A NEW APPROACH TO IDENTIFY THE EXPECTED CRASH PATTERNS BASED ON SIGNALIZED INTERSECTION SIZE AND ANALYSIS OF VEHICLE MOVEMENTS

by

HARI PRASAD SALKAPURAM
B.Tech., Indian Institute of Technology Madras, 2004

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of science
in the Department of Civil and Environmental Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2006

ABSTRACT

Analysis of intersection crashes is a significant area in traffic safety research. This study contributes to the area by identifying traffic-geometric characteristics and driver demographics that affect different types of crashes at signalized intersections. A simple methodology to estimate crash frequency at intersections based on the size of the intersection is also developed herein.

First phase of this thesis used the crash frequency data from 1,335 signalized intersections obtained from six jurisdictions in Florida, namely, Brevard, Seminole, Dade, Orange, and Hillsborough Counties and the City of Orlando. Using these data a simple methodology has been developed to identify the expected number of crashes by type and severity at signalized intersections. Intersection size, based on the total number of lanes, was used as a factor that was simple to identify and a representative of many geometric and traffic characteristics of an intersection. The results from the analysis showed that crash frequency generally increased with the increased size of intersections but the rates of increase differed for different intersection types (i.e., Four-legged intersection with both streets two-way, Four-legged intersection with at least one street one-way, and T-intersections). The results also showed that the dominant type of crashes differed at these intersection types and severity of crashes was higher at the intersections with more conflict points and larger differential in speed limits between major and minor roads. The analysis may potentially be useful for traffic engineers for evaluating safety at signalized intersections in a simple and efficient manner. The findings in this analysis provide strong evidence that the patterns of crashes by type and severity vary with the

size and type of intersections. Thus, in future analysis of crashes at intersections, the size and type of intersections should be considered to account for the effects of intersection characteristics on crash frequency.

In the second phase, data (crash and intersection characteristics) obtained from individual jurisdictions are linked to the Department of Highway Safety and Motor Vehicles (DHSMV) database to include characteristics of the at-fault drivers involved in crashes. These crashes are analyzed using contingency tables and binary logistic regression models. This study categorizes crashes into three major types based on relative initial movement direction of the involved vehicles. These crash types are, 1) Initial movement in same direction (IMSD) crashes. This crash type includes rear end and sideswipe crashes because the involved vehicles for these crashes would be traveling in the same direction prior to the crash. 2) Initial movement in opposite direction (IMOD) crashes comprising left-turn and head on crashes. 3) Initial movement in perpendicular direction (IMPD) crashes, which include angle and right-turn crashes. Vehicles involved in these crashes would be traveling on different roadways that constitute the intersection. Using the crash, intersection, and at-fault driver characteristics for all crashes as inputs, three logistic regression models are developed. In the logistic regression analyses total number of through lanes at an intersection is used as a surrogate measure to AADT per lane and also intersection type is introduced as a 'predictor' of crash type. The binary logistic regression analyses indicated, among other results, that at intersections with one-way roads, adverse weather conditions, older drivers and/or female drivers increase the likelihood of being at-fault at IMOD crashes. Similar factors associated with other groups of crashes (i.e., IMSD and IMPD) are also identified. These findings from the study may be used to develop specialized training programs by zooming in onto problematic intersections/maneuvers.

ACKNOWLEDGMENTS

I would firstly like to thank my advisor, Dr. Mohamed Abdel-Aty, for being helpful and supportive throughout my Masters study at UCF. I would also like to thank Dr. Essam Radwan and Dr. Anurag Pande for serving in my committee. I thank Dr. David Cooper for his support. I thank Piyush Nawathe and Dr Chris Lee for their help from the beginning of Masters. I thank Abdul Rawoof and Shasank for being ready to discuss any aspect of my thesis at any point of time. My friend circle at UCF, Albinder, Nezam, Jai, Ravi, Raja, Sai, Xuesong, Subash, Diwakar, Praveen, and Kishore has been a continuous source of encouragement. Finally, I am indebted to my parents for encouraging me to perform at the best of my potential and to never give up.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

Traffic crashes are the most undesirable transportation outcome. They cause loss of wages, time, productivity, and especially loss of human lives, for which value cannot be estimated. Intersection-related crashes make up a very high percentage of the total crashes in the roadway system. According to Federal Highway Administration, Intersection Safety Facts and Statistics, 2004, over 9,117 Americans lost their lives as a result of intersection-related crashes. It also states that each year more than 2.7 million intersection crashes occur (over 45% of all reported crashes) and more than one intersection fatality occurs every hour. The cost to society for intersection-related crashes is approximately $40 billion every year. These statistics indicates that there is a tremendous need for improving traffic safety especially at intersections, where crashes happen more frequently as compared to roadway segments.

Figure 1.1 presents the national statistics for crashes by location and crash severity for 1999. For all fatal crashes, 22.98% occurred at intersections or intersection-related locations. Among all traffic crashes, 44.69% occurred at intersections or intersection-related locations. For injury crashes, the percentage is close to 50%, while for property damage only (PDO) crashes it is over 42%. The main reason for the high percentage of crashes at intersection (or intersection related) is that the intersections are areas shared by two or more roads, where roadway users including vehicle drivers, cyclists, and pedestrians have to make a decision or are confronted with many choices to make, whether to stop or keep going, go left, right or straight, etc. The complexity of movements of vehicles at an intersection is the basic problem for intersections resulting in too many conflict points. Usually, if a traffic conflict is not avoided, traffic crash will occur.

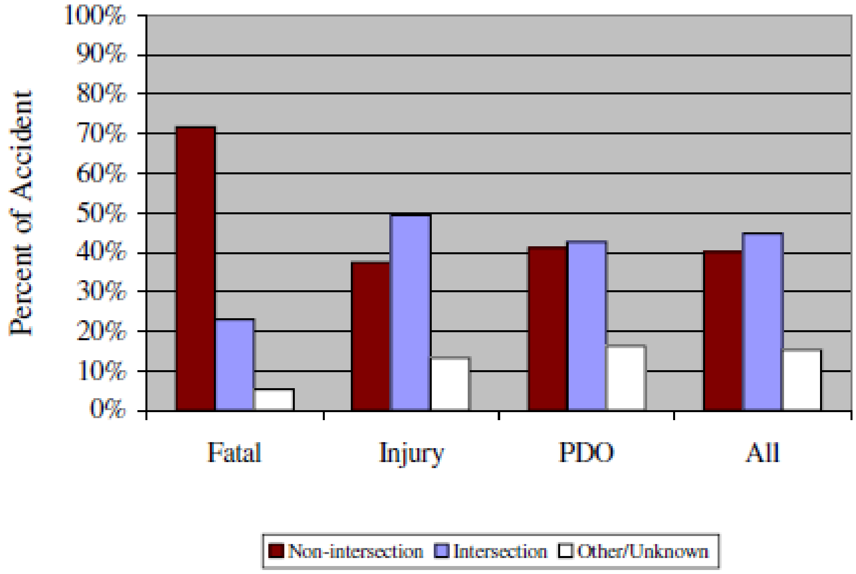


Figure 1.1 National Statistics of Accidents by Location and Severity
(Source: Traffic Safety Facts 1999)

According to Federal Highway Administration's issue briefs, Human factors issues in intersection safety (2004), driver error account for approximately 90 percent of all crashes. It also states that while advances in automotive safety and highway design continue to improve, the one component that has not changed is the driver. Understanding how drivers and all roadway users interact within an intersection environment is fundamental to improve roadway safety and save lives.

Negotiating through intersections is one of the most complex and demanding tasks a driver faces. To successfully execute a vehicle maneuver through an intersection, the driver must assimilate the information, make a decision and execute the desired action. One of the critical limitations is that human brains are serial processors and the cognitive task-load at intersections can be quite larger than segments. Common items a driver must consider when approaching an intersection include (but not limited to); Monitoring and adjusting speed, maintaining lane position, being aware of other vehicles, attending to signals or signs, scanning for pedestrians, bicyclists, people in wheelchairs and blind or visually impaired people, decelerating for a stop, searching for path guidance, and selecting proper lane.

Given the short time drivers have to process a large amount of information, it is imperative that designers and engineers provide clear and accurate information to drivers to help them navigate an intersection. Vision is the most important information reception characteristic of drivers. Perceptual failures also account for a large portion of driver errors. These can include such items as “looked but failed to see,” visual obstructions, reduced visibility due to environmental factors, poor judgment of speed and/or distance and low conspicuity of target. However, distractions, misinterpretation of information and driver impairment are also major contributing factors. Intersections themselves present their own unique set of driver errors, depending on the type of intersection at hand.

A valid approach to address safety at intersection is through analysis of intersection crash patterns with the idea of exploring different crash patterns at intersections and their relationship to involved drivers and intersection characteristics.

1.2 Research Objectives

The main aim of this thesis is to analyze the crash characteristics at signalized intersections with respect to characteristics of the intersections and at-fault drivers. The objectives of this study towards that aim are as follows:

1. To review previous studies on intersection safety and document methodologies used by them.
2. Identification of major intersection types and explore differences in the crash patterns at those intersection types
3. To develop a simple methodology to identify the crash frequencies based on the intersection's size.
4. To find factors (driver and intersection related) affecting different types of crashes.
5. To estimate conditional probability occurrence of crashes belonging to three major crash types. In this study crashes are categorized into three types based on relative direction of the initial (pre-crash) direction of the involved vehicle at an intersection. Hence, the analyses of crashes include initial movement in same direction (IMSD) crashes, initial movement in opposite direction (IMOD) crashes, and initial movement in perpendicular direction (IMPD) crashes. IMPD crashes include crashes between vehicles two different roadway.

To achieve these research objectives, previous studies dealing with intersection safety are rigorously reviewed. Preliminary analysis, based on simple linear regression is used to explore differences among the three major intersection types and relationship of the crash frequencies with total number of lanes at an intersection. Contingency tables and binary logistic regression models are employed to achieve the final objective listed above.

1.3 Organization of Thesis

In the next chapter various studies performed in the areas of intersection safety and the analysis methods used by them are reviewed. Next, in the Methodology chapter, models that are used in the present research are discussed in detail. In Chapter 4 data preparation steps are listed and a methodology is developed based on three major intersection types and total number of lanes at an intersection. The fifth chapter, binary logistic regression models developed for three crash types using data of at-fault drivers at an intersection. Conclusions are presented in the sixth and final chapter.

2 LITERATURE REVIEW

2.1 Introduction

Prior to the analysis of crash data for the present work, existing research exploring relationship between intersection crashes and independent variables, such as intersection geometry and driver related factors, has been reviewed in this chapter. This chapter is divided into two sections. First, this chapter reviews past studies in the area of intersection safety to identify the significant variables to be included in the data preparation and modeling exercise. Next, this chapter elaborates on the statistical methodologies adopted by various researchers.

2.2 Studies Related To Factors Effecting Crash Type

2.2.1 Studies associating geometry of the intersection with crashes

Geometric and traffic characteristics of intersections generally affect traffic delay and crash risk. These intersection characteristics are often represented by the size of intersection such as the number of lanes, the directions of travel such as one-way or two-way roads, and the connection with crossroads such as the four-legged or T-intersections. It is generally believed that increasing the size of intersections increases capacity and thereby alleviates congestion and improves safety. In fact, the number of lanes is closely associated with capacity and average traffic volume at intersections. Mucsi and Khan (2003) demonstrated that increasing the number of lanes has some marginal effects on increasing capacity and reducing delay at signalized intersections. However, the effect of increasing the number of lanes on crash risk is more complex. For instance, additional number of lanes may potentially reduce the risk of head-on crashes at intersections but may also result in many drivers traveling at higher speeds, leading to more rear-end crashes. Also, the other intersection characteristics may influence driver behavior and cause different crash patterns at different types of intersections.

The number of lanes on an intersection approach is determined primarily by traffic demand and the desired level of service. Intuitively, one might assume that the number of accidents is proportional to the number of lanes (i.e., as the number of lanes increases so would the total number of accidents, since the potential number of conflicts would appear to increase). However, Bauer and Harwood (1996) found that for unsignalized intersections in both rural and urban areas, the number of accidents tended to be higher on facilities with one approach lane and accidents tended to be lower at intersections with two or more approach lanes. The opposite appears to be the case for urban, four-leg, signalized intersections.

Pernia et al.(2002) examined the relationship between the average number of crashes at intersections and the number of lanes. They observed that the average number of crashes increased as the number of lanes on major roads increased. Porter and England (2000) concluded that more red-light running tended to occur at intersections with more lanes in both roads, which could imply that the likelihood of a crash at larger intersections is greater. However, their results are based on the crashes at only six intersections and they did not clearly define the size of intersections in terms of the number of lanes.

In another study, Harwood et al. (2000) developed algorithms to predict the expected safety performance of intersection. The prediction algorithms combined elements of historical accident data, predictions from statistical models, results of before-after studies, and expert judgments made by experienced engineers. As part of the research, an expert panel of safety researchers developed accident modification factors (AMFs) for specific geometric design and traffic control features. AMFs are used in the accident prediction algorithms to represent the effects of safety of the respective features. The panel estimated that installation of a left-turn lane along one major approach reduces intersection-related accidents by 18 to 24 percent, depending upon the type of traffic control and the number of legs, and installation of left-turn lanes along both major approaches to a four-leg intersection reduces intersection-related accidents by 33 to 42 percent, depending upon the type of traffic control.

Gluck et al. (1999) reported accident rate reductions ranging from 18 to 77 percent due to the installation of left-turn lanes, based on the review of work by the New Jersey Department of Transportation.

Poch and Mannering (1996) found that approach grades, the number of approach and opposing lanes, approach left-turn or right-turn volumes, the type of left-turn approach and the type of signal control have different effects on accident causation by maneuver type. For example, the permissive left-turn control tends to increase total and approach-turn accidents while the restrictive left-turn tends to decrease approach-turn accidents.

Not all studies, however, have shown that left-turn lanes reduce accidents. Bauer and Harwood (1996) found that left-turn lanes were associated with higher frequencies of both total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents. However, this result was not advanced by the authors as a basis for policy because the directions of specific effects in predictive models often represent the surrogate effects of other variables, rather than the true effect of the variable of interest.

A potential problem in installing left-turn lanes at intersections is that vehicles in opposing turn lanes on the major road may block drivers' views of approaching traffic. This can lead to collisions between vehicles turning left from the major road and through vehicles on the opposing major-road approach. To reduce the potential for crashes of this type, the left-turn lanes can be offset by moving them laterally so that vehicles in opposing lanes no longer obstruct the opposing driver. Lau and May (1988) reviewed the differences between conventional and offset four-leg intersections and between T and Y three-leg intersections. They found that these differences are statistically significant in modeling of injury accidents at both signalized and unsignalized intersections, but their classification and regression tree (CART) analysis results are difficult to interpret as a specific effect of these factors.

Bauer and Harwood (1996) find that crash rates increase with increasing design speed on four-legged rural intersections. Vogtmd Bared (1998) found the same for posted speeds on three-legged and four legged intersections. Pickering, Hall, and Grimmer (1986) observe that higher operating speeds at three-legged intersections are associated with more right-turn crashes, but with fewer crashes of other types.

2.2.2 Studies associating environmental characteristics with crashes

Sabey (1991) suggested that roadway characteristics, such as geometric design elements, traffic control measures, and traffic demand patterns contribute to about 30% of all traffic accidents, either alone or in combination with human, vehicular or environmental factors. Rumar (1985), citing previous studies, concluded that roadway characteristics alone contribute to only a small proportion of traffic accidents and that accident occurrence was mainly through the errors in road user interaction with other factors, especially those of the road environment.

The road environment conditions that could play a significant role in intersection accidents and they may contain all kinds of non-driver related factors such as lighting conditions, roadway surface conditions, weather conditions, and so on.

Bad weather is recognized as a contributing factor to crashes. Shankar, Mannering, and Barfield (1996) call attention to the interaction of extreme weather and extreme alignment. Vogtand Bared (1998), using a regional, but not particularly local weather variable in Minnesota, find that weather conditions do not have a strong effect on crashes.

Adverse weather conditions contribute to crashes by impairing visibility, reducing stability and decreasing controllability. According to a report on crashes on U.S. highways, over 22% of the total crashes in 2001 were weather-related (Goodwin, 2003).

Bauer and Harwood (1996) find that the absence of lighting contributed significantly to the number of injury crashes at rural three-legged and four-legged intersections. A study by Blower, Campbell, and Green (1993) indicated that truck crashes in Michigan are more frequent at night and in rural settings; the combination of the two is deemed to imply inadequate lighting.

2.2.3 Studies related to driver characteristics

Numerous studies have indicated that older drivers have traffic safety problems and higher crash rates, but the exact reasons for these crash rates are unclear. For example, Stamatiadis et al. (1991) point out that driver older than age 65 and younger than age 25 were more likely than other age groups to be involved in crashes at both signalized and nonsignalized intersections. In another study, McKelvey and Stamatiadis (1989) reported that older driver crash rates were higher than young- or middle-aged drivers, primarily because of angle collisions and crashes at intersections controlled by stop signs rather than by traffic signals. In addition, they revealed that older drivers are penalized more often than other drivers for failing to yield the right-of-way. A subsequent study also confirmed that older drivers are more likely to be involved in angle crashes, rear-end crashes, and head-on crashes while turning left (Villalba, Kirk, & Stamatiadis, 2001). Left-turning maneuvers require a driver to consider two directions of travel and to cross oncoming traffic. This places high demands on the driver's visual recognition and search capabilities. It tests the driver's ability to estimate the speed of vehicles through depth perception, and requires a high cognitive demand to evaluate a situation involving many complex actions. Therefore, increased involvement for the senior citizen in this type of crash may indicate a deficiency in any of these capabilities, and it is not surprising to find these crashes are more common among the older drivers. The occurrence of rear-end crashes is directly affected by the drivers' abilities. This crash type is related to the depth perception and visual recognition abilities of the driver along with the driver's reaction time.

Past research has shown that older drivers experience higher crash involvement rates at night, during inclement weather, and at signalized intersection areas (McKelvey & Stamatiadis, 1989; Stamatiadis & Deacon, 1995). The day versus night factor is examined in this analysis because older drivers have difficulties driving at night due to their diminished visual ability at night. Research has shown that glare recovery had a marginally significant relationship with driving performance for those older than 54. This study accounts for whether some of these confounding effects relate to the impact of passengers.

The driver age and gender were considered as main driver characteristics that might be associated with the rear end characteristics. There is general consensus among the researchers that older drivers tend to process the information slowly than the younger driver. Slower reaction times of older drivers versus younger drivers contribute to a disproportionately heightened degree of risk especially when older drivers face with two or more choices of action (Staplin et al., 1998).

According to Mathews and Moran (1986), young drivers (26 year old or younger) tend to view the chances of a crash happening to themselves as about the same as they saw the chances of older drivers. Moreover, young drivers viewed the chances of a crash as being much higher for other young drivers than for themselves. As results, young drivers clearly underestimated the potential risks associated with certain behavior and situations. They were overconfidence and they overestimated their abilities.

According to Evans (1991), as driver age, various capabilities relevant to driving decline, thus, crash rates increased. The elderly drivers' crashes are more likely to be side impact and multiple vehicle crash. Their crashes are less likely to be a rollover, involve alcohol, or occur at night. Drivers from about 30 to 60 year-old have the lowest involvement rates. As age decreases below 30, crash rate increases rapidly. For age greater than about, crash rate increases somewhat, but much less quickly than as one approaches younger ages. Evans also discusses some risks in perspective increases by a factor of 3, such as (1) traveling 80 km/hr compared with traveling 60 km/hr increase the fatality risk by factor of 3, (2) an unbelted driver in a small car compared with belted driver in large car increase fatality by a factor of 3, (3) overall fatality rate was a factor of 3 higher 30 years ago than it is today, (4) many counties today have fatality rates more than a factor of 3 times the present U.S rate, and (5) driving 300km generates 3 times the fatality risk than driving 100km.

Referring to the analysis of the distributions of legal responsibility, Hakamies-Blomquist (1993) mentioned that older drivers (age of 65 and more) considered legally responsible for causing a collision 74.1 % of their crashes, the comparison group (age of 26-40) in 39.0%. 91.1 % of the crashes for both groups took place in rural area. Older drivers collided most often (54.9%) with a vehicle having a crossing direction; 38.1 % were head on collisions and 7.1 % rear-end collisions. the causation of crashes for elderly drivers were 57.7 % of general or specific inattention and faulty or lacking perception (observation error). For error in handling the vehicle (driving errors), elderly drivers were 24.1 % smaller than for the control group. The risk of causing a fatal crash was found to increase with age.

In addition to the age effect on driver performance, there are a number of time-related changes that are impacting the overall number of older driver crashes. These temporal effects that influence the crash propensity of drivers include the increase in older licensed drivers, both in numbers and age; the characteristics of each cohort, which carries the generational effects of the traffic environment changes in vehicle capability; and the roadway types, including increasing congestion. There are also societal temporal changes that affect older drivers, including the increased reliance on the automobile. Older drivers have different attitudes and behaviors that come with age, including a reduction in the use of alcohol, an increased awareness of safety, and value for life. Many of these factors affect the crash experience of all drivers, but they have their largest impact on this special population (Stamatiadis & Deacon, 1995).

It is well known that male and female differ in their driving behaviors and in driving experience. The result of Dejoy study (1992) indicated that the relationship of optimism to the excess involvement of young males in crashes. They possess an exaggerated sense of their own driving skill and they perceive less risk in a variety of dangerous driving behaviors. According to this study, males and females held similar perception concerning frequencies and apprehension likelihood of risky behavior, but males perceived the behaviors as less serious and less likely to result in crash.

According to Massie and Compbell (1993), women in the group of 25 and over had higher rate of involvement in non-fatal crashes per mile driven than do men. Women higher rate of non-fatal involvement compared with men was entirely by their crash experience in the daytime, not at night.

Alcohol is one of the main factors contributing to traffic accident occurrence. Evans (1991) estimates that about 10% of property damage, 20% of injuries and 47% of fatalities from traffic accidents are attributable to alcohol. The study by Abdel Aty (1999) examines the differences in alcohol-related accident involvement among the different groups considered in this study are: age, gender, race, and residency of the driver of a motor vehicle involved in an accident while under the influence of alcohol, drugs, or alcohol and drugs. Conditional probability results have indicated that larger percentages of young (20–24 years old) and middle age (25–64) drivers are involved in traffic accidents while under the influence of alcohol.

2.3 Statistical Methods Adopted for Analyzing Intersection Crashes

The most common regression method is conventional regression analysis (CRA), either linear or nonlinear, when the response variable is continuous (iid). However, when the outcome (the response variable) is discrete, CRA is not appropriate. Among several reasons, the following two are the most significant:

1. The response variable in CRA must be continuous, and
2. The response variable in CRA can take nonnegative values.

These two primary assumptions are not satisfied when the response variable is categorical.

Regression methods have become an integral component of any data analysis concerned with the relationship between a response variable and one or more explanatory variables. Jovanis and Chang (1986) found a number of problems with the use of linear regression in their study applying Poisson regression as a means to predict accidents. For example, they discovered that as vehicle kilometers traveled increases, so does the variance of the accident frequency. Thus, this analysis violates the homoscedasticity assumption of linear regression.

Researchers have attempted three approaches to relate accidents to geometric characteristics and traffic related explanatory variables: Multiple Linear regression, Poisson regression and Negative Binomial regression. However, recent research shows that multiple linear regression suffers some undesirable statistical properties when applied to accident analysis, some of which have been discussed by Jovanis and Chang (1986). To overcome the problems associated with multiple linear regression models, Jovanis and Chang proposed Poisson regression for modeling accident frequencies. They argued that Poisson regression is a superior alternative to conventional linear regression for applications related to highway safety. In addition, it could be used with generally smaller sample sizes than linear regression.

