Spring 2013

An Investigation of Pavement Distress Variables on Crash Outcomes Using Hierarchical Generalized Linear Regression Modeling

Robert Alan Morgan
Old Dominion University

AN INVESTIGATION OF PAVEMENT DISTRESS VARIABLES ON CRASH OUTCOMES USING HIERARCHICAL GENERALIZED LINEAR REGRESSION MODELING

by

Robert Alan Morgan, P.E.
B.S.E.T. December 1981, Old Dominion University
M.S.C.E. June 1994, Virginia Polytechnic Institute and State University

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

CIVIL AND ENVIRONMENTAL ENGINEERING

OLD DOMINION UNIVERSITY
May 2013

Approved by:

Dr. Asad J. Khattak (Chair)

Dr. Mecit Cetin (Member)

Dr. ManWo Ng (Member)

ABSTRACT

AN INVESTIGATION OF PAVEMENT DISTRESS VARIABLES ON CRASH OUTCOMES USING HIERARCHICAL GENERALIZED LINEAR REGRESSION MODELING

Robert Alan Morgan, P.E.

Old Dominion University, 2013

Chair: Dr. Asad J. Khattak

Previous macro-level studies show that the condition of the pavement is associated with roadway safety. However, this premise has not been examined on a micro-level using detailed pavement distress variables (PDVs). The main focus of this research will be to expand the use of the PDV data to further understand the safety risks associated with the type of pavement and each individual PDV by analyzing the likelihood of having a rear-end and/or injurious crash, given that a crash occurs.

Sample crash data and PDV data from the Commonwealth of Virginia for the years 2007 and 2008 were used to produce a dataset that provides crash and pavement condition, type and ride quality data at the crash site and specific intervals upstream for three types of pavement. Binary logistic regression statistical modeling was used to determine if PDVs have an association with rear-end crashes and crashes with injuries. By investigating this relationship at the crash site and specific intervals upstream of the crash, this study provides valuable insights into the spatial component of pavement safety.

Additionally, there is an a priori reason that the morphology of the built-up environment could influence the severity of an accident. By including social/economic factors and applying hierarchical generalized linear modeling this study will use the hierarchical nature of crash data to examine socio-economic characteristics of the locality where the crash occurs.

The analysis of rear-end and injurious crashes resulted in PDVs that are associated with an increased risk of these types of crashes on each of the three pavement types. While this association is weak for injurious crashes, the results indicate the critical location to be upstream of the crash; for rear-end crashes, the critical location was at the crash site. It was determined the type of pavement is not significant for crashes with injuries, but it is for rear-end crashes.

The results indicate there is little benefit to using HGLM to model crashes with injuries, but the variability in rear-end crashes can be explained by the nested structure. The two socio-economic factors that reduce the odds of a rear-end crash are the average age of the driver and unemployment percentage.

DEDICATION

In memory of my Mother, Agnes M. Morgan for her love and support

and

Father-in-law, Dr. Howard L. Sparks, Ed.D for my inspiration

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Dr. Asad J. Khattak, my advisor and chairman of my committee. His patience and encouragement allowed me to persevere. Every insightful comment or suggestion challenged me to better understand the concepts behind this topic and the paradigms of quality research, and for that I am truly thankful.

I thank also Dr. Mecit Cetin and Dr. ManWo Ng for being on my dissertation committee. They had many helpful comments and brought to my attention key aspects of my research topic which provided richness and substance. My appreciation also goes to Dr. A. Osman Akan my original advisor. He helped me in many ways, but most of all he kept me focused, and was always there to offer support and guidance.

There are many people at the Virginia Department of Transportation I would like to thank, but especially my supervisors Mark Cacamis and Michael Davis. Without their understanding and encouragement, conducting this research would have been much more difficult. To the construction managers Tom Rakowski, Michael Johnson, Mitch Layton and Blaine Tudor that took care of business. To Xin Wang, Tom Tate and Raja Shekharan for helping me obtain the crash and pavement condition data and to the folks at the VDOT Research Library who help me obtain references for the literature review and continued to renew my overdue books.

