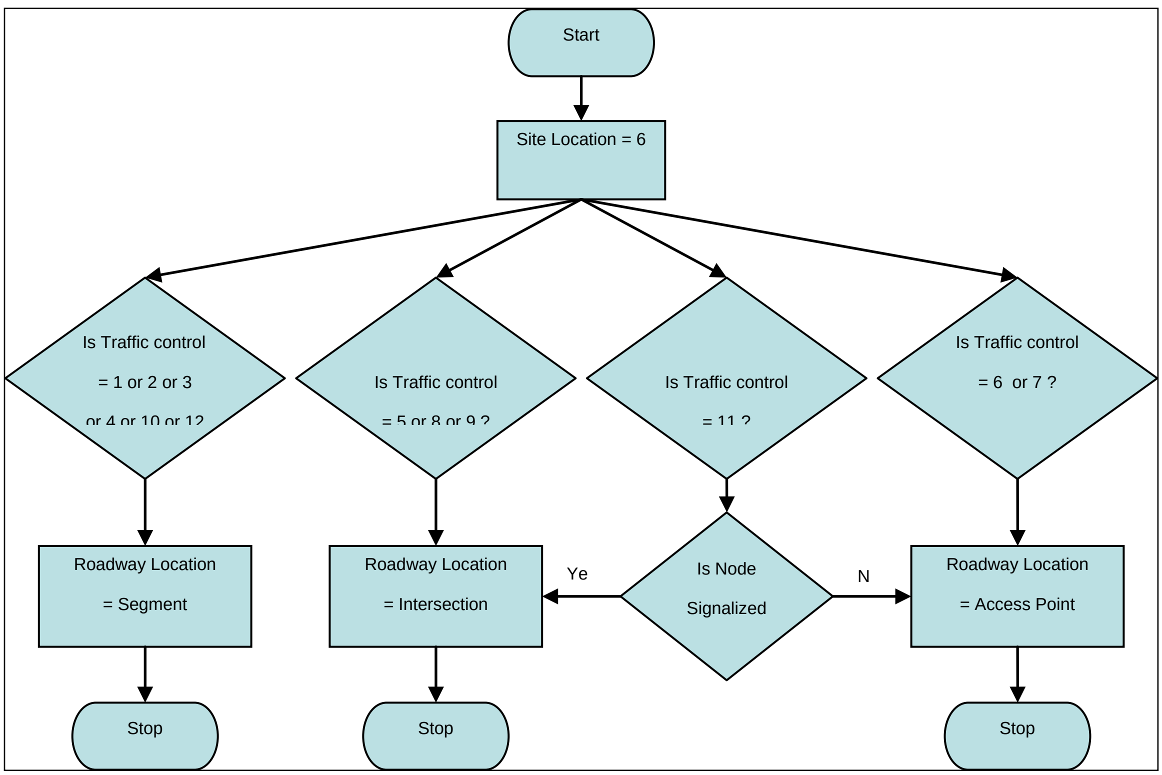


Figure 4-16: Rules for Assigning Crashes to Roadway Locations, Based on Site Location = 6

4.8 Site Locations 7 / 8: Entrance / Exit Ramp

The site locations for entrance and exit ramps are essentially intersections which could be signalized or unsignalized. The traffic control will be used along with the node check procedure in order to assign crashes correctly, whether these are signalized intersection-related or access-related. Figure 4-17 below shows the flow chart with the corresponding rules.

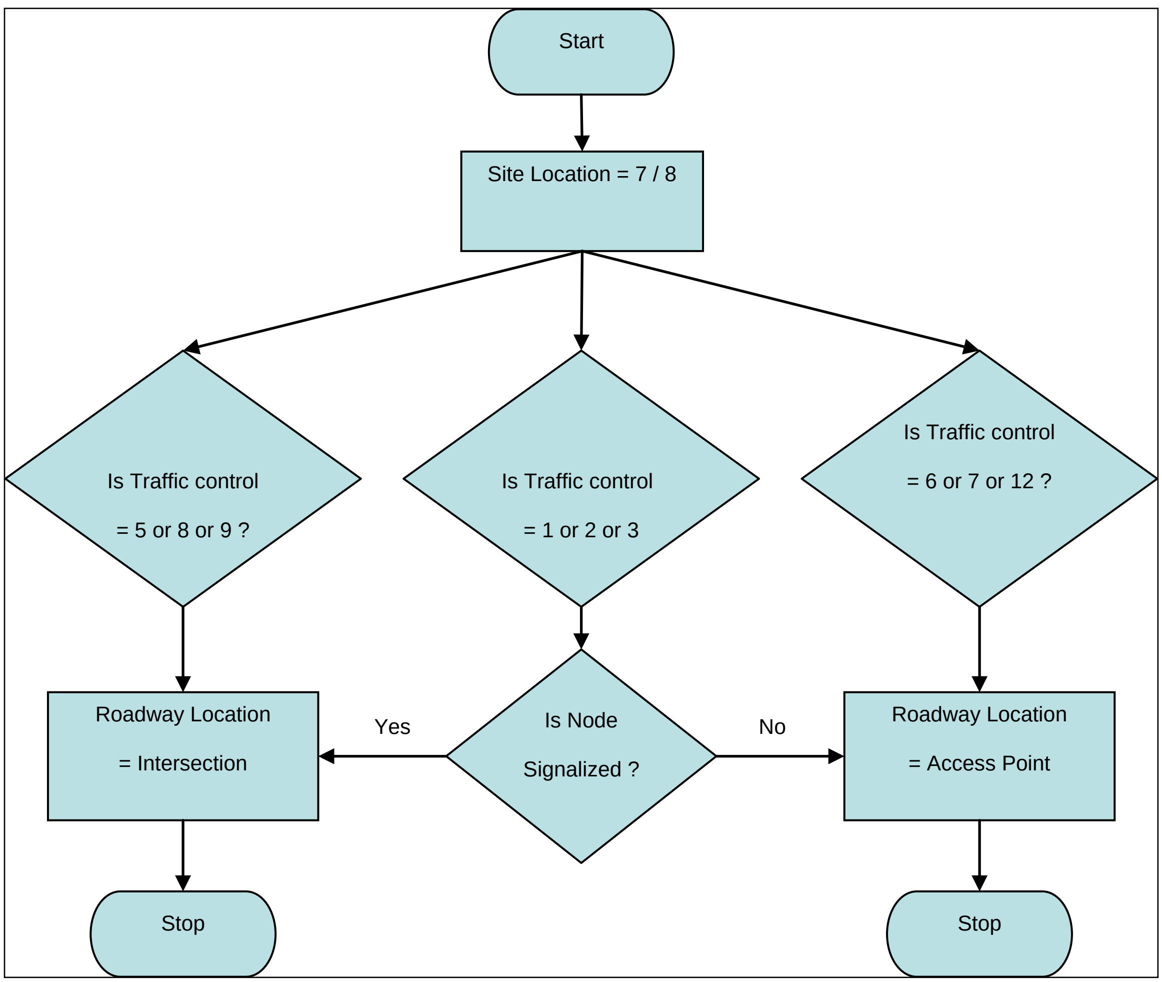


Figure 4-17: Rules for Assigning Crashes to Roadway Locations, Based on Site Location = 7 / 8

4.9 Site Location 13: Public Bus Stop Zone

The site location value for public bus stop zone correspond to segment-related crashes, some being intersection or access-related. The corresponding rules are provided in Figure 4-18 below.

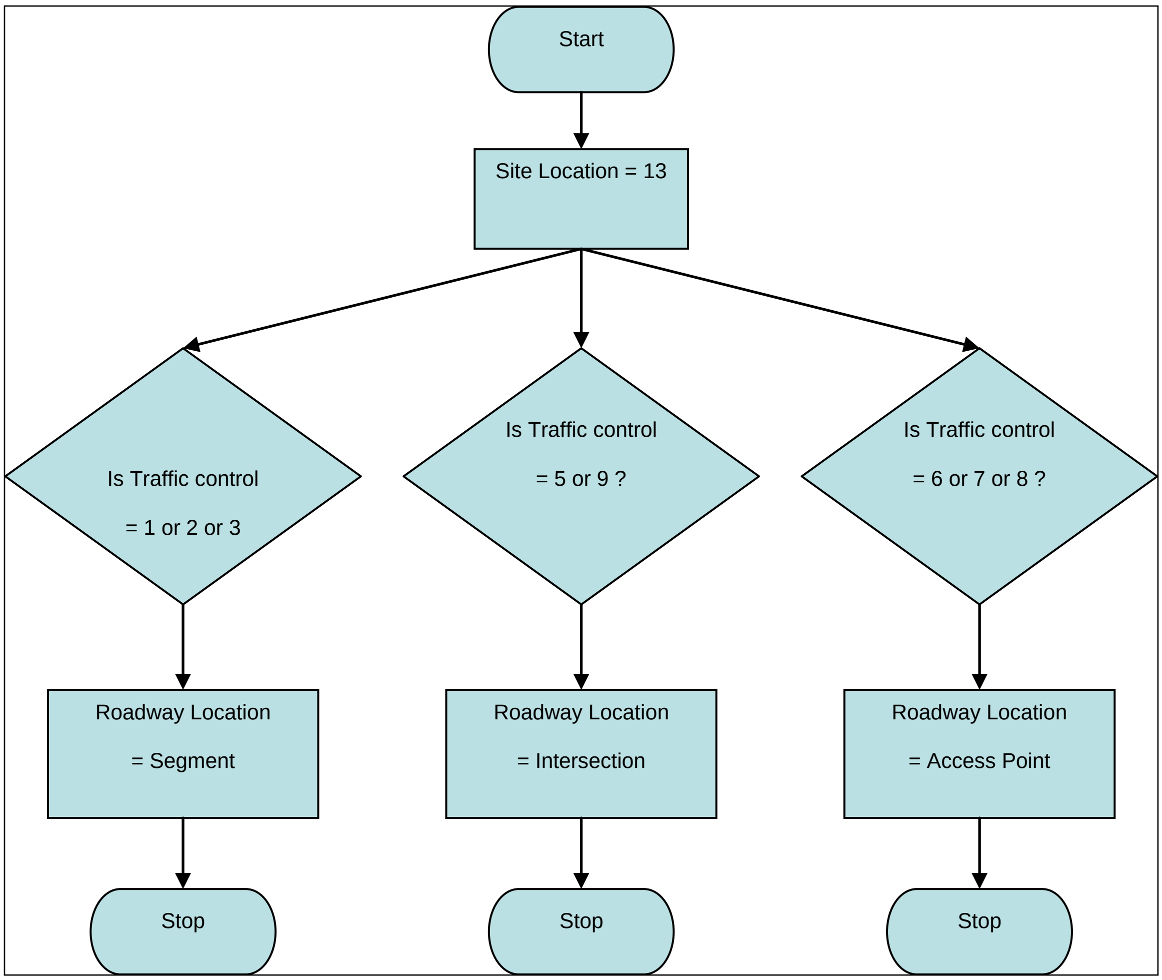


Figure 4-18: Rules for Assigning Crashes to Roadway Locations, Based on Site Location = 13

4.10 Quantitative Validation of the Rules

As mentioned earlier, all these rules have been based on a very careful examination of crash reports, considering different combinations for site location and traffic control. Though these rules have not been developed through any statistical process, a quantitative validation is required not only for evaluating how good they perform but also for having an estimate of how much better they are provided the site location is taken into consideration. The crash reports were thoroughly scrutinized and each crash was assigned to one of the three roadway locations defined earlier: segments, signalized intersections and access points. Recalling the crash reports, these rules were developed by analyzing 96 of these, trying different combinations of site location and traffic control; this essentially constituted a training set of crash reports. The rules were then validated using other 281 crash reports. Therefore, a total of 377 crash reports were examined in order to come up with a complete set of rules.

Out of the first 96 crash reports, the assigning accuracy without the rules (i.e. by using only the site location) was 53.13% whereas the accuracy using the rules was 87.5%. Even at the training or development stage of the rules, a considerable improvement in the crash assignment to the correct roadway element was observed. The validation crash reports gave an assigning accuracy of 59.43% without the rules, and with the rules this accuracy improved to 95.73%. Overall, the accuracy for all 377 crash reports was 57.82% without the rules and it improved to 93.63% with the rules. Therefore, it can be observed that approximately 36% more crashes are assigned correctly by using the rules than by not using them.

CHAPTER 5. DATABASES AND CLUSTERING

5.1 Existing Databases

The FDOT has two very comprehensive resources, namely the Crash Analysis and Reporting (CAR) System and the Roadway Characteristics Inventory (RCI). Since the focus of this study is to investigate severe injury/fatality crashes occurring on Florida's state roads, the aforementioned databases have been used for this purpose since they provide all the necessary traffic- and geometric-related variables.

5.1.1 CAR

The CAR system database has records for all crashes that occurred in the state of Florida, particularly those that required a Florida Traffic Crash Report Long Form to be filled out. The crash records provide information at the crash, vehicle, person and citation levels; this makes this database to be a very exhaustive resource. These records can be viewed online by authorized users and can also be downloaded in text format. For this study, the particular datasets from CAR that were used are as follows: 1) Augmented Detail Extract, and 2) Vehicle – Driver – Passenger Extract; the former dataset has information on the crash characteristics associated with roadway geometry and environmental conditions, whereas the latter option has driver and passenger information for all vehicles involved in a crash. Each of these datasets provides information for 86 variables. In addition to these datasets, CAR also counts with statistical reports for *high crash* roadway segments across Florida. These crash locations (segments) are termed *high crash*, and have been statistically confirmed to be problematic areas, considering certain confidence level and a minimum number of crashes; the default values for the confidence level and minimum number of crashes are 99% and eight crashes, respectively. The present study investigates corridors for the entire state; therefore, the setting of values for the confidence level and minimum number of crashes requires making some assumptions which may prove or not to be incorrect later on. Based on the latter statement, it was decided to generate those reports with all values set to zero; this also helped to study the crash information on all roadway segments across the state. Figure 5-1 below provides a snapshot of the report generated for roadway segments.

REPORT..CARPJ66- FLORIDA - DEPARTMENT OF TRANSPORTATION PAGE NO 1
DATE....07/19/2007 C A R - CRASH ANALYSIS REPORTING SYSTEM
TIME....16:17:23 REFERENCE CRASH ROADWAY SEGMENTS FOR 2005 USING: STATE AVERAGES AS OF: 08/11/2006
COMMENT: ______________ USERID: SF945AP DISTRICT:

NUMB	COSECSUB	BMP	EMP	STROAD	LENGTH	CC	CRASHES	ADT	ACTUAL	AVERAGE	CONLV	FTL	INJ	PRTY	CL-1	CL-2	CL-3
4708	75000001	0.000	2.200	CR 436A	2.200	UNKN	0	N/A	*******	N/A	*****	0	0	0			
4709	75000012	0.000	7.844	CR OFF	7.844	UNKN	0	N/A	*******	N/A	*****	0	0	0			
4710	75000151	0.000	1.520	CR OFF	1.520	UNKN	0	N/A	*******	N/A	*****	0	0	0			
4711	75000154	0.000	2.821	CR OFF	2.821	UNKN	0	N/A	*******	N/A	*****	0	0	0			
4712	75000200	0.000	0.402	CR OFF	0.402	UNKN	0	N/A	*******	N/A	*****	0	0	0			
4713	75000227	0.000	6.350	CR OFF	6.350	UNKN	0	N/A	*******	N/A	*****	0	0	0			
4714	75000260	0.000	1.260	CR OFF	1.260	UNKN	0	N/A	*******	N/A	*****	0	0	0			
4715	75000267	0.000	0.835	CR OFF	0.835	UNKN	0	N/A	*******	N/A	*****	0	0	0			
4716	75001500	0.000	2.608	CR 532	2.608	UNKN	0	N/A	*******	N/A	*****	0	0	0			
25	75002000	0.000	0.220	SR 482	0.220	U-6DR	44	38,497	14.233	3.338	99.99	0	62	14			
233	75002000	0.220	0.520	SR 482	0.300	U-4DR	29	39,625	6.683	2.699	99.99	0	28	13			
642	75002000	0.520	0.682	SR 482	0.162	U-4DR	13	40,498	5.428	2.699	99.50	1	10	8			
2642	75002000	0.682	0.957	SR 482	0.275	S-4DR	3	40,498	0.738	1.461	25.24	0	2	1			
3223	75002000	0.957	1.367	SR 482	0.410	U-4DR	4	40,496	0.660	2.699	12.22	1	4	0			
4717	75002000	1.367	1.474	SR 482	0.107	S-4DR	0	40,499	0.000	1.461	0.00	0	0	0			
3089	75002000	1.474	2.110	SR 482	0.636	U-4DR	9	46,760	0.829	2.699	15.35	0	7	5			
1659	75002000	2.110	2.510	SR 482	0.400	S-4DR	11	47,500	1.586	1.461	50.00	0	5	6			
4718	75002000	2.510	2.810	SR 482	0.300	S-4DR	0	47,499	0.000	1.461	0.00	0	0	0			
54	75002000	2.810	3.280	SR 482	0.470	S-4DR	47	45,588	6.009	1.461	99.99	0	48	26			

Figure 5-1: Snapshot of the *High Crash* Reference Report for Roadway Segments

The column *NUMB* in the report is a reference number that denotes segments from highest to lowest based on a specific criterion. In addition, this report provides information for *adt*, number of crashes, actual crash rate, average crash rate, fatalities, injuries, as well as property damage crashes along the particular length of the selected segment. The actual crash rate (crashes per million vehicle miles) can be determined by dividing the number of crashes along the segment for a certain time span by the total vehicle miles for the segment in that particular time span. Furthermore, information on the roadway design type (urban, suburban, and rural) is also provided. The information gathered will be used later for combining continuous roadway segments.

5.1.2 RCI

The RCI database contains all the essential roadway traffic- and geometric-related information maintained by the state; a detailed list of all the aforementioned roadway features and characteristics can be found in the RCI Office Handbook (FDOT, 2007). This database has 107 roadway characteristics for each roadway segment, and its data can be downloaded from the FDOT mainframe. Among the roadway features considered for this study's purposes are: functional classification, curvature of roadway segments, intersection type, etc. The features obtained from RCI will be integrated with the variables from CAR for the analyses.

In addition to roadway characteristics, RCI also provides a plethora of reports online with information that could be used in specific ways. For this study, lists of roadway segments that are part of multilane arterial segments were downloaded; also, the respective list of signalized intersections was also retrieved from the RCI website.

5.2 Data Preparation

First, it was critical to have a clear definition for a corridor. The FDOT does not have an exact definition, so it was critical that the analysis started by defining a corridor. There were several parameters on which to base the definition of a corridor; still, the most important requirement for a defining parameter is that it should be able to make corridors homogenous in one way or the other.

A representative state road is comprised of different roadway segments, which are typically a representation of administrative boundaries. Any change in the administrative boundaries is bound to affect the length of the corridors being studied; thus, the choice of using these managerial roadway segments is ruled out as the homogeneity will not be consistent.

Another parameter to consider was the median type. There are two main types of physical medians: divided and undivided. It has to be noted that several of the selected corridors were less than 1 mile in length; this can be explained by the fact that as an arterial winds its way through the geographical area, cutting across various residential areas, the median type changes very frequently, hence, the very large number of smaller corridors. Though the method could provide a homogenous section, it was found unacceptable because of the reason just mentioned.