Miaou et al., (1993) used a Poisson regression model to establish the empirical relationship between truck accidents and highway geometric on a rural interstate in North Carolina. The estimated Poisson model suggested that Average Annual Daily Traffic (AADT) per lane, horizontal curvature, and vertical gradient were significantly correlated with truck accident likelihood. During their work, a limitation of the Poisson model was uncovered. Using the Poisson model necessitates that the mean and variance of the accident frequency variable (the dependent variable) be equal. In most accident data, the variance of the accident frequency exceeds the mean and, in such case, the data would be over dispersed. In a well-summarized review of models predicting accident frequency, Milton and Mannering (1996) state: ‘‘the use of Poisson regression models is inappropriate for making probabilistic statements about the occurrences of vehicle accidents on the road.’’ They showed that the negative binomial regression is a powerful predictive tool and one that should be increasingly applied in future accident frequency studies.

Miaou (1994) studied the relationship between highway geometric and accidents using Negative Binomial regression. In this study, he evaluated the performance of the Poisson regression, zero-inflated Poisson regression, and Negative Binomial regression. Maximum likelihood was used to estimate the coefficients of the models. As an initial step in developing a model and suggested that the Poisson regression model should be used to establish the relationship between highway geometric and accidents. If over dispersion exists and is found to be moderate or high, both the Negative Binomial and zero inflated Poisson regression models can be explored. He suggested that the zeroinflated Poisson regression model appears to be appropriate when the data exhibits a high number of zero frequency observations.

Contingency-table analyses have been used in analyzing categorical or qualitative response variables for their statistical relationship. This type of analysis is usually limited to two variables (two-way table) at a time. For tables with an order greater than two-way, iterative numerical procedures are utilized that are time-consuming since calculations must be carried out to several decimal places to ensure reasonable accuracy of the estimates. Nevertheless, today's computer and software capabilities are able to manipulate sophisticated models, such as log-linear models, logistic regression models to analyze categorical data with more than two variables (Lum, 1989). A log-linear model is a generalized linear model (GLM) for Poisson-distributed data; it specifies how the size of a cell count depends on the levels of the categorical variables for that cell. The nature of this specification relates to the association and interaction structure among the variables. A log-linear model describes the association and interaction patterns among a set of categorical variables (Agresti, 1990). An SAS program procedure, CATMOD, can be used to fit a log-linear model (SAS Institute, 2000).

Kim et al. (1996) developed a logistic model and used it to explain the likelihood of motorists being at fault in collisions with cyclists. Covariates that increase the likelihood of motorist fault include motorist age, cyclist age (squared), cyclist alcohol use, cyclists making turning actions, and rural locations.

Kim et al. (1994) attempted to explain the relationship between types of crashes and injuries sustained in motor vehicle accidents. By using techniques of categorical data analysis and comprehensive data on crashes in Hawaii during 1990, a model was built to relate the type of crash (e.g. rollover, head-on, sideswipe, rear-end, etc.) to a KABCO injury scale. They also developed an 'odds multiplier' that enabled comparison according to crash type of the odds of particular levels of injury relative to noninjury. The effects of seatbelt use on injury level were also examined, and interactions among belt use, crash type, and injury level were considered. They discussed how loglinear analysis, logit modeling, and estimation of 'odds multipliers' may contribute to traffic safety research.

Kim et al. (1995) built a structural model relating driver characteristics and behavior to type of crash and injury severity. They explained that the structural model helps to clarify the role of driver characteristics and behavior in the causal sequence leading to more severe injuries. They estimated the effects of various factors in terms of odds multipliers — that is, how much does each factor increase or decrease the odds of more severe crash types and injuries. Nassar et al. (1997) developed an integrated accident risk model (ARM) for policy decisions using risk factors affecting both accident occurrences on road sections and severity of injury to occupants involved in the accidents. Using negative binomial regression and a sequential binary logit formulation, they developed models that are practical and easy to use. Mercier et al. (1997) used logistic regression to determine whether either age or gender (or both) was a factor influencing severity of injuries suffered in head-on automobile collisions on rural highways.

A binary logistic regression is proper to use when the dependent variable is a dichotomy (an event happened or not) and can be applied to test association between a dependent variable and the related potential factors, to rank the relative importance of independents, and to assess interaction effects. Binary logistic regression is used in this study since the dependent variable Y (accident classification) can only take on two values: $Y = 1$ for through movement crashes and $Y = 0$ for other crashes.

2.4 Summary

The rigorous review of the past studies reveals that the number of lanes, exclusive left-turn lanes, posted speed limit, weather, surface condition, lighting, driver age, gender, and alcohol/drugs usage significantly affect intersection crashes. However, there are some limitations to the studies documented in this chapter. First, many of the studies have considered only the number of lanes in the road where crash occurred, thereby not accounting for the effect of traffic approaching from opposite direction and/or cross roads at the same intersection.

None of the studies have included intersection type (e.g. Four-legged two-way intersection and T-intersections) as an independent parameter in the models. The discussion on joint contribution of explanatory variables (such as geometric characteristics, environmental factors and driver characteristics) on crash type remains relatively unexplored. This present work attempts to overcome the aforementioned limitations.

The methodologies used in the studies documented in this chapter have proved to be an invaluable tool in establishing relationship(s) between crashes and factors related to the crash occurrence. The overall conclusion from the literature review is that categorical data analysis techniques may be used to establish relationships between crashes and relevant independent variables. Therefore, in the following chapter categorical data analysis techniques (e.g., contingency tables and logistic regression models) are described in detail.

3 METHODOLOGY

3.1 Categorical Data Analysis

Categorical data, such as crash variables used in the analysis, consist of frequency counts of observations occurring in the response categories. Significant association between the categorical variables is determined by rejecting the null hypothesis for the χ^2 (Chi-square) test of independence.

3.2 Conditional Probabilities

The relationship between crash type and the variables affecting crash type may be investigated using conditional probabilities. Let *X* and *Y* denote two categorical variables, X having *I* levels and *Y* having *J* levels. The *I*J* possible combinations of outcomes could be displayed in a rectangular table having *I* rows for the categories of *X* and *J* columns for the categories of *Y*. The cells of the table represent the *I*J* possible outcomes. A table of this form, in which the cells contain frequency counts of various outcomes, is called a ''contingency table.''

Let $P_{ij} = P(X = i, Y = j)$ denote the probability that (*X*, *Y*) fall in the cell belonging to row *i* and column *j*. The probabilities $\{P_{ij}\}$ form the joint distribution of *X* and *Y*. These are the cell proportions. They satisfy the constraint $\sum P_{ij} = 1$, Agresti (1996).

The marginal distributions are the row and column of the joint probabilities. These are denoted by $\{P_{i1}\}$ for the row variable and $\{P_{1j}\}$ for the column variable, where the subscript ''1'' denotes the sum over the index it replaces. For instance, for 2*2 contingency tables

$$P_{1+} = P_{11} + P_{12} \text{ and } P_{+1} = P_{11} + P_{21} \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

The cell counts are denoted by $\{n_{ij}\}$, with $n = \sum n_{ij}$ denoting the total sample size. The cell proportions and cell counts are related by $P_{ij} = n_{ij} / n$. The marginal frequencies are the row totals $\{n_{i1}\}$ and the column totals $\{n_1\}$, Agresti (1996).

Contingency-table type of analysis is usually limited to two variables (two-way table) at a time. For tables with an order greater than two-way, iterative numerical procedures are utilized, which are time-consuming since calculations must be carried out to several decimal places to ensure reasonable accuracy of the estimates.

3.2.1 Significant association

To compare the 'strength' of one association with another we need to quantify the strength of each association. The idea here is to find some re-parameterization of χ^2 which maps it into some convenient interval, like (0, 1) where the result is not dependent on the quantity of data that we happen to sample, but rather depends only on the underlying population from which the data were drawn. While there are several different parameters that achieve this, one of the more commonly used is the *contingency coefficient, C*, defined as follows (Kendall and Stuart, 1979):

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Where N is the Sample size

It lies between zero and one, but (as is apparent from the formula) it can never achieve the upper limit. While it can be used to compare the strength of association of two contingency tables with the same *I* and *J*, its upper limit depends on *I* and *J*. Therefore it can not be used to compare tables of different sizes.

3.3 Correlation

The most common measure of "correlation" or "predictability" is Pearson's coefficient of correlation. Pearson's ρ, as it is often symbolized, can have its value anywhere between -1 and 1. The larger the magnitude of ρ, ignoring sign, the stronger the association between the two variables and the more accurately you can predict one variable from knowledge of the other variable. At its extreme, a correlation of 1 or -1 means that the two variables are perfectly correlated, meaning that you can predict the values of one variable from the values of the other variable with perfect accuracy. At the other extreme, $\rho = 0$ implies an absence of correlation indicating no relationship between the two variables. This implies that knowledge of one variable gives you absolutely no information about what the value of the other variable is likely to be. The sign of the correlation implies the "direction" of the association. A positive correlation means that relatively high scores on one variable are paired with relatively high scores on the other variable, and vice versa. The correlation coefficient of a set of observations $\{(x_i,y_i): i=1,..,n\}$ is given by the formula, Agresti (1996),

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad \ldots\ldots\ldots\ldots\ldots(3)$$

3.4 Binary Logistic Regression Model

The goal of logistic regression analysis is the same as that of any model-building technique used in statistics: to find the best fit and the most parsimonious one. What distinguishes a logistic regression model from, say, a linear regression model is the response variable. In a logistic regression model, the response variable is binary or dichotomous. The difference between logistic and linear regression is reflected both in the choice of a parametric model and in the assumptions. Once this difference is accounted for, the methods employed in an analysis using logistic regression follow the same general principles used in linear regression analysis. According to (Agresti, 1996), any regression analysis the key quantity is the mean value of the response variable given the values of the independent variable,

$$E(y/x) = \beta_0 + \beta_1 x \text{ ……………. (4)}$$

Where *Y* denotes the response variable, *x* denotes the independent variable, and the β_0 and β_1 values denote the model parameters. The quantity is called the conditional mean or the expected value of *Y* given the value of *x*. Many distribution functions have been proposed for use in the analysis of a dichotomous response variable (Hosmer and Lemeshow (1989) , Agresti, 1996).

The specific form of the logistic regression model is

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \text{ …………………… (5)}$$

The transformation of the $\pi(x)$ logistic function is known as the logit transformation

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \text{ ………... (6)}$$

The importance of this transformation is that $g(x)$ has many of the desirable properties of a linear regression model. The logit, $g(x)$, is linear in its parameters, may be continuous, and may range from minus infinity to plus infinity, depending on the range of x.

Hosmer and Lemeshow (1989) summarized the main features in a regression analysis, when the response variable is dichotomous, as follows:

1. The conditional mean of the regression equation must be formulated to be bounded between zero and 1.
2. The binomial, not the normal, distribution describes the distribution of the errors and will be the statistical distribution upon which the analysis is based.
3. The principles that guide an analysis using linear regression will also apply for logistic regression.

In linear regression the method used most often for estimating unknown parameters is least squares, in which the parameter values are chosen to minimize the sum of squared deviations of the observed values of Y from the estimated values. Under the assumptions of linear regression, the method of least squares yields estimators with a number of desirable statistical properties. Unfortunately, when the method of least squares is applied to a model with a dichotomous outcome, the estimators no longer have these same properties. The general method of estimation that leads to the least squares function under the linear regression model (when the error is normally distributed) is called maximum likelihood. This method provides the foundation for estimating the parameters of a logistic regression model. A brief review of maximum likelihood estimation method for the logistic regression model is provided in following section.

3.4.1 Maximum likelihood estimation

If Y is coded as 0 or 1 (a binary variable), the expression $\pi(x)$ given in Eq. (5) provides the conditional probability that Y is equal to 1 given x, denoted as $P(Y=1/x)$. It follows that the quantity $1-\pi(x)$ gives the conditional probability that Y is equal to zero given x, $P(Y=0/x)$. Thus, for those pairs (x_i, y_i) where $y_i=1$, the contribution to the likelihood function is $\pi(x_i)$, and for those pairs where $y_i=0$, the contribution to the likelihood function is $1-\pi(x_i)$, where the quantity $\pi(x_i)$, denotes the values of $\pi(x)$ computed at x_i. A convenient way to express the contribution to the likelihood function for the pair (x_i, y_i) is through the term, (Hosmer and Lemeshow, 1989),

$$\zeta(x_i) = \pi(x_i)^{y_i}[1 = \pi(x_i)]^{1-y_i} \ldots\ldots\ldots\ldots (7)$$

Since x_i-values (observations on the independent variable) are assumed to be independent, the product for the terms given in the foregoing equation gives the likelihood function as follows:

$$l(\beta) = \prod_{i=1}^{n} \zeta(x_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

(Hosmer and Lemeshow, 1989) states that it is easier mathematically to work with the log of Eq. (8), which gives the log likelihood expression:

$$L(\beta) = \ln[l(\beta)]$$

$$= \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1-y_i)\ln[1-\pi(x_i)]\} \ldots\ldots\ldots (9)$$

Maximizing the above function with respect to β and setting the resulting expressions equal to zero will produce the following values of β, (Hosmer and Lemeshow, 1989),

$$= \sum_{i=1}^{n} [y_i - \pi(x_i)] = 0$$

$$= \sum_{i=1}^{n} x_i [y_i - \pi(x_i)] = 0 \text{ ……………….. (10)}$$

These expressions are called likelihood equations, (Hosmer and Lemeshow, 1989). An interesting consequence of Eq. (10) is

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \ \pi(x_i)] \text{ ……………….. (11)}$$

That is, the sum of the observed values of y is equal to the sum of the expected (predicted) values. This property is especially useful in assessing the fit of the model (Hosmer and Lemeshow, 1989). After the coefficients are estimated, the significance of the variables in the model is assessed. If y_i denotes the observed value and $\hat{y}_i$ denotes the predicted value for the i th individual under the model, the statistic used in the linear regression is, (Hosmer and Lemeshow, 1989),

$$SSE = \sum_{i=1}^{n} [y_i - \hat{y}_i]^2 \text{ …………….... (12)}$$

The change in the values of *SSE* is due to the regression source of variability, denoted *SSR*, (Hosmer and Lemeshow, 1989):

SSR=Total Sum of Squares (SS) −Sum of Squares of Error term (SSE)

$$= \sum_{i=1}^{n} [y_i - \overline{y}_i]^2 - \sum_{i=1}^{n} [y_i - \hat{y}_i]^2 \text{ ……… . (13)}$$

where y_i is the mean of the response variable , (Hosmer and Lemeshow, 1989). Thus, in linear regression, interest focuses on the size of *R*. A large value suggests that the independent variable is important, whereas a small value suggests that the independent variable is not useful in explaining the variability in the response variable. The principle in logistic regression is the same. That is, observed values of the response variable should be compared with the predicted values obtained from models with and without the variable in question. In logistic regression this comparison is based on the log likelihood function defined in Eq. (9). Defining the saturation model as one that contains as many parameters as there are data points, the current model is the one that contains only the variable under question. The likelihood ratio is as follows, (Hosmer and Lemeshow, 1989):

$$D = -2\log_n \left[\frac{\text{likelihood of the current model}}{\text{likelihood of the saturated model}} \right] \ldots\ldots\ldots\ldots\ldots \text{(14)}$$

Using Eqs. (5) and (7), the following test statistic can be obtained:

$$D = -2\sum_{i=1}^{n} \left[y_i \ln\left(\frac{\pi_i}{y_i} \right) + (1-y_i)\ln\left| \frac{1-\hat{\pi}_i}{1-y_i} \right| \right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(15)}$$

$$where, \hat{\pi}_i = \hat{\pi}\ (x_i)$$

The statistic *D* in Eq. (14), for the purpose of this study, is called the deviance, and it plays an essential role in some approaches to the assessment of goodness of fit. The deviance for logistic regression plays the same role that the residual sum of squares plays in linear regression (i.e. it is identically equal to *SSE*).For the purpose of assessing the significance of an independent variable, the value of *D* should be compared with and without the independent variable in the model. The change in *D* due to inclusion of the independent variable in the model is obtained as follows, (Hosmer and Lemeshow, 1989):

G=D(for the model without the variable) −D(for the model with the variable) ……...(16)

3.5 Model Assessment Procedures

3.5.1 The Hosmer-Lemeshow goodness-of-fit test

Sufficient replication within subpopulations is required to make the Pearson and deviance goodness-of-fit tests valid. When there are one or more continuous predictors in the model, the data are often too sparse to use these statistics. Hosmer and Lemeshow (1989) proposed a statistic that was shown, through simulation, to be distributed as chi-square when there is no replication in any of the subpopulations. This test is only available for binary response models.

First, observations will be sorted in increasing order of their estimated event probability. The event is the response level, identified in the "Response Profiles" table as "Ordered Value 1." The observations are then divided into approximately ten groups according to the following scheme, (Hosmer and Lemeshow, 1989).

$$M = 0.1 \times N + 0.5 \ldots\ldots\ldots\ldots\ldots\ldots(17)$$

Where, N is the total number of subjects and, M is the target number of subjects for each group

Suppose if there are n_1 subjects in the first block and n_2 subjects in the second block. The first block of subjects is placed in the first group. Subjects in the second block are added to the first group if

$$n_1 < M \text{ and } n_1 + [0.5 \times n_2] < M \ldots\ldots(18)$$

Otherwise, they are placed in the second group.

Note that the number of groups, g, may be smaller than 10 if there are fewer than 10 patterns of explanatory variables. There must be at least three groups in order for the Hosmer-Lemeshow statistic to be computed.

The Hosmer-Lemeshow goodness-of-fit statistic is obtained by calculating the Pearson chi-square statistic from the 2×g table of observed and expected frequencies, where g is the number of groups. According to (Hosmer and Lemeshow, 1989) statistic is calculated as

$$\chi^2_{HL} = \sum_{i=1}^{g} \frac{\left(O_i - N_i\bar{\pi}_i\right)^2}{N_i\bar{\pi}_i(1-\bar{\pi}_i)} \quad \ldots\ldots\ldots\ldots\ldots.... (19)$$

Where N_i is the total frequency of subjects in the i th group, O_i is the total frequency of event outcomes in the i th group, and $\bar{\pi}_i$ is the average estimated probability of an event outcome for the i th group. The Hosmer-Lemeshow statistic is then compared to a chi-square distribution with (g-n) degrees of freedom, where the value of n can be specified in the LACKFIT option in the MODEL statement in SAS (SAS Institute, 2001). The default value is n=2. Larger value of χ^2_{HL} (and small p-values) indicates a lack of fit of the model.

3.5.2 Receiver Operating Characteristic (ROC) Curve

Since the primary purpose of the logistic regression model is binary classification, another way to assess the model would be based on the classification accuracy. Two indices are used to evaluate the accuracy of a test that predicts dichotomous outcomes (e.g., logistic regression) – sensitivity and specificity (Le, 1998). Sensitivity is the proportion of true positives or the proportion of cases correctly identified by test as meeting a certain condition. Specificity - the proportion of true negatives or the proportion of cases correctly identified by the test as not meeting a certain condition

ROC curve is a graphical representation of the trade off between the false negative and false positive rates for every possible cut off. By convention, the plot shows the false positive rate (1-specificity) on the X axis and the true positive rate (sensitivity or 1 - the false negative rate) on the Y axis. The accuracy of a test (i.e. the ability of the test to correctly classify cases with a certain condition and cases without the condition) is measured by the area under the ROC curve. An area of 1 represents a perfect test. Statistically, more area under the curve means that it is identifying more true positives while minimizing the number/percent of false positives.

3.6 Summary

This chapter describes data analysis techniques, such as, Contingency Tables and Binary Logistic regression models, used in this research. As we shall observe in the following chapters, variables affecting three categories of crashes (separated based on the direction of initial movement of the involved vehicles) are determined using Contingency Table analysis. Significant relationship between response variable and explanatory variable is determined by lower p-value (Variables with p-values less than 0.05) and higher contingency coefficients. Significant variables are also tested for correlation using Pearson correlation coefficient.

Binary logistic regression models are developed to estimate the relative likelihood of crash occurrence (of specific types) at a signalized intersection. Goodness-of-fit for the models is determined using Hosmer-Lemeshow goodness-of-fit test, while prediction accuracy is assessed based on ROC curve (sensitivity and specificity analysis).

Next chapter, Data preparation and preliminary analysis, explores differences among intersection types. The description of the data used in this study is also provided in the next chapter.

4 DATA PREPARATION AND ESTIMATION OF CRASH FREQUENCIES BASED ON THE INTERSECTION SIZE

4.1 Introduction to Preliminary Analysis

One of the limitations in the existing literature described at the end of chapter 2 was that most of the studies only account for the number of lanes on the major road of the intersection. To overcome this limitation and enhance the literature we proposed to account for number of lanes at both major and minor roads as well as intersection type. Relevant analysis to account for these factors is provided in this chapter.

The size is represented by the sum of the total number of lanes on all approaches including the through, left and right turning lanes. Using this simple and easy to identify measure, the size of the intersection, many geometric and traffic features of the intersection may be captured, since the number of lanes (size) could be a potential surrogate to the traffic volume, crossing width, signal cycle length, phasing, etc. Therefore, this chapter investigates the effects of the number of lanes (size of intersection) on the expected number of crashes per intersection per year (crash frequency) by type and severity. This chapter also develops a simplistic method to assist traffic engineers in identifying the expected crash patterns at an intersection, which would help them to conduct a quick and efficient safety evaluation of signalized intersections.

4.2 Data for Preliminary Analysis

Crash data for this research were collected from six counties, namely, Brevard, City of Orland, Dade, Hillsborough, Orange, and Seminole. Each county provided a database of crash reports for intersection related crashes that occurred in their respective counties during the three year period of analysis, from 1999 through 2001. There were a total of 26,603 crashes at 1,335 intersections in the six counties over the three year period. Crashes from each of these are put together in a “master database”. Further details on data collection effort may be found in Nawathe (2005).

Intersection characteristics such as number of through, left-turn, and right-turn lanes for each approach, presence of median on each approach, speed limit on major road, traffic volume (AADT), crash characteristics such as collision type (Rear-end, Angle, Left turn, Right Turn, Sideswipe crash, and Head-on), level of injury severity and other driving environment conditions such as surface condition, lighting conditions, weather conditions, time of the day, and date on which the crash occurred were also available for these crashes. Crashes were categorized into the following eight types: rear-end, side-swipe, head-on, angle, left-turn, right-turn, pedestrian-bicycle and other crashes. Injury severity of crashes is categorized into fatal, incapacitating, non-incapacitating evident, possible, and no injuries. Detailed CAD drawings of the intersections were also obtained from the respective counties/city. From the drawings of intersections, the detailed road geometric features (such as the number of through lanes, exclusive left-turn lanes and channelized right-turn lanes, the presence of medians, and the speed limits) were identified by Nawathe (2005). Note that the information about the drivers involved was not available in this database. The information about the involved drivers would be obtained from the DHSMV database. However, in the following section we are not analyzing any of the drivers characteristics, therefore, the details on DHSMV database would be provided later in the chapter (Section 4.6).

The classification of intersections is typically based on the number of lanes on major and minor roads. For example, if the numbers of through lanes on major and minor roadways of one intersection are 4 and 2, respectively, the intersection is classified as "4 × 2". However, this way of classification cannot clearly distinguish if one or both roadways are one or two-way roads or the exact configuration of the intersection based on the number of exclusive left and right turn lanes. It is an inconsistent way to identify the size and configuration of the intersection. For example, a 4x2 intersection with one exclusive left turn lane at all approaches would be in the same category as a 4x2 intersection without any left turn lanes, even though the first has a total of 10 lanes and the second has only 6 lanes. The size and the expected crash patterns would be different for two categories.

Therefore, intersections were classified into the following three types – 1) four-legged two-way intersections; 2) four-legged one-way intersections; and 3) T-intersections (i.e. the three-legged intersections). The four-legged two-way intersections include only intersections where both the major and minor roads are two-way. The four-legged one-way intersections are those intersections where either major, minor or both roadways are one-way. The T-Intersections are those intersections with two-way major and minor roads. Among 1,335 intersections, there are 1,001 four-legged two-way intersections, 132 four-legged one-way intersections, and 202 T-intersections.