To my Dad who is the nicest person I know, love, respect, and aspire to be. To my boys Nate and Michael, the two reasons for much of my happiness. But above all, there is no one that I want to thank more than my wife, Amy. Many folks helped make this research less difficult, but she made it possible. She kept me working hard, and each time I got discouraged, she picked me up in her very caring ways. She is the love of my life and best friend, thank you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background

Safety is an important quality of life issue and drivers expect roadways to be safe, but they also want them to be smooth, energy efficient, well drained and quiet. Because there is no absolute safety, modern studies speak in terms of acceptable risks and, more recently, available resources. Finding countermeasures to reduce traffic crashes has always been a primary focus in transportation research and the recent release of the Department of Transportation's Strategic Plan once again puts improving safety as its top priority.

However, second on the list is maintaining and modernizing our current critical infrastructure. With limited transportation funding, states across the nation are struggling to set spending priorities on maintenance projects that primarily deal with pavement preservation and rehabilitation. Collecting data and providing quality information on the condition of our nation's highways is becoming more and more critical to the transportation agency's decision-makers at all levels so that money can be wisely spent. When and how to spend the limited capital investment is becoming the critical question.

In 1965, Congress mandated that the condition of the nation's highways be reported every two years. The Federal Highway Administration's (FHWA) Highway Performance Monitoring System (HPMS) was developed in 1978 as a national highway transportation

system database to replace the uncoordinated annual state data condition assessment reports. Its primary mission is to provide a database and analysis process for assessing and reporting the extent, condition and performance of the nation's highway system (FHWA, 2008). It provided its first report to Congress in 1981. HPMS Reassessment 2010+ program designates improvements to the pavement condition assessment data collection process one of its top priorities.

Along with the pavement condition assessment data, the HPMS provides national-level highway inventory information to national, state, and metropolitan planning organizations and regional agencies. In addition, this program provides information on traffic operations, performance and highway safety statistics. These data are used by transportation agencies to calculate performance measures with two of the most common ones being fatality rates and pavement smoothness. From these various performance measures highway investment needs can be forecasted and the limited capital resources put towards the most efficient projects. It is clear from the emphasis placed on the need to collect quality robust pavement conditions and safety data that these two highway attributes are important in the decision-making process, but there exists very little research that correlates the two.

1.2 Contributing Theory

Roadway crashes are considered unique random events caused by multiple factors many times preceded by a person(s) not being able to properly cope with the environment (Noyce, etal, 2007). While there are numerous factors affecting general road safety, research in roadway safety is categorized into three main categories; vehicle, driver and environment, each with their own set of sub-variables. The complexity of this has been traditionally explained with the Haddon matrix that was developed by Dr. William Haddon in 1970 (See Figure 1).

	Human	Vehicle	Environment	Socio/Economic
Pre-Crash	Poor vision or reaction time, speeding or alcohol	Bad brakes, car lighting or warning systems	Narrow lanes or shoulders, poor pavement condition	Cultural norms permitting speeding, age of driver
Crash	No seat belt worn	Seat belt no working properly, no air bag	Substandard guardrails	No vehicle design regulations for safety
Post-Crash	High susceptibility, alcohol	Poorly designed fuel tanks	Poor emergency response	Poor local hospitals

Figure 1. Haddon's Road Safety Matrix

Pavement surface texture properties contribute to skid resistance which is the friction force developed at the tire-pavement contact point. It is the force that resists skidding and sliding on pavement surfaces, and it is an essential component of traffic safety because it provides the grip that a rubber tire needs to maintain vehicle control and for stopping in the proper distance in case of emergency. Skid resistance has two major

components (adhesion and hysteresis) and is related to the two key properties of asphaltic pavement surfaces, that is microtexture and macrotexture. (Noyce, etal, 2007). These key properties of asphalt can also be related to similar properties of concrete pavements.