Other parameter considered was the design type. For the roadway design of arterials, the following three design types were considered: urban, suburban and rural. The features that distinguish these three types are the drainage type and city limits. For example, urban roads have a curb and gutter design within city limits or urban residential areas, whereas roads with open drainage but within city limits are categorized as suburban, and roads with open drainage and outside the city limits are categorized as rural. The resulting corridors had a higher number of longer homogenous sections; however, a large number of these roads were still less than 1 mile long. Therefore, a refinement was made based on the design and the city limits: the roadways with urban/suburban design were combined together –thus giving rise to section within city limits–, whereas the rural roads –outside the city limits– were combined together. After this procedure, the number of sections with length of less than 1 mile was reduced. These sections with very short length were later removed from further analysis; the reasons to drop them were twofold: 1) the sectional characteristics will not change much for such short lengths, and 2) the total number of severe crashes for most of those corridor sections was very small.

5.3 Clustering

Having the corridors grouped according to roadway design and city boundaries, the next task to undertake was the analysis. Note that the corridor lengths varied from 1 mile to 78 miles; this wide variation in section length justified a clustering of the corridors based on the length itself. In addition, note that corridors with similar length are more likely to have similar properties; thus, the variations will be similar (i.e. the heterogeneity will be minimized).

First, there is a need to fulfill one of the more difficult tasks in cluster analysis: to find the optimum number of clusters to which the corridors will be assigned. In this study, the Partitioning Around Medoids (PAM) algorithm has been used in order to find the optimum number of clusters; this algorithm operates based on the average dissimilarity. According to Kaufman and Rousseeuw (1990), the *medoid* is an object within a cluster having a minimal average dissimilarity to all the objects in that cluster. Once the medoids are identified, all the objects are assigned to the nearest medoid. The objective function is the sum of the dissimilarities of all objects to the nearest medoid. The algorithm terminates when the interchange of an unselected object with an already selected object no longer minimizes the objective function. In order to find the optimum number of clusters and to differentiate a bad cluster from a good one, a set of values called *silhouettes* are computed (UNESCO, 2007). Following is an algorithm showing how one would calculate the silhouette value.

Consider any object k in the data and let it be assigned to a cluster X. Let x(k) be the average dissimilarity of the object k to all other objects in cluster X. For any other cluster Y different from X, let d(k, Y) be defined as the average dissimilarity of object k to all objects in Y. d(k, Y) for all clusters Y not equal to X is computed, and the smallest is computed. If the minimum is attained in cluster Z, then d(k, Z) = z(k) and Z is the neighbor of object k. Thus, the silhouette value s(k) is defined as:

$$s(k) = (z(k) - x(k)) / \min(z(k), x(k)) \quad (5.1)$$

A silhouette with a value close to 1 suggests that in-cluster dissimilarity is less than the between dissimilarity, whereas a silhouette with a value of 0 suggests that the object could have belonged to either cluster. Also, negative values of a silhouette, especially those that are close to -1, suggest that the clustering has been poorly done. Then, the silhouette values computed can be used for indentifying the optimal number of clusters. In this study, the optimal number of clusters by using the PAM algorithm was found to be four (see Table 5-1 below). Once the optimal number of clusters has been defined, the actual clustering can be done.

Table 5-1: Clusters and their Respective Ranges

Cluster	**Range (in Miles)**
1	1.009 – 2.89
2	2.898 – 5.729
3	5.762 – 10.556
4	10.644 – 78.293

CHAPTER 6. SEVERE CRASH PATTERNS AT SEGMENTS OF MULTILANE ARTERIALS WITH PARTIALLY LIMITED ACCESS

6.1 Introduction

Multilane arterials consist of signalized/unsignalized intersections joined by mid-block segments. Safety assessment for multilane arterials (or any roadway for that matter) is generally based on two broad criteria: 1) crash counts or crash rate (i.e. counts normalized by *vmt*), and 2) crash injury severity. Regarding crash counts or crash rates, these are traditionally estimated using negative binomial regression models (e.g. Knuiman et al., 1993; Abdel-Aty and Radwan, 2000); since crash counts might not be a linear function of traffic flow and section length, crash frequency models having *adt* and section length as independent variables are more appropriate (Caliendo et al., 2007). With regards to severity-based analysis, it classifies crash outcomes in terms of injury severity levels (e.g. Abdel-Aty, 2003; Yau, 2004); injury severity outcomes may be formulated as binary (severe vs. non-severe) (e.g. Yau, 2004), ordinal (e.g. Abdel-Aty, 2003), or multinomial target variable (e.g. Shankar and Mannering, 1996). Based on the aforementioned statements, this study aims to identify the traffic and highway design parameters that are significantly associated with severe crashes on the *segments* of multilane arterials. For this purpose, an alternative to the traditional approaches is proposed, which involves a binary classification relying on the comparison of *crash* and *non-crash* cases; the binary target variable takes a value of 1 for crash cases and a value of 0 for the non-crash cases.

The need for an alternative method for safety assessment of multilane arterial segments arises due to some concerns related to the traditional approaches. As mentioned by Golob et al. (2004) and Abdel-Aty and Pande (2007), crash frequency analysis is a collective way to look at crash data, having the dependent variable (i.e. frequency of crashes) calculated by aggregating them over specific time periods (e.g. months or years) and locations. With regards to locations, signalized and unsignalized intersections are well defined entities within the roadway infrastructure; hence, individual intersections constitute logical units for aggregating crash data in the form of crash frequencies (Wang and Abdel-Aty, 2006). For the case of roadway segments, crash frequency analysis requires the aggregation of crash data over segment(s) of certain length(s). For example, Caliendo et al. (2007) divided each direction of a four-lane divided arterial into segments with constant horizontal curvature and longitudinal slope; also, Donnel and Mason (2006) analyzed the crash frequencies for half-mile segments. As it can be observed, the segment(s)' length selection process for aggregating crash data is arguably arbitrary. The comparisons of non-crash data with crash data proposed in this study allow for using crashes themselves as the unit of analysis for assessing arterial safety as a function of geometric design, time of day, etc.

It should be noted that using crashes themselves as units of analysis is not a new approach in the field of traffic safety. Previous studies that analyzed the severity outcomes of crashes, such as those by Abdel-Aty (2003) and Yau (2004), have used the aforementioned approach for assessing –given a crash has occurred– how severe would a crash be. However, comparative analysis between severe and non-severe crashes can be affected by underrepresentation of the least severe crashes in the documented crash data (Abdel-Aty and Keller, 2005). Furthermore, as Milton et al. (2008) have pointed out, the insights provided by these models have limited application in safety improvement programs, especially since they require *event-specific* explanatory variables for producing useable estimates of injury severity outcomes. The approach proposed herein has the advantages of the methodology used by Milton et al. (2008), since it uses *non-event (i.e. crash)-specific* factors affecting severe crashes on roadway sections. Furthermore, underrepresentation of non-severe crashes within the database is not an issue for the proposed approach, as it relies on comparisons between *severe* crashes and *non-crash* data. Of course, no aggregation of crash data over any segment is necessary since individual crashes, as well as non-crash cases, act as the unit of analysis.

The analysis presented here is based on 6,857 crashes (year 2006) corresponding to 151 multilane arterial corridors in Florida, with lengths ranging from 5 to 15 miles. These corridors are composed of signalized intersections and access points without signal control (i.e. unsignalized intersections). Specifically, the analysis focuses on segment crashes that are not affected by the intersecting traffic streams and that may be attributed only to the segments of the corresponding roadways. These crashes have been identified based on an extensive review of crash reports and by using the following information obtained from the crash database: type of crash, traffic control device, site location and contributing cause. The comparison group for these crashes, used to identify the significant factors associated with their occurrence, is a sample of non-crash cases generated through a random selection of milepost location, time of day, and day of week combination corresponding to these arterials. These randomly selected locations and time related to the arterials, considering no crash was observed, are later used for building matched strata of crash and non-crash cases for each of the 151 arterials.

The preliminary analysis based on simple models (i.e. with one covariate at a time) relies on the following comparison cases: 1) c*rash* vs. non-crash and 2) *severe crash* vs. non-crash. Severe crashes are the crashes involving an incapacitating injury and/or fatal injury; on the other hand, non-severe crashes are those involving the rest of injury severity levels, including possible and non-incapacitating injuries. The analysis is extended by developing separate multi-covariate models by individually comparing four different types of severe crashes with a sample of non-crash cases; the aforementioned crash types are defined by the *first harmful event*, as provided in the crash reports. Logistic regression, using within stratum matched sampling of crash and non-crash cases, is the statistical tool of choice.

6.2 Data Extraction and Exploration

As mentioned earlier, this investigation focuses on the crashes attributed to arterial segments. These segment-related crashes are those not dealing with the traffic on the intersecting streets (i.e. the vehicles involved in the crash were neither coming from nor going to the intersecting roads/driveways); therefore, turning volume count data are not of interest in this study. First, the crashes with first harmful event characterized as *Collision with Motor Vehicle in Transport (Left turn)* and *Collision with Motor Vehicle in Transport (Right-turn)* were eliminated from the sample. Next, from the sample of remaining crashes, those that may be attributed to arterial segments only (i.e. those not attributed to signalized/unsignalized intersections) had to be identified. A thorough review of crash reports revealed that the *site location* parameter was a weak indicator by and for itself; for example, it was observed that it is possible for a crash to not be attributed to a signalized intersection even if it may have occurred very close to one. As a way to overcome the latter issue, both *traffic control* and *site location* parameters were combined, resulting in a successful attempt for attributing crashes to one of the three roadway elements of interest (i.e. segments, signalized intersections and unsignalized intersections) associated with the respective crash event. In addition, the crashes with first harmful event characterized as *Collision with Motor Vehicle in Transport (Angle)* were also excluded from the sample provided the contributing cause for the crash was classified as *Improper turn* (i.e. vehicles making improper right/left turns) or *Failed to yield Right of Way* (i.e. vehicles failing to yield the right of way to vehicles going through). Regarding the crashes remaining in the database, these are not attributed to signalized/unsignalized intersections and may be solely attributed to segments of multilane highways.

The segment crash data consisted of 6,857 events with 10.69% of them resulting in fatal or incapacitating injury; the remaining 89.31% of these crashes were non-severe crashes. For the 6,208 crashes counting with information on the type of crash (i.e. first harmful event), the breakdown of severe vs. non-severe crashes by type, along with their respective share within the crash data, is shown in Figure 6-1 below. Again, it must be emphasized that one may expect the share of non-severe crashes within each of the crash types to be even higher than the one shown in Figure 6-1, which can be explained by the well-documented problem of crash underreporting (Abdel-Aty and Keller, 2005).

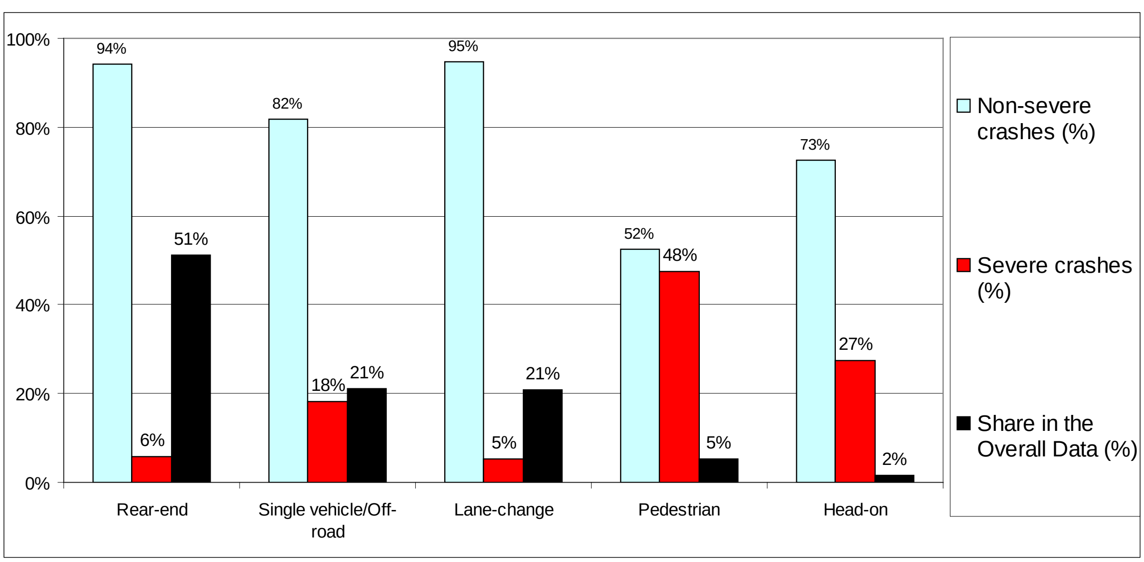


Figure 6-1: Severity of Crashes by Collision Type and their Share in the Overall Crash Data

As shown in Figure 6-1, page 72, the crash data are divided into five collision types: rear-end, single-vehicle/off-road, lane-change-related, pedestrian, and head-on. This categorization is obtained by logically combining the categories of *first harmful event* contained in the crash database; for example, crashes with first harmful events characterized as *Motor vehicle ran into Ditch/Culvert* and *Ran off-road into water* were assigned to the single-vehicle/off-road crash type. These five types of crashes have been arranged from left to right in Figure 6-1 by descending share in the overall crash data. Note that head-on collisions are rare on these multilane highways and constitute only 2% of the data, even though 27% of them are severe. In addition, collisions involving pedestrians have the highest percentage of severe crashes, followed by head-on and single-vehicle/off-road crashes. With regards to lane-change-related and rear-end collisions, both of these have the least percentages of severe crashes. Lane-change-related crashes consist of crashes with first harmful event characterized as *Collision with Motor Vehicle in Transport (Sideswipe)* and *Collision with Motor Vehicle in Transport (Angle)*, where the contributing cause is neither *Improper turn* nor *Failed to yield Right of Way*; therefore, only the angle crashes attributed to arterial segments (i.e. crashes not affected by traffic streams intersecting roadways, either from or turning on to) have been considered. The authors postulated that the latter crashes would never be of the right angle type; therefore, the crashes for which the first harmful event has been characterized –by the law enforcement personnel at the crash site– as *Collision with Motor Vehicle in Transport (Angle)* are essentially lane-change-related crashes. This postulation was verified with a meticulous and manual review of 70 randomly selected crash reports for such crashes.

6.2.1 Extraction of Non-crash Cases

A sample of non-crash cases has been used in the analysis, which acts as control within strata defined by the corridors. These non-crash cases were randomly drawn from each corridor. For this analysis' purposes, the year 2006 was divided into 35,040 15-minute periods (four 15-minute periods per hour * 24 hours * 365 days = 35,040 15-minute periods), which would be the number of options available for choosing the *time of the non-crash* case. Similarly, the group of possible milepost locations for each corridor consisted of mileposts starting at beginning milepost and culminating at the ending milepost, with an increment of 0.001 miles. For example, for a corridor with beginning milepost 0.0 and ending milepost 6.0, there would be 210,240,000 options to select (35,040 * (6.0/0.0001) = 210,240,000 options); these options may include day, time, and location of the non-crash cases. From these non-crash cases, 0.5% of them were randomly drawn from the available options to select for each corridor. The selected non-crash cases for each of the 151 corridors were then matched with the crash cases from the same corridors in order to create 151 strata for within stratum matched sampling framework. The details of this methodology are described in forthcoming sections of this document. Note that it is possible to improve the resolution for the time of non-crash cases in order to make it more precise than at the 15-minute level. However, as it will be noted later, the time information was used for creating broad categories for time of day in the form of peak and off-peak periods; therefore, the 15-minute resolution was sufficient for the time of non-crash cases.

6.2.2 Traffic/Geometric Information for Crash and Non-crash Cases

The next step was to extract geometric design features (e.g. curvature, median type, sidewalk, etc.) based on the milepost locations. These relevant variables for crash and non-crash cases were extracted from the Roadway Characteristics Inventory (RCI) (2001) database. The extraction of traffic/geometric information was based on both the milepost locations and roadway IDs of the arterial corridors. For the crash cases, these were assigned using the actual milepost location of the crash from the FDOT's crash database; for the non-crash cases, these were assigned using the procedure described in the last section.

RCI is a computerized database maintained by the FDOT, which provides basic information on highway design and roadside features for roadways maintained by the state of Florida; this information is indexed by data segments. Also, the aforementioned roadway-related features are listed in the RCI Handbook, also published by the FDOT; Table 6-1, page 77, details the variables relevant to this study which were extracted from the database. Note that the form of most of the tabulated variables is not as the one originally contained in the RCI database; the latter resulted from the combination of the original variable categories contained in the database, which was done with the purpose of obtaining a representative sample size of crash and non-crash data.

Continuous variables such as *adt*, T-factor and K-factor were also categorized since their relationships with severe crash occurrence are not expected to be monotonous in nature. As it will be observed later, the results obtained support the reasoning behind this categorization process. For example, the categorization of the continuous variable *adt* is such that the four resulting categories have the same number of observations; similarly, both the T-factor and K-factor variables resulted in three categories having the same number of observations.

With regards to time of crash (and non-crash cases) and day of week, these two were combined into a single variable representing both of them at the same time. Four categories resulted from this combination: weekday morning peak hour, weekday afternoon peak hour, Friday/Saturday night, and other off-peak periods; it has to be noted the Friday/Saturday night category was separated from the other off-peak periods category because of the increased likelihood of alcohol-impaired driving.

The RCI database also provides pavement condition information in the form of Present Serviceability Rating (PSR) based on the AASHTO Road Test. PSR is based on passenger interpretations of ride quality and is represented with the following categories: 1) 1.00-1.90 Very poor (75% or more deteriorated), 2) 2.00-2.90 Poor (large potholes and deep cracks with discomfort even at slow speeds), 3) 3.00-3.90 Fair (rutting, map cracking with extensive patching), 4) 4.00-4.90 Good (first class ride with only slight surface deterioration), and 5) 5.00 Very good (only new or nearly new pavements). The final variable used for pavement condition had the following three categories: 1) Very poor/Poor pavements, 2) Fair pavements, and 3) Good/Very good pavements.

Presence of horizontal curvature, roadside parking and crash attenuators were also used in the analysis in the form of three binary variables, along with type of median and presence/width of sidewalk. It also has to be noted that information on attenuator type was also available in the RCI database; however, sample sizes for different attenuator types (e.g. Quadguard, DragNet, etc.) were not enough for estimating their effects. Similar information on type of parking was also available but could not be used due to limited sample sizes for individual categories (e.g. angle one side etc.). Median types were consolidated into nine categories as shown in Table 6-1, page 77; with regards to the median width, this variable was sufficiently represented within the median type variable since the former is dependent on the latter. In addition, the presence/width of sidewalk was represented by the variable *sidewalk*. A detailed list of the categories for all of the aforementioned variables is provided in Table 6-1; the variables in this table are not event-specific characteristics (e.g. driver characteristics, seat belt use, etc.) which, as Milton et al. (2008) argued, allows for a more general, non-event-specific interpretation of factors.