Table 4.1 Description of crash and intersection data

Type of intersection	Number of intersections	Number of crashes	Average number of crashes per intersection	Range of total number of lanes at intersections
Four-legged two-way intersections	1,001	21,406	21.4	4~21
Four-legged one-way intersections	132	2,495	18.9	3~12
T-intersections	202	2,702	13.4	4~16

In the determination of size of intersections, the total number of lanes was calculated as the sum of the number of lanes for through, exclusive left-turn, and channelized right-turn lanes. The larger the total number of lanes, the larger the size of intersections. The range of the total number of lanes for each intersection type is also presented in Table 4.1.

4.3 New Approach for Determining the Expected Crashes at Intersections

4.3.1 Expected crash frequency and type

Using the crash frequency data for each intersection the frequency/intersection/year may be defined as total number of crashes divided by total number of intersections and total number of years of the data. Simple regression equations may be estimated for different crash types for three major intersection types. The various regression models that are used to fit the data are linear model, logarithmic model, exponential model, polynomial model and power model. The regression model having higher R-square values is chosen to represent corresponding crash type at an intersection. In the regression equations presented in the next section, total number of lanes at an intersection (size of an intersection) is represented by x.

It is found that exponential models are found to be good fit for four-legged two-way intersections and polynomial models are good for four-legged one-way intersections and T-intersections.

4.3.2 Four-legged two-way intersections

An equation was fitted for the total average crashes and all crash types.

$$Y_{Total} = 5.2289.e^{(0.107x)} \quad (1)$$

$$Y_{\mathrm{Re}\,arend} = 1.4652.e^{(0.1533x)} \quad (2)$$

$$Y_{Angle} = 1.2973.e^{(0.1533x)} \quad (3)$$

$$Y_{Leftturn} = 0.839.e^{(0.0906x)} \quad (4)$$

$$Y_{Sideswipe} = 0.4772.e^{(0.0936x)} \quad (5)$$

The fitted equations for the expected total number of crashes, as a function of the total number of lanes (x) are presented in Eq. 1 through 5 and R-square values are 0.8931, 0.948, 0.7087, 0.789, and 0.6156, respectively.

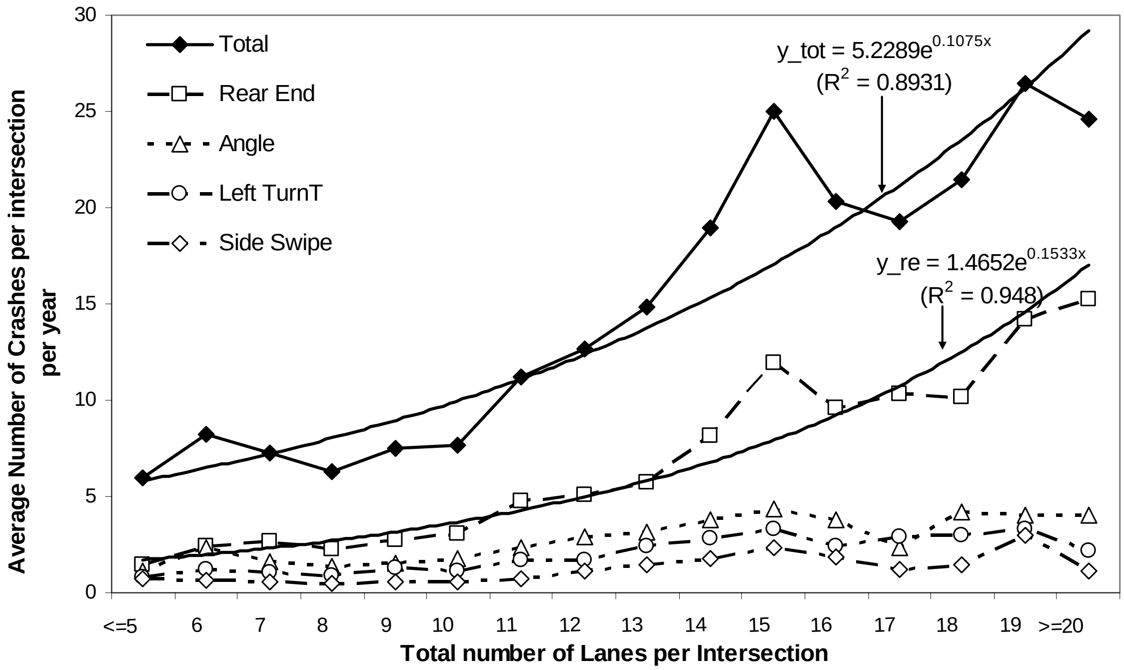


Figure 4.1 Relationship between average number of crashes and total number of lanes by crash types (four-legged two-way intersections)

The R-square values indicate a very good fit of the models particularly for the four-leg two-way intersections. In comparing the types of crashes that occurred in each intersection type, rear-end crash was the dominant type of crashes followed by angle, left-turn, side-swipe, right-turn, pedestrian/bicycle, and head-on crashes at the four-legged two-way intersections (Figure 4.1).

The exponential increase of crash frequency with the total number of lanes at the four-legged two-way intersections implies higher crash risk with the increased size of intersections at this intersection type compared to the other intersection types. This may be because more directions of travel and vehicle movement at the four-legged two-way intersections increased the number of conflict points and consequently, the likelihood of crash occurrence.

4.3.2.1 Four-legged one-way intersections

At the four-legged one-way intersections, the angle crash was the dominant type of crashes followed by rear-end, side-swipe, left-turn, pedestrian/bicycle, right-turn, and head-on crashes (Figure 4.2). The fitted polynomial equations for other crash types at four-legged one-way intersection are as follows.

$$Y_{Total} = 0.1407x^2 - 0.1973x + 7.2178 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

$$Y_{Angle} = 0.0114x^2 + 0.3275x + 1.6237 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

$$Y_{\mathrm{Re}\,arend} = 0.0596\mathrm{x}^2 - 0.152\mathrm{x} + 1.5178 \ldots\ldots\ldots\ldots\ldots\ldots . (8)$$

$$Y_{Leftturn} = 0.0256\mathrm{x}^2 - 0.1208\mathrm{x} + 0.814 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

$$Y_{Sideswipe} = 0.0295\mathrm{x}^2 - 0.1706\mathrm{x} + 1.368 \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

The R-square values for the above models are 0.4745, 0.4135, 0.5954, 0.3433, and 0.2433, respectively. Though R-square values do not represent an excellent fit of the model, it may still be observed that the dominant type of crashes at the four-legged one-way intersections are the angle crashes unlike the other types of intersections where the dominant type of crashes is the rear-end crashes.

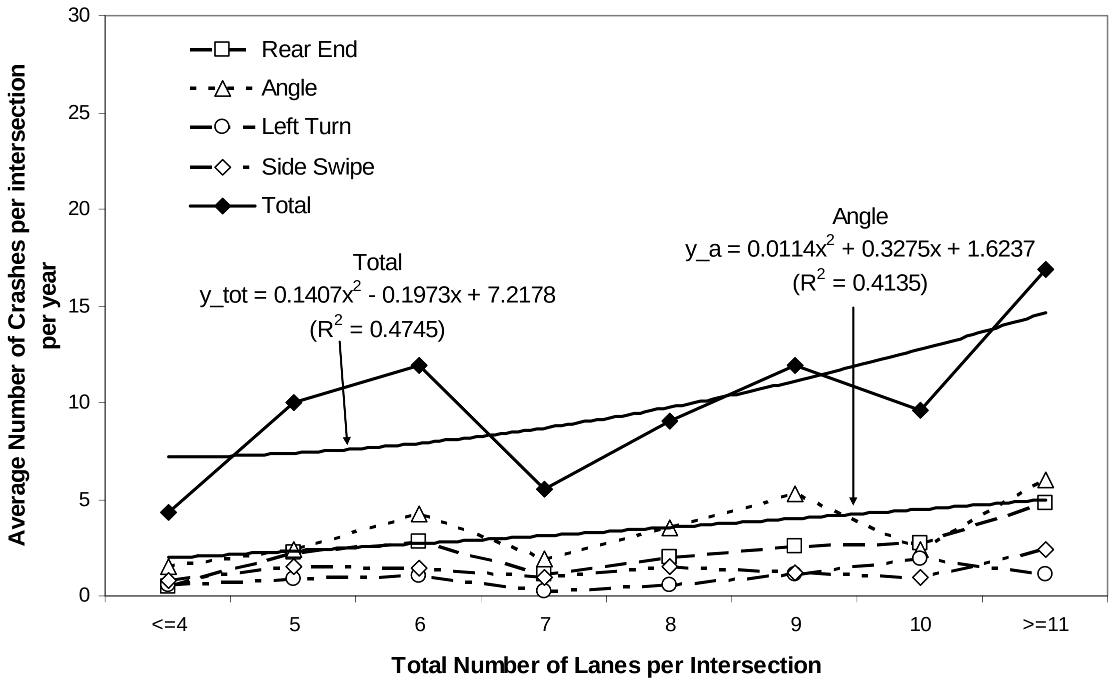


Figure 4.2 Relationship between average number of crashes and total number of lanes by crash types (four-legged one-way intersections)

4.3.2.2 T-intersections

At the T-intersections, the rear-end crashes were dominant similar to the four-legged two-way intersections. The fitted equations for all crash types at four-legged one-way intersection are as follows.

$$Y_{Total} = 0.1374x^2 - 0.8652x + 6.7224 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

$$Y_{\mathrm{Re}arend} = 0.0319x^2 + 0.0056x + 2.0266 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

$$Y_{Angle} = -0.0027x^2 + 0.2096x + 0.0335 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

$$Y_{Leftturn} = -0.0071x^2 + 0.1858x + 0.1089 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

$$Y_{Sideswipe} = 0.0089x^2 - 0.0177x + 0.3282 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

The corresponding R-square values for the above models are 0.5151, 0.4997, 0.5907, 0.7249, and 0.5596, respectively.

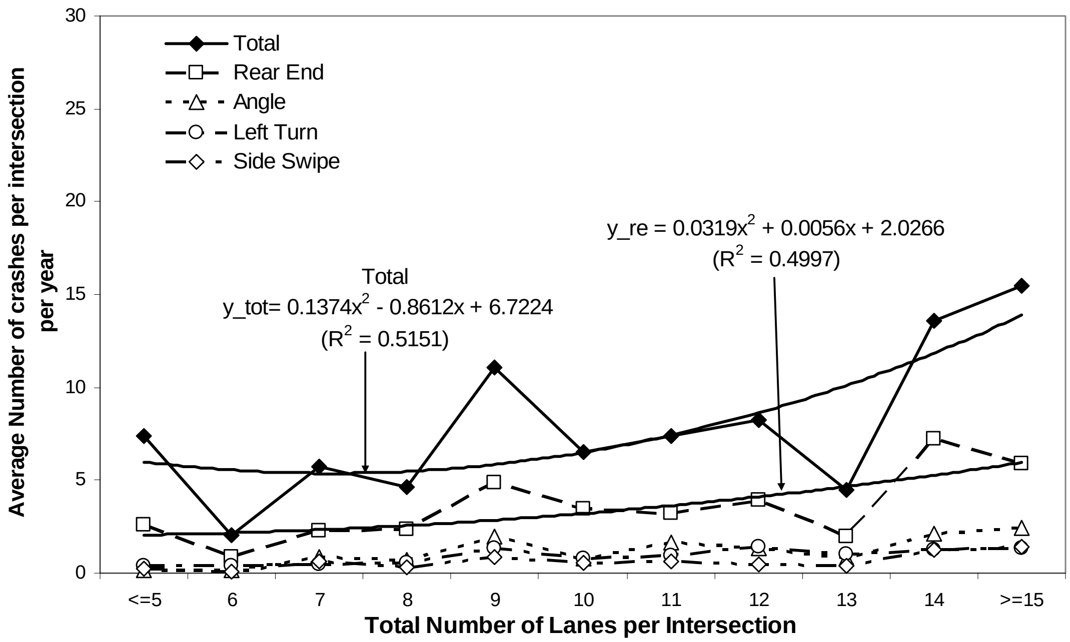


Figure 4.3 Relationship between average number of crashes and total number of lanes by crash types (T-intersections)

4.3.2.3 Comparison across three intersection types

Comparing across intersection types, crash frequency varies exponentially with the total number of lanes at the four-legged two-way intersections whereas the frequency varies in the function of a second-order polynomial at the four-legged one-way and T-intersections. The exponential increase of crash frequency with the total number of lanes at the four-legged two-way intersections implies higher crash risk with the increased size of intersections at this intersection type compared to the other intersection types.

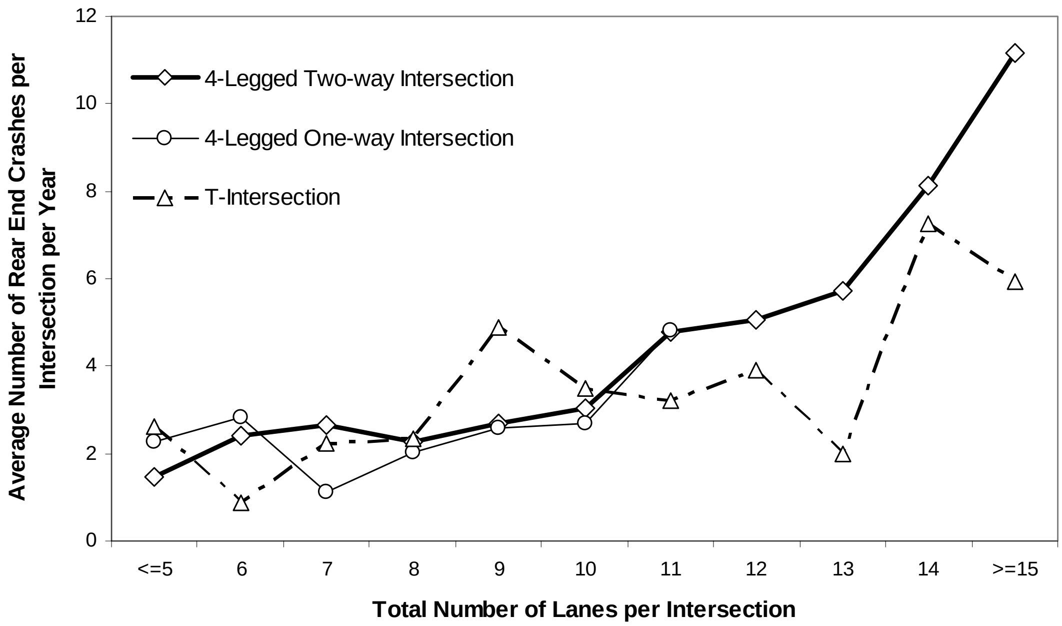


Figure 4.4 Comparison of average number of rear-end crashes among the intersection types

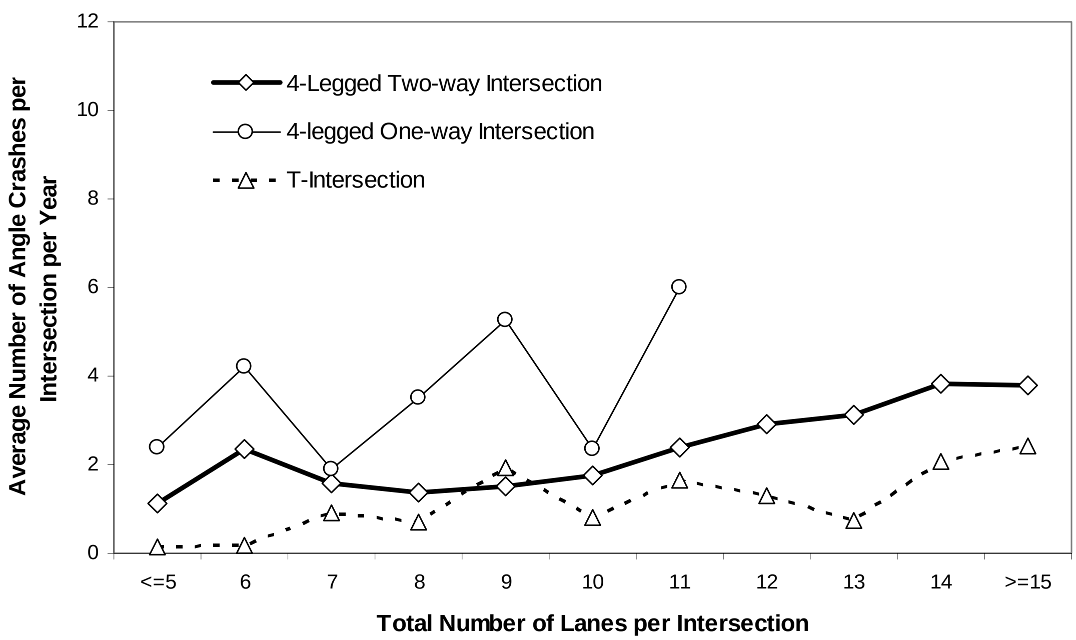


Figure 4.5 Comparison of average number of angle crashes among the intersection types

To investigate the difference in predominant crash types, the information on first contributing causes for the crashes was utilized. A majority of the first contributing cause for the crashes at all intersection types was “careless driving”. However, it was found that the percentage of “disregarded traffic signal” was relatively higher (15.5%) at the four-legged one-way intersections than the other intersection types (8.1% at the four-legged two-way intersections and 6.94% at the T-intersections). Also, there is a possibility that drivers make improper turns to one-way street by misjudging the direction of travel. Therefore, we can speculate that higher rate of driver's disobeying traffic signals causes higher chances of crashes between vehicles traveling in different directions (i.e. angle crashes).

4.3.3 Injury Severity

It was also checked if the injury severity increases across different severity levels with increase in the total number of lanes at an intersection. The Figure 4.6 shows that the crash rate increases exponentially as the number of lanes per intersection increases at four legged 2-way intersections. Though there is clear increasing trend for four legged 2-way intersections, four legged one-way intersection and T-Intersection’s injury severity is not clearly increasing with the total number of lanes.

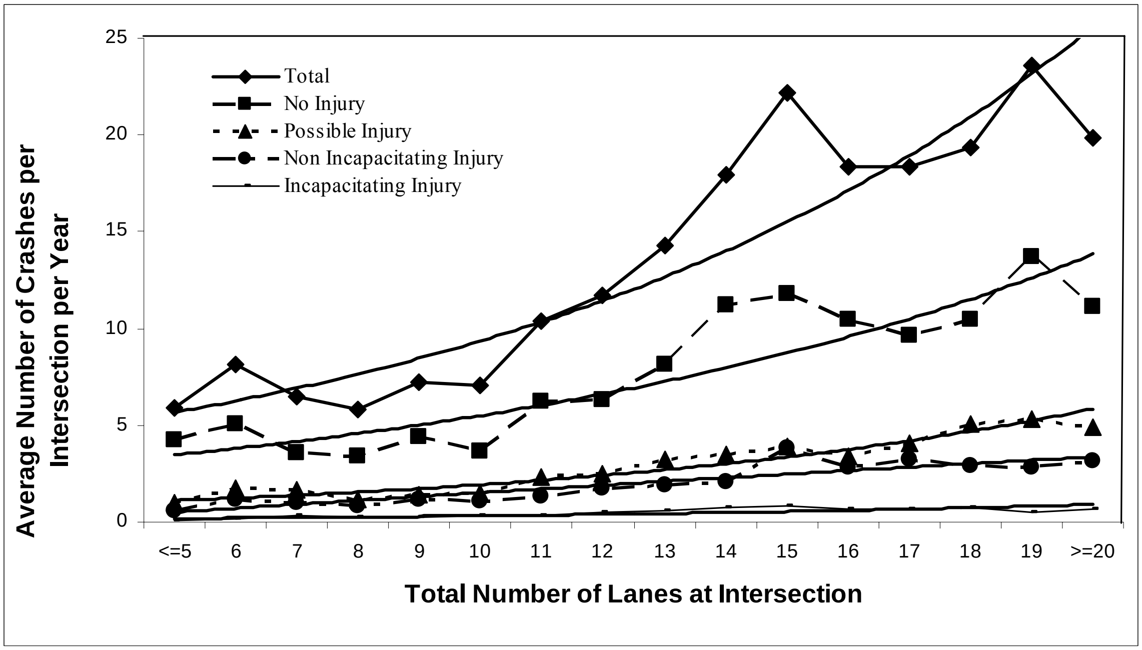


Figure 4.6 Relationship between average number of injury crashes and total number of lanes (four-legged 2-way intersections)

An equation was fitted for the average number of injury crashes with total number of lanes at the intersection. The equations for four legged 2-way intersections are as follows.

$$Y_{Total} = 5.1161e^{0.1007x} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(16)}$$

$$Y_{No-Injury} = 3.1635e^{0.0922x} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(17)}$$

$$Y_{Possible-Injury} = 0.9939e^{0.1107x} \ldots\ldots\ldots\ldots\ldots\ldots \text{(18)}$$

$$Y_{Non-Incapacitating-Injury} = 0.6778e^{0.1095x} \ldots\ldots\ldots \text{(19)}$$

$$Y_{Incapacitating-Injury} = 0.1722e^{0.1037x} \ldots\ldots\ldots\ldots \text{(20)}$$

The corresponding R-square values for the above models are 0.8643, 0.8022, 0.8989, 0.8556, and 0.7433, respectively. In comparing injury severity of crashes by intersection types, the percentage of high injury crashes (i.e. fatal, incapacitating and non-incapacitating evident injuries) in total crashes was higher at the four-legged two-way intersections than the four-legged one-way intersections as shown in Figure 4.6. This is mainly due to more conflict points and the increased chances of severe impact of collisions due to high speeds at large intersections.

On the other hand, the proportion of high injury crashes was higher at the T-intersections than the four-legged one-way intersections in spite of the less conflict points as shown in Figure 4.7.

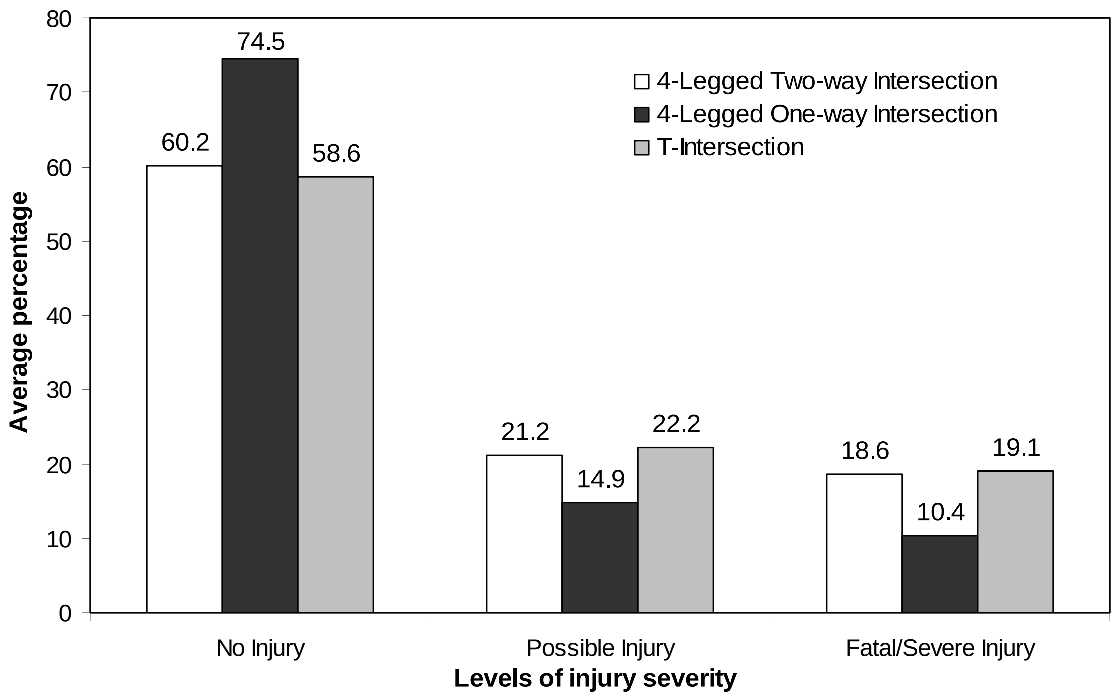


Figure 4.7 Comparison of injury severity among the intersection types

At the T-intersections, priority (e.g. higher speed limit, longer green-time) is usually given to major roads with higher traffic volume. To verify this characteristic, the average difference in speed limits between major and minor roads per intersection was calculated at the intersections in all counties (except Dade and Brevard counties where the speed limit data were not available for the minor roadways). It was found that the average speed limit difference was relatively higher at the T-Intersections (12.55 mph or 20.2 km/hour) than the four-legged two-way intersections (10.57 mph/17.0 km/hour) and the four-legged one-way intersections (6.7 mph/10.8 km/hour). Thus, higher severity at the T-intersections may be because i) large speed difference between the major and minor roads might cause more severe impact of crashes on human bodies when vehicles from different roads collide; and/or ii) the duration of red-light phase on minor roads is generally longer than the major roads and drivers on minor roads tend to be more impatient and rush into major roads. Such sudden violation movement is likely to reduce the amount of time to accelerate to meet the large increase in speed limit on the crossroads. For the same reason, smaller speed limit difference between major and minor roads at the four-legged one-way intersections might have resulted in relatively lower injury crashes compared to the other intersection types.