In addition, there are two other pavement characteristics that may be less significant to skid resistance but are still key components of the overall quality of the roadway surface; megatexture and roughness (Noyce, etal, 2007). Megatexture is larger irregularities that can result from rutting, potholes, patching, alligator cracking, spalling and other major cracking of asphalt and concrete pavements. These are some of the PDVs that will be explored in this research, and it is the primary hypothesis based on this contributing theory that individual PDVs have an association with crash risk, given that a crash occurs.

Well documented in the literature, surface irregularities have a direct effect on the environmental aspects of roadway safety. This study focused on the safety aspect of the roadway performance, its relationship to the environment in which the roadway crashes can occur and, in particular, the condition and roughness of the riding surface (i.e. pavement). It specifically adresses the performance of the pavement as expressed by a complex set of predictor pavement distress variables (PDVs) and its roughness computed in the International Roughness Index (IRI) from collected wheel path profile data.

The driver's ability to collect information and carry out intended maneuvers is greatly compromised due to the vibrations encountered on rough roadways (TRB, 2009). Studies

have shown that roughness affects safety in many ways, particularly the ability of the driver to steer and brake (Burns, 1981). It is widely understood that severe wash-boarding surfaces and repeated cycling undulations of the surface can potentially shake a loaded truck so much that it looses part, or all, of its load, causing catastrophic multi-vehicle crashes. It has been shown that only seven to ten percent of roadway crashes can be attributed to the road and its environment, but when combined with human error, more than twenty percent of crashes can be ascribed to this combination (TRB, 2009).

Additionally, in a face-to-face interview study, Vanlaar and Yannis asked active licensed car drivers to rate 15 causes of road crashes and concluded drivers expect road surfaces to be in good condition (i.e. taken for granted). Respondents interviewed placed "poorly maintained roads" in the lower left quadrant ranking them low in prevalence and a low risk (Vanlaar and Yannis, 2006). So when the condition or roughness of the roadway violates this driver expectancy, particularly when transitioning from a smooth surface to a rough one, crashes are likely to occur (TRB, 2009).

Therefore, based on the underlying premise that road conditions affect critical reaction times for drivers, it is important to explore where within the footprint of the crash is the condition of the pavement critical. This lead to the second hypothesis for this study: the condition and/or ride quality of the pavement may not be critical at the crash site but a certain distance upstream. By examining individual PDVs at the crash site and specific intervals upstream, this research explores the spatial aspect of this relationship to evaluate

where within the footprint of the crash is the most critical location with respect to pavement condition and/or ride quality.

The theories behind the causal mechanisms of highway crashes have morphed from being strictly random events to the modern principles entrenched in the study of human behavior. Nonetheless, no matter which theory is subscribed to, road planners and designers should design roads in such a way that the highest possible level of safety is inherently built into them and ensure no easily preventable traffic hazards exist before a roadway is open to the public (Elvik and Vaa, 2004). It is incumbent on transportation officials to ensure that highway infrastructure minimizes all the risks of predictable/simple mistakes, and if a crash occurs, the highway should be maintained in such a forgiving condition that serious injuries or death does not occur. Every causal relationship, no matter how insignificant, should be thoroughly examined to ensure the proper, and in these times, most cost effective solutions are pursued.

1.3 Problem Statement

According to the auto insurance industry, 1.2 million people die worldwide each year in automobile crashes. In the United States (U.S.) 33,963 people died last year in car crashes. That is 25% of all accidental deaths per year. Almost 2.7 million injuries were related to the 6.1 million highway crashes reported (Noyce, et al., 2007). While this has decreased substantially in recent years due to the strong emphasis on road safety at the state and federal level, 94 people still die every day in the United States due to car crashes. That alarming indicator equates to 1 person every 15 minutes. Because these

statistics remain at an unacceptable level, traffic safety research must continue searching for ways to improve the safety of our highways.