Table 6-1: Variables Used in the Analysis

Variable Description	Categories
Posted speed limit	Speed limit <40 mph, 40<= Speed limit <50 mph, 50<= Speed limit <60 mph, and Speed limit >=60 mph
***adt* (annual daily traffic)**	*adt* < 14,900, 14,900 <= *adt* < 26,500, 26500 <= *adt* < 40,000, and *adt* >= 40,000
Average K-factor	K-factor < 9.35, 9.35<= K-factor < 10.14, and K-factor >= 10.14
Average T-factor	T-factor < 4.84, 4.84 <= T-factor < 8.75, and T-factor >= 8.75
Combination of day of week and time of day	Afternoon Peak Weekday Friday or Saturday Night Morning Peak Weekday Off-peak
Pavement condition (PSR)	PSR < 3.0 Very poor/poor 3.00 <= PSR < 3.90 Fair, and PSR >=4.00 Good/very good
Median type	Two-way left turn lane (TWLTL), Grass/lawn, Guardrail, Barrier other than guardrail, Canal or Ditch, Curb < 6 inches, Curb >=6 inches, Paved not for travel, No Median
Sidewalk	No Side Walk, Side Walk <= 6 ft, and Side Walk > 6 ft
Presence of traffic crash attenuators	Yes or No (Binary)
Presence of on-street parking	Yes or No (Binary)
Presence of horizontal curvature	Yes or No (Binary)

One of the variables considered, but not included in the analysis, was sun glare; the variable was available for crash cases from the event reports but was missing for the non-crash cases. The presence/possibility of sun glare variable could be derived from the location and time of day for the non-crash cases, but it was observed that the total number of crashes for which *Glare* was noted as a vision obstruction was only 19; thus, such derivation may not be recommended. Therefore, in order to further investigate this factor, the crash reports with the vision obstruction variable characterized as *Other (Explain)* were examined one by one; it was found that from 99 of such crashes only two were affected by Glare, but to use a sample size of 21 (out of a total of 6,857 crashes) was not sufficient to examine sun glare as a factor.

6.3 Modeling Methodology

Within stratum matched case-control sampling is a recommended approach for the modeling of crash vs. non-crash cases. The purpose of this matched sampling-based analysis is to explore the effects of the variables of interest while controlling for the confounding variables through the design of the study (Abdel-Aty et al., 2004). This approach was used in this study for comparing sample of crash with non-crash cases within the data stratified by the corridors.

Under the matched study design, crash and non-crash cases from each of the 151 arterials of interest form an individual stratum; each stratum consists of crashes and non-crash cases from the corresponding corridor. This sampling procedure is referred as m:n matching, and each corridor (i.e. stratum) can have a different number of crash (m) and non-crash cases (n). Differences between characteristics of crash and non-crash cases from within stratum may then be utilized for the estimation of statistical model(s) for the binary target variable, the latter being 1 for crash cases and 0 for non-crash cases.

Harb et al. (2008) used the m:n matched sampling procedure for comparing work zone with non-work zone crashes. For the present analysis, there would be 151 strata with m crash cases and n non-crash cases within stratum j (j=1,2,……,N). Lets stipulate $p_j(x_{ij})$ to be the probability that i^{th} observation in the j^{th} stratum is a crash with $x_{ij} = (x_{1ij}, x_{2ij},\ldots\ldots x_{kij})$ being the vector of k variables $x_1, x_2,\ldots\ldots x_k$; i = 0, 1, 2,…..m+n-1; and j = 1, 2,……N. The probability $p_j(x_{ij})$ of an observation being a crash may be modeled as follows:

$$logit\ (p_j(x_{ij})) = \alpha_j + \beta_1\, x_{1ij} + \beta_2\, x_{2ij} + \ldots + \beta_k\, x_{kij} \tag{6.1}$$

The intercept term α_j varies per different strata. It summarizes the effect of variables used to form strata on the probability of crash. In order to account for the stratification within the analysis of the observed data, one constructs a conditional likelihood. The likelihood function $L(\beta)$ is independent of the intercept terms $\alpha_1, \alpha_2,\ldots\ldots\ \alpha_N$; therefore, the effects of matching variables cannot be estimated, and Equation 6.1 cannot be used for the estimation of crash probabilities. However, the values of the β parameters that maximize the likelihood function are in fact the estimates of β coefficients in Equation 6.1. Further details on the derivation of the maximum likelihood function may be found in the study done by Collett (1991).

6.4 Preliminary Analysis: Simple Models

The first step in the analysis was to estimate two sets of simple (i.e. with only one covariate) models: 1) one with the binary target variable representing *crash* vs. *non-crash* cases, and 2) other with the binary target variable representing *severe crash* vs. *non-crash* cases. In addition, the purpose of this analysis is twofold: to provide preliminary information on the factors that may be significantly related with crashes and specifically severe crashes, and to focus on the differences, if any, between these two sets of results. Note also that the categorized variables based on the ranges of T-factor, K-factor and *adt*, along with speed limits, width of side walk and pavement surface conditions are not used as ordinal variables but as nominal variables; this nominal scale ensures that one is able to capture the non-monotonous nature of the relationship between these variables and crash occurrence. The simple (one covariate) models are estimated using the TPHREG procedure in SAS (SAS Institute, 2003). Table 6-2, page 82, shows the coefficients of the simple models for each of the variables along with the corresponding p-values. The first two columns depict the results from the set of models when *all crashes* are compared with non-crash cases, whereas the last two columns depict results from the models' when only *severe crashes* are compared with non-crash cases. In addition, the significantly positive (**p-value in bold font**), significantly negative (***p-value in bold-italic font***) and statistically insignificant coefficients (p-value in regular font) are distinguished in the table.

It may also be observed in Table 6-2 that the sections with speed limits less than 60 mph are more likely to have crashes compared to sections with speed limits greater than 60 mph. However, the results from comparing severe crashes with non-crash cases show that the only sections with *40 <=speed limit*< 50 *mph* are more likely to have severe crashes as compared to sections with a speed limit greater than 60 mph (i.e. the base case); these results justify the use this variable on a nominal scale and not on a continuous or ordinal scale. Similarly, while sections with lower *adt* are less likely to have crashes –as expected–, severe crash segments with *adt* ranging from 26,500 through 40,000 are not significantly different from segments with *adt* greater than 40,000. In the first set of models, the positive coefficients for the two levels of the nominal variable representing the K-factor (compared to the base case *K-factor >=10.14%*) suggest that highways with a lower K-factor are more likely to have a crash. It should be noted that while it is not a factor when comparing all severe crashes with non-crash cases, it is possible that it relates with certain types of severe crashes; this issue would be addressed in the next section. Also, based on the coefficients for the two levels of the nominal variable representing the T-factor (compared to the base case *T-factor>= 8.75*), it can be stated that segments with T-factor<4.84% are more likely to have a severe crash; this may be related to the fact that multilane arterials with a low T-factor are expected to have higher pedestrian traffic, which increases the likelihood of pedestrian-related crashes (which in turn have a disproportionately high severe to non-severe crash ratio). However, this inference further highlights the need for segregating the data by crash types. In addition, it has to be noted that while Friday/Saturday nights are barely significant (compared to other off-peak periods, which constitute the base case) at a 90% confidence interval for the overall crash occurrence, their corresponding coefficient is very significant for severe crashes; in fact, in terms of severe crashes, Friday/Saturday nights have the most significant coefficient.

While the presence of horizontal curvature decreases the likelihood of crash occurrence (negative coefficient with p-value<.001), it is not a statistically significant factor when considering only severe crashes (p-value=0.6423); this indicates that if the data were analyzed with the *given a crash has occurred* approach (e.g. severe vs. non-severe crash analysis, or analysis with ordered injury severity levels as the target), one might observe horizontal curvature as a factor that significantly increases severity. The latter inference is actually consistent with findings from Abdel-Aty (2003), who used ordered probit models for analyzing the injury severity of crashes on arterial segments with partial access controls.

Table 6-2: Results from the Preliminary Analysis (One Covariate)

Objective / Variables	Crash vs. non-crash comparisons			Severe crash vs. non-crash comparisons		
	Parameter	Standard error	p-value	Parameter	Standard error	p-value
Traffic-related Parameters						
Speed limit<40 mph	0.877	0.088	**<.0001**	0.154	0.253	0.5425
40<= Speed limit<50 mph	1.074	0.078	**<.0001**	0.629	0.217	**0.0038**
50<= Speed limit<60 mph	0.482	0.080	**<.0001**	0.340	0.216	0.116
Speed limit>=60 mph	.	.	.	.		.
adt <14900	-1.889	0.086	***<.0001***	-1.041	0.228	***<.0001***
14900<=*adt*<26500	-1.089	0.055	***<.0001***	-0.651	0.154	***<.0001***
26500<=*adt*<40000	-0.404	0.042	***<.0001***	-0.187	0.124	0.1336
adt >=40000	.	.	**.**	.		**.**
K-factor<9.35	0.750	0.113	**<.0001**	0.466	0.287	0.1053
if 9.35<= K-factor < 10.14	0.472	0.101	**<.0001**	0.200	0.286	0.4851
K-factor >= 10.14	.	.	**.**	.		**.**
T-factor<4.84	0.491	0.062	**<.0001**	0.383	0.176	**0.0295**
4.84<= T-factor < 8.75	0.357	0.055	**<.0001**	0.186	0.160	0.2435
T-factor>=8.75	.	.	**.**	.		**.**
Afternoon Peak Weekday	0.952	0.039	**<.0001**	0.497	0.121	**<.0001**
Friday or Saturday Night	0.108	0.057	**0.0586**	0.700	0.122	**<.0001**
Morning Peak Weekday	0.422	0.046	**<.0001**	0.049	0.147	0.7386
Off-peak period	.	.	**.**	.		**.**
Highway design-/Pavement-related Parameters						
Two-way left turn lane	1.157	0.102	**<.0001**	1.536	0.358	**<.0001**
Lawn	0.854	0.099	**<.0001**	1.847	0.345	**<.0001**
Guardrail	0.993	0.254	**<.0001**	2.466	0.816	**0.0025**
Barrier other than guardrail	1.462	0.158	**<.0001**	2.185	0.492	**<.0001**
Canal or Ditch	0.935	0.366	**0.0107**	2.769	0.759	**0.0003**
Curb < 6 in	1.296	0.098	**<.0001**	1.937	0.348	**<.0001**
Curb > = 6 in	1.410	0.101	**<.0001**	1.900	0.358	**<.0001**
Paved not for travel	0.958	0.120	**<.0001**	1.478	0.408	**0.0003**
No Median	.	.	**.**	.		**.**
No Side Walk	-0.502	0.053	***<.0001***	-0.293	0.154	***0.0566***
Side Walk <= 6 ft	-0.039	0.047	0.4071	0.008	0.143	0.9573
Side Walk > 6 ft	.	.	**.**	.		**.**
Good/very good pavement	-0.492	0.091	***<.0001***	-0.908	0.285	***0.0014***
Fair pavement	0.035	0.040	0.3819	-0.169	0.117	0.1502
Poor/very poor pavement	.	.	**.**	.		**.**
Horizontal curvature	-0.661	0.124	***<.0001***	-0.142	0.306	0.6423
Attenuators	0.450	0.123	**0.0003**	0.448	0.348	0.1977
Presence of on-street Parking	0.443	0.040	**<.0001**	0.324	0.111	**0.0035**

Regarding corridors without a sidewalk, these are less likely to have crashes as well as severe crashes. The results also show that improved riding quality (represented by the levels of pavement condition) does improve safety with pavements classified as very good/good, having a significantly negative coefficient when compared to poor/very poor pavements. Note that for all crashes, fair pavements are not significantly better than poor pavements (p-value=0.3819). Overall, these results indicate that improving pavement conditions may be a good countermeasure for avoiding segment-related crashes on arterials. Furthermore, even though multilane arterials with attenuators are more likely to experience a crash, at a 90% confidence interval attenuator presence has no significant effect on the occurrence of *severe* crashes; the results indicate that the attenuators are placed at high crash risk locations and indeed reduce the severity of crashes, which is reflected by their statistically insignificant coefficient for severe crash occurrence.

Median types are divided into nine types with the no median as the base case. In the *all crash* model, two-way left turn lane (TWLTL), curb less than 6 inches, and curb more than 6 inches are the three most significantly different from the sections without a median. In terms of severe crashes, the coefficient for grassed median is almost as significant as the curbs (i.e. rural sections with grass medians are more likely to have severe crashes), while in terms of all crashes, median types found in urban areas are the most significant. It is interesting to note that the p-values for the parameters in the all crash models are lower, when compared to the p-values for the same parameters in the severe crash models. The latter could be due to the large sample for all crashes; however, it is also possible that the highway design factors being considered play a more significant role in estimating where a crash (any crash, regardless of severity) is more likely to occur. Crash severity, on the other hand, might be more related to event-specific factors such as speeding, alcohol use etc.; this could lead to higher p-values (i.e. less significance) of highway design parameters for severe crashes. The next section outlines some of the issues with this preliminary analysis, and then expands on it in order to estimate logistic regression models having multiple covariates for severe crash data segregated by type; this extended analysis is useful in drawing more precise inferences.

6.5 Analysis with Multiple Covariates

Insightful as it was, there are two obvious concerns with the preliminary analysis: 1) the crash vs. non-crash comparison may be unreliable since least severe crashes are usually underrepresented (it is well documented that PDO crashes are rarely completely reported), and 2) the crash sample consisted of various crash types that may in fact have different traffic/geometric design variables associated with them.

The analysis detailed in this section focuses on severe crashes only; furthermore, this analysis is carried out for severe crashes being segregated by crash type. Among severe crashes, single-vehicle/off-road crashes were in a plurality with more than 35% of the data followed by rear-end, pedestrian-related, and lane-change-related crashes. The head-on collisions were less than 4% of the severe crashes; also, from Figure 6-1, page 72, it can be observed that head-on crashes constitute only 2% of the overall crash data. Therefore, head-on crashes were not used in the analysis since the absolute number of severe crashes for this type was too low for providing any meaningful sample size to be analyzed. In the end, the analysis was limited to four different severe crash models of four different types individually compared with non-crash cases; the crash types considered were rear-end, lane-change-related, pedestrian-related, and single-vehicle/off-road crashes.

Backward variable selection procedure was used for identifying the most significant variables from the initial set of potential independent variables, which are shown in Table 6-1, page 77; the results of this backward variable selection procedure with within stratum matched case-control logistic regression are provided in Table 6-3, page 86. First, the parameters for the complete model (i.e. having all potential variables) are estimated. Then, the results of the Wald test for individual parameters are examined, and the least significant effect (i.e. effect not meeting the p-value criterion for keeping a variable in the model) is removed; the removed effect is permanently excluded from the model. This process is repeated until no other effect may be removed based on the p-value threshold (p-value > 0.15) (Breiman, 2001). Thus, it can be concluded that the backward selection model is preferred because it starts with the complete set of variables included in the model. For more details on the backward variable selection and its advantages one may consult the work by Vittinghoff et al. (2005).

The results in Table 6-3 are tabulated in eight columns; from these, the four crash types are assigned two columns each (one for the coefficient and another for the corresponding p-value). Note that the rows corresponding to some of the variables show X since these variables (or any of the categories of nominal variables) were not found to be significantly associated with the corresponding type of severe crash occurrence. The likelihood ratio test statistic and corresponding p-values have also been provided as goodness of fit measures, which indicate the statistical significance of all four models; these can be found at the bottom of the table. Based on the p-values for likelihood ratio test, it seems that the model explaining severe lane-change-related crashes has the least explanatory power; this suggests that (compared to the other three groups of severe crashes) severe lane-change-related crashes might be more dependent on event-specific factors (e.g. careless driving) rather than on the highway design-related parameters explored here.

Table 6-3: The Parameters Significantly Affecting Severe Crashes of Different Types

Type of crashes	*Rear-end*		*Lane-change*		*Pedestrian*		*Single vehicle*	
Variable	Parameter	p-value	Parameter	p-value	Parameter	p-value	Parameter	p-value
Speed limit<40 MPH	0.526	0.4367	X	X	X	X	X	X
40<= Speed limit<50 MPH	0.990	0.0837	X	X	X	X	X	X
50<= Speed limit<60 MPH	0.327	0.5299	X	X	X	X	X	X
Speed limit>=60 MPH	.	.	X	X	X	X	X	X
ADT<14900	-0.949	0.0975	-4.898	0.0004	X	X	X	X
14900<=ADT<26500	-1.305	0.0002	-1.245	0.0321	X	X	X	X
26500<=ADT<40000	-0.533	0.0305	-0.833	0.0621	X	X	X	X
ADT>=40000	.	.	.	.	X	X	X	X
K-Factor<9.35	1.884	0.0018	-3.914	0.0496	X	X	X	X
9.35<= K-Factor < 10.14	0.213	0.7201	-3.636	0.0188	X	X	X	X
K-Factor >= 10.14	.	.	.	.	X	X	X	X
T-Factor<4.84	X	X	-0.426	0.5061	0.783	0.0524	X	X
4.84<= T-Factor < 8.75	X	X	-1.160	0.0484	0.294	0.4501	X	X
T_factor>=8.75	X	X	.	.	.	.	X	X
Afternoon Peak Weekday	1.060	<.0001	X	X	-0.281	0.4179	0.344	0.1327
Friday or Saturday Night	0.196	0.5372	X	X	0.582	0.0225	1.061	<.0001
Morning Peak Weekday	0.588	0.0141	X	X	-1.397	0.0170	-0.101	0.7168
Off-peak period	.	.	X	X	.	.	.	.
TWLTL	1.147	0.3164	X	X	X	X	2.250	0.0360
Lawn	3.096	0.0051	X	X	X	X	3.721	0.0003
Guard rail	3.189	0.0552	X	X	X	X	4.298	0.0013
Barrier other than guard rail	1.556	0.3128	X	X	X	X	4.565	<.0001
Canal or Ditch	4.339	0.0067	X	X	X	X	6.040	0.0034
Curb<6 in	2.837	0.0104	X	X	X	X	3.016	0.0039
Curb >=6 in	2.715	0.0153	X	X	X	X	3.040	0.0038
Paved not for travel	1.907	0.1191	X	X	X	X	2.992	0.0065
No Median	.	.	X	X	X	X	.	.
No Side Walk	X	X	X	X	-0.947	0.0059	X	X
Side Walk <= 6ft.	X	X	X	X	-0.190	0.4758	X	X
Side Walk > 6ft.	X	X	X	X	.	.	X	X
Good or very good	-0.653	0.2955	-2.045	0.1215	X	X	X	X
Fair	-0.575	0.0168	-0.910	0.0548	X	X	X	X
Poor or very poor pavement	.	.	.	.	X	X	X	X
Presence of horizontal curvature	-2.613	0.0551	X	X	0.751	0.0880	X	X
Attenuators	X	X	X	X	1.219	0.0471	X	X
Parking	X	X	X	X	0.521	0.0385	0.673	0.0027
Likelihood ratio test for goodness-of-fit	*119.086*	*<.0001*	*34.592*	*0.0005*	*57.354*	*<.0001*	*88.849*	*<.0001*
Base case categories	Binary variable	Statitically significant at 90% C.I.			X=Variable not included based on Backward Selection			

The results tabulated in Table 6-3 above are discussed in the following two subsections. First, the significant factors associated with each of the four crash types are discussed individually, followed by a discussion on the differences in significance of factors among these crash types.