4.4 Data Preparation for Crash Type Analysis

As mentioned earlier the information related to driver and vehicle movements were not available in the database assembled by Nawathe (2005). We therefore referred to the DHSMV database to obtain that information. The number of crashes for which DHSMV information was available was significantly less than the crashes in the database maintained by the counties. The reason for the same is that the DHSMV database only maintains crashes that are reported on long-form and hence involve either a hit-and-run case or at least a possible injury. Abdel-Aty et al (2005) investigated the differences in the important crash-related factors between models based solely on crashes reported on long forms and models based on crashes reported on both long- and short-forms (referred as models based on restricted and completed datasets, respectively). They found that the estimated effects of important factors are fairly consistent between the models based on restricted and complete datasets. This result indicates that models based on complete and restricted datasets for these types of crashes are roughly equivalent. Hence, crashes from DHSMV database can be used for the analysis.

The DHSMV constitutes a relational database that includes seven files. These files are Events file, Vehicle File, Driver File, Pedestrian File, Violation File, Passenger File and Department of Transportation (DOT) Site Location file. Each file deals with a specific feature of a traffic accident. These files can be linked using a unique crash identification number (Crash ID), which is common for a crash in all the seven files.

The "master database" hence was expanded to "combined database" to include driver and vehicle characteristics of a crash for the purpose of analyzing the crash types. Out of the seven files in DHSMV database, only 4 files were used for the present study. They are 1) Events file (containing the characteristics and environment of the accident), 2) Drivers file (contains drivers characteristics), 3) Vehicle file (contains information about the vehicles' characteristics and vehicles actions in traffic crashes), and 4) Violation file (contains information about who is at-fault).

In order to retain intersection characteristics obtained from six counties "master database" are linked to four files in DHSMV for the years 1999-2001 by using the "Crash ID". Therefore, crashes in master database are linked to four files in DHSMV. A total of 8761 of master database crashes had all the information such as driver, vehicle and driving environment in DHSMV database. The information about vehicle involvement in these crashes in provided in Table 4.2. There were 18554 drivers involved in these 8761 crashes and in 8540 of the crashes one unique driver was identified to be at-fault by the responding police officer on the scene. In remaining crashes police officer could not identify and/or could not uniquely identify at-fault drivers in all the crashes. These crashes occurred at the same 1047 intersections of the six counties that were reported earlier.

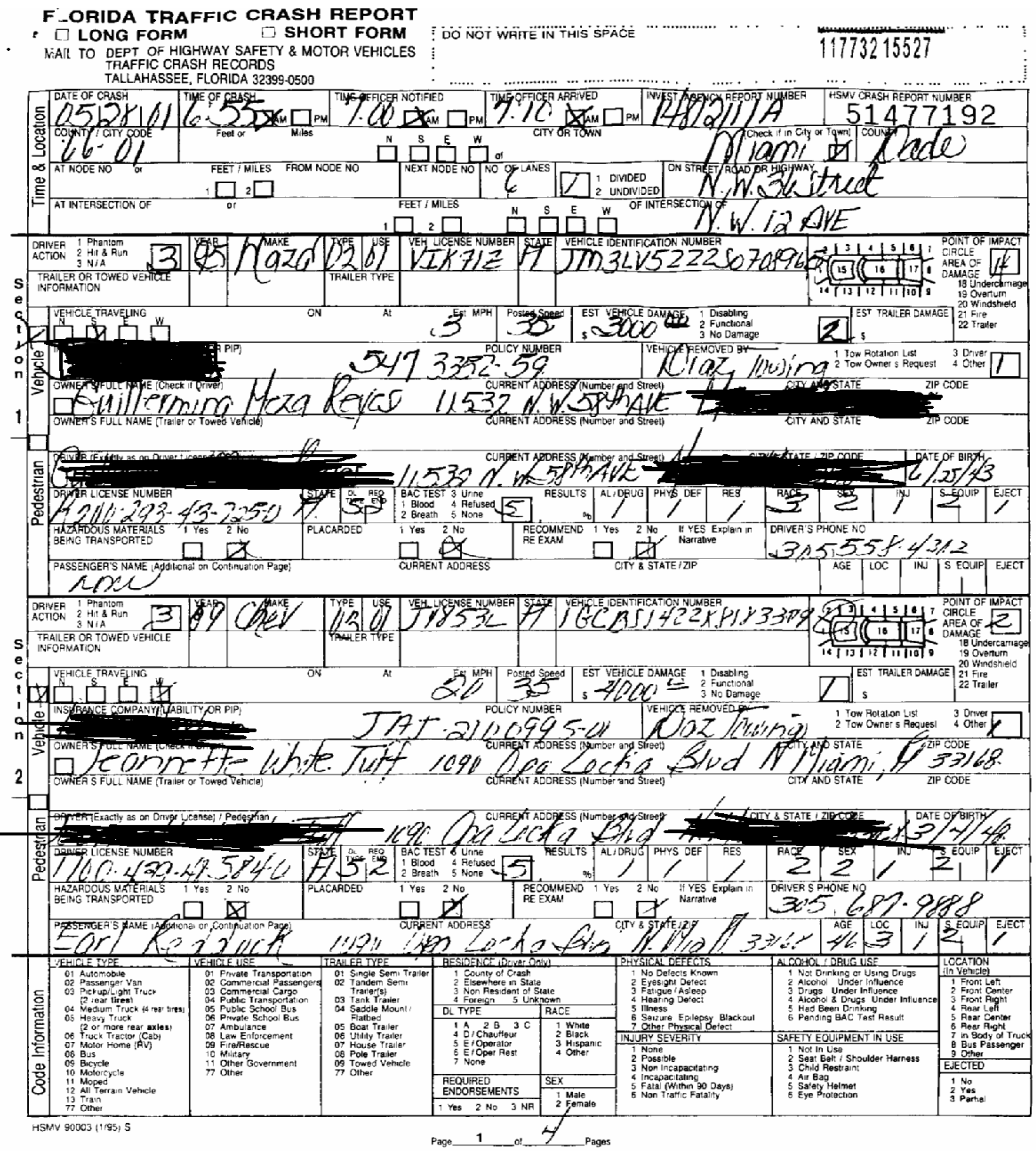
Table 4.2 Crash data composition

Crash Classification by No. of. Vehicles Involved	No. of. Crashes
Single Vehicle Crashes	161
Two Vehicle Crashes	7,534
Three Vehicle Crashes	867
Four Vehicle Crashes	175
More than Four Vehicle Crashes	23
Grand Total	**8,761**

4.4.1 Data rearrangement for vehicle movement type analysis

It is assumed that analyzing only at-fault drivers from driver's dataset will increase our understanding of drivers causing crashes and crash types. Hence, this section details the process of extracting relevant information on the at-fault drivers. Violation file of the DHSMV records provides this information. The variable, "Section Number" in the violation file indicates which driver among the drivers involved in a crash was identified to be at-fault.

In most of the crashes (8540 of 8761 crashes as mentioned in the previous section) a single driver is at-fault. But, in some crashes it is possible to have two or more drivers at fault. Similarly, at other time it is difficult to determine fault and none of the drivers are designated at-fault. Since the police officer is not always able to identify the at-fault driver at the crash location, it is always worth to check the contributing causes of other drivers involved in crash to check if other driver(s) is/are at-fault. For example, following Figures 4.8 through 4.11 is a crash report police could not identify which driver is at-fault but using the contributing cause of the drivers we can conclude that both the drivers were at-fault. In this crash both the drivers had a contributing cause of “Failed to Yield Right of Way”. It should be noted that accident is an angle crash.



FLORIDA TRAFFIC CRASH REPORT
LONG FORM SHORT FORM
MAIL TO DEPT OF HIGHWAY SAFETY & MOTOR VEHICLES TRAFFIC CRASH RECORDS TALLAHASSEE, FLORIDA 32399-0500
DO NOT WRITE IN THIS SPACE
11773215527

DATE OF CRASH	05 28 01
TIME OF CRASH	6:55 AM
TIME OFFICER NOTIFIED	7:00 AM
TIME OFFICER ARRIVED	7:10 AM
INVEST AGENCY REPORT NUMBER	14802111A
HSMV CRASH REPORT NUMBER	51477192
COUNTY / CITY CODE	26-01
CITY OR TOWN	Miami
COUNTY	Dade
NO OF LANES	6
ON STREET, ROAD OR HIGHWAY	N.W. 36 Street
OF INTERSECTION	N.W. 12 AVE

Section 1 – Vehicle / Pedestrian

DRIVER ACTION	3
YEAR	95
MAKE	Mazda
TYPE / USE	02 01
VEH LICENSE NUMBER	VIK71Z
STATE	FL
VEHICLE IDENTIFICATION NUMBER	JM3LV5222S0708968
POINT OF IMPACT	14
EST MPH	3
POSTED SPEED	35
EST VEHICLE DAMAGE	$3000
INSURANCE COMPANY	[redacted]
POLICY NUMBER	547 3352-59
VEHICLE REMOVED BY	Diaz Towing
OWNER'S FULL NAME	Guillermina Meza Reyes
CURRENT ADDRESS	11532 N.W. 58th AVE [redacted]
DRIVER	[redacted]
CURRENT ADDRESS	11532 N.W. 58th AVE [redacted]
DATE OF BIRTH	6/25/43
DRIVER LICENSE NUMBER	R200-293-43-725-0
STATE	FL
BAC TEST	5
RACE / SEX	3 2
DRIVER'S PHONE NO	305 558-4312
PASSENGER'S NAME	none

Section 2 – Vehicle / Pedestrian

DRIVER ACTION	3
YEAR	89
MAKE	Chev
TYPE / USE	12 01
VEH LICENSE NUMBER	JYY53L
STATE	FL
VEHICLE IDENTIFICATION NUMBER	1GCBS14Z2KP1Y3374
POINT OF IMPACT	2
EST MPH	20
POSTED SPEED	35
EST VEHICLE DAMAGE	$4000
INSURANCE COMPANY	[redacted]
POLICY NUMBER	JAT-2110995-01
VEHICLE REMOVED BY	Diaz Towing
OWNER'S FULL NAME	Jeannette White Tuff
CURRENT ADDRESS	1090 Opa Locka Blvd N Miami, FL 33168
DRIVER	[redacted]
CURRENT ADDRESS	1090 Opa Locka Blvd [redacted]
DATE OF BIRTH	[redacted] 3/4/46
DRIVER LICENSE NUMBER	T100-450-47-584-0
STATE	FL
BAC TEST	5
RACE / SEX	2 2
DRIVER'S PHONE NO	305 687-9888
PASSENGER'S NAME	Earl Kendrick
PASSENGER'S CURRENT ADDRESS	1090 Opa Locka Blvd N Miami FL 33168
AGE / LOC	46 3

Code Information

HSMV 90003 (1/95) S
Page 1 of 4 Pages

Figure 4.8 Crash report page 1 (Some information is intentionally blacked out to not reveal the driver/pedestrian details)

Section 3 — Vehicle / Pedestrian

DRIVER ACTION 1 Phantom 2 Hit & Run 3 N/A	YEAR	MAKE	TYPE	USE	VEH LICENSE NUMBER	STATE	VEHICLE IDENTIFICATION NUMBER	POINT OF IMPACT CIRCLE - AREA OF DAMAGE 16 Undercarriage 19 Overturn 20 Windshield 21 Fire 22 Trailer
TRAILER OR TOWED VEHICLE INFORMATION			TRAILER TYPE					

VEHICLE TRAVELING N S E W — ON — AT — Est MPH — Posted Speed — EST VEHICLE DAMAGE $ — 1 Disabling 2 Functional 3 No Damage — EST TRAILER DAMAGE $

INSURANCE COMPANY (LIABILITY OR PIP) — POLICY NUMBER — VEHICLE REMOVED BY — 1 Tow Rotation List 2 Tow Owner's Request 3 Driver 4 Other

OWNER'S FULL NAME (Check if Driver) — CURRENT ADDRESS (Number and Street) — CITY AND STATE — ZIP CODE

OWNER'S FULL NAME (Trailer or Towed Vehicle) — CURRENT ADDRESS (Number and Street) — CITY AND STATE — ZIP CODE

DRIVER (Exactly as on Driver License) / Pedestrian — CURRENT ADDRESS (Number and Street) — CITY & STATE / ZIP CODE — DATE OF BIRTH

DRIVER LICENSE NUMBER — STATE — DL TYPE — REQ END — BAC TEST 3 Urine 1 Blood 4 Refused 2 Breath 5 None — RESULTS % — AL/DRUG — PHYS DEF — RES — RACE — SEX — INJ — S EQUIP — EJECT

HAZARDOUS MATERIALS BEING TRANSPORTED 1 Yes 2 No — PLACARDED 1 Yes 2 No — RECOMMEND RE EXAM 1 Yes 2 No — If YES, Explain in Narrative — DRIVER'S PHONE NO ( )

PASSENGER'S NAME (Additional on Continuation Page) — CURRENT ADDRESS — CITY & STATE / ZIP — AGE — LOC — INJ — S EQUIP — EJECT

#	PROPERTY DAMAGED OTHER THAN VEHICLES	EST AMOUNT	OWNER'S NAME	ADDRESS	CITY	STATE	ZIP
1		$					
2		$					

CONTRIBUTING CAUSES DRIVER / PED (1: 13, 2: 03, 3: /)
01 No Improper Driving / Action
02 Careless Driving
03 Failed to Yield Right of Way
04 Improper Backing
05 Improper Lane Change
06 Improper Turn
07 Alcohol Under Influence
08 Drugs Under Influence
09 Alcohol & Drugs Under Influence
10 Followed Too Closely
11 Disregarded Traffic Signal
12 Exceeded Safe Speed Limit
13 Disregarded Stop Sign
14 Failed to Maintain Equip. / Vehicle
15 Improper Passing
16 Drove Left of Center
17 Exceeded Stated Speed Limit
18 Obstructing Traffic
19 Improper Load
20 Disregarded Other Traffic Control
21 Driving Wrong Side / Way
22 Fleeing Police
23 Vehicle Modified
77 All Other (Explain)

VEHICLE DEFECT (1: 01, 2: 01, 3: /)
01 No Defects
02 Def. Brakes
03 Worn / Smooth Tires
04 Defective / Improper Lights
05 Puncture / Blowout
06 Steering Mech
07 Windshield Wipers
08 Equipment / Vehicle Defect
77 All Other (Explain in Narrative)

VEHICLE MOVEMENT (1: 01, 2: 01, 3: /)
01 Straight Ahead
02 Slowing / Stopped / Stalled
03 Making Left Turn
04 Backing
05 Making Right Turn
06 Changing Lanes
07 Entering/Leaving Parking Space
08 Properly Parked
09 Improperly Parked
10 Making U Turn
11 Passing
12 Driverless or Runaway Veh
77 All Other (Explain in Narrative)

VEHICLE SPECIAL FUNCTIONS (1: /, 2: /, 3: /)
1 None
2 Farm
3 Police Pursuit
4 Recreational
5 Emergency Operation
6 Construction / Maintenance

LOCATION ON ROADWAY (1: /, 2: /, 3: /)
1 On Road
2 Not On Road
3 Shoulder
4 Median
5 Turn Lane / Safety Zone

PEDESTRIAN ACTION (1: /, 2: /, 3: /)
01 Crossing Not at Intersection
02 Crossing at Mid block Crosswalk
03 Crossing at Intersection
04 Walking Along Road With Traffic
05 Walking Along Road Against Traffic
06 Working on Vehicle in Road
07 Other Working in Road
08 Standing/Playing in Road
09 Standing in Pedestrian Island
77 All Other (Explain)
88 Unknown

LOCATION TYPE (/)
1 Primarily Business
2 Primarily Residential
3 Open Country

FIRST - SUBSEQUENT HARMFUL EVENT (F: 13, S: /)
01 Collision With MV in Transport (Rear end)
02 Collision With MV in Transport (Head on)
03 Collision With MV in Transport (Angle)
04 Collision With MV in Transport (Left Turn)
05 Collision With MV in Transport (Right Turn)
06 Collision With MV in Transport (Sideswipe)
07 Collision With MV in Transport (Backed Into)
08 Collision With Parked Car
09 Collision With MV on Other Roadway
10 Collision With Pedestrian
11 Collision With Bicycle
12 Collision With Bicycle (Bike Lane)
13 Collision With Moped
14 Collision With Train
15 Collision With Animal
16 MV Hit Sign/Sign Post
17 MV Hit Utility Pole/Light Pole
18 MV Hit Guardrail
19 MV Hit Fence
20 MV Hit Concrete Barrier Wall
21 MV Hit Bridge/Pier/Abutment/Rail
22 MV Hit Tree/Shrubbery
23 Collision With Construction Barricade/Sign
24 Collision With Traffic Gate
25 Collision With Crash Attenuators
26 Collision With Fixed Object Above Road
27 MV Hit Other Fixed Object
28 Collision With Moveable Object On Road
29 MV Ran Into Ditch/Culvert
30 Ran Off Road Into Water
31 Overturned
32 Occupant Fell From Vehicle
33 Tractor/Trailer Jackknifed
34 Fire
35 Explosion
77 All Other (Explain)

ROAD SYSTEM IDENTIFIER (15)
01 Interstate
02 U S
03 State
04 County
05 Local
06 Turnpike / Toll
07 Forest Road
77 All Other

LIGHTING CONDITION (01)
01 Daylight
02 Dusk
03 Dawn
04 Dark (Street Light)
05 Dark (No Street Light)
88 Unknown

ROAD SURFACE / CONDITION (01)
01 Dry
02 Wet
03 Slippery
04 Icy
77 All Other (Explain)

WEATHER (01)
01 Clear
02 Cloudy
03 Rain
04 Fog
77 All Other (Explain)

ROAD SURFACE TYPE (01)
01 Slag / Gravel / Stone
02 Blacktop
03 Brick / Block
04 Concrete
05 Dirt
77 All Other (Explain)

CONTRIBUTING CAUSES ROAD (77, /)
01 No Defects
02 Obstruction With / Without Warning
03 Road Under Repair / Construction
04 Loose Surface Materials
05 Shoulders Soft / Low / High
06 Holes / Ruts / Unsafe Paved Edge
07 Standing Water
08 Worn / Polished Road Surface
77 All Other (Explain)

CONTRIBUTING CAUSES ENVIRONMENT (01, /)
01 Vision Not Obscured
02 Inclement Weather
03 Parked / Stopped Vehicle
04 Trees / Crops / Bushes
05 Load on Vehicle
06 Building / Fixed Object
07 Signs / Billboards
08 Fog
09 Smoke
10 Glare
77 All Other (Explain)

TRAFFIC CONTROL (77, /)
01 No Control
02 School Zone
03 Traffic Signal
04 Stop Sign
05 Yield Sign
06 Flashing Light
07 Railroad Signal
08 Officer / Guard / Flagman
09 Posted No U-Turn
10 Special Speed Zone
11 No Passing Zone
77 All Other (Explain)

SITE LOCATION (02)
01 Not At Intersection / RR X'ing / Bridge
02 At Intersection
03 Influenced By Intersection
04 Driveway Access
05 Railroad Crossing
06 Bridge
07 Entrance Ramp
08 Exit Ramp
09 Parking Lot Public
10 Parking Lot Private
11 Private Property
77 All Other (Explain)

TRAFFICWAY CHARACTER (/)
1 Straight Level
2 Straight Upgrade / Downgrade
3 Curve-Level
4 Curve Upgrade / Downgrade

TYPE SHOULDER (/)
1 Paved
2 Unpaved
3 Curb

VIOLATOR	FL STATUTE NUMBER	
1	2	[redacted]
2	[redacted]	

Page 2 of 4 Pages

Figure 4.9 Crash report Page-2(Some information is intentionally blacked out to not reveal the driver/pedestrian details)

FLORIDA TRAFFIC CRASH REPORT
NARRATIVE / DIAGRAM
MAIL TO DEPT OF HIGHWAY SAFETY & MOTOR VEHICLES
TRAFFIC CRASH RECORDS
TALLAHASSEE, FLORIDA 32399-0500

DO NOT WRITE IN THIS SPACE

11773215528

EMS INFO FATALS ONLY	TIME EMS NOTIFIED AM PM	TIME EMS ARRIVED AM PM	COUNTY / CITY CODE	DATE OF CRASH	INVEST. AGENCY REPORT NUMBER	HSMV CRASH REPORT NUMBER
			06-01	5-28-01	148121114	51477192

NARRATIVE / ADDITIONAL PASSENGERS

Veh#1 was traveling Southbound on NW 12 AV / 36 St.
Veh #2 was traveling Westbound on NW 36 St / 12 AVE.

Traffic Lights were inoperable completely.
Both drivers claimed they Stopped however
the vehicles collided.