As mentioned, the link between crashes and pavement performance (condition and ride quality) is rarely considered a primary contributing factor when planning pavement maintenance. However, a recent study found poor road conditions contribute to over half -52.7 percent- of those deaths (Miller and Zaloshnja, 2009). This is greater than speeding, non-use of seat belts or alcohol related vehicle deaths, and in terms of severity, "it is the single most lethal contributing factor" (Miller and Zaloshnja, 2009). With these staggering statistics one can certainly see that poor road conditions are placing a huge burden on society and the quality of life for many individuals. These poor road conditions across the nation are not going unnoticed by the public. A recent poll conducted by the AAA Foundation for Traffic Safety found that while the death rate caused by highway crashes has decline substantially across the Country, many Americans feel less safe driving today than they did five years ago (AAA, 2009). When asked if they think driving feels safer only 12% said yes, and when asked why, approximately 13 % indicated road safety issues. In order to make the nation's highways safe, transportation officials in a position to make critical decisions must be able to identify all the safety risks including those associated with the condition of the pavement at each stage of its life cycle so meaningful and cost effective remedies can be implemented.

It is well documented in the literature that various pavement surface characteristics can influence the risk of crashes occurring. However, most studies to date have been on the

macro-level and focused on specific crash types related to pavement surface characteristics such as texture, skid resistance, rutting, wet and icy conditions, luminance and reflectivity. While most states now spend significant resources (time and money) to collect an enormous amount of pavement condition data in the form of PDVs, this researcher found no studies where it is used for pavement/traffic safety. States use International Roughness Index (IRI) to determine a pavement's ride quality along with a complex set of PDVs to assess the severity of distresses (i.e. condition) in their pavement assets solely for pavement maintenance programming. No examples were found where states use this data to help them better understand the relationship between the condition/ride quality profiles and roadway safety was found in the literature. For example, while road roughness (IRI) was the first surface condition measure to be collected on a routine basis, it was used primarily to model its impact on constant speed and fuel consumption (McLean and Foley, 1998). The sole purpose of establishing this relationship was so highway agencies could translate percent increase in car fuel consumption to a unit increase in IRI. The pavement condition assessment is focused primarily on the smoothness of the ride and the noise level of the pavement for fuel economy and drive quality, but this is only one small part of the driver's expectation equation. Safety is assumed to be only inherent in a certain pavement condition rating. There is no direct consideration given to safety. Attention is needed to better understand the link between these pavement distress elements (PDVs) and safety of the pavement/roadway so transportation officials can use this detailed knowledge for developing better road designs, pavement management strategies and in detecting safety related hot spots.

1.4 Motivation

The poor state of our nation's pavement infrastructure and the current funding limitations state DOTs are witnessing coupled with the federal mandates to develop programs to maintain the safety of the nations rapidly deteriorating roadways motivate this research and adds to the timeliness attribute. It was found in a recent study conduct by the Pew Center on the States and the Rockefeller Foundation that in fiscal year 2010 $131 billion of taxpayer money was spent on transportation without knowing the outcomes/results of this enormous investment. Over the past 15 years, an increasing number of states have dedicated the resources to gather information on the benefits but only 13 states (Virginia being one) have set goals and have the data/tools to measure their performance so the right decisions are made to ensure safety and mobility, but the rest, while they recognize the need to evaluate the return on investment, still do not have the data to make informed decisions on allocated the limited/ prized ever decreasing transportation funds.

Additionally, the PEW study listed safety as fundamental and the first policy goal in measuring performance for the states and infrastructure preservation as the last policy goal. It also recognizes the federal government's mandate for safer roads. The Centers for Disease Control and Prevention reports that motor vehicle-related injuries are the leading cause of death for people ages one to 44, and automobile accidents sent more than two million people to emergency rooms in 2009.103 Transportation-related accidents also generate high economic costs in lost wages, medical bills and traffic delays (PEW, 2011). It is also important to note the substantial improvement in road safety due to the emphasis it has received at both the state and federal level, but there is still a lot of

work that remains particularly in reducing the number of severe injuries and property damage.