6.5.1 Significant Variables for Each Crash Type

For *severe rear-end crashes* speed limit, *adt*, K-factor, time of day/day of week, median type, pavement condition, and presence of horizontal curvature were significant. Severe rear-end crashes are more likely to occur on sections with *40 <=speed limit< 50 mph*. On multilane arterial sections with speed limit less than 40 mph speeds are likely too low to have *severe* rear-end crashes, while on sections with speed limit greater than or equal to 50 mph rear-end crashes are less likely to occur. Compared to the base case (*adt>=40,000*) the sections with lower *adt* are more likely to observe severe rear-end crashes. However, the coefficient magnitudes (even though they are negative for all three categories with lower *adt*) show that sections with *adt<14,900* (category with lowest *adt*) are in fact more likely to experience a severe rear-end crash compared to sections with *14,900<adt<=26,000*. It indicates that occurrence of severe rear-end crash occurrences are not directly related with exposure. The results once again highlight the importance of measuring these variables on a nominal scale. The sections with *K-factor<9.35* are more likely to have severe rear-end crashes. With a lower K-factor, relatively less traffic is served during the design peak hour and thus the vehicles are more likely to interact during off-peak hours, possibly at higher speeds, leading to an increased likelihood of severe rear-end crashes. In terms of severe rear-end crashes Friday and Saturday nights are statistically no different than other off-peak periods. Severe rear-end crashes are less likely to occur on the fair pavements compared to poor pavements. Note that the coefficient for good/very good pavement is also negative when compared to the base case (poor/very poor pavements) even though it is not statistically significant. Presence of horizontal curvature was negatively associated with likelihood of *severe* rear-end crashes. It indicates that slower speeds of most drivers on the curved sections result in reduced likelihood of *severe* rear-end crashes. Median types related with likelihood of severe rear-end crash occurrence are discussed in the next section since it requires some context from their relationship (or lack thereof) with occurrence of severe crashes of other types.

The variables *adt*, K-factor, T-factor, and pavement condition are significantly related with *severe lane-change-related crashes*. As such a monotonic trend may be observed in the coefficients for the three classes of the *adt* variable. It indicates that the severe lane-change-related crashes on arterial segments may be explained in terms of exposure. Segments with K-factor greater than 10.14% are more likely to have severe lane-change-related crashes. Severe lane-change-related crashes are also more likely to occur on sections with *4.84 <=T-factor< 8.75*. The sections with even lower T-factor also have a negative coefficient compared to sections with *T-factor >=8.75*, but it is not statistically significant. The results indicate that arterial sections with higher percentage of trucks are more likely to have severe lane-change-related crashes. It is an expected result since the lane-change-related collisions involving large trucks are likely to be more severe; therefore, lane-change-related warnings on sections with high truck traffic may be an effective countermeasure for such crashes. Fair pavements reduce the likelihood of severe lane-change-related crashes and the coefficient for good/very good pavement is also negative with a p-value only slightly higher than 0.10. It shows that improving pavement condition can lead to reduction in severe lane-change-related crashes.

For *pedestrian-related severe crashes* on arterial segments T-factor, time of day/day of week, along with presence of sidewalk, attenuators, and roadside parking were significant factors. Segments with very low truck traffic are more likely to have pedestrian-related crashes, since the sections with high pedestrian traffic are expected to have very little to no large trucks. This is why the relationship between crash likelihood and T-factor is not monotonous. The corridors with *4.84 <= T-factor < 8.75* and *T-factor>=8.75* are not statistically different from each other, while the corridors with *T-factor <4.84* are significantly more likely to have pedestrian-related severe crashes. Sidewalks greater than or less than 6 ft are not statistically different in terms of occurrence of severe pedestrian crashes. It indicates that widening the sidewalk may not lead to risk reduction for severe pedestrian-related crashes; however, the roadways with no sidewalk are in fact less likely to have these crashes likely due to low pedestrian traffic. Presence of roadside parking is significantly related to increased likelihood of pedestrian-related crashes. With roadside parking one expects a significant number of mid-block road crossings/pedestrian traffic and thus an increased likelihood of severe pedestrian-related crashes. As expected, pedestrian-related crashes are also likely to occur during Friday/Saturday nights. Significantly positive coefficients for presence of attenuators and horizontal curve are explained in the next section since the relevant discussion requires the context of insignificance of these parameters in the other three models.

For *severe single-vehicle/off-road crashes* time of day/day of week, median type, and presence of parking are significant. Friday/Saturday nights have significantly higher likelihood of severe single vehicle crashes compared to other periods of the day. All eight median types are significantly more crash prone compared to no-median. Canal or ditch as median increase the likelihood of severe single vehicle crashes and have the largest coefficient. *Barrier other than guardrail* has the most significant coefficient in terms of the (smallest) p-value. It indicates that the presence of median barriers other than guardrail may increase the likelihood of severe crashes. The finding appears to be consistent with Elvik (1995) who noted that median barriers (other than guardrail) lead to a 30% increase in crash rate without a corresponding reduction in severity given a crash has occurred. It is also worth mentioning that while presence of attenuators is not a significant factor for severe single-vehicle/off-road crashes, it had a significantly positive effect when all (not just severe) single-vehicle/off-road crashes were compared with non-crash cases. A significant proportion of severe single-vehicle crashes involve hitting road side signs, and roadside objects. In the area where roadside parking is present one is more likely to find and hit such objects; therefore, presence of roadside parking is positively related with the severe crashes involving single vehicle.

6.5.2 Differences among Crash Types

It is interesting to note that none of the factors are significantly associated with all of the four crash groups. Even the categories of the variable day of week/time of day, which is significantly related to three types of crashes, have widely varying coefficients. It indicates that examining crashes by type was indeed a better approach. In this section we discuss the differences in coefficients of the same variables for different crash types. Figure 6-2, page 91, through Figure 6-5, page 94, compare different crash types by plotting a combination of the sign (positive above the x-axis and negative below) and strength of the coefficients (in terms of the chi-square test statistic value corresponding to the coefficients) for each variable.

In Figure 6-2, page 91, it may be observed that speed limit on the arterial segments is not a significant factor for any group of severe crashes except for the rear-end crashes. The *adt* is not a significant factor for pedestrian-related and single-vehicle/off-road crashes. Since *adt* has an effect on inter-vehicle interactions it is reasonable that this variable only affects the severe crash types involving more than one moving vehicles. The coefficients for three categories of the *adt* provide an interesting contrast between rear-end and lane-change-related crashes. The contrast is clearly visible in this figure. With increasingly negative coefficients for the three categories with lower *adt*, it is apparent that severe lane-change-related crashes on arterial segments are actually better explained by exposure compared to the severe rear-end crashes.

Most drivers drive slower on the curved sections, which leads to the presence of horizontal curve either not being significant (for severe lane-change-related and single-vehicle crashes) or even negatively related with likelihood of severe rear-end crashes; however, slower speeds do not reduce the severity of pedestrian crashes. Therefore, presence of horizontal curvature is positively related with likelihood of severe pedestrian-related crashes.

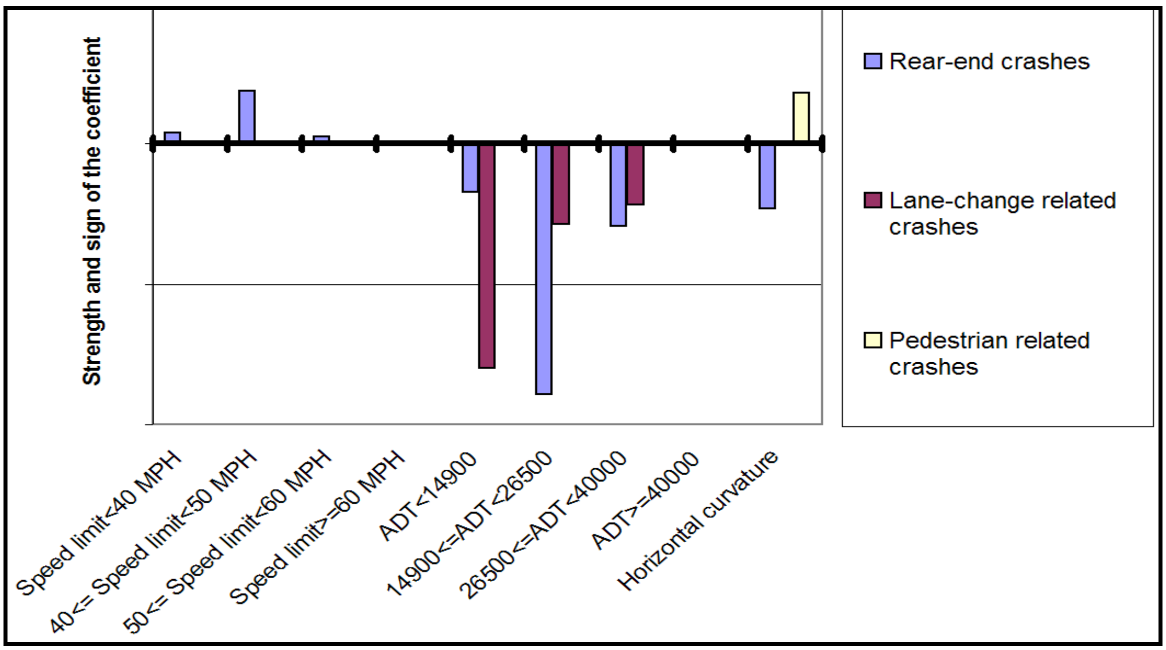


Figure 6-2: Comparison between Crash Types, Based on Chi-Square Statistic Corresponding to Coefficients for *Speed limit, adt* and *Presence of Horizontal Curve*

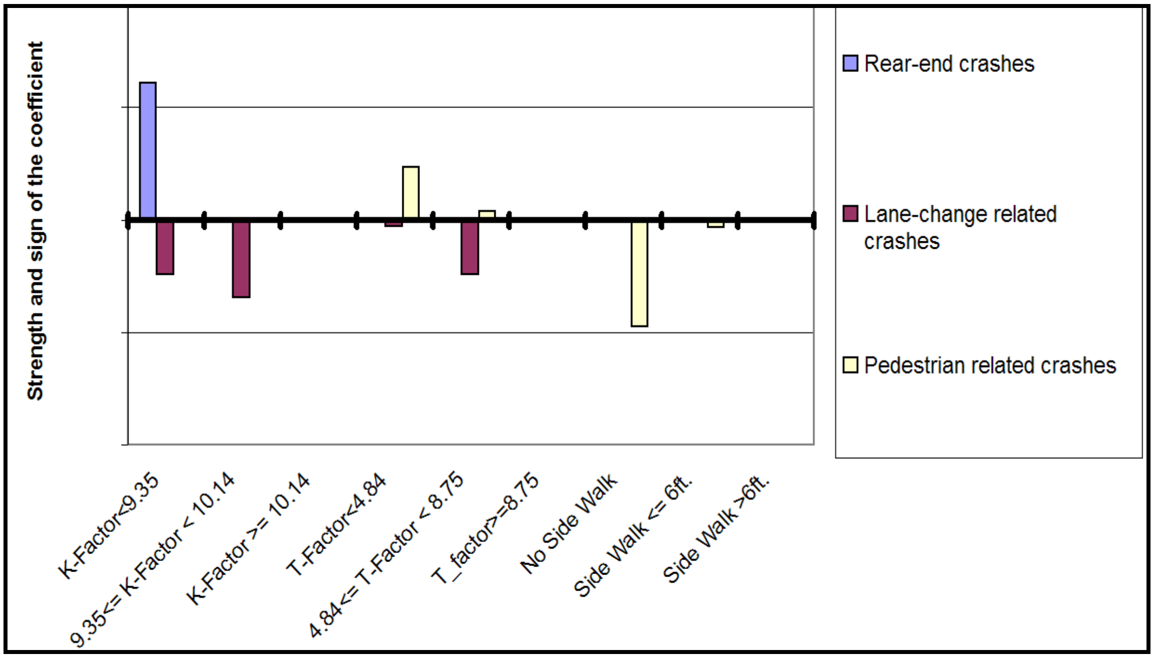


Figure 6-3: Comparison between Crash Types, Based on Chi-Square Statistic Corresponding to Coefficients for *K-factor, T-factor* and *Presence/Width of Sidewalk*

The contrast between coefficients of T-factor for pedestrian and lane-change-related severe crashes is interesting (see Figure 6-3 above). While corridor sections with lowest percentage of trucks (T-factor < 4.84%) are more likely to have severe pedestrian-related crashes, the sections highest percentage of trucks (T-factor >= 8.75%) are more likely to have severe lane-change-related crashes. The former may just be related with higher pedestrian exposure on arterials with low T-factor, while the later could be the basis for warning motorists about being cautious while changing lanes on sections with high T-factor.

Note that with poor/very poor pavements as the base case both remaining categories have a negative coefficient indicating that improving pavement condition may actually reduce the likelihood of both severe rear-end and lane-change-related crashes (see Figure 6-4, page 93). One may suspect that improved ride quality would increase the travel speed thereby increasing the likelihood of severe crashes. That concern, however, is somewhat alleviated by the fact that

that pavement condition is not significant for severe single-vehicle/off-road crashes. Presence of crash attenuators is a significant factor associated with severe pedestrian-related crashes; it had a significantly positive coefficient when all crashes were compared to non-crash cases (in the preliminary analysis). It indicates that attenuators are installed at high crash risk locations; but since these can only reduce the severity of impact for the vehicles, the attenuators are unlikely to reduce the severity of pedestrian-related crashes. It explains their significantly positive association with likelihood of severe pedestrian crashes while no significant association with severe crashes of other three types. The results show that in addition to the crash attenuators some countermeasures for pedestrian-related crashes also need to be considered.

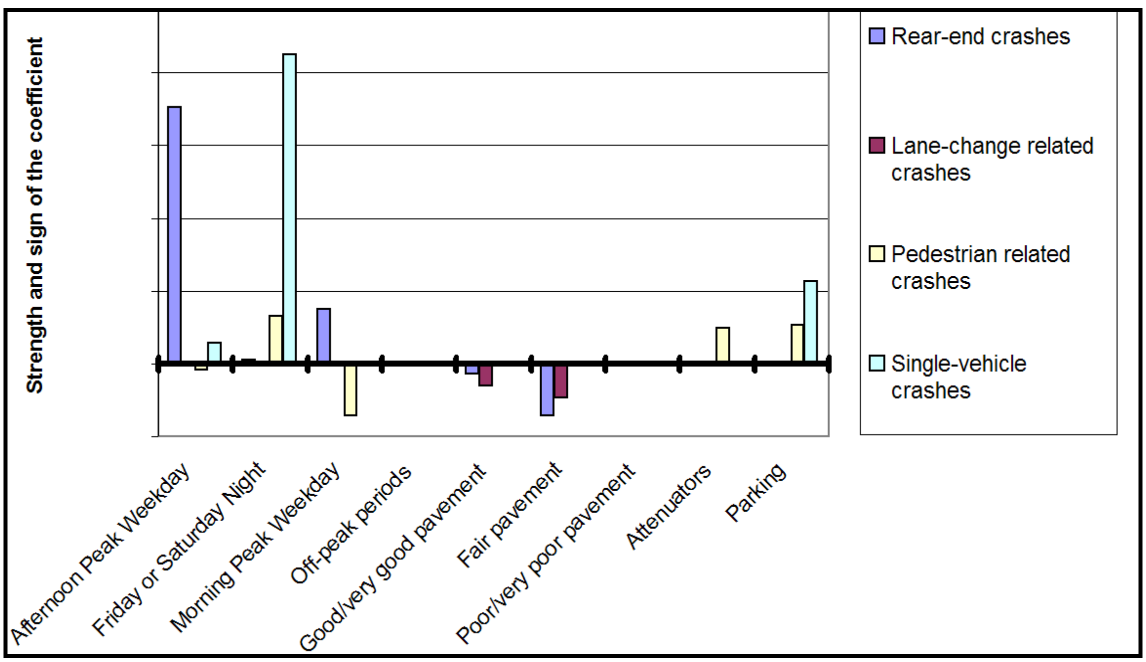


Figure 6-4: Comparison between Crash Types, Based on Chi-Square Statistic Corresponding to Coefficients for *Day of Week/Time of Day*, *Pavement Condition*, *Presence of Attenuators* and *Roadside Parking*

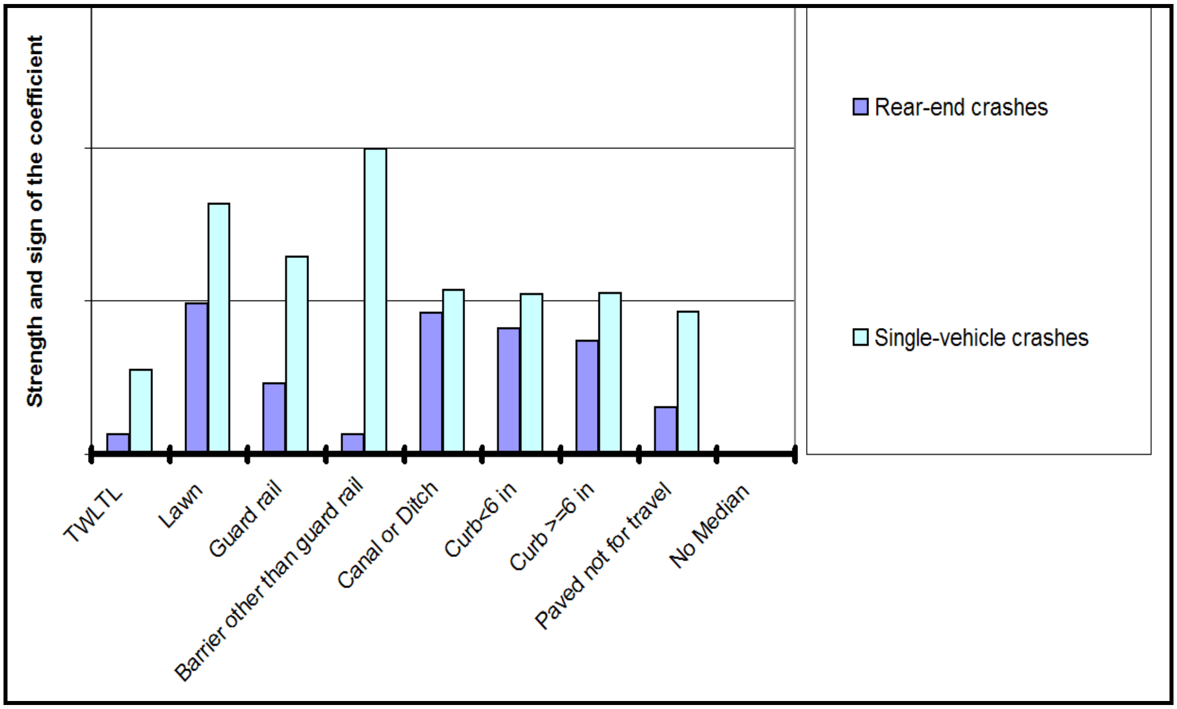


Figure 6-5: Comparison between Crash Types, Based on Chi-Square Statistic Corresponding to Coefficients for Median Type

Median type is not a significant factor for severe lane-change-related or pedestrian crashes, which is why Figure 6-5 above only shows bars corresponding to rear-end and single-vehicle/off-road crashes. Among the severe rear-end crashes, barrier other than guardrail are not significantly different from the roadway sections without a median. However, for severe single-vehicle crashes, barrier other than guardrail is the most significant category for separating crashes from non-crash cases. Sections with paved median not for travel and TWLTL are not significantly associated with severe rear-end crashes but are more likely to have severe single-vehicle/off-road crashes. Sections with lawn/grass median and with canal and ditch are significantly associated with both of these groups of severe crashes. Their association with severe rear-end crashes is explained by the fact that these medians are generally found on sections with high travel speeds where drivers are likely to be caught unaware of the traffic ahead. On the other hand, these sections are also prone to collisions of the severe single-vehicle/off-road type, either because of drivers trying to avoid a rear-end collision and/or losing control of the vehicles due to excessive speeds.