Rescue 6 transported unk passengers to J.M.H.
Responded to J.M.H to locate. other passenger but
was not able to locate

SEC #	PASS #	PASSENGER NAME	ADDRESS	CITY & STATE	ZIP	Age	Loc	Inj	Safety Equip	Eject
									|	
									|	
									|	
									|	
									|	
									|	

VIOLATOR	FL STATUTE NUMBER	NAME	CHARGE	CITATION #

VIOLATOR	FL STATUTE NUMBER	NAME	CHARGE	CITATION #

WITNESS NAME	ADDRESS	CITY & STATE	ZIP
1			
2			

FIRST AID GIVEN BY NAME: Rescue 6 — 1 Physician or Nurse 4 Certified 1st Aider 2 Paramedic or EMT 5 Other 3 Police Officer: 2 — INJURED TAKEN TO: J.M. ld — BY NAME

WAS INVESTIGATION MADE AT SCENE? 1 YES 2 NO ☒ — WHERE? — IS INVESTIGATION COMPLETE? 1 YES 2 NO ☒ — WHY? — DATE OF REPORT: 05/28/01 — PHOTOS TAKEN? 1 YES 2 NO ☒ 3 INVEST AGENCY 4 OTHER

INVESTIGATOR RANK & SIGNATURE: P/O J. Minaya — ID / BADGE NUMBER: 1376 — DEPARTMENT: City of Miami — FHP SO CPD ☒ OTHER

HSMV 90005 (Rev 1/95) S — Page____ of____ Pages

Figure 4.10 Crash report Page-3

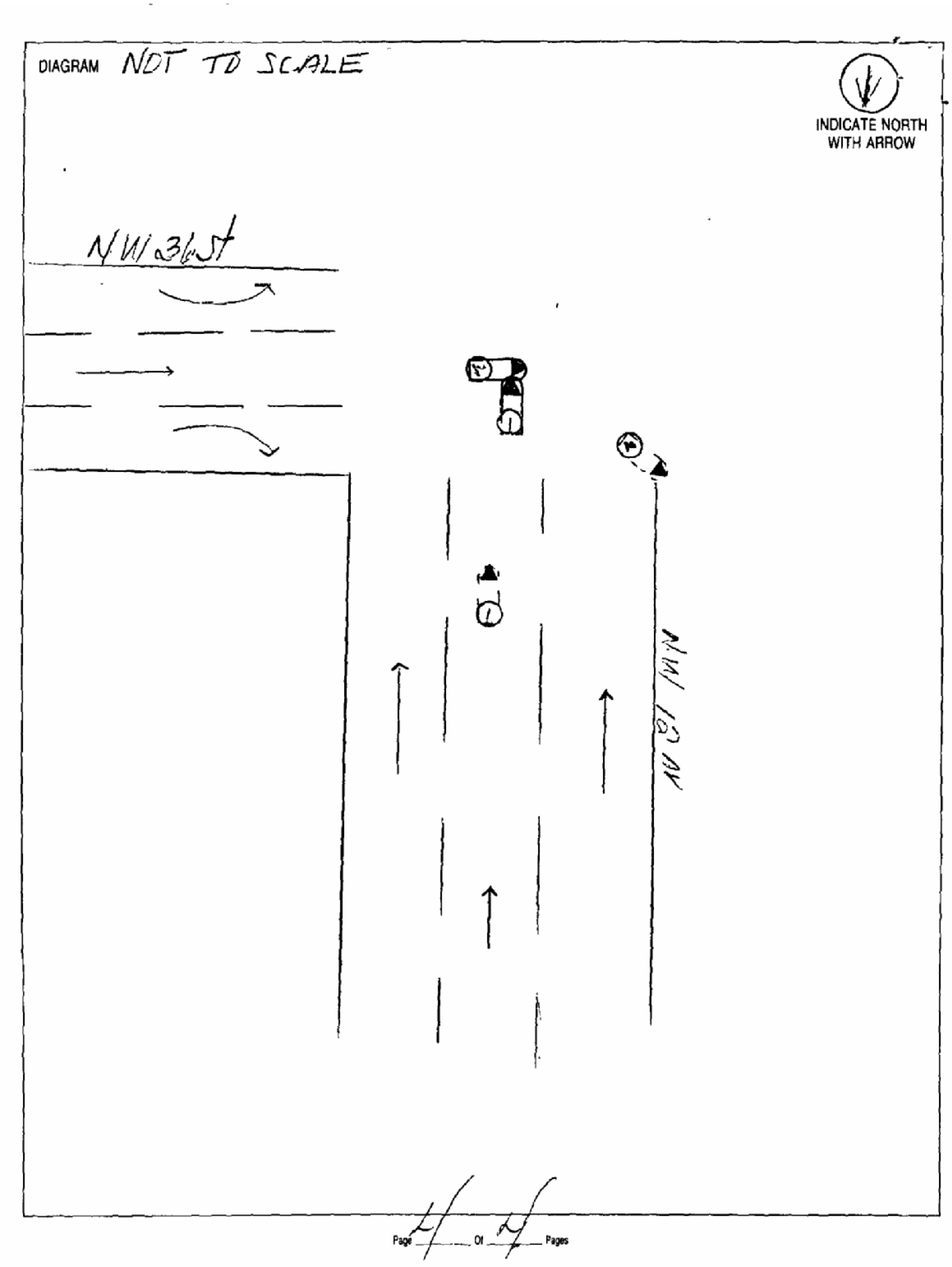


Figure 4.11 Crash report Page-4

With this simple example in perspective, it was decided to identify “First Driver Contributing Cause” of all the drivers involved in the crash (available in the Driver’s file of the DHSMV database). This variable may be used to assign if the driver is at-fault in addition to the “Section Number” of Violation file. The Figure 4.12 is a list of driver contributing causes available in DHSMV files.

01 No Improper Driving/Action
02 Careless Driving
03 Failed To Yield Right-Of-Way
04 Improper Backing
05 Improper Lane Change
06 Improper Turn
07 Alcohol - Under Influence
08 Drugs - Under Influence
09 Alcohol & Drugs - Under Influence
10 Followed Too Closely
11 Disregarded Traffic Signal
12 Exceeded Safe Speed Limit
13 Disregarded Stop Sign
14 Failed To Maintain Equipment/Vehicle
15 Improper Passing
16 Drove Left Of Center
17 Exceeded Stated Speed Limit
18 Obstructing Traffic
19 Improper Load
20 Disregarded Other Traffic Control
21 Driving Wrong Side/Way
22 Fleeing Police
23 Vehicle Modified
24 Driver Distraction
77 All Other
88 Unknown and/or Dummy Record

Figure 4.12 List of contributing causes of drivers involved in crash

If driver’s first contributing cause is other than “01-No Improper Driving/Action”, “77-All Other” and “88-Unknown and/or Dummy Record” then the driver is assigned at-fault and he/she is included in at-fault driver database. Based on this careful examination of ‘fault’, 520 additional drivers were added to the at-fault drivers list. The contributing causes belonging to these additional drivers are listed in Table 4.3

Table 4.3 Contributing causes of additional at-fault drivers

First Contributing Cause of Driver	Frequency	Percentage
Careless Driving	208	40.0%
Failed to Yield Right of Way	144	27.7%
Improper Lane Change	23	4.4%
Improper Turn	26	5.0%
Alcohol Under Influence	10	1.9%
Alcohol & Drugs Under Influence	1	0.2%
Followed Too Closely	24	4.6%
Disregard Traffic Signal	60	11.5%
Exceed Safe Speed Limit	10	1.9%
Disregard Stop Sign	1	0.2%
Improper Passing	2	0.4%
Exceeded Stated Speed Limit	5	1.0%
Obstructing Traffic	2	0.4%
Improper Load	1	0.2%
Disregard Other Traffic Control	2	0.4%
Driver Distraction	1	0.2%
Grand Total	520	100.0%

Careless driving, failed to yield right of way and disregard the traffic sign appear to be the major contributing causes for the additional drivers at-fault. If we add these 520 drivers to the uniquely at-fault driver's the number of observations in the at-fault drivers database is increased to 9,060 drivers.

Table 4.4 The contributing causes of driver before adding 520 additional at-fault drivers to the combined database.

First Contributing Cause	**Frequency**	**Percentage**
No Improper Driving Action	526	6.16%
Careless Driving	3,656	42.82%
Failed To Yield Right of Way	2,092	24.49%
Improper Backing	5	0.06%
Improper Lane Change	235	2.75%
Improper Turn	386	4.52%
Alcohol Under Influence	85	1.00%
Drugs Under Influence	2	0.02%
Alcohol & Drugs Under Influence	10	0.12%
Followed too closely	155	1.81%
Disregard Traffic Signal	1,032	12.08%
Exceed Safe Speed Limit	39	0.46%
Disregard Stop Sign	21	0.25%
Failed to Maintain Equipment/Vehicle	28	0.33%
Improper Passing	18	0.21%
Drove Left off Center	4	0.05%
Exceeded Stated Speed Limit	0	0.00%
Obstructing Traffic	4	0.05%
Improper Load	0	0.00%
Disregard Other Traffic Control	14	0.16%
Driving Wrong Side/way	6	0.07%
Fleeing Police	5	0.06%
Driver Distraction	3	0.04%
Other	214	2.51%
Grand Total	**8,540**	100.00%

The frequency of various "Driver Contributing Causes" before addition of 520 additional at-fault drivers to the database is shown in Table 4.4. From the Table 4.4, we can see that major contributing cause for a crash is "careless driving", which is 42.82% of total crashes. "Failed to yield Right of Way" and "Disregard Traffic Signal" are next two major contributing causes of the drivers involved in crash, which are 24.49% and 12.08% respectively. Similar information after the addition of the 520 drivers is shown in Table 4.5.

Table 4.5 Contributing causes of at-fault drivers after adding 520 additional drivers

First Contributing Cause	Frequency	Percentage
No Improper Driving Action	526	5.8%
Careless Driving	3,864	42.7%
Failed To Yield Right of Way	2,236	24.7%
Improper Backing	5	0.1%
Improper Lane Change	258	2.8%
Improper Turn	412	4.5%
Alcohol Under Influence	95	1.0%
Drugs Under Influence	2	0.0%
Alcohol & Drugs Under Influence	11	0.1%
Followed too closely	179	2.0%
Disregard Traffic Signal	1,092	12.1%
Exceed Safe Speed Limit	49	0.5%
Disregard Stop Sign	22	0.2%
Failed to Maintain Equipment/Vehicle	28	0.3%
Improper Passing	20	0.2%
Drove Left off Center	4	0.0%
Exceeded Stated Speed Limit	5	0.1%
Obstructing Traffic	6	0.1%
Improper Load	1	0.0%
Disregard Other Traffic Control	16	0.2%
Driving Wrong Side/way	6	0.1%
Fleeing Police	5	0.1%
Driver Distraction	4	0.0%
Other	214	2.4%
Grand Total	**9,060**	**100.0%**

From the Tables 4.4 and 4.5, we can see that major contributing cause after adding 520 additional drivers has not changed and it is still "careless driving". "Failed to yield Right of Way" and "Disregard Traffic Signal" are next two major contributing causes of the drivers involved in crash, which are 24.7% and 12.1% respectively. Chi-squared test was performed on the frequency distribution of the "First contributing cause" to verify if the addition of 520 drivers caused a significant change to this distribution. Chi-square test resulted in p-value of 0.97 indicating that the relative distributions of contributing causes have not changed significantly following the addition of 520 at-fault drivers. Therefore, it was concluded that for the purpose of this study we may use driver information only from the 8540 crashes in which at-fault drivers were uniquely identified. This conclusion completes the data preparation phase of this study. In the next chapter, logistic regression analysis of crashes, categorized based on the relative direction of initial movement of involved vehicles, is provided.

4.5 Summary

This chapter first examined the variation of crash frequency and severity with the size and type of intersections as represented by total number of lanes, including left- and right-turn lanes. The appeal of using the approach developed in this research to use the intersection size, represented by the total number of lanes from all approaches, is the simplicity to apply and therefore identify the crash profile by type and severity. It would be easy for a traffic engineer to apply the simple equations developed here by knowing only the number of lanes, x in the equations (an approach similar to the ITE trip generation manual). The engineer can simply calculate the expected crash profiles (by type and severity), and therefore determine if a specific type or severity at a specific intersection is above these expected profiles and hence warrant a treatment. It in fact could be a good surrogate measure of various traffic and geometric features of the intersection including both major and minor roadways, traffic volume, cycle length, number of phases, pedestrian crossing width, among other factors (i.e., larger intersections usually have more traffic, longer cycles and more phases).

The types of intersections in this study were classified based on the directions of travel and the connection between major and minor roads. The results showed that the expected crash frequency expressed as the average number of crashes per intersection per year generally increased as the total number of lanes increased at all types of intersections. However, the rates of increase were different - crash frequency increased with the size of intersections at higher rate for the four-legged two-way intersections than the other intersection types. This result suggests that as the intersection geometry becomes more complex, increasing the size of intersections has more impact on increasing the number of crashes. The results also showed that the dominant crash types were different at different intersection types.

It is worth mentioning here that though the R-square values for the best fitting regression equations are low for four-legged one-way and T-Intersections, the underlying idea was to bring forward the differences in crash patterns at different intersection types.

Angle crashes were dominant at the four-legged one-way intersections unlike the other intersection types where rear-end crashes were dominant. This suggests the possibility that drivers on one-way roads tend to disobey traffic signals and collide with vehicles on crossroads. On the other hand, higher severity was observed at the four–legged two-way intersections than the four-legged one-way intersections due to more conflict points at arguably higher speeds.

The findings in this study provide strong evidence that the patterns of crashes by type and severity vary with the size and type of intersections. This indicates that the characteristics of intersections not only change traffic patterns and driver behavior but also the patterns of crash occurrence. Thus, it is recommended that the size and type of intersections should be considered to account for the effects of intersection characteristics on crash risk.

Besides this preliminary analysis based on the crash data obtained from 6 Florida counties the chapter also described the effort to match the data with the DHSMV database which provides information related to Driver demographics. The final database assembled for the study includes information on crash characteristics as well as characteristics of at-fault drivers. The database is used in the next chapter for final analysis.

5 CATEGORIZATION AND ANALYSIS OF CRASHES BASED ON THE DIRECTION OF INITIAL MOVEMENT OF INVOLVED VEHICLES

5.1 Introduction

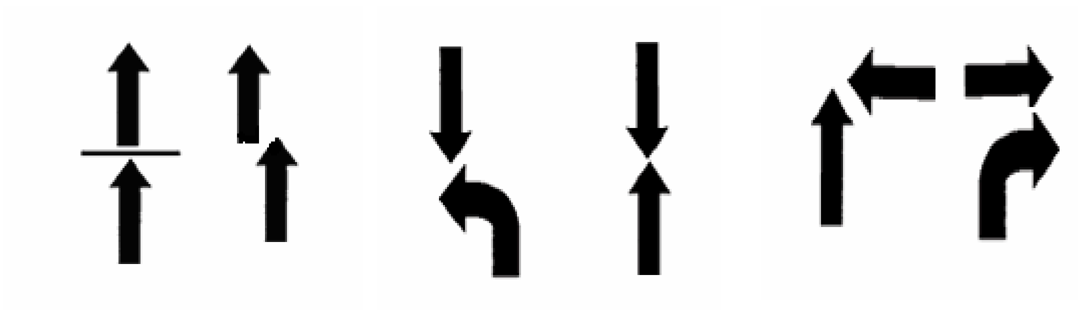
In the previous chapter a simple methodology to identify the expected crash frequency based on the intersection size was presented. The analysis in this chapter is aimed at the identification of the relationship between major crash types and the characteristics of drivers who were at-fault in (and hence caused) those crashes along with the characteristics of the intersection at which these crashes were observed. The crashes in this chapter are categorized based on the direction of the initial movement of the vehicles involved. The approach for categorization of crashes is explained in the next section. Once the categorization is obtained, binary logistic regression models are employed for identification of driver/intersection characteristics critically associated with each of the three groups of crashes.

5.2 Rationale

A crash can occur between vehicles traveling on same approach or on different approaches (legs) at an intersection. A very common approach adopted by traffic engineers for diagnosing possible engineering crash-causes for an intersection is Collision Diagram. It uses representative symbols to illustrate crash patterns. According to Hwang and Wei (2005) the approach of categorizing crashes based on the relative direction of initial movements of the involved vehicles would be consistent with the application of Collision Diagrams in crash analysis.

The rationale behind this approach is that the vehicle's initial moving direction and driver errors associated with the movements are critical for crash occurrence. There are many possible movements at an intersection and a crash can occur during any movement. Categorizing crashes based on the relative direction of initial movement of the vehicles involved in a crash would be better than just knowing the first harmful event initiating the crash, such as, Rear end, Sideswipe, Left-turn, etc. The reason for the same is that, for example, sometimes driver can see the vehicles going in their direction clearly but a vehicle coming in opposite direction or vehicle coming from the perpendicular intersecting street might be difficult to spot given the complexity of movements at an intersection. Such differentiation can only be accounted for by dividing the crash data based on the relative initial movement direction of the vehicles involved.

As mentioned earlier, the objective of this chapter is to categorize the crashes based on the relative direction of initial movement at an intersection and identify the relationship between crashes and the characteristics of drivers who were found to be at-fault for these crashes along with the characteristics of the intersections where these crashes occurred. The three crash types based on the initial movement are shown in Figure 5.1. It is clear that the vehicles involved in rear-end and sideswipe crashes would be traveling in the same direction with respect to each other hence these crashes are named "Initial Movement in Same Direction (IMSD)" or IMSD crashes. Similarly, prior to the crashes initiated by left-turn or head-on collisions the vehicles involved would be traveling in opposite direction with respect to each other. Hence, these crashes are referred to as "Initial Movements in Opposite Direction (IMOD)" or IMOD crashes. The right-turn and right-angle crashes would include vehicles initially traveling on intersecting roadways. Hence, the direction of their initial movement may be considered perpendicular to each other. These crashes are termed “Initial Movements in Perpendicular Direction (IMPD)” or IMPD crashes.



IMSD (Initial Movement in Same Direction) Crashes (Rear-end and Sideswipe Collisions)

IMOD (Initial Movement in Opposite Direction) Crashes (Head-on and Left-turn Collisions)

IMPD (Initial Movement in Perpendicular Direction) Crashes (Right-Angle and Right-turn Collisions)

Figure 5.1 Crashes by Direction of Initial Vehicle Movement

The sample compositions by the “first harmful event” for these three groups of crashes are shown in Figure 5.2. It is clear that for signalized intersections the composition of the three categories (IMSD, IMPD, and IMOD) is largely uniform with respect to the “first harmful event”. For example, IMOD cashes consist of mostly left-turn crashes, and signalized intersections have very few head-on collisions. Similarly, rear-end and angle crashes, respectively, dominate the sample of IMSD and IMPD crashes.

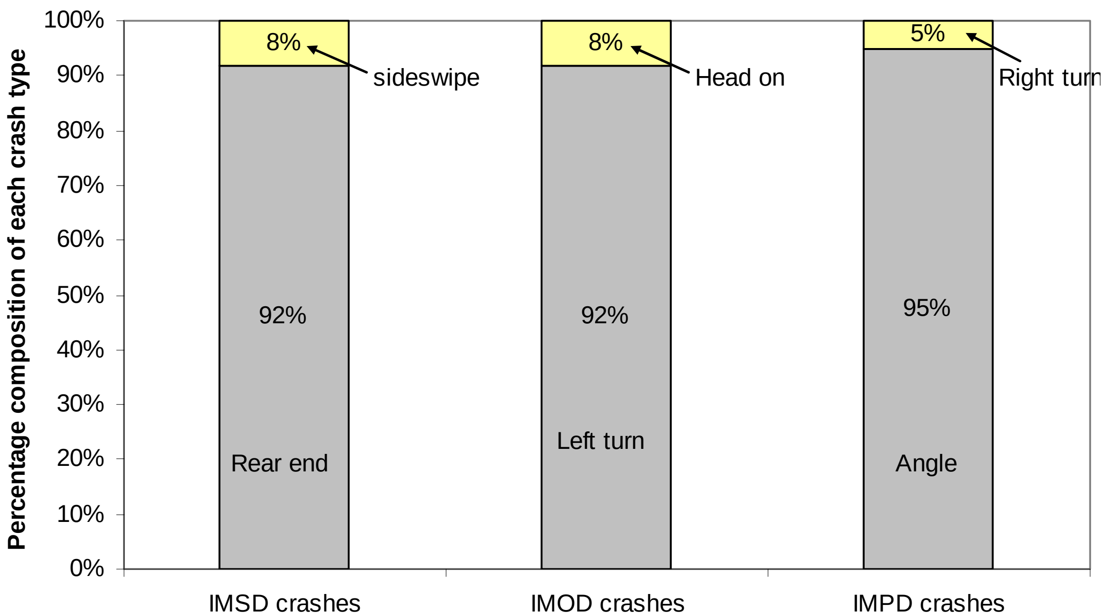


Figure 5.2 Composition of the crash data by "first harmful event" for the three groups of crashes

According to Tijerina et al. (1994) angle crashes on intersections occur when the path of a vehicle without the right-of-way intersects with the path of another vehicle traveling on the intersecting roadway (which does have the right-of-way). Possible sources of problems associated with such crashes are inaccurate detection or interpretation of the signal status, time-estimation errors associated with signal status, lack of detection of the cross traffic movements, and problematic visual obstructions. These driver and/or intersection related problems could potentially cause right-turn or right-angle collisions which in this study are categorized as IMPD crashes.

FHWA (2004) states that failure to judge speeds of closing vehicles correctly, obstruction of driver's view, failure to perceive opposing vehicle are the sources of errors for left-turn and head-on crashes. A major cause of left-turn crashes is the permitted left-turn on small intersections. Such intersections are also less likely to have traffic divided by a physical barrier (Median) resulting higher chances of Head-on crashes. Based on the involved vehicles' direction of initial movement being opposite for head-on and left-turn collisions, these crashes are categorized as the IMOD crashes. The frequency of the corresponding harmful events, along with their categorization, in the combined database is shown in Table 5.1.

Table 5.1 Crash type composition in combined database

Crash Type	Frequency	Percentage	Crash Categorization
Rear End	3,786	44%	IMSD
Head On	168	2%	IMOD
Angle	2,276	27%	IMPD
Left Turn	1,847	22%	IMOD
Right Turn	122	1%	IMPD
Sideswipe	341	4%	IMSD
Grand Total	**8540**	**100%**	-

5.3 Contingency Table Analysis

Data preparation for the analysis was presented in Section 4.4 of the previous chapter. This section summarizes the contingency table analysis to identify significant variables affecting the three groups of crashes presented in Figure 5.1. The purpose of the contingency table analysis is to find the variables that should be included in the final models. The analysis also summarizes general pattern of variables towards corresponding crash types.

The Statistical Analysis Software (SAS Institute, 2001) is used in for the conditional probability analysis. The procedure "FREQ" is used for the cross tabulation. The tables presented in this section are only for those variables which are found to be significant at the 95% level. The contingency coefficient is used as a parameter to indicate significance of the factors. The larger the contingency coefficient the stronger will be the association between the row and column variables.

5.3.1 Significant variables

One of the ideas presented in this study is to use number of lanes at an intersection as a surrogate measure to AADT. This is because, as discussed in chapter 4, size of an intersection represents various traffic and geometric features of the intersection including both major and minor roadways, traffic volume, cycle length, number of phases, pedestrian crossing width, among other factors (i.e., larger intersections usually have more traffic, longer cycles and more phases). Abdel-Aty and Radwan (1999) have examined effects of AADT and accident occurrence and found that AADT per lane has the greatest relative effect on the accident frequency among all the independent variables considered in the study. Since most of the intersections in the combined database do not have AADT information on minor road, AADT on major road is used. AADT per lane is obtained as AADT on the major road divided by total number of lanes (thorough lanes+ exclusive left turn lanes+ right turn lanes). It was also found that there is a linear relationship between AADT per lane of the major road and number of lanes at the intersection.

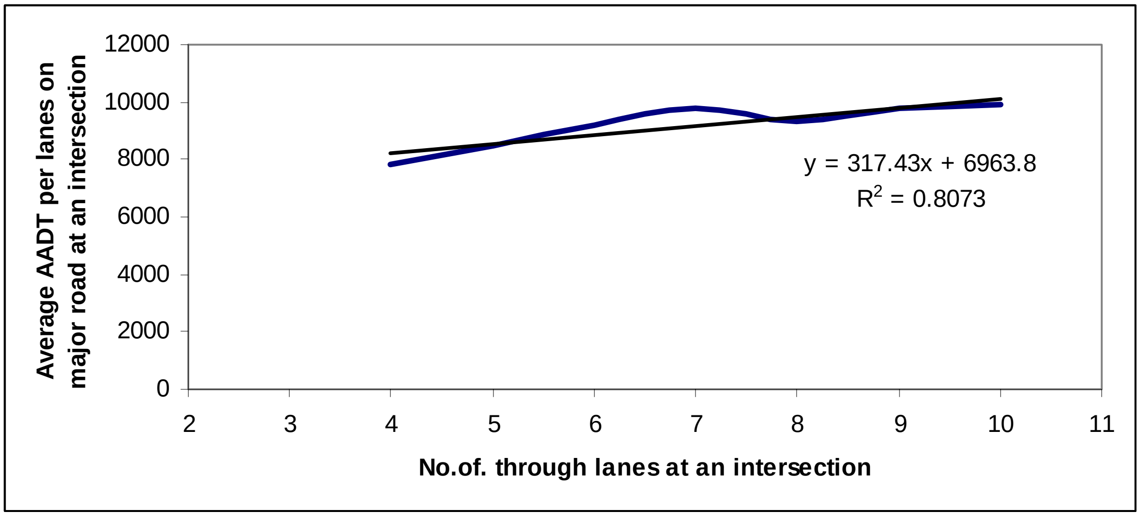


Figure 5.3 Relationship between average AADT per lane and number of through lanes at an intersection.