With respect to infrastructure preservation, while the majority of the nation's roads are in good or fair condition, almost 25% of the interstates and principal arterials have pavement in poor condition. The study considers this a safety hazard, but as a result puts more emphasis on the costs associated with vehicle operation costs rather than driving on bad roads. There is no link between the data gathered for assessing pavement condition/ride quality and the data gathered for evaluating the safety of roads.

In these times of limited available resources, there is an underlying need for "strategic and data-driven deliberation" (PEW, 2011). The number one policy option listed in the study is the development of processes that make the best use of available data. To ensure they meet this safety mandate while efficiently allocating the limited infrastructure improvement funds, there is a need to arm state transportation agencies with the best available information so they can make critical pavement maintenance programming decisions.

The results found in a survey of the literature are varied and inconclusive. Far more ambiguity than agreement exists when determining the relationship between pavement condition (assessed in a set of distress variables) and ride quality (IRI) and the crash outcomes. We do not know if the condition of the pavement as assessed by states alone, or in combination with other specific variables (e.g. IRI), increases the odds of certain

consequences resulting from a crash. Unfortunately, very little is known about this association and what impacts certain pavement attributes have on crash characteristics (FHWA, 2006). There is no agreement on whether these variables increase or reduce these crash characteristics.

Therefore, the relationship between safety and pavement condition, type and ride quality remains unclear. Exacerbating the issue is the dynamic nature of pavement construction materials and mixture design properties, construction methods and standards, vehicle and tire characteristics, traffic volumes and speeds, etc. (AASHTO, 2007-2008). Lastly, the need for more comprehensive research is suggested in a majority of the previous studies. This research is aimed at completing the driver's expectation equation by determining if the pavement condition and/or ride quality are contributing risk factors in the severity of roadway crashes (i.e., safety). A commensurate objective is to quantify this risk so transportation agencies can determine if countermeasures are worthy of the limited available resources.

1.5 Research Objectives

In spite of their limitations, statistical models have been proven invaluable in estimating the significance of the relationship between the various casual factors mentioned above and vehicle crashes. This research will focus on crash types/severity and measure their significance to determine if these potentially safety critical parameters are substantial.

The primary objective of this study is to expand the use of already collected detailed pavement performance data by linking it to pavement/roadway safety to assist transportation officials in developing strategies for pavement maintenance that focus on reducing the number of pavement-related crashes. To this end, the main focus of this research will be to expand the use of the PDV data to further understand the safety risks associated with the type of pavement and each individual PDV by analyzing the likelihood of having a rear-end and/or injurious crash, given a crash occurs.

This study shall increase the body of knowledge of the spatial and causal mechanism of pavement-related roadway crashes for use by highway agencies in developing pavement management strategies to improve traveler safety. In essence, by conducting this in-depth study to better understand this important relationship, the consequences of crashes may be reduced by changes in how roads are maintained.

With the addition of a spatial component, this research will explore in more detail the affects of the condition of pavement and location of the crash by including socioeconomic factors. This will provide unique and valuable insights so effective site-specific countermeasures may be identified.

As suggested by the Haddon matrix, there is an a-priori reason that the morphology of the built-up environment could affect accident occurrence and wanting to provide additional insights into the nature of the spatial and casual mechanism of various crash characteristics. This study will use multilevel modeling techniques to analyze the

hierarchical nature of crash data and examine the socio-economic characteristics that may influence crash outcomes.

To fully exploit the clustering nature of road crash data and develop a comprehensive predictive model, unique datasets will be developed by combining pavement condition and ride quality (IRI roughness index) along with roadway environmental and local socioeconomic data. This data will be aggregated with available crash data to determine which of these parameters are important in explaining the outcomes of vehicle crashes.

The following are the three main hypotheses for this study along with the objectives to support them:

1. Hypothesis: the condition, type and ride quality of the pavement has an association with rear-end crashes and crashes with injuries.

   Objective: Given a crash occurs, evaluate the relationship between pavement condition, type of pavement and ride quality on crash outcomes by estimating the strength and significance of various pavement condition distress parameters along with the type of pavement and ride quality as expressed by IRI.