6.6 Concluding Remarks

This study provides a new approach for identifying significant factors related with risk of severe crashes on segment (or mid-blocks) of multilane arterials with partially limited access. The fundamental difference between this approach and crash frequency analysis is that crashes themselves are used as units of analysis rather than crash counts over roadway segments of arbitrary lengths. Traditionally, in traffic safety studies for multilane arterials (with partially limited access) where crashes are used as units of analysis, the objective is to assess the severity of crash given a crash has occurred. It is not the best way to analyze the factors influencing crash severity since non-severe crashes are usually underreported and thus underrepresented in the data sources (Abdel-Aty et al., 2004). Under the proposed methodology, samples of m crashes and n non-crash cases (with m and n varying for each corridor depending on the number of crashes) were generated for 151 multilane arterial corridors in Florida. The data were analyzed using within stratum matched logistic regression models with each stratum defined by a single corridor; thus, the sampling scheme implicitly controls for the factors that may vary among corridors.

The proposed methodology addresses the following issues associated with traditional approaches for analysis of severe crashes on arterial segments:

- It does not use information on non-severe crashes and relies solely on the severe crash data. Therefore, well-documented underreporting of least severe crashes does not affect the analysis.
- The variables used in the analysis are general roadway information as opposed to event-specific factors used in the severe vs. non-severe crash comparisons. Therefore, the proposed approach may be more useful in examining engineering solutions for the safety concerns.
- No aggregation of crash data over arbitrary segment lengths is necessary as is the case with analysis based on crash frequency or crash rates.

It is worth mentioning that this approach is limited in that it is not suitable for comparing intersections' crash patterns. Reason for that is that there is no way of assigning non-crash cases to an intersection. For example, comparisons between selected non-crash cases with the intersection-related crashes (i.e. signalized or unsignalized) would yield information that would mostly reflect the characteristics belonging to *locations of the signalized intersection*. However, with segment crashes, the comparisons yield important geometry-/traffic- related parameters that are significantly related to crash occurrence on the segments. Since individual intersections provide logical units for aggregating the crash data, a frequency approach is still best suited for analysis of intersection crashes. It also has to be noted that the underreporting of least severe crashes may have affected the simple models (in the preliminary analysis) comparing all crashes vs. non-crash cases; however, the most critical part that relates to severe crashes is not affected by the aforementioned issue and therefore is the focus of this study.

The analysis yielded some interesting relationships between severe crash occurrence and presence of crash attenuators, times of day/day of week, and horizontal curvature. The relationship between exposure (represented by *adt*) on arterial sections and severe lane-change-related crashes was found to be more apparent compared to relationship between *adt* and severe rear-end crashes. The information not used explicitly for the analysis is the driver-related factors and a within stratum matched sampling technique was used to implicitly control for these factors. One way to account for these factors is to use induced exposure to derive the driver-related factors for crash and non-crash locations and then include them as independent variables. Once those factors are considered explicitly, the crash vs. non-crash classification approach may be also suitable for a data mining-type of analysis.

CHAPTER 7. SEVERITY ANALYSIS OF CRASHES ON MULTILANE ARTERIALS USING CONDITIONAL INFERENCE FORESTS

7.1 Background

One of the objectives of this study is to identify contributing factors related to severe/fatal crashes occurring on the high-speed (speed limit greater than 45 mph), multilane (more than one lane in each direction of travel) corridors in the state of Florida. Several safety studies deal with identifying contributing factors and use various modeling techniques for the same. Improvements in modeling methodology lead to better detection of causal factors. In this study, the authors have not only introduced certain new variables (i.e. data improvement), but have also adopted new data mining methodology for a better understanding.

Approaches to safety on multilane corridors have traditionally been twofold. Brown and Tarko (1999), Abdel-Aty and Radwan (2000), as well as Rees (2003), all treated the corridors entirely; on the other hand, Milton and Mannering (1998), as well as Miaou and Song (2005), divided the corridors into segments and intersections. Abdel-Aty and Wang (2006) have shown a spatial correlation between crash patterns of successive signalized intersections, which may be attributed to the characteristics of the segments joining them.

Though both approaches have worked well for investigation purposes, the issue that still remains is how to assign crashes to the segments and the intersections. There is no uniformity in the influence area of an intersection among the states. For example, in Florida, all the crashes occurring within 250 ft from the center of an intersection are categorized as intersection-related crashes, as has been reported by Abdel-Aty and Wang (2006) and Wang et al. (2006). Recently, Das et al. (2008) showed that proximity only is not the best way to assign crashes. Wang et al.

(2008) used frequency modeling for crashes with fixed as well as varying influence distance and found different sets of significant factors. Apart from the aforementioned research, it is also well-known that the way the crashes are reported varies among different administrative units. The authors investigated several crash reports and came up with an innovative approach to assign crashes, the details of which are given in the next section which explains the data used in the study.

As previously mentioned, it is not only important to find the contributing factors but also to improve the methodology adopted. In their work on association rules, Pande and Abdel-Aty (2008), point out that data mining techniques remain underutilized for crash analyses. This underutilization is especially noteworthy since most studies use observational data collected outside the scope of experimental design. Simple data mining tools like classification and regression trees have traditionally been used to identify variables of importance in safety studies (Pande and Abdel-Aty, 2008). A decision tree, with all its simplicity and handling of missing values, can be very unstable; however, if a forest (i.e. a robust ensemble of decision trees) is used, outputs can become much more stable. Therefore, using forests can be a better choice than using single decision trees. This raises the topic of random forests, developed by using the Classification and Regression Trees (CART) algorithm, and that have recently been used by the authors (Abdel-Aty et al., 2008) for identifying significant variables and for developing neural network classifiers. However, the method has been shown to have selection bias as shown by Strobl et al. (2007). The selection bias is in favor of variables which are continuous or have higher number of categories. At the root of this selection bias is the application of the *Gini* index criterion to split a node (while building the tree) as well as for variable selection (generally based on the frequency a variable was chosen for the split). Details of the Gini index criterion and the resulting bias are provided in the Modeling Methodology section. Therefore, for this study's variable selection purposes, the conditional inference trees and forest developed by Hothorn et al. (2006) have been used. The authors are confident on the application of this new methodology for the improvement of traffic safety research. Details of how this algorithm is different (and better suited for the application at hand) than the CART are presented in the methodology section.

The authors included new variables like 'element' in this study, which assigns crashes to segments, intersections or access points based on the information from site location, traffic control and presence of signals. The authors were able to identify roadway locations where severe crashes tend to occur. Failure to use safety equipment by all passengers as well as presence of driver/passenger in the vulnerable age group (older than 55 years or younger than 3 years) were two new variables also included within the data. The details of how the inclusion helped in a better understanding of the severity aspect are discussed in the Analysis and Results section of the report.

Crash data from the high-speed multilane arterials with partial access control in Florida have been collected. These arterials have been divided into groups based on their lengths and roadway design standards (urban/suburban and rural). The following section will focus on the details of the data collection and aggregation processes. It is followed by the methodology section where conditional inference trees and forests will be discussed. The results and analysis section will explain the results from the conditional inference trees and the forests. While the random forests provide a more robust set of variables associated with severe/fatal crashes, individual tree helps in making relevant inferences about the relationship.

7.2 Data Collection and Preparation

7.2.1 Study Area and Available Data

The crash data available were from the Crash Analysis and Reporting (CAR) system of the FDOT. The Roadway Characteristics and Inventory (RCI) data were also made available to the authors through the FDOT. The data used are for the years 2004 through 2006 for all the Florida state roads. The datasets have information regarding traffic, roadway geometric and driver-related factors. The datasets were merged and the parameters were modified to suit the data mining methodology being implemented in the study. The corridors, which were originally divided according to administrative units (i.e. roadway IDs based on county boundaries), were logically combined to form continuous sections based on design standards. The details of the applied design standards are given in the next sub-section.

7.2.2 Data Preparation

As mentioned earlier the corridors available for the study were logically combined into continuous sections based on their design. Corridors with continuous urban/suburban design were grouped together as well as the ones with rural design. However, it should be noted that in the present study the authors focus only on the urban/suburban corridors. Since the corridors are of variable lengths, it was logical to cluster them based on the same parameter before further analysis on severity could take place. The optimum number of clusters was found based on the partitioning around the medoids (*pam*) algorithm proposed by Kaufman and Rousseeuw (1990). In the *pam* algorithm, which operates on the average dissimilarity, a *medoid* is an object of the cluster whose average dissimilarity to all the objects in the cluster is minimal. Once the medoids are identified, all the objects are assigned to the nearest medoid. The objective function is the sum of the dissimilarities of all the objects to the nearest medoid. The algorithm terminates when the interchange of an unselected object with an already selected object no longer minimizes the objective function. The optimum number of clusters was found to be four. The corresponding corridor lengths per cluster are: Cluster 1 (1.009 – 2.89 miles), Cluster 2 (2.898 – 5.729 miles), Cluster 3 (5.762 – 10.556 miles) and Cluster 4 (10.644 – 78.293 miles) (refer to Table 5-1, page 67).

Different types of crashes occur on the corridors and the contributing causes for the different types also vary. Even though the overall safety of the corridor is being analyzed, the approach to investigate different crash types separately would shed more light. The crashes were grouped into six major types as follows: 1) rear-end, 2) head-on, 3) angle/turning, 4) sideswipe, 5) crashes involving slow moving vehicles (e.g. cycles, mopeds, etc.), and 6) crashes involving single vehicles; the total number of crashes in each of the aforementioned categories is 5,536, 1,264, 6,234, 2,207, 1,305 and 2,407, respectively.

The conditional inference trees used in this study helps us in identifying the contributing factors associated with the severity of the crashes that occurred along a corridor. However, too many parameters lessen the discriminating ability of the models, as the overall degrees of freedom available for the model development decrease; therefore, only a subset of the available factors should be chosen for model development. Milton et al. (2008) have also pointed out that event-specific variables are the least desirable when developing injury severity models; hence, an educated variable selection (i.e. by using engineering judgment) was made for the analysis, taking into account that event-specific factors are not in use to a relatively large extent. The variables were broadly based on two different categories: 1) environmental and road geometric factors, and 2) driver- and vehicle-related factors. The variables used in the study are described in Table 7-1 below. They have been derived directly from the datasets or a combination of parameters. Both these sets of parameters have their application values.

Table 7-1: Dependent/Independent Variables Used in the Analysis

Variable Name	Variable Description	Urban / Suburban
Target or Dependent Variable		
Sev	Severity	Binary (1 = incapacitating injuries/ fatalities; 2 = possible/ non-incapacitating injuries)
Environmental and Roadway Geometric Parameters		
pavecond	Pavement condition	4 levels (poor, fair, good and very good)
surf_type	Type of surface	Binary (1 = black top surface; 2 = other)
surface_width	Surface width	Continuous
shld_t	Type of shoulder	Binary (1 = paved; 2 = unpaved)
max_speed	Maximum posted speed limit	Continuous
park	Presence of parking	Binary (1 = no; 2 = yes)
skid_f	Friction resistance	Skid <= 34 34 < skid <= 38 Skid > 38
median	Types of median	9 levels (0 = no median; 1 = painted; 2 = median curb <= 6”; 3 = median curb > 6”; 4 = lawn; 5 = paved; 6 = curb <= 6” and lawn; 7 = curb> 6” and lawn; 8 = other)
ACMANCLS_num	Type of median openings	7 levels (0 = no opening; 2 = restrictive opening w/ service roads; 3 = restrictive opening; 4 = non-restrictive opening; 5 = restrictive opening with shorter directional openings; 6 = non-restrictive opening with shorter signal connection; 7 = both median types)
road_cond	Road condition at time of crash	Binary (1 = no defects; 2 = defects)
vision	Vision obstruction	Binary (1 = no; 2 = yes)
shld_side	Shoulder + sidewalk width	Continuous
curvclass	Horizontal degree of curvature	6 levels (curve < 4’; 4 <= curve <= 5’; 5 < curve <= 8’; 8 < curve <= 13’; 13 < curve <= 27’; curve > 27’
surf_cond	Surface condition	Binary (1 = dry; 2 = other)
light	Daylight condition	Binary (1 = daylight; 2 = other)

Variable Name	Variable Description	Urban / Suburban
Environmental and Roadway Geometric Parameters (continued)		
adt	Annual daily traffic	*adt* <= 31000 31000 < *adt* <= 40000 40000 < *adt* <= 52500 *adt* > 52500
t_fact	Average T-factor	t_fact <= 4.05 4.05 < t_fact <= 5.895 t_fact > 5.895
k_fact	Average K-factor	k_fact <= 9.85 k_fact > 9.85
dayandtime	Combination of the day of week and time of day	Afternoon Peak Weekday Morning Peak Weekday Friday or Saturday Night Off-peak
trfcway	Vertical curvature	Binary (1 = level; 2 = upgrade/ downgrade)
element/ element 1	Assignment of crashes to roadway elements	Ternary (1 = segment; 2 = intersections; 3 = access points) / Binary (1 = segments/ access points; 2 = intersections)
LIGHTCDE	Street lighting	Ternary (Y = full lighting; N = no lighting; P = partial lighting)
Driver- and Vehicle-related Parameters		
age_gr	Age group of the at fault driver	Age<=25; 25<age<=35; 35<age<=45; 45<age<=55; 55<age<=65; 65<age<=75; Age>75
veh_type1	At-fault type of vehicle	4 levels (1 = automobiles; 2 = light trucks; 3 = heavy vehicles; 4 = light slow moving vehicles)
alcohol_use	Alcohol/ drug use of the at-fault driver	3 level (1 = non-use; 2 = use; 3 = no info.)
vuln_age	Presence of vulnerable age group passengers in the vehicle (age<5 or age>55)	Binary (1 = yes; 2 = no)
more	Presence of more than 5 passengers inside either of the involved vehicles	Binary (Y = yes; N = no)
sfty	Use of safety equipment in the vehicle by driver/passengers	Binary(1 = yes; 2 = no)
gender	Gender of the at-fault driver(s)	3 levels (1 = male; 2 = female; 3 = both)
veh_move1	Vehicle movement of the at-fault vehicle	4 levels (1 = straight ahead; 2 = turning movements; 3 = changing lanes; 4 = other)

The variables illustrated in Table 7-1, pages 103 and 104, are mostly derived from the RCI database. As it can be seen, most of these variables, when in raw form, have too many categories; therefore, a level reduction is critical for the variables, thus making them more readily explainable as well as simplifying the model. For example, vehicle movement, vehicle type, roadway conditions, vision obstruction, surface condition, surface type and type of median are some of the variables with many categories. For example, the proposed methodology (conditional inference trees/forests) uses chi-square test statistic to identify the relationship between a particular parameter and target variable. Each category of the variable should have a sufficient number of observations in the contingency table for the chi-square to be evaluated as discussed by Das et al. (2008). Continuous variables like *adt*, T-factor (percentage of trucks), K-factor (design hour volume as a percentage of *adt*) and skid (friction resistance multiplied by a factor of 100) were also categorized. Their relationships with severe/fatal crash occurrence may not be monotonous in nature. Time of crash, along with day of week, were combined into one variable representing day of week and time of day. The weekend night times were not treated as off-peak hours as there may be higher instances of alcohol-impaired driving.

The authors have introduced some new variations to the traditional parameters. Traditionally, the site location variable has been used by researchers to assign crashes to the three roadway elements (segments, intersections and access points). However a detailed review of several hundred crash reports, suggested that the site location variable was a weak indicator by itself. For example, it was observed that it is possible for a crash to be not attributed to a signalized intersection even if it may have occurred very close to one. In fact, traffic control in combination with the site location along with the information of the presence or absence of signal, did an excellent job in attributing crashes to one of the three roadway elements. Based on these three independent parameters, a variable *element* was created to assign the crashes to the three roadway elements, namely segments, intersections and access points. However, it was also observed and verified through the study of crash reports that to distribute crashes to the three roadway elements works fine with all crash types except with angle-/turning-related crashes. Most of such crashes occur at the signalized intersections. The crashes which occur on the segments were observed to have occurred mostly on auxiliary lanes (right/left turning lanes); thus, these could be attributed to either the segment or to the access points. Therefore, for angle-/turning-related crashes the ternary variable element takes the form of binary *element1* where the crashes either belong to the signalized intersection or to segment/ access points. This new variable appears in certain tree results (developed along with conditional inference forests for relevant inference) and is a positive contributor to model development in the forests.

Zhang et al. (2000) found the non-use of seat belts to increase the risk of severe injuries. In this study, the parameter for safety equipment in use is for all the passengers. This is different from the traditional approach as it is more useful to look at the overall safety of all the passengers rather than just focusing on the safety equipment use of the driver. The importance lies in the fact that there are many crashes in which the drivers may not be injured at all. The vulnerable age group binary variable points out the presence of children or elderly passengers inside the vehicle. The physical fragility of the people belonging to these age groups described in Table 7-1, pages 103 and 104, makes it an interesting variable and the results also show an interesting pattern related to severity.

The median types were combined into 9 levels. It does the twofold job of not only giving a sense of the median obstruction imposed, but also gives an idea as to how far apart the opposing directional roads could be. The authors observed that median width was a variable that is really dependent on the median type; hence, the median width was sufficiently represented within the variable median type. A new variable called *shld_side* has been created which simply represents the total width of the outside shoulder and the sidewalk. This variable gives a more realistic idea of the side space available for the vehicles traveling in the outer lane, especially in the urban areas where the shoulder width sometimes is negligible as compared to those available in rural settings. For this reason, the original information on shoulder width and sidewalk width were replaced with this new variable.

The target variable of severity is binary. The first level represents fatalities and incapacitating injuries; these are combined into one level for two reasons. First, fatal crashes have relatively small frequencies compared to other injury severity levels; for example, the chi-square tests may not be valid due to low expected cell-frequency. Second, the crashes that involve incapacitating injury could easily have been fatal, and vice versa, possibly due to vulnerability of the subjects involved (Das et al., 2008). The second level includes crashes with possible injuries and non-incapacitating injuries. The crashes with no injuries were not included as these are likely to be incomplete. This issue has been well investigated and documented by Abdel-Aty and Keller (2005). Yamamoto et al. (2008) have also discussed the issue of possible underreporting of such crashes and the bias resulting from it. Therefore, the authors have included in this study the crashes with an injury severity level of having at least one possible injury or higher.

It should be noted here that the conditional inference forests, which have been used to calculate the variable importance score, do not accept missing values; hence, the dataset has no missing data. Therefore, the introduction of random parameters in order to account for missing data, as done by Milton et al. (2008), is not required for this study.