Figure 5.4 Relationship between average AADT per lane and total number of lanes at an intersection.

The parameter 'y' in the functions shown in Figures 5.3 and 5.4 represents average AADT per lane on major road. The parameter 'x' represents total number of *through* lanes at the intersection in Figure 5.3 and represents total number of lanes at the intersection in Figure 5.4.

From Figures 5.3 and 5.4 it may be argued that number of through lanes better represents AADT than the total number of lanes at an intersection (because of higher R-Square value). Hence, number of through lanes at an intersection is used to represent the AADT.

The other variables that are investigated in the contingency analysis fall into 3 major types. These are “Driving Environment” related, “Intersection” related and “Driver” related. Let *X* and *Y* denote two categorical variables, with X being the row variable and Y the column variable with *I* and *J* categories, respectively. The *I**J* possible combinations of outcomes could be displayed in a rectangular table. A sample 2 X 2 cross-tabulation is shown in Table 5.2. Each cell in the table presents Frequency, Row Percent and Column Percent of that I*J combination.

Table 5.2 A sample 2 X 2 cross-tabulation

	Column Variable Y1	**Column Variable Y2**	**Total**
Row Variable X1	Frequency Percent Row Percent Column Percent	Frequency Percent Row Percent Column Percent	Total column frequency for row variable X1 Average column percentage for row variable X1
Row Variable X2	Frequency Percent Row Percent Column Percent	Frequency Percent Row Percent Column Percent	Total column frequency for row variable X2 Average column percentage for row variable X2
	Row frequency for Column Variable Y1 Average Row percentage for Column Variable Y1	Row frequency for Column Variable Y1 Average Row percentage for Column Variable Y1	**Total**

This cross table can be used in two ways. We can either compare across columns or we can compare across rows. It essentially means that either we can compare row percentages against average row percentage for that column, or column percentages against average columns for that row variable.

5.3.1.1 Driving environment related factors

(i) Lighting Conditions

The light conditions were originally classified into five categories. These are Daylight, Dusk, Dawn, Dark (With street lights), and Dark (no street lights). All categories except for daylight conditions were combined into the second category, as the lighting conditions will be almost dark during the other cases. According to Traffic Safety Handbook by Minnesota DOT (2001) most of the intersection crashes occur during the day, on dry roads and during good weather conditions. It indicates that the crash frequency is a function of exposure. This implies that people drive more in daylight conditions than in dark, hence, high number of crashes in daylight. There were 5,852 crashes occurred in daylight conditions and 2,909 in the dark conditions.

Table 5.3 Lighting condition and crash type contingency table

Lighting Condition	**Crash Type**			**Total**
	IMPD Crashes	**IMOD Crashes**	**IMSD Crashes**	
Dark	874	758	1277	2909
	9.98	8.65	14.58	33.2
	30.04	26.06	43.9	
	35.53	36.78	30.12	
Daylight	1586	1303	2963	5852
	18.1	14.87	33.82	66.8
	27.1	22.27	50.63	
	64.47	63.22	69.88	
Total	2460	2061	4240	8761
	28.08	23.52	48.4	100

Statistic	DF	Value	p-value
Chi-Square	2	36.0715	<.0001
Likelihood Ratio Chi-Square	2	36.1112	<.0001

Contingency Coefficient 0.0640

As we can see from Table 5.3, in daylight conditions, overall highest proportion crashes (33.82 %) are IMSD crashes. Also, in dark conditions higher proportions of crashes are IMOD crashes (36.78 % vs. average 33.2 %) and in daylight conditions higher proportion of crashes (69.88 % vs. average 66.8 %) are IMSD crashes. Comparing across crash types we can see from Table 5.3 that most of IMSD crashes happen in daylight driving conditions, where as IMOD and IMPD crashes happen in dark driving conditions.

From an application perspective crashes in dark conditions (only conditions with or without traffic lights and not those that occurred during dusk/dawn) are further investigated. It is found that 18.2 % of crashes occurred when there are no street lights at intersections. In dark conditions and when street lights are not present high proportions of crashes tend to be IMOD crashes. From Table 5.4 it can be said that street lights tend to decrease IMOD crashes. On the other hand, intersection with street light experience high proportions of IMPD crashes.

Table 5.4 Dark driving conditions and crash type contingency table

Lighting Condition	Crash Type IMPD Crashes	IMOD Crashes	IMSD Crashes	Total
Dark (Street Lights)	694 25.77 31.5 84.22	554 20.57 25.15 78.81	955 35.46 43.35 81.9	2203 81.8
Dark (No Lights)	130 4.83 26.53 15.78	149 5.53 30.41 21.19	211 7.84 43.06 18.1	490 18.2
Total	824 30.6	703 26.1	1166 43.3	2693 100

Statistic	DF	Value	p-value
Chi-Square	2	7.4955	0.0236
Likelihood Ratio Chi-Square	2	7.4486	0.0241
Contingency Coefficient		0.0527	

(ii) Road Surface Conditions

Surface conditions were originally available from DHSMV database as Dry, Wet, Slippery, Icy and Other. In Florida road surface conditions are rarely Icy (There were only 33 intersection crashes listed in combined database for Icy conditions over the three year period). As road surface conditions Wet, Slippery, Icy and Other represent non dry surfaces, these conditions are combined into Non-Dry road surface category. Hence, road surface conditions are brought down to two categories, Dry and Non-Dry conditions.

Table 5.5 Surface conditions and Crash type contingency table

Surface Conditions	Crash Type IMPD Crashes	IMOD Crashes	IMSD Crashes	Total
Not Dry	347	224	763	1334
	3.96	2.56	8.71	15.23
	26.01	16.79	57.2	
	14.11	10.87	18	
Dry	2113	1837	3477	7427
	24.12	20.97	39.69	84.77
	28.45	24.73	46.82	
	85.89	89.13	82	
Total	2460	2061	4240	8761
	28.08	23.52	48.4	100

Statistic	DF	Value	p-value
Chi-Square	2	57.8997	<.0001
Likelihood Ratio Chi-Square	2	59.6059	<.0001
Contingency Coefficient		0.0810	

From Table 5.5 it can be seen that less percentage (15.23 %) of crashes occurred in non-dry surface conditions while high percentage (84.77 %) of crashes occurred when surface is dry. It can be argued that when surface is not dry driver needs more stopping sight distance for a vehicle to come to stop. Braking performance of vehicle is substantially affected and deceleration capacity may decrease in wet surface condition. Therefore, reduction in braking capacity of the vehicle may be the cause of high proportions (18.00% vs. the average 15.23%) of crashes are IMSD crashes in non dry conditions.

(iii) Weather conditions

Weather conditions were originally classified as 5 categories, which include clear, cloudy, Rain, Fog and all other conditions. The weather conditions except clear conditions are combined as "not clear" weather conditions.

Table 5.6 Weather conditions and Crash type contingency table.

Weather Conditions	Crash Type			Total
	IMPD Crashes	IMOD Crashes	IMSD Crashes	
Not Clear	634 7.24 24.64 25.77	535 6.11 20.79 25.96	1404 16.03 54.57 33.11	2573 29.37
Clear	1826 20.84 29.51 74.23	1526 17.42 24.66 74.04	2836 32.37 45.83 66.89	6188 70.63
Total	2460 28.08	2061 23.52	4240 48.4	8761 100

Statistic	DF	Value	p-value
Chi-Square	2	55.5541	<.0001
Likelihood Ratio Chi-Square	2	55.5769	<.0001
Contingency Coefficient		0.0794	

Adverse weather can reduce visibility and road surface friction and thus increase likelihood of crashes. Table 5.6 indicates that if weather is not clear, high percentages of crashes are IMSD crashes (33.11 % vs. average 29.37 %).

iv) Rural-Urban conditions

The Table 5.7 indicates that high percentage (62.58%) of crashes occurred in rural intersection while 37.42 % occurred in urban intersections. High proportions of crashes in rural intersection are tend to be IMSD crashes (64.08%, average 62.58%) where as intersections in urban areas tend to have IMPD crashes.

Table 5.7 Rural/Urban and Crash type contingency table.

Rural/ Urban Environment	Crash Type			Total
	IMPD Crashes	IMOD Crashes	IMSD Crashes	
Rural	1448	1318	2717	5483
	16.53	15.04	31.01	62.58
	26.41	24.04	49.55	
	58.86	63.95	64.08	
Urban	1012	743	1523	3278
	11.55	8.48	17.38	37.42
	30.87	22.67	46.46	
	41.14	36.05	35.92	
Total	2460	2061	4240	8761
	28.08	23.52	48.4	100

Statistic	DF	Value	p-value
Chi-Square	2	20.2498	<.0001
Likelihood Ratio Chi-Square	2	20.1040	<.0001
Contingency Coefficient		0.0480	

Federal Highway Administration (FHWA) statistics (2002) of rural/urban crashes states that the rural roads in most of the cases evolved from farm roads upgraded to accommodate increased traffic volumes and vehicle size. In many areas, farmers, commuters, school buses, trucks, and tourists share roads with narrow lanes, limited sight distance, less enforcement, and lack of clear roadsides. In rural areas, legal speeds on collector and local roads are often higher than their urban counterparts. On rural roads, unlike urban roads, traffic is not often slowed by frequent traffic signals, stop signs, and traffic congestion. Hence, this reason can substantiate that the intersection in rural roads tend to have IMSD crashes.

5.3.1.2 Intersection related factors

i) Intersection type

Underlined idea of this study is to introduce the concept of the intersection type in intersection safety analysis. This is because in the preliminary analysis it is shown that four legged intersections with two-way road in both directions and three legged Intersections (T-Intersections) had predominantly rear-end crashes where as four legged intersections with one-way road in one of the road at an intersection had predominantly angle crashes. In other words, we can say that these three major intersection types experience different crash patterns. Hence, intersection types have to be considered in the intersection safety analysis.

Table 5.8 Intersection type and Crash type contingency table

Intersection type	Crash Type			
	IMPD Crashes	IMOD Crashes	IMSD Crashes	Total
Four legged Two-way Intersections	2078	1797	3677	7552
	23.72	20.51	41.97	86.2
	27.52	23.8	48.69	
	84.47	87.19	86.72	
T-Intersections	165	170	415	750
	1.88	1.94	4.74	8.56
	22	22.67	55.33	
	6.71	8.25	9.79	
Four legged One-way Intersections	217	94	148	459
	2.48	1.07	1.69	5.24
	47.28	20.48	32.24	
	8.82	4.56	3.49	
Total	2460	2061	4240	8761
	28.08	23.52	48.4	100

Statistic	DF	Value	p-value
Chi-Square	4	105.5837	<.0001
Likelihood Ratio Chi-Square	4	99.1054	<.0001
Contingency Coefficient		0.1091	

Table 5.8 indicates that at T-intersections high proportions of the crashes are IMSD crashes while at four legged One-way Intersections high proportion of the crashes are IMPD crashes. This result again reiterates the findings in the previous chapter.

ii) Speed limit on major road

Speed limits are originally 25 mph, 30mph, 35mph, 45mph, 50mph and 55 mph on the major roads. There were only 104 drivers involved in crashes at speeds lesser than

30mph. Hence, drivers involved in crashes 25 mph and 30 mph are combined into one category, which is speed limits less than or equal to 30 mph.

Table 5.9 Speed limits on major road and Crash type contingency table.

Speed Limit on Major Road (mph)	**Crash Type**			**Total**
	IMPD Crashes	**IMOD Crashes**	**IMSD Crashes**	
≤30	405 4.62 37.29 16.46	216 2.47 19.89 10.48	465 5.31 42.82 10.97	1086 12.4
35	364 4.15 34.02 14.8	241 2.75 22.52 11.69	465 5.31 43.46 10.97	1070 12.21
40	554 6.32 28.42 22.52	521 5.95 26.73 25.28	874 9.98 44.84 20.61	1949 22.25
45	985 11.24 24.36 40.04	941 10.74 23.27 45.66	2117 24.16 52.36 49.93	4043 46.15
50	67 0.76 22.56 2.72	77 0.88 25.93 3.74	153 1.75 51.52 3.61	297 3.39
55	85 0.97 26.9 3.46	65 0.74 20.57 3.15	166 1.89 52.53 3.92	316 3.61
Total	2460 28.08	2061 23.52	4240 48.4	8761 100

Statistic	DF	Value	p-value
Chi-Square	10	119.0191	<.0001
Likelihood Ratio Chi-Square	10	116.4732	<.0001
Contingency Coefficient		0.1158	

Table 5.9 indicates that at intersection having major roads with high speed limits have high percentage of drivers at-fault in IMSD crashes. While, at low speeds drivers are at-fault in IMPD crashes. The possible reason for this phenomenon could be that at high speeds drivers require a larger distance to stop the car. If a minimum distance is not maintained a IMSD crash is likely to occur.

iii) Divided/Undivided Road

Crash trend on divided/undivided highway is confounded by the traffic volume effect because the intersections with high heavy traffic tend to have more lanes and separated by medians.

Table 5.10 Divided-Undivided road and crash type contingency table.

Divided/ Undivided	**Crash Type**			**Total**
	IMPD Crashes	**IMOD Crashes**	**IMSD Crashes**	
Divided	1667	1432	3038	6137
	19.03	16.35	34.68	70.05
	27.16	23.33	49.5	
	67.76	69.48	71.65	
Undivided	793	629	1202	2624
	9.05	7.18	13.72	29.95
	30.22	23.97	45.81	
	32.24	30.52	28.35	
Total	2460	2061	4240	8761
	28.08	23.52	48.4	100

Statistic	DF	Value	p-value
Chi-Square	2	11.6241	0.0030
Likelihood Ratio Chi-Square	2	11.5930	0.0030
Contingency Coefficient		0.0364	

Table 5.10, indicates that high proportions (71.65 %, average 70.05%) of crashes are IMSD crashes on divided roads. This can be attributed to the fact that intersections with high traffic volumes in most cases are divided and high volumes indicate that high IMSD crashes. Hence, divided roads could have high proportions of IMSD crashes.

5.3.1.3 Driver related factors

i) Driver Age Group

Teenage driver (15-19) group are separated and then 10-year interval was chosen to group drivers by age. (Yan et.al 2005). From the Table 5.11, we can see that driver age group 25-34 is the largest single group at-fault in crashes with 22.78% of crashes. The number (and/or percentage) of drivers at-fault in a crash tend to decreases steadily as driver gets older after a consistent increase in percentages of drivers at-fault till 25-34 age category. High proportions of drivers of age between 19 and 55 are at-fault in IMSD crashes. High proportions of teenage drivers and drivers of age above 55 are tend to be at-fault in IMPD and IMOD crashes. Drivers being at-fault in IMOD crashes tend to increase consistently as driver grows older after 24 years. Similar increasing trend is observed for IMPD crashes for drivers older than 34 years. Driver’s inability to judge the speed of on coming vehicle and decline in perception reaction times of old age groups could be reason for involvement in IMPD and IMOD crashes.

Table 5.11 Age group and Crash type and contingency table

Driver Age Group	Crash Type: IMPD Crashes	IMOD Crashes	IMSD Crashes	Total
15-19	337 3.95 28.2 14.05	316 3.7 26.44 15.68	542 6.35 45.36 13.13	1195 13.99
20-24	390 4.57 28.55 16.26	302 3.54 22.11 14.99	674 7.89 49.34 16.33	1366 16
25-34	525 6.15 26.99 21.89	389 4.56 20 19.31	1031 12.07 53.01 24.98	1945 22.78
35-44	446 5.22 26.27 18.6	352 4.12 20.73 17.47	900 10.54 53 21.81	1698 19.88
45-54	288 3.37 27.88 12.01	237 2.78 22.94 11.76	508 5.95 49.18 12.31	1033 12.1
55-64	179 2.1 31.51 7.46	162 1.9 28.52 8.04	227 2.66 39.96 5.5	568 6.65
65-74	131 1.53 31.41 5.46	132 1.55 31.65 6.55	154 1.8 36.93 3.73	417 4.88
≥75	102 1.19 32.08 4.25	125 1.46 39.31 6.2	91 1.07 28.62 2.2	318 3.72
Total	2398 28.08	2015 23.59	4127 48.33	8540 100

Statistic	DF	Value	p-value
Chi-Square	14	145.6384	<.0001

Likelihood Ratio Chi-Square	14	144.5509	<.0001
Contingency Coefficient		0.1295	

ii) Gender

Over all, male drivers are at-fault in most of the cases than female driver. In most of the crashes, male drivers tend to be at-fault in IMSD crashes (65.83%, average 61.99%) and female drivers are at-fault in IMOD crashes (44.37% average 38.01%). Compared to male drivers, female drivers tend to be at-fault in IMPD and IMOD crashes.

Table 5.12 Gender and Crash type contingency table

Gender	**Crash Type**			**Total**
	IMPD Crashes	**IMOD Crashes**	**IMSD Crashes**	
Female	942 11.03 29.02 39.28	894 10.47 27.54 44.37	1410 16.51 43.44 34.17	3246 38.01
Male	1456 17.05 27.50 60.72	1121 13.13 21.17 55.63	2717 31.81 51.32 65.83	5294 61.99
Total	2398 28.08	2015 23.59	4127 48.33	8540 100.00

Statistic	DF	Value	p-value
Chi-Square	2	62.1016	<.0001
Likelihood Ratio Chi-Square	2	61.8373	<.0001
Contingency Coefficient		0.0850	

iii) Race

Table 5.13 indicates that high percentages of Caucasian drivers are tend to be at-fault in IMSD crashes. African-American drivers are highly at-fault in IMPD crashes.

Table 5.13 Race and Crash type contingency table

Race	Crash Type			Total
	IMPD Crashes	IMOD Crashes	IMSD Crashes	
Caucasian	2036 23.84 28.15 84.9	1743 20.41 24.1 86.5	3454 40.44 47.75 83.69	7233 84.7
African-American	281 3.29 27.55 11.72	201 2.35 19.71 9.98	538 6.3 52.75 13.04	1020 11.94
Hispanic	55 0.64 27.36 2.29	44 0.52 21.89 2.18	102 1.19 50.75 2.47	201 2.35
Others	26 0.3 30.23 1.08	27 0.32 31.4 1.34	33 0.39 38.37 0.8	86 1.01
Total	2398 28.08	2015 23.59	4127 48.33	8540 100

Statistic	DF	Value	p-value
Chi-Square	6	16.6919	0.0105
Likelihood Ratio Chi-Square	6	16.8846	0.0097
Contingency Coefficient		0.0442	

iv) Alcohol/drugs in use

In the database, only 893 of 8540 were at-fault found to be under alcohol/drugs influence while driving. Table 5.14 indicates that driver using alcohol/drugs tend to be more at-fault in IMSD crashes.

Table 5.14 Alcohol/Drugs use and Crash type contingency table

Alcohol/Drugs	Crash Type IMPD Crashes	 IMOD Crashes	 IMSD Crashes	Total
Not in use	2207 25.84 28.86 92.04	1873 21.93 24.49 92.95	3567 41.77 46.65 86.43	7647 89.54
In Use	191 2.24 21.39 7.96	142 1.66 15.9 7.05	560 6.56 62.71 13.57	893 10.46
Total	2398 28.08	2015 23.59	4127 48.33	8540 100

Statistic	DF	Value	p-value
Chi-Square	2	83.6177	<.0001
Likelihood Ratio Chi-Square	2	84.5293	<.0001
Contingency Coefficient		0.0985	

5.4 Summary

As discussed earlier, contingency analysis is used to find the significant variables, which influence each crash type. The significant factors from the preliminary analysis are presented in Table 5.15. Hence, from Table 5.15 it is evident (based on contingency coefficient as well as corresponding p-value) that Age groups, speed limits on major road and intersection type have strong influence on the crash types. Though the magnitudes of the contingency coefficients are low for the significant variables their high significance is indicated by near zero p-values.

Table 5.15 Significant categorical variables after preliminary analysis

Variable	Contingency Coefficient	p-value
Lighting Conditions	0.0640	<0.0001
Surface Conditions	0.0810	<0.0001
Weather conditions	0.0794	<0.0001
Rural / Urban Area Type	0.0480	<0.0001
Intersection type	0.1091	<0.0001
Divided/Undivided	0.0364	<0.0001
Speed Limit on Major Road	0.1158	<0.0001
Age Group	0.1295	<0.0001
Gender	0.0850	<0.0001
Race	0.0442	<.0105
Alcohol/drugs in use	0.0985	<0.0001

Many people think that most crashes occur at night and/or during bad weather. However, most crashes occur during the day, on dry roads and during good weather conditions because of more exposure in these conditions. At four-legged two-way intersections and three legged intersections drivers predominantly found to be at-fault in IMSD crashes. Whereas, at four-legged one-way intersection drivers tend to make mistakes predominantly in IMPD crashes. At high speed limit, high percentages of drivers are at-fault in IMSD crashes. Teenage drivers and Old drivers tend to be at-fault in IMOD and IMPD crashes. Male drivers tend to be at-fault in most of the crashes than female drivers. Male drivers tend to be at-fault in IMSD crashes and female drivers tend to be at-fault in IMOD crashes.

Using contingency tables significant variables are determined. However, there are two limitations to the contingency table analysis:

1. It is difficult to know how significantly greater are the cell percentages to the corresponding average.
2. It is also difficult to assess the joint significance of explanatory variables on a crash type.

To overcome these limitations of contingency table analysis a binary logistic regression models are employed, to estimate the maximum likelihood of crash occurrence to the corresponding vehicle movement. But, before developing binary logistic regression models, variables are cross checked if there was any correlation between the variables. Therefore, next step is to find if there are any correlated variables using Pearson correlation.

5.5 Pearson Correlation between Significant Variables

The Pearson correlation describes the strength of the linear association between the row and column variables. It is generally believed that rural environment is associated with high speeds and road surface condition correlated with weather conditions. Hence, this section investigates if there is any correlation between driving environment related variables such as weather and surface conditions, speed limit and rural/urban driving conditions, lighting conditions and weather.

Table 5.16 Correlation between driving environment related variables.

Prob > |r| under H0: Rho=0						
	Lighting Conditions	**Surface Conditions**	**Weather Conditions**	**Rural/ Urban**	**Speed Limit on major road**	**Divided/ Undivided**
Lighting Conditions	1	0.04456 <.0001	-0.0055 0.6066	-0.0163 0.1268	0.0382 0.0003	0.00226 0.8324
Surface Conditions	0.04456 <.0001	1	0.60214 <.0001	-0.015 0.1599	0.00833 0.4358	0.00731 0.4936
Weather Conditions	-0.0055 0.6066	0.60214 <.0001	1	0.06044 <.0001	-0.03555 0.0009	0.06436 <.0001
Rural/ Urban	-0.01631 0.1268	-0.01502 0.1599	0.06044 <.0001	1	-0.23218 <.0001	0.18601 <.0001
Speed Limit on major road	0.0382 0.0003	0.00833 0.4358	-0.03555 0.0009	-0.2322 <.0001	1	-0.15851 <.0001
Divided/ Undivided	0.00226 0.8324	0.00731 0.4936	0.06436 <.0001	0.18601 <.0001	-0.15851 <.0001	1

Table 5.16 shows that the Pearson correlation coefficient between surface conditions and weather is relatively high (0.60214). Despite their high correlation, weather and road surface can be included separately in crash type models because rain can significantly

reduce visibility, whereas the road surface can stay wet long after the rain. Pearson correlation between speed limit on major roads and rural/urban is also relatively high (0.22613). Hence, for the logistic regression models rural/urban variables is dropped as it is correlated with speed limits on major roads.

5.6 Logistic Regression for three categories of crashes

As discussed earlier, the final objective of this chapter is to analyze three groups of crashes separated based on the relative direction of the initial movement of the vehicles involved. Three binary logistic regression models are developed in this section. Each of the models is created to separate one specific crash type (e.g., IMSD) from the other two types of crashes (i.e., IMPD and IMOD) based on intersection, driving environment and at-fault driver characteristics. The outcome of the model is the conditional probability of crash occurrence of a specific type (IMSD or IMPD or IMOD) given a crash has occurred.