7.3 Modeling Methodology

7.3.1 Conditional Inference Trees

The modeling approach adopted here in is the conditional inference trees and the forests developed from them. The focus of the study is to find out parameters that are related to the injury severity. The trees not only give the variables of importance, but also contribute to a better interpretation of the results. For severity analysis specifically, the advantage in using trees is that it helps to determine the values of the parameters that contribute the most to crash severity. From a safety perspective this is critical since it can help to determine what changes need to be made in the design and/or policies for improving safety. Conventional classification and regression trees have always been used to select variables of importance. According to Strobl et al. (2007), the CART trees have a variable selection bias towards variables which are continuous or with higher number of categories. The most common splitting criterion in the CART tree is the Gini index to find a favorable split. The Gini index checks for the purity of the resulting daughter nodes in the tree. According to Breiman et al. (1984), for a given node t with estimated class probabilities p(j|t), (j=1,2,…..,J), the node impurity i(t) is given by:

$$i(t) = \Phi(p(1|t),......, p(J|t)) \tag{7.1}$$

A search is made for the most favorable split, one that reduces the node or equivalently tree impurity. If the adopted form is Gini diversity index, then i(t) takes the form:

$$i(t) = \sum_{j \neq i} p(j|t)\, p(i|t) \tag{7.2}$$

The Gini index considered as a function $\Phi(p_1,……,p_J)$ of the $p_1,……,p_J$ is a quadratic polynomial with non-negative coefficients; therefore, for any split s: $\delta(s, t) \geq 0$. Since the criteria looks for a favorable split, the chances to find a good split increases if the variable is continuous or has more categories; thus, even if the variable is not informative, it could be located higher up on the tree's hierarchical structure. For this reason, in this study the researchers have used conditional inference trees (Hothorn et al., 2006) where the node split is selected based on how good the association is. The resulting node should have a higher association with the observed value of the dependent variable. The conditional inference tree uses a chi-square test statistic to test the association; therefore, it not only removes the bias due to categorization, but also chooses those variables which are informative.

The key to this recent algorithm is the separation of variable selection and splitting procedure. The recursive binary partitioning which is the basis of the framework is defined in the following paragraph.

The response Y comes from sample space Y, which may be multivariate. The m-dimensional covariate vector X = (X_1,….,X_m) is taken from a sample space X = X_1,*……*X_m. Both the response variable and the dependent variables may be measured at any arbitrary scale. The conditional distribution of the response variable given the covariates depends on the function of the covariates.

$$D(Y \mid X) = D(Y \mid X_1,....., X_m) = D(Y \mid f(X_1,....., X_m)) \tag{7.3}$$

For a given learning sample of n independent and identically distributed (iid) observations, a generic algorithm can be formulated using non-negative integer valued case weights w = (w_1,….,w_n). Each node of a tree is represented by a vector of case weights having non-zero elements when the corresponding observations are elements of the node and are zero otherwise. The generic algorithm is given as follows:

1) For case weights w, the global null hypothesis of independence between any of the covariates and the response is tested. The step terminates if the hypothesis cannot be

rejected at a pre-specified nominal level α. Otherwise, the j^{th} covariate X_j with the strongest association to the response variable is selected.

2) Set $A \subset X_j$, is chosen to split X_j into two disjoint sets. The case weights w_{left} and w_{right} determine the two subgroups with $w_{left,i} = w_i I(X_{ji} \in A)$ and $w_{right,i} = w_i I(X_{ji} \notin A)$ for all i=1,….,n and I( ) denotes the indicator function, which indicates the membership of an element in a subset.

3) Recursively repeat the steps 1 and 2 with modified case weights w_{left} and w_{right}, respectively.

The separation of variable selection and splitting procedure is essential for the development of trees with no tendency towards covariates with many possible splits. For more details of the algorithm, the reader may refer to the paper by Hothorn et al. (2006).

7.3.2 Conditional Inference Forest

Forests, which are a collection of multiple tree classifiers, are used for variable selection. A decision tree, with all its simplicity and handling of missing values, can be very unstable (i.e. small changes in the input variables might result in large changes in the output). In this regard, forests are more robust variable selection tool. The Random Forests' algorithm was developed by Breiman (2001) which works in the framework of the classification and regression trees, but instead of having one tree, they have multiple trees. The forests are most important in calculating the variable importance measure. Recent research in transportation by Abdel-Aty et al. (2008) and Harb et al. (2008) used the random forests algorithm to determine the variables of importance. However, Strobl et al. (2007) showed that the bootstrapping method (i.e. sampling with replacement) and the use of the Gini index leads to a biased selection of variables of importance. The Gini index shows a strong preference for variables with many categories or for the ones which are continuous. Variables with more potential cut off points are more likely to produce a good criterion value by chance. This variable selection bias which occurs in each individual tree also has an effect on the variable importance measure. In the previous sub section it was mentioned that the algorithm for recursive binary partitioning uses the association tests like chi-square test to select informative variables. Therefore, bootstrap sampling with replacement induces bias because the cell counts in the contingency table are affected by observations that are either not included or are multiplied in the bootstrap sample; hence the forests that we have used in this study comprise of the trees that have developed in the conditional inference framework. The next subsection describes the variable importance computation process.

7.3.3 Variable Importance

The basis of the variable importance in forests is as follows. By first randomly permuting the predictor variable X_j, the original association with the response variable Y is broken. When the permuted variable is used along with other non-permuted variables for predicting the response for the out-of-bag observations, the classification accuracy decreases substantially if the permuted variable is associated with the response. Therefore, the importance of a variable is the difference in the prediction accuracy before and after the permutation of variable X_j, averaged over all trees. Out-of-bag observations are those that the method excluded while developing the trees; they form an internal test dataset and there is no need to allocate a test dataset separately. Let $B^{(t)}$ be the out-of-bag sample for a tree t with $t \in \{1,.....,ntree\}$. Then, the variable importance of one tree is given by the following:

$$VI^{(t)}(x_j) = \frac{\sum_{i \in B^{(t)}} I(y_i = \hat{y}_i^{(t)})}{|B^{(t)}|} - \frac{\sum_{i \in B^{(t)}} I(y_i = \hat{y}_{i,\pi_j}^{(t)})}{|B^{(t)}|} \tag{7.4}$$

where $\hat{y}^{(t)} = f^{(t)}(x_i)$ is the predicted classes for observation i before and $\hat{y}_{i,\pi_j}^{(t)} = f^{(t)}(x_{i,\pi_j})$ is the predicted classes for observation i after permuting its value of variable. The raw variable importance score for each variable is then computed as the mean importance over all trees and is given by:

$$VI(x_j) = \frac{\sum_{t=1}^{ntree} VI^{(t)}(x_j)}{ntree} \tag{7.5}$$

Since the individual importance scores $VI^{(t)}(x_j)$ are computed from ntree independent bootstrap samples, a simple test for the relevance of variable X_j can be constructed based on the central limit theorem for the mean importance of $VI^{(t)}(x_j)$. If individual importance has a standard deviation σ, then the mean importance from ntree replications has a standard error of $\sigma / \sqrt{ntree}$.

The next section emphasizes on the results of the random forests results for the various severity models developed on the urban/suburban and rural corridors according to the various crash types.

7.4 Analysis and Results

7.4.1 Conditional Inference Forest Variable Importance Results

In the present study, the conditional inference forests generated for the models –with the binary severity variable as the target– give the variable importance score for all the variables in the model. The sign (positive/negative) of the importance score indicates whether the presence or absence of a variable in the model will improve or lower the efficiency of the model. In other words, it is an indicator of how well the parameters are associated with the target variable. Table 7-2 below and Table 7-3, page 114, show the conditional inference forest result for the angle/turning movement crashes belonging to Cluster 3. Results in Table 7-2 are for the model with only environmental and roadway geometric factors and those in Table 7-3 are for the driver and vehicle-related characteristics' model. As a reminder to the reader, Table 7-1, pages 103 and 104, has the explanation of the variables.

Table 7-2: Conditional Inference Forest Sample Result for Environmental and Roadway Geometric Factors

Variable Name	Variable Importance Score
Shoulder + Side	0.000358
Pavement condition	0.00026
Median Openings	0.000163
Median type	0.000163
T-factor	0.00013
Vision obstruction	6.50E-05
Skid (friction resistance)	6.50E-05
Roadway condition	0
Horizontal Degree of Curvature	0
Surface condition	0
Parking type	0
Traffic-way character	0
Surface width	-9.76E-05
K-factor	-6.50E-05
Day of the week and time of the day	-6.50E-05
Surface type	-3.25E-05
Daylight condition	-3.25E-05
Roadway element	-0.00013
Maximum posted speed limit	-0.00026
adt	-0.00029
Shoulder type	-0.00036

Table 7-3: Conditional Inference Forest Sample Result for Driver- and Vehicle-Related Factors

Variable Name	Variable Importance Score
Alcohol use	0.004544
Age group	0.004488
Vehicle movement	0.000309
Safety equipment use	0.00014
Vehicle type	5.61E-05
At fault driver gender	2.81E-05
Vulnerable age group	2.81E-05
Presence of more than 5 persons	0

The variables with a positive variable importance score are the most important for the severity model developed here in the example. Their association with the target variable is the maximum and their absence would decrease the model performance. The variables with zero importance score are believed to have no effect on the severity model, while the ones with negative importance (refer to highlighted items in Table 7-2, page 113) are the ones decreasing the model performance. Researchers may be inclined to remove the variables with negative importance score and recalculate the scores again. This iterative process can go on until we get a subset of the original variables and all of them have positive importance scores. This may not be the best method to follow, as the number of variables left will be minimal. It is important to distinguish the significant variables from the insignificant ones. Variables which are negatively or neutrally associated with the severity help the analysts to draw appropriate results. As the dataset changes (i.e. a new model is being developed) the importance score may also change. A corridor group may have a particular variable that positively associates with severity, but this association/relationship may vary for another corridor group. All the conditional inference forests results were developed at 90% confidence level.

Table 7-4 and Table 7-5, page 116, display the random forests results developed for all severity models in the study. For certain crash types (namely: head-on, sideswipe, single vehicle involved, slow moving vehicles involved), the total number of crashes corresponding to urban Clusters 1 and 2 was not sufficient for the trees to develop; therefore, Clusters 1 and 2 were combined for these crash types. All the results were developed with the use of the R statistical software package; the package *party* developed by Hothorn et al. (2008) was used to generate the conditional trees and forests results. The key for Table 7-4 and Table 7-5, both in page 116, is as follows:

- +: variables which increase the model efficiency,
- –: variables which decrease the model efficiency, and
- 0: variables which are neutral to model efficiency.

Table 7-4: Severity Models' Conditional Inference Forests Results for Urban Clusters with Environmental and Roadway Geometric Factors

	Cluster 1		Cluster 2		Cluster 1 & 2				Cluster 3						Cluster 4					
variable	angle	rearend	angle	rearend	headon	sideswipe	single	slow	angle	rearend	headon	sideswipe	single	slow	angle	rearend	headon	sideswipe	single	slow
surface_width	+	0	-	-	0	0	-	+	-	0	+	0	-	+	+	-	+	0	+	+
max_speed	-	0	+	-	-	0	-	-	-	+	-	0	-	-	+	0	0	0	-	+
LIGHTCDE	+	0			0	0	0	0							0	0	0	-	0	0
ACMANCLS_num	+	0	+	0	+	-	-	0	+	0	-	0	-	0	+	-	+	0	0	+
road_cond	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	0	0
vision	0	0	0	0	0	0	0	0	+	0	+	0	0	+	+	0	-	-	0	0
shld_side	+	0	+	0	0	+	-	-	+	0	0	0	0	+	+	-	0	0	+	+
curvclass	-	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0
surf_cond	+	0	-	0	-	-	-	-	0	0	-	0	-	+	+	0	0	0	0	-
light	-	0	0	0	0	0	+	0	-	0	0	0	-	+	0	0	0	+	0	+
ADT	+	0	+	+	-	0	-	+	-	+	+	0	-	+	+	-	0	-	-	+
t_fact	+	0	-	0	-	0	-	+	+	+	-	0	-	+	-	+	+	0	+	+
k_fact	+	0	+	0	-	-	0	0	-	0	-	0	+	-	+	0	0	0	-	+
dayandtime	-	0	+	0	0	-	+	0	-	-	-	0	-	0	+	-	0	0	+	+
trfcway	0	0	-	-	0	0	+	0	0	0	+	0	-	+	+	0	0	0	0	+
pavecond	+	0	-	0	+	0	-	+	+	0	0	0	+	0	+	-	+	-	+	-
park	+	0	0	0	0	0	0	0	0	0	0	0	0	-	0	-	0	0	0	+
surf_type	0	0	-	0	-	0	0	0	-	0	0	0	0	+	0	+	0	0	0	+
skid_f	+	0	+	0	+	0	+	+	+	-	+	0	0	+	+	0	+	0	-	+
median	+	0	+	-	+	0	0	+	+	0	+	0	-	+	+	+	+	-	+	+
element1	+	0	+	0	-	0	-	0	-	+	+	0	+	+	-	+	+	0	-	+
shld_t	+	0	-	0	-	0	+	0	-	0	-	0	-	+	+	+	0	-	-	+

Table 7-5: Severity Models' Conditional Inference Forests Results for Urban Clusters with Driver- and Vehicle-Related Factors

	Cluster 1		Cluster 2		Cluster 1 & 2				Cluster 3						Cluster 4					
variable	angle	rearend	angle	rearend	headon	sideswipe	single	slow	angle	rearend	headon	sideswipe	single	slow	angle	rearend	headon	sideswipe	single	slow
age_gr	-	0	+	0	-	0	-	+	+	+	+	-	-	+	+	0	+	+	-	+
veh_type1	+	-	+	0	0	0	+	-	+	0	+	-	+	+	+	0	-	-	+	+
alcohol_use	-	-	+	0	0	0	+	0	+	-	+	0	+	+	+	0	+	0	+	-
more	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sfty	+	0	+	0	0	0	+	+	+	0	-	0	+	+	-	0	0	0	-	+
gender	+	0	+	0	+	0	-	0	0	+	+	0	+	-	-	0	-	+	+	0
veh_move1	-	0	+	0	-	0	+	+	-	+	0	-	-	+	+	+	-	-	+	+
vuln_age	0	0	+	0	0	0	0	0	-	0	0	0	-	0	+	0	0	0	0	-

As mentioned earlier, the variables with + sign in the boxes are the variables with more importance than others. The ones with 0 indicate that they are neutral for the severity model. The variables with – are the ones with least effect on the corresponding model. Also, it has to be noted that the + sign need does not necessarily mean that a variable is positively associated with severity. For a better interpretation of the variable's influence on the severity, single conditional inference trees were developed for the models; depending on how the variables split, the approach to severe/fatal crashes would be clearer.

7.4.2 Conditional Inference Tree Results

The conditional inference trees are critical to observe which parameters are related more with severity. Figure 7-1 below and Figure 7-2, page 118, are illustrative examples of individual conditional inference trees developed. The trees are for angle/turning movement crashes corresponding to Cluster 1 corridors.

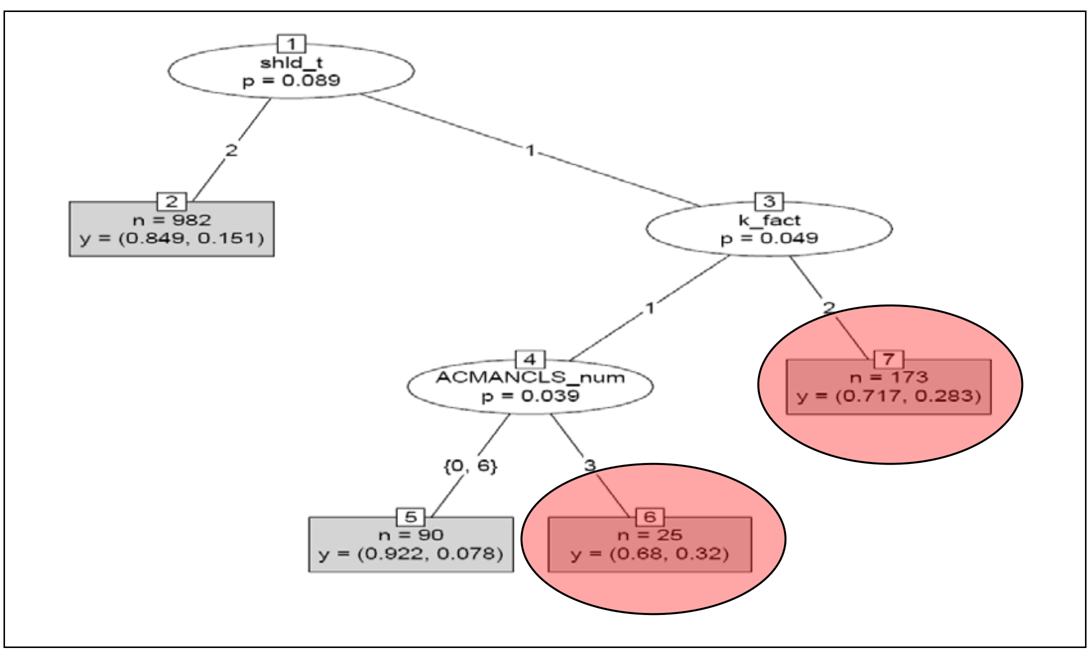


Figure 7-1: Conditional Inference Tree Sample Result for Environmental and Roadway Geometric Factors

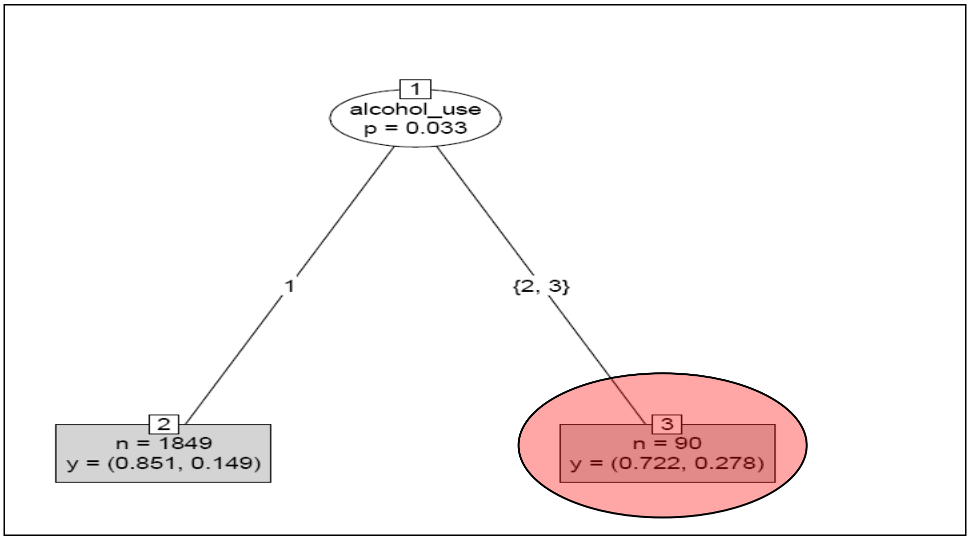


Figure 7-2: Conditional Inference Tree Sample Result for Driver- and Vehicle-Related Factors

All the trees were developed at 90% confidence level. In Figure 7-1, page 117, and Figure 7-2 above, the terminal nodes or the leaves of the trees that are enclosed in red oval shapes are the ones with higher proportion of severe crashes as compared to the proportion of severe crashes for a particular type in a certain cluster. The number of crashes in the particular leaf is represented by n, whereas the proportion of non-severe and severe crashes is indicated by the first and second values of y enclosed within parenthesis, respectively. The path taken to reach the node with higher proportion of severe crashes is critical and the factors on the path are the subject of the discussion here. These variables and their split throw light on to the safety issues of the corridors under study.