This analysis uses significant categorical variables obtained from the contingency table analysis (Table 5.15). Apart from the categorical variables discussed in contingency table analysis, continuous variables such as total number of through lanes at the intersection, number of exclusive left turning lanes on major road, total number of exclusive left turning lanes on minor road, and total number of channelized (exclusive) right turn lanes at the intersection are also included in models as inputs. Binary logistic regression models are developed using PROC LOGISTIC procedure in SAS (SAS Institute, 2001).

5.6.1 Logistic regression models for IMSD crashes

The first binary logistic regression model is developed to estimate the conditional probability of IMSD crashes (probability of occurrence of an IMSD crash given a crash has occurred). The crash, intersection, and at-fault driver characteristics for all 8040 crashes are used as inputs to develop this model. The response variable, Y, is equal to 1 for crashes in the dataset that belong to IMSD category (i.e., rear-end and sideswipe crashes) and is 0 for the crashes that do not belong to IMSD crashes. The database comprises of 48% IMSD crashes (Y=1) as opposed to 52% other two crash types (Y=0).

The coefficients of the logistic regression model are presented in Table 5.17 and the Goodness of fit statistics is provided in Table 5.18. Figure 5.5 depicts the ROC curve for the model indicating the accuracy of the model in terms of separating IMSD crashes from the other two types of crashes. A detailed discussion on these results is provided in Section 5.6.4 onwards, after the depiction of all three logistic regression models.

Table 5.17 IMSD crash type model (Y=1 if crash type is IMSD and Y=0 for other two crash types)

Variable	**Estimate**	**Odds Ratio**	**Standard Error**	**Pr > ChiSq (p-value)**
Intercept	-0.63		0.1069	<.0001
Light (1=Daylight, 0=Dark)	0.229	1.58	0.0252	<.0001
Weather (1=Not clear, 0=Clear)	0.0888	1.19	0.0308	0.0039
Surface (1=Not Dry, 0=Dry)	0.1588	1.37	0.0393	<.0001
Intersection Type (1=Four legged Two-way intersections, 0= Four legged One-way intersections)	0.1898	1.46	0.0571	0.0009
Intersection Type (1=T-Intersections, 0= Four legged One-way intersections)	0.3586	2.05	0.0672	<.0001
Thru_lanes (Total no. of through lanes)	0.0958	1.10	0.013	<.0001
TotExLTLsMN(Total no. of exclusive left-turn lanes on minor road)	0.0633	1.07	0.023	0.006
Speed (35mph Vs ≤30 mph)	-0.1206	0.96	0.0578	0.0371
Speed (40mph Vs ≤30 mph)	-0.1024	0.98	0.0464	0.0275
Speed (45mph Vs ≤30 mph)	0.1575	1.27	0.0384	<.0001
Speed (50mph Vs ≤30 mph)	0.1469	1.26	0.0714	0.0398
Speed limit ≤30 mph (Base case)	-	-	-	-
Alc_drugs (1=In use, 0=Not in use)	0.3482	2.01	0.0394	<.0001
Age_grp (20-24 Vs 15-19)	0.2271	1.14	0.0559	<.0001
Age_grp (25-34 Vs 15-19)	0.3566	1.3	0.0493	<.0001
Age_grp (35-44 Vs 15-19)	0.3272	1.26	0.0516	<.0001
Age_grp (45-54 Vs 15-19)	0.1912	1.1	0.0621	0.0021
Age_grp (55-64 Vs 15-19)	-0.1897	0.75	0.0811	0.0193
Age_grp (65-74 Vs 15-19)	-0.3269	0.66	0.0941	0.0005
Age_grp (≥74 Vs 15-19)	-0.6783	0.46	0.1133	<.0001
Age_grp 15-19(Base Case)	-	-	-	-
Gender(1=Male, 0=Female)	0.1612	1.38	0.0233	<.0001

Table 5.18 Partition for the Hosmer and Lemeshow test for IMSD crashes

Partition for the Hosmer and Lemeshow Test					
		Y=1 (IMSD crashes)		Y = 0	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	855	232	224.48	623	630.52
2	854	270	295.32	584	558.68
3	858	351	338.36	507	519.64
4	854	353	369.1	501	484.9
5	856	402	399.73	454	456.27
6	856	437	429.18	419	426.82
7	854	460	458.42	394	395.58
8	854	498	489.35	356	364.65
9	856	548	531.36	308	324.64
10	843	576	591.71	267	251.29

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
9.1288	8	0.3315

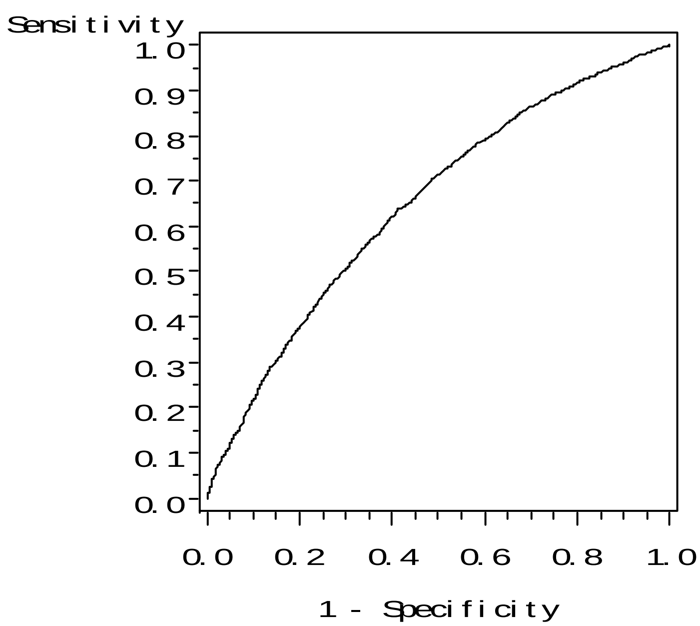


Figure 5.5 ROC curve for IMSD crashes.

5.6.2 Logistic regression models for IMOD crashes

To estimate the conditional probability of IMOD crashes (probability of occurrence of an IMOD crash given a crash has occurred), the second binary logistic regression model is developed. The response variable Y=1 for crashes in the dataset that belong to IMOD category (i.e., for left-turn and head-on crashes) and Y=0 for other crashes. The database comprises of 24% IMOD (Y=1) crashes against 76% other two crash types (Y=0).

The coefficients of the logistic regression model are presented in Table 5.19 and the Goodness of fit statistics is provided in Table 5.20. Figure 5.6 depicts the ROC curve for the model indicating the accuracy of the model in terms of separating IMOD crashes from the other two types of crashes. A detailed discussion on these results is provided in Section 5.6.4 onwards, after the depiction of all three logistic regression models.

Table 5.19 IMOD crash type model (Y=1 if crash type is IMOD and Y=0 for other two crash types)

Parameter	Estimate	Odds Ratio	Standard Error	Pr > ChiSq (p-value)
Intercept	-0.8258		0.1263	<.0001
Light (1= Daylight, 0=Dark)	-0.1888	0.685	0.0283	<.0001
Surface(1= Not Dry ,0= Dry)	-0.2527	0.603	0.0401	<.0001
Intersection Type (1=Four-legged two-way intersections , 0= Four-legged two-way intersections and T-Intersections)	0.0907	1.199	0.0646	0.012
Total no. of through lanes at an intersection	-0.0871	0.917	0.0143	<.0001
Speed (35 mph Vs ≤30 mph)	-0.0352	1.180	0.0704	0.0133
Speed (40 mph Vs ≤30 mph)	0.2012	1.494	0.067	0.5598
Speed (45 mph Vs ≤30 mph)	0.0119	1.237	0.0525	<.0001
Speed (>45 mph Vs ≤30 mph)	0.0227	1.250	0.0443	0.9182
Speed ≤30 mph (Base case)	-	-	-	-
Alc_drugs (1=In use, 0=Not in use)	-0.2876	0.563	0.0502	<.0001
Age_grp (20-24 Vs 15-19)	-0.2249	0.833	0.0645	0.0005
Age_grp (25-34 Vs 15-19)	-0.3244	0.755	0.0578	<.0001
Age_grp (35-44 Vs 15-19)	-0.2712	0.796	0.0602	<.0001
Age_grp (45-54 Vs 15-19)	-0.1519	0.896	0.0713	0.0329
Age_grp (55-64 Vs 15-19)	0.1475	1.209	0.0868	0.0877
Age_grp (65-74 Vs 15-19)	0.2849	1.388	0.097	0.0037
Age_grp (≥74 Vs 15-19)	0.5826	1.869	0.1054	<.0001
Age_grp 15-19 (Base case)	-	-	-	-
Gender (1=Male, 0=Female)	-0.1613	0.724	0.0265	<.0001

Table 5.20 Partition for the Hosmer and Lemeshow test for IMOD crashes

Partition for the Hosmer and Lemeshow Test					
		Y=1 (IMOD crashes)		Y = 0	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	854	99	102.17	755	751.83
2	854	134	134.21	720	719.79
3	855	152	153.12	703	701.88
4	864	187	171.99	677	692.01
5	854	183	185.14	671	668.86
6	837	188	198.59	649	638.41
7	838	217	215.75	621	622.25
8	871	250	246.3	621	624.7
9	852	266	269.69	586	582.31
10	861	339	338.04	522	522.96

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
2.6959	8	0.952

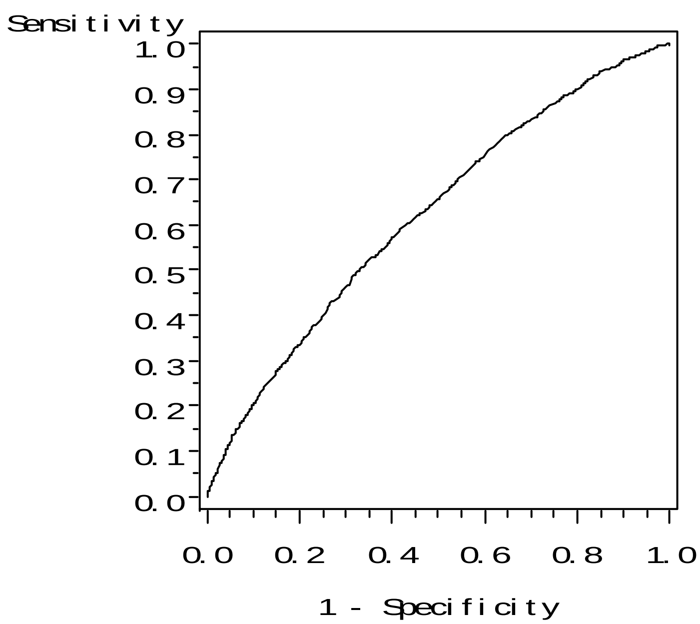


Figure 5.6 ROC curve for IMOD crashes

5.6.3 Logistic regression models for IMPD crashes

A binary logistic regression model is also developed to estimate the conditional probability of IMPD crashes (probability of occurrence of an IMPD crash given a crash has occurred). The crash, intersection, and at-fault driver characteristics for all 8040 crashes are used for as input to develop this model. The response variable Y=1 for crashes in the dataset that belong to IMPD category (i.e., for angle and right-turn crashes) and Y=0 for the crashes that do not belong to IMPD crashes. The database comprises of 28% IMPD (Y=1) crashes against 72% other two crash types.

The coefficients of the logistic regression model are presented in Table 5.21 and the Goodness of fit statistics is provided in Table 5.22. Figure 5.7 depicts the ROC curve for the model indicating the accuracy of the model in terms of separating IMPD crashes from the other two types of crashes. A detailed discussion on these results is provided in Section 5.6.4 onwards, after the depiction of all three logistic regression models.

Table 5.21 IMPD crash type model(Y=1 if crash type is IMPD and Y=0 for other two crash types)

Parameter	**Estimate**	**Odds Ratio**	**Standard Error**	**Pr > ChiSq (p-value)**
Intercept	-0.3067		0.1135	0.0069
Light (1= Daylight, 0= Dark)	-0.0994	0.819	0.0268	0.0002
Weather(1=Not Clear,0=Clear)	0.1089	1.243	0.0275	<.0001
Speed (35 mph Vs ≤30 mph)	0.1762	0.947	0.060	0.0033
Speed (40 mph Vs ≤30 mph)	-0.0463	0.758	0.0501	0.3550
Speed (45 mph Vs ≤30 mph)	-0.1839	0.660	0.0423	<0.0010
Speed (>45 mph Vs ≤30 mph)	-0.1769	0.665	0.0797	0.0263
Speed ≤30 mph (Base case)	-	-	-	-
Intersection Type (1=Four-legged two-way intersections, 0=T-Intersections)	0.1691	1.402	0.0481	0.0004
Intersection Type (1= Four-legged one-way intersections, 0=T-Intersections)	0.4593	2.506	0.0685	<.0001
Total no. of through lanes at an intersection	-0.0353	0.965	0.0142	0.0131
TotExLTLsMN (Total no. of exclusive left-turn lanes at an intersection)	-0.085	0.919	0.0252	0.0007
Alc_drugs (1=In use, 0=Not in use)	-0.1942	0.678	0.0451	<.0001
Age_grp (25-34 Vs 15-24)	-0.1216	0.924	0.054	0.0242
Age_grp (35-34 Vs 15-24)	-0.1281	0.918	0.0567	0.024
Age_grp (45-34 Vs 15-24)	-0.0628	0.98	0.0672	0.3501
Age_grp (55-34 Vs 15-24)	0.1003	1.154	0.0835	0.2301
Age_grp (65-34 Vs 15-24)	0.1202	1.177	0.0957	0.209
Age_grp (≥74 Vs 15-24)	0.1349	1.195	0.1079	0.2111
Age_grp 15-24 (Base case)*	-	-	-	-
Gender (1=Male, 0=Female)	-0.0458	0.912	0.0253	0.0698

* Note that the base age group is 15-24 instead of 15-19 since there was no significant difference between age groups 15-19 and 19-24.

Table 5.22 Partition for the Hosmer and Lemeshow Test for IMPD crashes

Partition for the Hosmer and Lemeshow Test					
		Y=1 (IMPD crashes)		Y = 0	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	853	149	150.62	704	702.38
2	853	183	178.83	670	674.17
3	855	192	195.89	663	659.11
4	853	201	209.09	652	643.91
5	855	225	223.21	630	631.79
6	856	231	237.38	625	618.62
7	854	261	252.39	593	601.61
8	855	277	271.86	578	583.14
9	854	306	299.94	548	554.06
10	852	373	378.79	479	473.21

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
1.8228	8	0.986

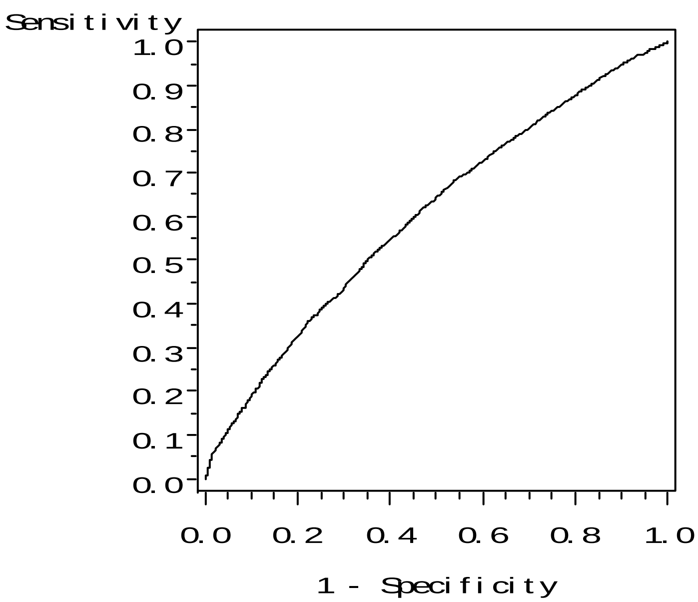


Figure 5.7 ROC curve for IMPD crash model

5.6.4 Brief summary of results

The results of crash type models for signalized intersections, provided in Tables 5.17, 5.19 and 5.21, include odds ratio for input parameters. Odds ratio is a measures of the odds of outcome (which is probability of crash occurrence corresponding to that vehicle movement) being increased if the value of that variable is subjected to a unit change. It is therefore a good indicator of the strength of that independent variable towards the conditional probability of the corresponding crash type. It could be inferred that the closer the odds ratio to 1.0, the less significant the variable, with odds ratio=1.0 representing full statistical independence (model coefficient=0). A summary of odds ratio for each independent variable in the three crash type models is provided in Table 5.23.

Table 5.23 Odds ratios across different levels of each variable for each crash type model

Variable	Odds ratio		
	IMSD Crashes	**IMOD Crashes**	**IMPD Crashes**
Light (1=Daylight, 0=Dark)	1.58	0.685	0.819
Weather (1=Not clear, 0=Clear)	1.19		1.243
Surface (1=Not Dry, 0=Dry)	1.37	0.603	
Intersection Type (1=Four legged Two-way intersections, 0= Four legged One-way intersections)	1.46	1.199*	1.402**
Intersection Type (1=T-Intersections, 0= Four legged One-way intersections)	2.05	-	2.506***
Total no. of through lanes	1.10	0.917	0.965
Total no. of exclusive left-turn lanes on minor road	1.07		0.919
Speed (35mph Vs ≤30 mph)	0.96	1.180	0.947
Speed (40mph Vs ≤30 mph)	0.98	1.494	0.758
Speed (45mph Vs ≤30 mph)	1.27	1.237	0.660
Speed (50mph Vs ≤30 mph)	1.26	1.250	0.665
Alc_drugs (1=In use, 0=Not in use)	2.01	0.563	0.678
Age_grp (20-24 Vs 15-19)	1.14	0.833	
Age_grp (25-34 Vs 15-19)	1.3	0.755	0.924
Age_grp (35-44 Vs 15-19)	1.26	0.796	0.918
Age_grp (45-54 Vs 15-19)	1.1	0.896	0.98
Age_grp (55-64 Vs 15-19)	0.75	1.209	1.154
Age_grp (65-74 Vs 15-19)	0.66	1.388	1.177
Age_grp (≥74 Vs 15-19)	0.46	1.869	1.195
Gender(1=Male, 0=Female)	1.38	0.724	0.912

* Intersection type (1=Four legged Two-way intersections, 0= Four legged One-way intersections and T- intersections)
** Intersection type (1=Four legged Two-way intersections, 0= T- intersections)
*** Intersection type (1=Four legged one-way intersections, 0= T- intersections)

Table 5.24 depicts the significant level for the model parameters in the three models. Drivers' race and total number of exclusive left-turning lanes on the major road were found to be insignificant in all the models. Total number of exclusive left-turns on minor road was found to be significant for IMSD and IMPD crashes. Age-group of the at-fault drivers was found significant at 6 degrees of freedom and 0.10 significant level for IMPD crashes but for other two crash types age groups was significant at 7 degrees of freedom and at 0.05 significant level.

Table 5.24 Summary of significant variables for each initial movement crash type model

Variable	IMSD Crashes		IMPD Crashes		IMOD Crashes	
	Degree of Freedom	**Pr > ChiSq (p-value)**	**Degree of Freedom**	**Pr > ChiSq (p-value)**	**Degree of Freedom**	**Pr > ChiSq (p-value)**
Lighting Conditions	1	<.0001	1	0.0002	1	<.0001
Weather Conditions	1	0.0039	1	<.0001		
Surface Conditions	1	<.0001			1	<.0001
Speed Limit on Major Roads	4	<.0001	4	<.0001	4	0.0007
Four-legged two-way intersections	1	0.0009	1	0.0004	1	0.0215
Four-legged one-way intersections			1	<.0001	1	0.012
T-Intersections	1	<.0001				
Total no. of through lanes at an intersection	1	<.0001	1	0.0131	1	<.0001
Total no. of exclusive left-turn lanes on minor road	1	0.006	1	0.0007		
Alcohol/drugs use	1	<.0001	1	<.0001	1	<.0001
Age group	7	<.0001	6	0.0797	7	<.0001
Gender	1	<.0001	1	0.0698	1	<.0001

5.6.4.1 Goodness of fit

Goodness-of-fit statistics examine the difference between the observed frequencies and the expected frequencies. The Hosmer-Lemeshow test is performed by dividing the predicted probabilities into deciles (10 groups based on percentile ranks) and then computing a Pearson chi-square that compares the predicted to the observed frequencies (in a 2*10 table). Large values of χ_{HL}^{2} (and small *p*-values) indicate a lack of fit of the model. As we can see from the Tables 5.18, 5.20, and 5.22 the χ_{HL}^{2} and p-values indicates good fit of all three models.

5.6.4.2 Prediction accuracy

Prediction accuracy can be estimated using 2*2 classification tables. Limitation to the classification table is that they ignore actual predicted probabilities and instead use dichotomized predictions based on a cutoff (e.g., 0.5). For instance, in binary logistic regression, the classification table does not reveal how close to 1.0 the correct predictions were nor how close to 0.0 the errors were. A model in which the predictions, correct or not, were mostly close to the 0.50 cutoff does not have as good a fit as a model where the predicted scores cluster either near 1.0 or 0.0. Hence, another model prediction measure receiver operating characteristic (ROC) curve is used to measure the prediction accuracy.

ROC curve (receiver operating characteristic) originates from signal detection theory that shows how the receiver operates the existence of signal in the presence of noise. It plots the probability of detecting true signal (sensitivity) and the false signal (1-

specificity) for an entire range of possible thresholds. The area under the ROC curve, which ranges from zero to one, provides a measure of the model's ability to discriminate between those subjects who experience the event versus those who do not. Sensitivity is a ratio consisting of the number of correctly classified events over the total number of events. Specificity is a ratio consisting of the number of correctly classified nonevents over the total number of nonevents.

SAS (SAS Institute, 2001) software is used to generate the ROC curves for the three crash type models. The areas under ROC curve are 0.642, 0.621 and 0.602 for IMSD crashes, IMOD crashes and IMPD crashes respectively. Figures from 5.5 through 5.7 show ROC curves for the three models.

According to Hanley and McNeil (1982), if the area under the ROC curve is between 0.5 to 0.75, the model prediction may be considered fair. Hence, it may be argued that the binary logistic regression models developed here perform fairly well.

5.6.5 Discussion of Results

5.6.5.1 Road environment conditions

From Tables 5.3 and 5.4 it can be seen that the effects of lighting conditions were not uniform over the crash types. Tables 5.23 shows that the odds ratio corresponding to lighting condition is greater than one for IMSD crashes while it is less than one for IMPD and IMOD crashes. In other words, IMOD and IMPD crashes are more likely to happen in dark driving condition possibly because speed of the vehicle in the opposite direction could be difficult to judge.

Contingency table analysis is also used to assess the relationship between lighting conditions and the driver age groups. It may be observed in Table 5.25 that in dark conditions there is an almost consistent decrease (column percentages) in drivers being at-fault in crashes as driver becomes older. This implies that older drivers tend to be less at-fault in dark driving conditions.

Table 5.25 Lighting Conditions and Driver age group contingency table

	Driver Age Group								**Total**
	15-19	**20-24**	**24-34**	**35-44**	**44-54**	**55-64**	**65-74**	**74+**	
Dark	443 5.19 15.62 37.07	531 6.22 18.72 38.87	689 8.07 24.29 35.42	534 6.25 18.82 31.45	342 4 12.05 33.11	157 1.84 5.53 27.64	90 1.05 3.17 21.58	51 0.6 1.8 16.04	2837 33.22
Day light	752 8.81 13.19 62.93	835 9.78 14.64 61.13	1256 14.71 22.02 64.58	1164 13.63 20.41 68.55	691 8.09 12.12 66.89	411 4.81 7.21 72.36	327 3.83 5.73 78.42	267 3.13 4.68 83.96	5703 66.78
Total	1195 13.99	1366 16	1945 22.78	1698 19.88	1033 12.1	568 6.65	417 4.88	318 3.72	8540 100

Statistic	DF	Value	p-value
Chi-Square	7	110.0755	<.0001
Likelihood Ratio Chi-Square	7	117.3167	<.0001
Contingency Coefficient		0.1128	

A contingency table analysis is performed to investigate relationship between driver age groups and road surface conditions. It can be seen from Table 5.26 that teen drivers tend to be at-fault in more crashes when surface is not dry and drivers older than 74 are least at-fault when surface is not dry.