For angle/turning movement crashes corresponding to Cluster 1 (1.009 – 2.89 miles) corridors, the severity is higher where the shoulders are paved and the K-factor is higher. The higher K-factor indicates that, higher the design hour volume (which is an indicator of the design peak hour volume) the higher risk it involves for such crashes. Cluster 1 comprises the shortest corridors and thus the higher K-factor for severe/fatal crashes is intuitive in nature. With lower K-factor but restrictive medians (with longer distance between openings), the severity of the crashes is found to be higher. For the same cluster, alcohol/drug use is also found to be associated with severe/fatal crashes. For Cluster 2 (2.898 – 5.729 miles) corridors, posted speeds greater than 40 mph seem to be associated with a higher risk. Along with that, failure to use safety equipment and alcohol/drug use also lead to severe/fatal crashes. For Cluster 3 (5.762 – 10.556 miles) corridors, posted speeds of greater than 50 mph along with the alcohol/drug use and non-use of safety equipment again lead to crashes which are more at threat to be severe. For Cluster 4 (10.644 – 78.293 miles) corridors, the parameters for severe/fatal crashes are higher *adt* and no daylight, but the results are at a lower confidence level (70% as compared to 90% for other clusters); thus, it may not be very practical to assume their inference. Therefore, it can be concluded from the results that angle/turning movement crashes are more severe under high speeds, non-use of safety equipment and DUI; the results are consistent with common perception.

With regards to rear-end crashes, they are more severe along Cluster 1 corridors where there is higher friction resistance (skid > 38). This is counterintuitive as higher friction should be better at preventing severe crashes. The results could provide insight to the phenomenon that when the friction is higher and the vehicles can brake within shorter distances, the internal movement could be sudden and any internal/secondary collision (i.e. passengers hitting something inside the vehicle) could lead to a severe injury. The severe injuries are also related to light trucks where the aforementioned explanation fits perfectly. However, the severe/fatal crashes are also found to be linked to light, slow moving vehicles such cycles, mopeds, etc. The higher severity level makes sense, as any crash with these vehicles will generally be more severe no matter what the friction resistance of the road is. For Cluster 2 corridors, posted speed limits greater than 50 mph lead to severe rear-end crashes. When speeds are less than 50 mph, crashes will be severe/fatal when the K-factor is high. For the same Cluster 2 corridors, alcohol/drug use leads to crashes which are severe/fatal. When there is no alcohol/drug use (responsible driver) the presence of a person in the vulnerable age group (> 55 yrs or < 3 yrs) makes the crash more severe in general. While the former is a case of irresponsible behavior, the latter is a clear case of physical fragility. Persons in the vulnerable age group always tend to experience severe injuries resulting out of a crash. For Cluster 3 corridors, lower *adt* and alcohol/drug use leads to severe/fatal crashes. Lower *adt* could mean higher speeds which more often than not lead to severe/fatal crashes. For Cluster 4 corridors, which comprise longer corridor groups, higher friction resistance (skid > 34) leads to severe rear-end crashes. The explanation has been given in the beginning of the paragraph. For lower friction resistance higher surface widths (corresponding to 3 or more lanes per direction) and the presence of median curb increases the severity level of the crashes. On the same corridor group, older drivers (> 55 yrs) are also involved in severe rear-end crashes. The longer the corridors the more the exposure of the driver and the older the driver the more prone is he/she to make an error.

For head-on crashes on corridors belonging to Clusters 1 and 2 combined, crashes on dry surface condition were found to be more severe/fatal. Changing lanes is also associated with severe crashes (although low confidence level of 70%). Dry surface conditions probably indicate acceptable weather and more vehicles on the road; thus, an improper lane change could result in a head-on collision especially when the highways are undivided. For Clusters 3 and 4, alcohol/drug use is the primary reason for head-on crashes. In sideswipe crashes, restrictive medians are more dangerous on shorter corridors, while on longer corridors straight ahead movement is crucial. For all other type of movements, slow moving vehicle types and light trucks lead to more severe sideswipe crashes. For severe/fatal crashes involving slow moving vehicles, alcohol/drug use and changing lanes are the significant parameters on longer corridors. These crashes are also found to be more severe when occurring at signalized intersections. They are also severe during no daylight hours. For crashes involving single vehicles, higher friction factor also leads to increased severity in crashes on shorter length corridors (i.e. Clusters 1 and 2 combined). In the same corridors, straight vehicle movement-related crashes are found to be more severe. For the same type of crashes occurring along Cluster 3 corridors, that are related to segments or access points, the crashes tend to be more severe at stretches where the posted speed limits are 40 mph or greater. Failure to use safety equipment in single and slow moving vehicles also leads to severe injuries in crashes. For Cluster 4 corridors, crashes are at a higher risk of being severe when the posted speed limit is greater than 50 mph. For this same cluster, slow moving single vehicles generally experience severe crashes even though safety inside the car was taken care of.

7.5 Conclusions

The application of conditional inference trees and forests leads to the identification of an unbiased set of variables significantly related with severity. The advantage of the new type of tree/forest development over the traditional CART tree/forest is that it prevents the uninformative variables from being identified as significant just by the virtue of having higher number of categories or being continuous in nature. The novel way of separating the split criteria from the variable importance selection while developing a tree is what makes the conditional trees unique. The chi-square tests are used to determine the strength of association with the target variable; in the present application it is the binary severity variable. Once a variable is selected at a particular tree level for split, the split can then be decided based on any criteria, including those used in the CART algorithm. The conditional inference forests on the other hand calculates individual variable importance of each variable for every tree by first breaking the association with permutation and then testing the tree with out-of-bag estimates. In the forests, the variable importance is based on the result from multiple trees thus avoiding the instability of individual trees.

Among the results from the analysis, alcohol/drug use is associated with increased severity of crashes irrespective of the length of the corridors or the type of crashes. Since the drivers are less likely to be in control, it invariably leads to severe crashes. Failure to use safety equipment has lead to increased severity of single vehicle as well as angle-/turning movement-related crashes. In this regard, conclusions drawn by Abdel-Aty and As-Saidi (2000), by analyzing the zip codes of the offenders for better targeting the education programs, may be of renewed interest. Older at fault drivers are found to be more at risk of getting involved in a severe crash especially in a rear-end collision on longer corridors. On similar corridors, a crash is more likely to have a severe injury where there is person in the vulnerable age group (more than 55 years or less than 3 years).

Slow moving vehicles like cycles and mopeds have been observed to be involved in severe injury crashes. Many of these severe crashes occur at signalized intersections. It indicates that the designs of the intersections need to improve with respect to the slow-moving vehicle and possibly even pedestrians. For shorter urban corridors, a higher K-factor is a significant parameter for increased severity crashes. A higher K-factor essentially means that the corridor is designed for handling higher volume during the peak hour. In turn, it has the potential not only to reduce rear-end crashes during the peak hour (due to improved congestion situation), but also to increase speeds due to better design during off-peak periods. Since rear-end crashes tend to be less severe, a higher K-factor leads to increased likelihood of severe crashes.

Along Cluster 3 corridors (i.e. longer corridors), the severity of rear-end crashes increases when the posted speed limit is greater than 50 mph. Lowering the posted speed limit may not be the best strategy from an operations point of view; however, it may lead to a reduction in crash severity. A lower *adt* also leads to severe rear-end crashes on certain corridors. Severe/fatal crashes involving single vehicles are more likely to be associated with access points on longer corridors. Regarding the latter issue, to reduce the number of access points may not be adopted; however, design changes –such as improved merging– may be a viable option.

Corridors of smaller lengths (generally less than 5 miles) have been observed to have problems of increased severity if crashes occur on corridors with high skid resistance values. Shorter corridors also have problems when the posted speed limit is greater than 40 mph. Since most of these small urban/suburban corridors are located between longer stretches of rural corridors, they have lower speed limits when compared to adjacent sections. However, since congestion is not high on the rural sections, some drivers will tend to speed and thus create a larger variation in prevailing speeds; this variation could lead to more severe crashes on shorter length corridors. Restrictive median openings on shorter corridors have also been found to be problematic. The variable indicating the presence of subjects from the vulnerable age group also came out significant for shorter corridors rather than for longer corridors. On longer length (greater than 5 miles approx.) corridors, speed limit of greater than 50 mph is a cause of concern. It is worth noting that the newly developed variable, *element* which assigns crashes was useful in identifying roadway elements on longer corridors where severe crashes tend to occur. This makes sense as in longer corridors we will have more number of occurrences of the segments, intersections and access points. In shorter sections, on other hand, the number of intersections and access points would be limited. Also, non-use of safety equipment highly contributes to severity on longer corridors. In a recent paper by Eluru and Bhat (2007) the question of the endogenous relationship between the use of seat belts and injury severity is raised. There is possibility of intrinsically unsafe drivers not using the seat belt to be the ones more likely to be involved in high injury severity crashes, mostly due to their unsafe driving habits. In the present study, however, the researchers observe the overall safety equipment in use in the vehicle. Results also show that non-use of seat belts in single vehicle crashes and crashes involving slow vehicle lead to higher severity crashes; thus, the present study is not only in line with concurrent research, but also goes a step further in identifying the type of crashes which are more likely to be affected by the underlying endogenous relationship.

Due to these observed differences, the decision to cluster the corridors has been justified. The subtle differences are highlighted when the groups are logically made. The clusters which were originally made based on the length actually shed light on the factors and several new significant variables come into the picture.

The results from the forest and the trees are intuitive and their association with severity may be explained. Certain known results about severity of crashes have been confirmed while some new information is discovered about others. Alcohol/drug use along with higher speed limits usually results in more severe/fatal crashes. The new variable called element, which uses information from site location, signal type information and traffic control, also provided insights for identifying the most critical locations from the severity point of view. Drivers of vehicles with passengers in the vulnerable age group range must also be more careful while driving, as the physical fragility of these subjects tends to make the injuries more severe. Furthermore, the authors used the safety information for all passengers seated in the vehicle; that particular variable was also significantly associated with crash severity. Therefore, it is critical that internal safety should be a concern for law enforcement agencies if they intend to reduce the occurrences of severe/fatal crashes on Florida arterials.

CHAPTER 8. RECOMMENDATION OF SOLUTIONS FOR CORRIDORS

8.1 Background on the Previous Results

This work has produced several results that have increased our understanding of severe crash occurrence on Florida's corridors. The investigators attempted several different approaches to identify the potential causes of the problem in mention; state of the art methodologies were applied for the purpose at hand. Furthermore, innovative additions to the data were done in order to improve their quality. The authors did not automatically use the available crash location information contained in the crash reports. A set of simple heuristic rules based on empirical evidence were laid out in order to assign crashes. The authors investigated the corridors in their entirety (i.e. the segments and the intersections were not treated separately). In addition, conditional inference trees were used to identify the significant parameters associated to crash severity modeling; this was done by trying to find which variable categories lead to a higher proportion of severe crashes. Crash data were categorized into six different types: 1) rear-end, 2) head-on, 3) angle/turning, 4) sideswipe, 5) crashes involving slow moving vehicles (e.g. cycles, mopeds, etc.), and 6) crashes involving single vehicles. As mentioned in CHAPTER 5, the corridors were grouped into four clusters based on the length; thus, this would facilitate review of the results and for suggesting the respective countermeasures.

8.2 Previous Results in Brief

All results and their discussion are detailed in CHAPTER 6 through CHAPTER 7. The authors find it appropriate to discuss certain results prior to detailing the recommendations. For rear-end crashes, high friction resistance is found to be associated with a higher proportion of severe crashes. Also, higher speed limits in urban settings have been found to worsen the severity of crashes. Alcohol/drug use, as well as to have anyone within the vulnerable age group (< 3 years or >55 years) in the vehicle, both can make crashes more severe in general. On longer length corridors, a lower *adt* also leads to severe crashes. Having a low K-factor is also associated with severe rear-end crashes. Higher surface widths and presence of median curb also increase the severity of crashes. Surface condition and degree of horizontal curve also affect the severity of crashes. The location of the crash also plays a significant role in crash severity. For head-on crashes, changing lanes is an important factor in severity increase. Alcohol/drug use has been associated with severity in all types of crashes. For angle/turning movements, the severity is higher with the presence of paved shoulders. Non-use of safety equipment is also a leading cause of increased severity. Pavement conditions and certain types of medians are also associated with the severity model. For sideswipe crashes, restrictive medians are associated to a higher crash risk. For crashes involving slow moving vehicles, improper lane change is a critical factor. Higher friction results in severe crashes involving single vehicles as well. Higher truck percentage is associated with severe lane change-related crashes. Good pavement conditions reduce the likelihood of lane change-related crashes. Crashes occurring on Friday/Saturday nights have a higher likelihood of severe single vehicle crashes. Crashes near or at access points are also found to be more severe in corridor sections where the speed limit is >40 mph.

8.3 Recommendations

The following recommendations are in line with the results from this project's work and have been prepared taking the 4 Es (Education, Enforcement, Engineering and Emergency Management) into consideration. Details on the reasoning behind them, as well as the respective advantages and disadvantages, are also presented.

1) Alcohol/drug use (i.e. DUI) has resulted in increased severity across all types of crashes. The only way to reduce the number of crashes resulting from DUI is to count with a stricter law enforcement and education; however, improved emergency response systems can also help reducing the severity of the resulting crashes.
2) Higher speed limits have been responsible in many severe rear-end crashes as well as severe single vehicle collisions. Crashes occurring at access points also result in severe injuries at corridors with higher posted speed limits. To reduce the speed limit may not be the best solution considering the operational aspect of transportation. The types of crashes that have been affected due to higher speed limits also reflect that the vehicle types involved in the crashes could be a controlling factor. For these reasons, improvements in access management and median types are recommended; for example, the installation of proper crash attenuators at high risk locations should help reduce certain types of severe crash occurrence.
3) Lower *adt* along with higher posted speed limits have resulted in an increased likelihood of severe crashes, particularly of the types rear-end, sideswipe and head-on. Lower *adt* could mean more speed variance and thus more interaction among the vehicles. A possible solution is to lower the speed limit on high risk corridors; also,

the safety along these corridor sections would be improved by having stricter law enforcement and an improved emergency response system.

4) Lane change-related crashes on corridors with high truck traffic have been found to result in a higher severity of injuries. For this reason, to implement lane change-related warning signs can be effective countermeasure.

5) Poor pavement conditions and road defects are also responsible for an increased likelihood of severe crashes along these corridors. Road surface improvements, as well as a redesign of such corridors, could ameliorate the severity situation.

6) Higher friction resistance has also resulted in crashes with a higher severity of injuries. This is counterintuitive as higher friction should be better at preventing severe crashes. The results could provide some insight to the phenomenon as follows: when the friction is higher and vehicles can brake within shorter distances the internal movement could be sudden, and any internal/secondary collision (i.e. passengers hitting something inside the vehicle) could lead to a severe injury. Mandatory use of safety equipment by all passengers could be one recommendation that may alleviate the situation.

7) Presence of driver/passenger in the vulnerable age group (< 3 years or >55 years) has been found to increase the severity of the resulting injury. Though proper safety equipments are available, there could still be a need for better designed child restraint seats. Again, mandatory use of safety equipment by all passengers could be one recommendation that may alleviate the situation.

8) Certain types of severe crashes are associated with higher degree of horizontal curvature. There may be a need to revisit the design documentation of those sections.

Increasing the radius (where possible) may avert accidental lane departures and reduce the likelihood of crashes.

9) Roadside parking has resulted in severe single vehicle-related crashes as well as pedestrian crashes. Increased roadside parking results in higher number of mid-block crossings, hence the above observation. Though roadside parking cannot be avoided altogether, especially in urban settings, appropriate measures should be considered for decreasing mid-block crossings. In addition, widening of roads may reduce interactions with parked cars and associated signage.

8.4 Conclusion

The recommendations presented are general in nature. The ones on law enforcement and improved emergency response systems could be taken care of almost immediately. On the other hand, the recommendations involving design and speed limit changes require further studies by location(s). For example, reduction in speed limit should be supported by a simulation study to make sure that transportation operations are not negatively affected; this type of studies are beyond the scope of this project.

LIST OF REFERENCES

Abdel-Aty, M. "Analysis of driver injury severity levels at multiple locations using probit models," Journal of Safety Research, Vol. 34, No. 5, 2003, pp. 597-603.

Abdel-Aty, M. and H. T. Abdelwahab. "Predicting injury severity levels in traffic crashes: a modeling comparison," Journal of Transportation Engineering, Vol. 130, No. 2, 2004, pp. 204-210.

Abdel-Aty, M. and A. H. As-Saidi. "Using GIS to locate the high risk driver population," Swedish National Road and Transport Research Institute, 2000, pp. 111-126.

Abdel-Aty, M. and J. Keller. "Exploring the overall and specific crash severity levels at signalized intersections," Accident Analysis & Prevention, Vol. 37, No. 3, 2005, pp. 417-425.

Abdel-Aty, M. and A. Pande. "Comprehensive analysis of relationship between real-time traffic surveillance data and rear-end crashes on freeways," Transportation Research Record 1953, 2006, pp. 31-40.

Abdel-Aty, M. and A. Pande. "Crash data analysis: Collective vs. individual crash level approach," Journal of Safety Research, Vol. 38, No. 5, 2007, pp. 581-587.

Abdel-Aty, M., A. Pande, A. Das and W. J. Knibbe. "Analysis of infrastructure based ITS data for assessing safety on freeways in Netherlands." Presented at the 87th Annual Meeting of the Transportation research Board, Washington, DC, 2008.

Abdel-Aty, M., C. Lee, X. Wang, P. Nawathe, J. Keller, S. Kowdla and H. Prasad. "Identification of Intersections' Crash Profiles/Patterns." Florida Department of Transportation, Tallahassee, 2006.

http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_SF/FDOT_BC355_10_rpt.pdf Accessed June 10, 2007.

Abdel-Aty, M., R. Pemmanaboina and L. Hsia. "Assessing crash occurrence on urban freeways by applying a system of interrelated equations," Transportation Research Record 1953, 2006, pp. 1-9.

Abdel-Aty, M. and A. E. Radwan. "Modeling traffic accident occurrence and involvement," Accident Analysis & Prevention, Vol. 32, No. 5, 2000, pp. 633-642.

Abdel-Aty, M., N. Uddin, F. Abdalla, A. Pande and L. Hsia. "Predicting freeway crashes based on loop detector data using matched case–control logistic regression," Transportation Research Record 1897, 2004, pp. 88-95.

Abdel-Aty, M. and X. Wang. "Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors." Presented at the 85th Annual Meeting of the Transportation research Board, Washington, DC, 2006.

Bedard, M., G. H. Guyatt, M. J. Stones and J. P. Hirdes. "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities," Accident Analysis & Prevention, Vol. 34, No. 6, 2002, pp. 717-727.

Bichler-Robertson, G., G. Laycock, R. Clarke, R. Sampson, G. Cordner, G. Saville, R. Glensor, M. Scott and N. L. Vigne (2001). "Excellence in Problem-Oriented Policing: The 2001 Herman Goldstein Award Winners." Police Executive Research Forum, U.S. Department of Justice, http://www.policeforum.org/upload/01Goldstein%5B1%5D_715866088_12302005112938.pdf Accessed March 17, 2007.

Bonneson, J. A. and P. T. McCoy. “Effect of median treatment on urban arterial safety: an accident prediction model,” Transportation Research Record 1581, 1997, pp. 27-36.

Bowman, B. L., R. L. Vecellio and J. Miao. “Vehicle and pedestrian accident models for median locations,” Journal of Transportation Engineering, ASCE, Vol. 121, No. 6, 1995, pp. 531-537.