Table 5.26 Surface conditions and Driver age group contingency table

Surface	**Driver Age Group**								**Total**
	15-19	**20-24**	**24-34**	**35-44**	**44-54**	**55-64**	**65-74**	**74+**	
Not Dry	223	211	303	262	147	70	54	31	1301
	2.61	2.47	3.55	3.07	1.72	0.82	0.63	0.36	15.23
	17.14	16.22	23.29	20.14	11.3	5.38	4.15	2.38	
	18.66	15.45	15.58	15.43	14.23	12.32	12.95	9.75	
Dry	972	1155	1642	1436	886	498	363	287	7239
	11.38	13.52	19.23	16.81	10.37	5.83	4.25	3.36	84.77
	13.43	15.96	22.68	19.84	12.24	6.88	5.01	3.96	
	81.34	84.55	84.42	84.57	85.77	87.68	87.05	90.25	
Total	1195	1366	1945	1698	1033	568	417	318	8540
	13.99	16	22.78	19.88	12.1	6.65	4.88	3.72	100

Statistic	DF	Value	p-value
Chi-Square	5	18.6442	0.0022
Likelihood Ratio Chi-Square	5	19.3288	0.0017
Contingency Coefficient		0.0453	

5.6.5.2 Intersection related factors

i) Intersection type

From the preliminary analysis it was shown that different intersections types have different predominant crash types. It was also recommended that intersection types should be used in the intersection safety analysis. According to the models presented herein (refer Table 5.23 for comparison) the chances of driver being at-fault and end up in IMSD crashes are higher at T-Intersection followed by four-legged two-way intersections compared to four-legged one-way intersections. At Four-legged two-way intersections more drivers are at-fault in IMOD crashes while at four-legged one way intersection drivers are at-fault in IMPD crashes.

In a separate study conducted by Salman and Al-Maita (1995) traffic conflicts and crashes were correlated at three-leg unsignalized intersection in Jordan. Their study concluded that traffic conflicts and collisions are correlated. Therefore, to investigate the possible reasons for the results obtained from the logistic regression models for intersection types, the conflict points at different groups of intersections are analyzed. It is known that the collisions associated with merging/diverging movements are rear-end crashes and sideswipe collisions, which are IMSD crashes while crossing conflicts could be associated with IMOD and IMPD crashes (right-angle, left-turn crashes).

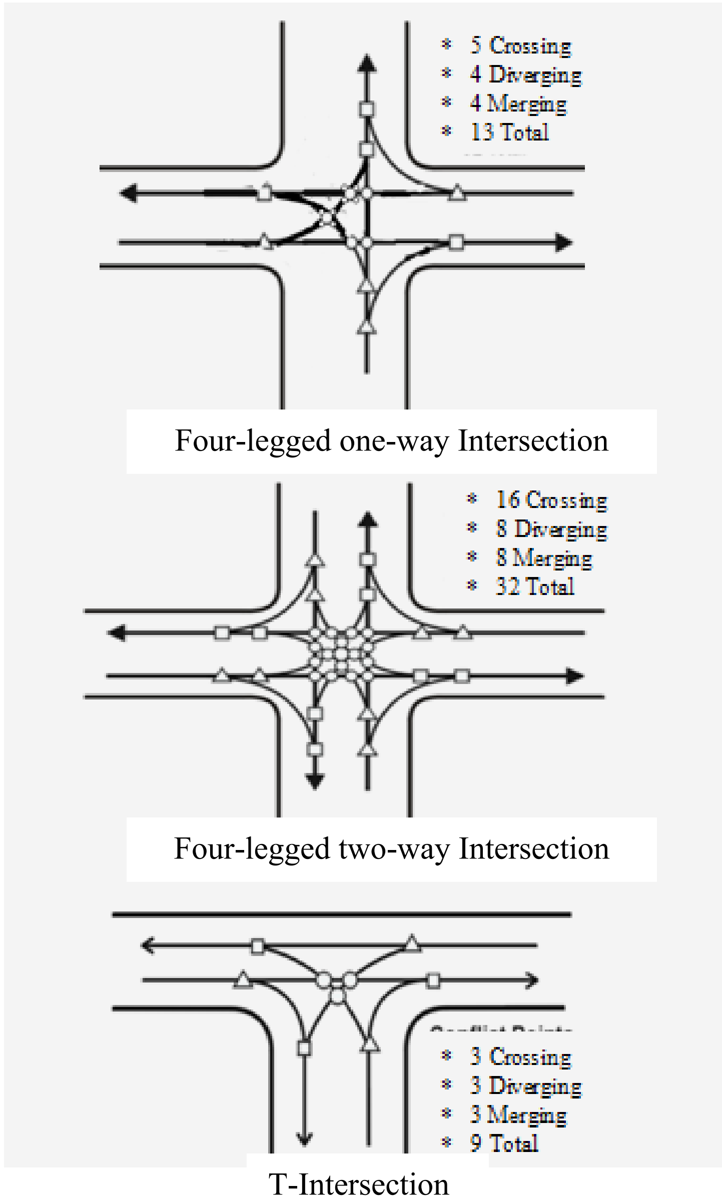


Figure 5.8 Intersection conflicts for three major intersection types.

Table 5.27 Merging and Diverging conflicts at three major intersection types

Conflicts	Four-Legged Two-way Intersection	Percent of Total Conflicts	Four-Legged one-way Intersections	Percent of Total Conflicts	T-Intersections	Percent of Total Conflicts
Crossing	16	50%	5	38%	3	33%
Merging + Diverging	16	50%	8	62%	6	67%
Total	32	100%	13	100%	9	100%

From Figure 5.8 and Table 5.27 it can be seen that T-Intersections have high percent of merging and diverging conflicts compared to four-legged two-way intersections. Hence, it is reasonable that IMSD crashes are more frequent at T-intersections.

Crossing conflicts are highest (50%) at four-legged two-way intersections compared to other two intersection types; this could be the potential reason for the drivers at these intersections being at-fault in IMOD crashes (consisting of predominantly left-turn crashes; Figure 4.1). On the other hand, even though the percentage of crossing conflicts are less for four-legged one-way intersections, it was found in the preliminary analysis (Section 4.3.1.4) that the percentage of "disregarded traffic signal" as a contributing cause was relatively higher (15.5%) at these intersections than the other intersection types (8.1% at the four-legged two-way intersections and 6.94% at the T-intersections). Also, there is a possibility that drivers make improper turns to one-way street by misjudging the direction of travel. Therefore, we can speculate that higher rate of drivers disobeying traffic signals renders higher chances of crashes between vehicles traveling in different directions (i.e., IMPD crashes).

ii) Number of lanes at an intersection

From Tables 5.24 it can be seen that total number of through lanes at an intersection found to be significant for all crash types. Total number of exclusive left turn lanes on minor road is found to be significant for the IMSD and IMPD crashes. The number of right-turn lanes, number of exclusive left-turn lanes on major road were found to be insignificant for all the three crash types.

Tables 5.23 shows that the odds ratio for total number of through lanes and total number of exclusive left turning lanes on minor road is greater than one for the IMSD crashes and less than one for IMPD crashes. These results indicate that as the number of lanes increases (through and or left turning lanes on minor road at an intersection) there will be increase in IMSD crashes. But increase in number of lanes (through and or left turning lanes on minor road at an intersection) has negative effect on IMPD crashes.

iii) Speed limit on major road

Results do not clearly indicate how certain speed limits will affect the crash occurrence of a particular type. However, Table 5.23 (based on odds ratio) does indicate that more drivers tend to be at-fault resulting in IMSD crashes as speed limit increases. It also indicates that drivers tend to be at-fault in IMPD crashes at intersections with low speed limits on major roads.

5.6.5.3 Driver related factors

Numerous studies have addressed the relationship between driver-age and crash involvement (Evans 1991; Massie Campbell 1993). Most of these studies have concluded that there exists a strong relationship between the age of the driver and crash involvement. Drivers' capabilities, experience, and perceptions vary substantially by drivers' age.

It can be seen from the Tables 5.23 that driver being at-fault in IMOD and IMPD crashes increases as drivers grows older. On the other hand, older drivers being at-fault in IMSD crashes is lesser compared to other age-groups. Misjudgment of gaps between vehicles, speed of on coming vehicle, decline is perception abilities could be the possible reasons for higher odds for these drivers being at-fault in IMOD and IMPD crashes. The results shows that female drivers are more likely to be at-fault in IMPD and IMOD crashes while male drivers are more likely to be at-fault in IMSD crashes (Table 5.23). From Table 5.23 it is also evident that drug/alcohol usage is positively associated with IMSD crashes (greater than 1 odds ratio).

5.7 Summary

This chapter has attempted to overcome limitations in the past studies discussed in the chapter 2 through detailed analysis based on the combined dataset assembled in Chapter 4. A signification portion of the chapter is devoted to the estimation of conditional probability of the crash occurrence of a specific type (given a crash has occurred). The type of crash is related to the relative direction of initial movement of the vehicles involved in the crash. Three such crash types are identified namely, IMSD

(Initial movement in same direction), IMPD (Initial movement in perpendicular direction), and IMOD (Initial movement in opposite direction) crashes. The geometric, driver and environmental characteristics are used as inputs to the three binary logistic regression models. This present analysis also introduces two variables (number of through lanes and intersection type) that have never been used before for such analysis.

First, 8761 crashes are used to determine significant variables affecting crash types using contingency analysis. Then, characteristics of at-fault drivers (belonging to 8540 crashes where such drivers were uniquely identified) are used to determine driver related significant variables. Finally, these two datasets are merged together to estimate the conditional probability of crashes belonging to each of three categories using binary logistic regression.

This analysis has found that, IMOD and IMPD crashes are more likely to occur in dark driving conditions. Results also indicate that the odds of drivers being at-fault in IMOD crashes and IMPD crashes increase as drivers become older. Frequency of older drivers being at-fault in IMSD crashes is very less. It might be due to the fact that due to their driving habits older drivers are less exposed to (congested) conditions when through-movement crashes are more frequent. It was also found that intersections with higher major road speed limits tend to experience more IMSD crashes.

Female drivers are more frequently at-fault in IMPD/IMOD crashes while male drivers are more frequently at-fault in IMSD crashes. Identification of these interesting combined effects of driver demographics and roadway/environmental factors on crashes associated with various maneuvers/intersection types is the most significant contribution of this study.

6 CONCLUSIVE REMARKS

6.1 Summary and Conclusions

This research delves into the crash patterns at signalized intersections. Statistics from various sources indicate that intersection-related crashes make up a very high percentage of the total number of crashes in the roadway system. Intersections can be made safer by analyzing different variables that affect crash occurrences and then controlling these variables during the design of the intersections such that they are less prone to crashes.

Before proceeding to analysis, previous studies related to intersection crashes and methodologies adopted by them were reviewed (Chapter 2). After careful review of the past studies, contingency table analysis and Binary logistic regression models were chosen for analysis. It was found that most of the past studies have not included intersection type in the intersection safety analysis. It is also found that joint contribution of driver, intersection and environment related independent variables in intersection analysis demands further attention. Present work attempts to overcome these limitations in past studies.

Modeling methodologies used in this work are discussed in Chapter 3. Contingency table analysis is used to make preliminary assessment of the factors affecting three major crash types (IMSD crashes, IMOD crashes, and IMPD crashes categorized based on the relative direction of initial movement of the involved vehicles) at an intersection. Binary logistic regression models are used to test the joint significance of variables found significant in contingency analysis. Hosmer-Lemeshow goodness-of-fit statistic and ROC curve are used for assessment of the model performance with respect to model-fit and classification accuracy.

Crash data obtained from six jurisdictions namely, Miami-Dade, Brevard, Seminole, Orange, City of Orlando, and Hillsborough are used in the first phase of the analysis. Inconsistent way of identifying intersection types are overcame in Chapter 4 by identifying three major intersection types – 1) four-legged two-way intersections; 2) four-legged one-way intersections; and 3) T-intersections (i.e. the three-legged intersections). This chapter also explored variation of crash frequency and severity with the size and type of intersections. The results showed that the expected crash frequency (expressed as the average number of crashes per intersection per year) generally increased as the total number of lanes increased at all types of intersections. However, the rates of increase were different - crash frequency increased with the size of intersections at higher rate for the four-legged two-way intersections than those for the other intersection types. The crash data obtained from six jurisdictions were expanded to include the at-fault driver characteristics for each crash. This is performed by linking the crash data to DHSMV file by crash identification number.

Chapter 5 dealt with two objectives. First, contingency analysis is performed to identify the significant variables individually affecting three major crash categories (i.e., IMSD crashes, IMOD crashes, and IMPD crashes). The dataset consisting of at-fault drivers from 8540 crashes, in which a unique driver was found to be at-fault, was analyzed along with the intersection characteristics. Following the contingency table based explorations, three binary logistic regression models, one each for every group of crashes, were developed to understand the joint contributions of intersection and driver related factors. The target or the dependent variable for the IMSD crash models was y, which was equal to 1 for IMSD crashes and 0 otherwise. Similarly, the dependent variable for the logistic regression model corresponding to IMOD crash model was equal to 1 for IMOD crashes and 0 otherwise.

From the results of binary logistic regression analysis it is observed that, IMOD and IMPD crashes occur more often in dark driving conditions. Drivers belonging to all groups are likely to make mistakes and 'cause' an IMSD crash in non-dry pavement surface and/or adverse weather conditions. Older drivers are less likely to make mistakes that result in IMSD crashes. For drivers, the likelihood of causing an IMOD/ IMPD crash due to their errors increases as they become older. At intersections with higher major-road speed limits speeds the chances of drivers being at-fault in IMSD and/or IMOD crashes are higher compared to intersections having lower major-road speed limits.

Among the key demographic factors, it was noticed that female drivers are more frequently at-fault in IMOD and IMPD crashes than their male counterparts. Similarly, male drivers are at-fault more often in IMSD crashes than females. Drivers impaired with alcohol and drugs more frequently at-fault in IMSD and IMOD crashes.

The analysis carried out in this thesis provides interesting insight into the joint effects of driver demographics and intersection related factors on crash occurrence of different types (categorized based on movement directions). The conclusions drawn from this study also provide interesting ideas for future research that are summarized in the next and final section.

Further research is required for IMPD and IMOD crashes in the area of combining left-turn crashes with head-on (IMOD crashes) and combining of right-turn crashes with angles crashes (IMPD crashes).

6.2 Future Scope

One of the finding from this study is that number of through lanes at an intersection is a surrogate measure for AADT. The total number of lanes was also associated with the average annual crash frequency. It will be interesting to see how parameters like the signal timing, etc. related to crash frequency. Such information can be used to further enhance the analysis presented here.

Based on the findings from this research, driver training programs may be developed that specialize in training certain groups of drivers in ‘key’ and ‘dangerous’ intersection maneuvers that are identified as more risk prone for that respective group. Prior to such programs further analysis should be done using the exposure of the drivers to these intersection maneuvers. The information on not-at-fault drivers for the crashes may be utilized to determine the exposure based on the principal of Quasi-induced exposure.

REFERENCES

Abdel-Aty et al. “Analysis of the types of crashes at signalized intersections using complete crash data and tree-based regression”, *Proceedings of the TRB 84th Annual Meeting* Washington, DC, January 9–13, 2005.

Abdel-Aty, M., Chen, C., Radwan, E., Brady. “Analysis of accident-involvement trends by drivers age in Florida”. ITE Journal on the Web (Feb. 1999), pp. 69–74.

Abdel-Aty, M., Chen, C., Radwan, E. “Using conditional probabilities to explore the driver age effect in accidents”. ASCE Journal of Transportation Engineering 125(6), 1999b.

Agresti, A. (1996). *An introduction to categorical data analysis. Wiley, New York.*

Bauer, K. M. and D. W. Harwood. “Statistical Models of At-grade Intersection Accidents”. Report FHWA-RD-96-125, Washington, DC, USA: Federal Highway Administration, 1996.

Evans, L., “Older-driver risks to themselves and to other road users”, Transportation Research Record 1325, 1991.

Foody, T. J., and W. C. Richardson. "Evaluation of Left Turn Lanes as a Traffic Control Device," Ohio Department of Transportation, 1973.

Gluck, J., H.S. Levinson, and V. Stover. Impacts of Access Management Techniques, NCHRP Report 420, Transportation Research Board, 1999

Hakamies-Blomquist, L. E. (1993). Fatal accidents of older drivers. Accident nalysis and Prevention, 25, 19-27.

Harwood, D.W., F.M. Council, E. Hauer, W.E. Hughes, and A. Vogt. “Prediction of the Expected Safety Performance of Rural Two-Lane Highways”, Report No. FHWA-RD-99-207, Federal Highway Administration, December 2000.

Hanley J. A., and B. J. McNeil, 1982: “The meaning and use of the area under a receiver operating characteristic (ROC) curve”. Radiology, 143, 2936.

Harwood, D. W., M. T. Pietrucha, M. D. Wooldridge, R. E. Brydia, and K. Fitzpatrick. Median Intersection Design, NCHRP Report 375, Transportation Research Board, 1995.

Hill, S. E. and J. K. Lindly. “Red-Light Running Prediction and Analysis”. UTCA Report 02112, Tuscaloosa, Alabama, USA: University Transportation Center for Alabama, 2003.

Hosmer, D.W., Lemeshow, S., 1989. Applied Logistic Regression. Wiley, New York.

Huang, H., C. Zegeer, R. Nassi and B. Fairfa. “The Effects of Innovative Pedestrian Signals at Unsignalized Intersections: A Tale of Three Treatments”. Report FHWA-RD-00-098, Washington, DC, USA: Federal Highway Administration, 2000.

Hwang, K. and Wei, K. ”An approach based crash analysis and appraisal model for safety design and Management”, Hournal of the Eastern Asia Society for Transportation Studies, vol.6,pp.3713-3723,2005

Jovanis, P. P., and Chang, H. L., "Modeling the Relationship of Accidents to Miles Traveled", Transportation Research Record, 1068, Transportation Research Board, National Research Council, Washington, D.C., pp. 42-51, 1986.

Kendall, M. and Stuart, A. (1979). The advanced theory of statistics, volume 2. Griffin, London.

Lau, M. Y., and A. D. May. “Accident Prediction Model Development: Signalized Intersections”, Research Report No. UCB-ITS-RR-88-7, Institute of Transportation Studies, University of California-Berkeley, December 1988.

Le,C.T (1998), *Applied categorical data analysis*. Wiley, New York.

Matthaws, M.L., and Moran, A.R., “Age differences in male driver’ perception of accident risk: The role of perceived driving ability”, Accident Analysis and Prevention, Vol.18. No.4,1986.

McCoy, P. T., and M. S. Malone. "Safety Effects of Left-Turn Lanes on Urban Four-Lane Roadways," Transportation Research Record 1239, Transportation Research Board, 1989.

Mannering, F.L., “Male/Female driver characteristics and accident risk: Some new evidence”, Accident Analysis and Prediction Vol. 285 Iss. 1 pp. 77-84, 1991.

Massie, D.L. & Campbell, K.L. (1993). “Analysis of Accident Rates by Age, Gender, and Time of Day Based on the 1990 Nationwide Personal Transportation Survey.” (Report No. UMTRI-93-7). Ann Arbor,MI: University of Michigan Transportation Research Institute.

McKelvey, F. X., & Stamatiadis, N. (1989). “Highway accident patterns in Michigan related to older drivers”. Transportation Research Record, 1210, 53– 57.

Miaou S., Hu P., Wright T., Rathi A., and Davis S., “Relationship between truck accidents and highway geometric design: A poisson regression approach”, Transportation Research Record1376, 1993.

Miaou, S., 1994. The relationship between truck accidents and geometric design of road section: Poisson versus Negative Binomial regression. Accident Analysis and Prevention 26(4).

Milton J. and Mannering F., “The relationship among highway geometric, traffic-related elements and motor-vehicle accident frequencies”, Presented at 75th Annual Meeting of Transportation Research Record 923, 1983.

Mohamedshah, Y., L. Chen and F. Council. “Association of Selected Intersection Factors with Red-Light Running Crashes.” Presented at the ITE 2000 Annual Meeting and Exhibit, Nashville, Tennessee, August 2000.

Mucsi, K. and A. M. Khan. “Effectiveness of Additional Lanes at Signalized Intersections.” ITE Journal, No. 73, No. 1, January 2003, pp. 26-30.

Nawathe, P. (2005). “Neural Network Trees And Simulation Databases: New Approaches for Signalized Intersection Crash Classification And Prediction”, Master Thesis, University of Central Florida.

Pernia, J., J. Lu, X. Xie, M. Weng and D. Synder. “Development of Models to Quantify the Impacts of Signalization of Intersection Crashes.” Presented at the 81st Annual Meeting of the Transportation Research Board, Washington, D.C., 2002.

Poch, M. and F. Mannering "Negative Binomial Analysis of Intersection Accident Frequencies." Journal of Transportation Engineering Vol. 122 Iss. 2 pp. 105-113 1996.

Porter, B. E. and K. J. England. “Predicting Red-Light Running Behaviour: A Traffic Safety Study in Three Urban Settings.” Journal of Safety Research, Vol. 31, Issue 1, 2000, pp. 1-8.

Ragland R. D and Zabyshny A.A (2003), "Intersection Decision Support Project: Taxonomy of Crossing-Path Crashes at Intersections Using GES 2000 Data" U.C. Berkeley Traffic Safety Center. Paper UCB-TSC-RR-2003-08.

Rumar, K. 1985. “The Role of Perceptual and Cognitive Filters in ObservedBehaviour.” In Human Behaviour and Traffic Safety, L. Evans and R.C. Schwing, eds. New York: Plenum Press.

Sabey, B. (1991). Human factors in accidents. PACTS conference on New insights into driver behavior. Crowthorne, England: Transport and Road Research Laboratory.

Salman. N and Al-Maita. M, (1995). "Safety evaluation at three-leg, unsignalized intersections by traffic conflict technique," Transportation Research Record, Vol. 1485.

Shankar, V., F. Mannering and W. Barfield "Statistical Analysis of Accident Severity on Rural Freeways." Accident Analysis and Prediction Vol. 28 Iss. 3 pp. 391-401 1996.

Stamatiadis, N., & Deacon, J. (1995). “Trends in highway safety: Effects of an aging population on accident propensity.” Accident Analysis and Prevention, 27(4), 443–459.

Stamatiadis, N., Taylor,W. C., & McKelvey, F. X. (1991). “Older drivers and intersection traffic control devices.” Journal of Transportation Engineering, 117(3), 311 –319.

Staplin, L., Lococo, K., Byington, S. Older driver highway design handbook. McLean: Office of Safety and Traffic Operations R&D., 1998.

Tijerina, L., Chovan, J., Pierowicz, J., and Hendricks, D. (1994). “Examination Of Signalized Intersection, Straight Crossing Path Crashes, And Potential IVHS Countermeasures.” (Report No. DOT-HS-808 143). Washington, DC: National Highway Traffic Safety Administration.

Traffic Safety Facts 1999. National Center for Statistic and Analysis. National Highway raffic Safety Administration. Washington, D. C.

Traffic Safety Facts 2004. Federal Highway Administration, Intersection Safety Facts and Statistics (http://safety.fhwa.dot.gov/intersections/inter_facts.htm)

Traffic Safety Fundamentals Handbook, Minnesota Department of Transportation, 2001.

Villalba, J., Kirk, A., & Stamatiadis, N. (2001). Effects of age and cohort on older drivers. Proceedings of the 2001 ATT Conference and Exhibition,Oct. 1-3, Barcelona, Spain, 1, 113– 118.

Vogt, A. and Bared, J. “Accident models for two-lane rural segments and intersections”. Transportation Research Record No. 1635 pp. 18-29. 1998.