Breiman, L. “Random forests,” Machine Learning, Vol. 45, No. 1, 2001, pp. 5-32.

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone (1984). *Classification and Regression Trees*, Wadsworth International Group, Belmont.

Breyer, J. P. and S. C. Joshua (1999). “Identifying and Implementing Corridor Safety Improvements: A Highway Safety Improvement Process and Safety Analysis Tools for Arizona.” Publication FHWA-AZ 99-458. Arizona Department of Transportation and FHWA, U.S. Department of Transportation.

Brijs, T., D. Karlis, F. Van den Bossche and G. Wets (2003). “A Bayesian model for ranking hazardous sites.” Flemish Research Center for Traffic Safety, Diepenbeek, Belgium.

Brown, H. C. and A. P. Tarko. “Effects of access control on safety on urban arterial streets,” Transportation Research Record 1665, 1999, pp. 68-74.

Caliendo, C., M. Guida and A. Parisi. “A crash-prediction model for multi-lane roads,” Accident Analysis & Prevention, Vol. 39, No.4, 2007, pp. 657-670.

Carlin, B. P. and T. A. Louis (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd Edition, Chapman & Hall, Boca Raton.

Categorical Dependent Variable Models Using SAS, STATA, LIMDEP, and SPSS, Research Technologies at Indiana University.

http://www.indiana.edu/~statmath/stat/all/cdvm/cdvm1.html#s12 Accessed June 4, 2007.

Chang, L. Y. and F. Mannering. “Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents,” Accident Analysis & Prevention, Vol. 31, No. 5, 1999, pp. 579-592.

Cheng, W. and S. P. Washington. “Experimental evaluation of hotspot identification methods,” Accident Analysis & Prevention, Vol. 37, No. 5, 2005, pp. 870-881.

Collett, D. (1991). *Modeling Binary data*, Chapman & Hall, London, UK.

Dahir, S. and L. G. Wade (1990). “Wet-Pavement Safety Programs. NCHRP Synthesis of Highway Practice 158”, Transportation Research Board, National Research Council, Washington, DC.

Das, A., A. Pande, M. Abdel-Aty and J. B. Santos. “Urban arterial crash characteristics related with proximity to intersections and injury severity.” Presented at the 87th Annual Meeting of the Transportation research Board, Washington, DC, 2008.

Davis, G. A. and S. Yang. “Bayesian identification of high-risk intersections for older drivers via Gibbs sampling,” Transportation Research Record 1746, 2001, pp. 84-89.

Donnell, E. and J. Mason. “Predicting the frequency of median barrier crashes on Pennsylvania interstate highways,” Accident Analysis & Prevention, Vol. 38, No. 3, 2006, pp. 590-599.

Drummond, K. P., L. A. Hoel and J. S. Miller (2002). “A simulation-based approach to evaluate safety impacts of increased traffic signal density.” http://www.virginiadot.org/vtrc/main/online_reports/pdf/02-r17.pdf Accessed January 30, 2007.

Drummond, K. P., L. A. Hoel and J. S. Miller. “Using simulation to predict safety and operational impacts of increasing traffic signal density,” Transportation Research Record 1784, 2002, pp. 100-107.

Dummeldinger, M., P. Henderson, B. G. Ward and V. Zambito (1994). "Corridor Safety Improvement Program: Impact Evaluation", University of South Florida, Center for Urban Transportation Research, Tampa, http://www.lib.usf.edu/cgi-bin/Ebind2h3.pl/cutr0253 Accessed March 15, 2007.

Duncan, C., A. Khattak and F. Council. "Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions," Transportation Research Record 1635, 1999, pp. 63-71.

Eluru, N. and C. R. Bhat. "A joint econometric analysis of seat belt use and crash – related injury severity," Accident Analysis & Prevention, Vol. 39, No. 5, 2007, pp. 1037-1049.

Elvik, R. "The safety value of guardrails and crash cushions: a meta-analysis of evidence from evaluation studies," Accident Analysis & Prevention, Vol. 27, No. 4, 1995, pp. 523-549.

Elvik, R. "Evaluations of road accident black spot treatment: a case of the iron law of evaluation studies?," Accident Analysis & Prevention, Vol. 29, No. 2, 1997, pp.191-199.

Flahaut, B., M. Mouchart, E. S. Martin and I. Thomas. "The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach," Accident Analysis & Prevention, Vol. 35, No. 6, 2003, pp. 991-1004.

"Florida Corridor/Community Traffic Safety Program", NHTSA, summer 1996, http://www.nhtsa.dot.gov/people/outreach/safedige/Summer96/FHWA/Florida.html Accessed February 18, 2007.

Fontaine, M. D. and S. W. Read (2006). "Development and Evaluation of Virginia's Highway Safety Corridor Program", Publication FHWA/VTRC 06-R30. Virginia Department of Transportation and FHWA, U.S. Department of Transportation.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin (2003). *Bayesian Data Analysis*, 2nd Edition, Chapman & Hall, Boca Raton.

Geurts, K., G. Wets, T. Brijs and K. Vanhoof. "Identifications and ranking of black spots: sensitivity analysis." Presented at the 83rd Annual Meeting of the Transportation research Board, Washington, DC, 2004.

Geurts, K. and G. Wets (2003). "Black Spots Analysis Methods: A literature review", Flemish Research Center for Traffic Safety, Diepenbeek, Belgium.

Golob, T., W. Recker and V. Alvarez. "Freeway safety as a function of traffic flow," Accident Analysis & Prevention, Vol. 36, No. 6, 2004, pp. 933-946.

"Governor's Task Force on Highway Safety", State of Ohio Government Info and Services, August 4, 2005, http://corridorsafety.ohio.gov/ Accessed March 2, 2007.

Greene, W. H. (2003). *Econometric Analysis,* 5th Edition, Pearson Education, USA.

Green, E. R. and K. R. Agent (2002). "Evaluation of High Traffic Crash Corridors", Publication KTC-02-8/SPR231-01-1F. Kentucky Department of Transportation.

Greibe, P. "Accident prediction models for urban roads," Accident Analysis & Prevention, Vol. 35, No. 2, 2003, pp. 273-285.

Hanley, K. E., A. R. Gibby and T. C. Ferrara. "Analysis of accident-reduction factors on California State Highways," Transportation Research Record 1717, 2000, pp. 37-45.

Harb, R., A. E. Radwan, X. Yan, A. Pande and M. Abdel-Aty. "Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression," Journal of Transportation Engineering, Vol. 134, No. 5, 2008, pp. 203-215.

Harb, R., X. Yan, A. E. Radwan and X. Su. "Crash avoidance analysis using classification trees and random forests." Presented at the 87th Annual Meeting of the Transportation research Board, Washington, DC, 2008.

Hauer, E. "On the estimation of the expected number of accidents," Accident Analysis & Prevention, Vol. 18, No. 1, 1986, pp. 1-12.

Hauer, E., D. W. Harwood, F. M. Council and M. S. Griffith. "Estimating safety by the empirical Bayes method: a tutorial," Transportation Research Record 1784, 2002, pp. 126-131.

"Highway 25 Safety Corridor". California Department of Transportation, November 13, 2003, http://www.dot.ca.gov/dist05/projects/hwy25/corridor.htm Accessed March 6, 2007.

Highway Data Collection / Quality Control Section of Transportation Statistics Office, FDOT (2007). "The RCI Office handbook", http://www.dot.state.fl.us/planning/statistics/rci/officehandbook/fullrpt.pdf Accessed April 14, 2007.

Hiselius, L. W. "Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows," Accident Analysis & Prevention, Vol. 36, No. 6, 2004, pp. 985-992.

Hothorn, T., K. Hornik and A. Zeileis. "Unbiased recursive partitioning: a conditional inference framework," Journal of Computational and Graphical Statistics, Vol. 15, No. 3, 2006, pp. 651-674.

Hothorn, T., K. Hornik and A. Zeileis (2008). "A laboratory for recursive partitioning", http://cran.r-project.org/web/packages/party/party.pdf Accessed February 28, 2008.

Hughes, R. G. (1999). "Truck Safety in North Carolina: Effectiveness of NCDMV Efforts in FY99", University of North Carolina, University of North Carolina Highway Safety

Research Center, Chapel Hill, http://www.hsrc.unc.edu/pdf/2000/cvspoo.pdf Accessed March 9, 2007.

Hunter-Zaworski, K. M. and N. T. Price (1998). "Evaluation of the Corridor Safety Improvement Program: Phase 1 Final Report", Publication FHWA-OR-RD-98-20. Oregon Department of Transportation and FHWA, U.S. Department of Transportation.

Jernigan, J. D. (1999). "Comparative case studies of corridor safety improvement efforts." http://www.virginiadot.org/vtrc/main/online_reports/pdf/00-r17.pdf Accessed February 3, 2007.

Jernigan, J. D. (1997). "Lessons learned from Virginia's pilot corridor safety improvement program", Publication VTRC 97-R11. Virginia Department of Transportation.

Jones, B., A. Griffith and K. Haas (2002). "Effectiveness of Double Fines as a Speed Control Measure in Safety Corridors: Final Report", Publication FHWA-OR-DF-03-10. Oregon Department of Transportation and FHWA, U.S. Department of Transportation.

Jorgensen, R. E. (1966). "Evaluation of criteria for safety improvements on the highway", Westat Research Analysts Inc., Gaithersburg, Maryland, USA.

Kaufman, L. and P. J. Rousseeuw (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, 9th Edition, Wiley-Interscience, New York.

Kim, K., L. Nitz, J. Richardson and L. Li. "Personal and behavioral predictors of automobile crash and injury severity," Accident Analysis & Prevention, Vol. 27, No. 4, 1995, pp. 469-481.

Knuiman, W., F. Council and D. Reinfurt. "Association of median width and highway accident rates," Transportation Research Record 1401, 1993, pp. 70-82.

Kockelman, K. M. and Y. J. Kweon. "Driver injury severity: an application ordered probit models," Accident Analysis & Prevention, Vol. 34, No. 3, 2002, pp. 313-321.

Levinson, H. S. (1999). "Access spacing and accidents – a conceptual analysis." http://onlinepubs.trb.org/Onlinepubs/circulars/ec019/Ec019_c1.pdf Accessed February 3, 2007.

Long, S. J. (1997). *Regression Models for Categorical and Limited Dependent Variables*, 1st Edition, Sage Publications, Inc., Thousand Oaks.

Ma, J. and K. M. Kockelman. "Anticipating injury & death: controlling for new variables on Southern California highways." Presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC, 2004.

McGuigan, D. R. D. "The use of relationships between road accidents and traffic flow in black-spot identification," Traffic Engineering and Control, Vol. 22, No. 8-9, 1981, pp. 448-453.

Martin, J. L. "Relationship between crash rate and hourly traffic flow on interurban motorways," Accident Analysis & Prevention, Vol. 34, No. 5, 2002, pp. 619-629.

Miaou, S. P. and J. J. Song. "Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence," Accident Analysis & Prevention, Vol. 37, No. 4, 2005, pp. 699-720.

Milton, J. C., V. N. Shankar and F. L. Mannering. "Highway accident severities and the mixed logit model: An exploratory empirical analysis," Accident Analysis & Prevention, Vol. 40, No. 1, 2008, pp. 260-266.

Milton, J. and F. Mannering. "The relationship among highway geometrics, traffic related elements and motor-vehicle accident frequencies," Transportation, Vol. 25, No. 4, 1998, pp. 395-413.

Mountain, L., B. Fawaz and D. Jarret. "Accident prediction models for roads with minor junctions," Accident Analysis & Prevention, Vol. 28, No. 6, 1996, pp. 695-707.

Mulinazzi, T. E. and H. L. Michael (1967). "Correlation of design characteristics and operational controls and accident rates on urban arterials", Joint Highway Research Project, Purdue University, Lafayette, Indiana.

National Highway Traffic Safety Administration, Traffic safety facts 2005: A compilation of motor vehicle crash data from the Fatality Analysis Reporting System and the General Estimate System, Washington, DC, 2006.

National Highway Traffic Safety Administration, Traffic safety facts 2006: A compilation of motor vehicle crash data from the Fatality Analysis Reporting System and the General Estimate System, Washington, DC, 2006.

Noland, R. B. and L. Oh. "The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data," Accident Analysis & Prevention, Vol. 36, No. 4, 2004, pp. 525-532.

O'Donnell, C. J. and D. H. Connor "Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice," Accident Analysis & Prevention, Vol. 28, No. 6, 1996, pp. 739-753.

"Oregon Safety Corridor Program Guidelines". Oregon Department of Transportation, December 2006, http://www.oregon.gov/ODOT/TS/docs/Roadway/2006Safety_Corridor_Guidelines.pdf Accessed March 10, 2007.

"Oregon Safety Corridors". Oregon Department of Transportation, January 19, 2007, http://www.oregon.gov/ODOT/TS/docs/Roadway/CorridorMasterList2007.pdf Accessed March 10, 2007.

Pande, A. and M. Abdel-Aty. "Discovering indirect associations in crash data using probe attributes." Presented at the 87th Annual Meeting of the Transportation research Board, Washington, DC, 2008. (Forthcoming in Transportation Research Record).

Papayannoulis, V., J. S. Gluck, K. Feeney and H. S. Levinson (1999). "Access spacing and traffic safety." http://onlinepubs.trb.org/Onlinepubs/circulars/ec019/Ec019_c2.pdf Accessed February 3, 2007.

Parker, M. R. (1983). "Design guidelines for raised and transversable medians in urban areas", Virginia Highway and Transportation Research Council, Charlottesville, VA.

Parker, M. R. (1990). "Simplified guidelines for selecting an urban median treatment – urban median information", Virginia Transportation Technology Transfer Center, Charlottesville, VA.

"Partitioning Around Medoids", UNESCO, http://www.unesco.org/webworld/idams/advguide/Chapt7_1_1.htm Accessed September 5, 2007.

Persaud, B. N., C. Lyon and T. Nguyen. "Empirical Bayes procedure for ranking sites for safety investigation by potential safety improvements," Transportation Research Record 1665, 1999, pp. 7-12.

Petritsch, T. A., S. Challa, H. F. Huang and R. Mussa (2007). "Evaluation of Geometric and Operational Characteristics Affecting the Safety of Six-Lane Divided Roadways", Florida Department of Transportation, http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_SF/FDOT_BD543_05_rpt.pdf Accessed March 7, 2007.

Quinlan, J. R. "Induction of decision trees," Machine Learning, Vol. 1, No. 1, 1986, pp. 81-106.

Quinlan, J. R. "C4.5: Programs for Machine Learning," Machine Learning, Vol. 16, No. 3, 1994, pp. 235-240.

RCI Features and Characteristics Handbook (2001), Florida Department of Transportation, Transportation Statistics Office, Tallahassee, FL.

Rees, J. (2003). "Corridor management: identifying corridors with access problems and applying access management treatments, a U.S. 20 study." http://www.ctre.iastate.edu/mtc/papers/2003/JREES.pdf Accessed February 10, 2007.

Saccomanno, F. F., R. Grossi, D. Greco and A. Mehmood. "Identifying black spots along highway SS107 in Southern Italy using two models," Journal of Transportation Engineering, Vol. 127, No. 6, 2001, pp. 515-522.

Sawalha, Z., T. Sayed and M. Johnson. "Factors affecting the safety of urban arterial roadways." Presented at the 79th Annual Meeting of the Transportation research Board, Washington, DC, 2000.

SAS Institute (2003). *SAS user's guide: Statistics, Version 9.1*, SAS Institute, Cary, NC.

Schlutler, P. J., J. J. Deely and A. J. Nicholson. "Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model," The Statistician, Vol. 46, No. 3, 1997, pp. 293-316.

Shankar, V. and F. Mannering. "An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity," Journal of Safety Research, Vol. 27, No. 3, 1996, pp. 183-194.

Spainhour, L. K., D. Brill, J. O. Sobanjo, J. Wekezer and P. V. Mtenga (2005). "Evaluation of Traffic Crash Fatality Causes and Effects: A Study of Fatal Traffic Crashes in Florida from 1998-2000 Focusing on Heavy Truck Crashes", Florida Department of Transportation, http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_SF/FDOT_BD050_rpt.pdf Accessed March 7, 2007.

Squires, C. A. and P. S. Parsonson. "Accident comparison of raised median and two-way left-turn lane median treatments," Transportation Research Record 1239, 1989, pp. 30-40.

Strobl, C., A. Boulesteix, A. Zeileis and T. Hothorn. "Bias in random forest variable importance measures: illustrations, sources and a solution," BMC Bioinformatics, Vol. 8, No. 25, 2007.

Tunaru, R. "Hierarchical Bayesian Models for Multiple Count Data," Austrian Journal of Statistics, Vol. 31, No. 2-3, 2002, pp. 221-229.

"The QLIM Procedure", SAS Institute Inc., Cary, North Carolina, USA, http://support.sas.com/rnd/app/papers/qlim.pdf Accessed June 6, 2007.

Virginia's Surface Transportation Safety Executive Committee (2006). "Commonwealth of Virginia's Strategic Highway Safety Plan", Virginia Department of Transportation, http://www.virginiadot.org/info/resources/Strat_Hway_Safety_Plan_FREPT.pdf Accessed March 10, 2007.

Vittinghoff, E. G., D. C. Glidden, S. C. Shiboski and C. E. McCulloch (2007). *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*, Springer, New York.

Walton, M. C., T. W. Horne and W. K. Fung (1978). "Design criteria for median turn lanes", Federal Highway Administration, Washington, DC.

Wang, X. and M. Abdel-Aty. "Temporal and spatial analyses of rear-end crashes at signalized intersections," Accident Analysis & Prevention, Vol. 38, No. 6, 2006, pp. 1137-1150.

Wang, X., M. Abdel-Aty and P. A. Brady. "Crash estimation at signalized intersections: significant factors and temporal effect." Presented at the 85th Annual Meeting of the Transportation research Board, Washington, DC, 2006.

Wang X., M. Abdel-Aty, A. Nevarez and J. B. Santos. "Investigation of safety influence area for four-legged signalized intersections: nationwide survey and empirical inquiry." Presented at the 87th Annual Meeting of the Transportation research Board, Washington, DC, 2008.

"Washington/Corridor Safety Project". NHTSA, spring 1997, http://www.nhtsa.dot.gov/people/outreach/safedige/spring1997/n4-24.html Accessed March 6, 2007.

"Washington Cape Horn Corridor Traffic Safety Project". NHTSA, Traffic Safety Digest, Vol. 2, 2004, http://www.nhtsa.dot.gov/people/outreach/safedige/Volume-2-2004/Vol2_04_W04_WA.htm Accessed March 16, 2007.

"Washington State Corridor Safety Program". Washington Traffic Safety Commission, 2006, http://www.corridorsafetyprogram.com/ Accessed March 1, 2007.

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 & Prevention, Vol. 40, No. 4, 2008, pp. 1320-1329.

Yau, K. “Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong,” Accident Analysis & Prevention, Vol. 36, No. 3, 2004, pp. 333-340.

Zhang, J., J. Lindsay, K. Clarke, G. Robbins and Y. Mao. “Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario,” Accident Analysis & Prevention, Vol. 32, No. 1, 2000, pp. 117-125.

Zogby, J. J., T. E. Bryer and J. Tenaglia (1991). “Pennsylvania Corridor Highway Safety Improvement Program”. TR News 154, May-June 1991, pp. 11-13.