2. Hypothesis: The condition and/or ride quality of the pavement may not be critical at the crash site but upstream.

   Objective: To explore the spatial aspect of this relationship, models will be developed to evaluate crash outcomes at the crash site and specific intervals upstream. These

models will be used to evaluate where within the footprint of the crash is the most critical location with respect to pavement condition and/or ride quality.

3. Hypothesis: The socio-economic characteristics of the county where the crash occurrs have a relationship to rear-end crashes and crashes with injuries.
   Objective: In conjunction with goals one and two, uniquely model the hierarchical nature of a crash to determine which regional socio-economic factors are related to traffic safety.

Determining the direct associations of crash risk is very complicated. Many engineering, environmental, and human behavior factors may directly or indirectly impact the outcome of a crash. Therefore, to minimize the impacts of these other factors, the analysis is limited to interstate routes where accurate data are available in Virginia and the roadways have similar/common features. Additionally, crash and pavement condition data for this study was captured from the higher volume roadways where the benefits may be greater.

The three pavement types considered are asphalt concrete (ACP), continuously reinforced concrete (CRCP) and jointed reinforced concrete (JRCP). ACP includes bituminous concrete over jointed reinforced concrete (BOJ) and bituminous concrete over continuously reinforced concrete (BOC). With concrete and asphalt pavement materials varying only slightly across the Mid-Atlantic, this study can be generalized for other Departments of Transportation in this region and the methodology applicable for similar

studies across the country. In order to account for the randomness of the occurrence of vehicle crashes, stochastic regression models were used to determine these relationships.

By providing information on the safety benefits of asphalt and concrete pavements the results of this study should enhance the decision-making process and help Departments of Transportation meet their goals for safer highways. Additionally, as the pavements age and safety decreases appropriate countermeasures could be employed in a timely manner to optimize the safety of the roadways throughout the pavement's life cycle.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Review of Pertinent Literature

A synthesis of the conclusions from previous studies within my interest area was performed to determine the gaps in understanding the relationship pavement condition and ride quality have on roadway safety. The literature review is summarized in five categories: various roadway surface distress characteristics (road condition) as factor, pavement ride quality as a factor, previous research on pavement type and safety, programming safety into pavement maintenance, and use of Hierarchical Linear Modeling in roadway safety studies.

2.1a Previous Research on Pavement Condition and Safety

There is a vast amount of literature analyzing the relationship and quantifying the importance of various roadway surface discontinuities on road safety. In particular, skid friction related to the texture for different types and its relationship to crashes on wet pavement has received much attention throughout the national and international safety research community since transportation officials first recognized this link back in the 1920s. The current perspective on pavement surface characteristics and their impact on safety focus on pavement texture as it relates to surface friction. Research indicates that approximately 14 percent of all crashes occur in wet weather, and 70 percent of these crashes can be attributed to the pavement texture/friction (ACPA, 2007). Roughness as measured in IRI is used solely for construction acceptance and system monitoring. From

a study conducted on mostly rural interstates and parkway roads in Kentucky (AASHTO, 2008-2007) wet crash rates decreased with increase in pavement skid/friction number (SN).

The Transportation Research Board Circular Number E-C134 reviewed and summarized the literature on this subject and placed the issues into four categories: tire-pavement available friction, potholes and roughness, wet pavement, and influence of pavement edge drop-offs. They, too, recognized there are limited construction and/or maintenance funds which precipitates the need to prioritize treatment projects for these conditions since they can never be completely eliminated on all roadways all of the time. In the chapter written by Council, et al., they make a direct point that it is "wishful thinking" to assume traffic collisions can be accurately predicted on the basis of one antecedent condition (TRB, 2009). For example, one cannot conclude that crashes that occur due to the lack of surface friction are solely attributed to the skid number assigned to the pavement. Rather, the tacit assumption uses wet surface crashes as the surrogate for inadequate surface friction because it is rare that crash reports list this poor surface friction as the primary cause, but the literature is quick to point out it is very difficult to use skid resistance as the sole predictor for wet condition crashes. Other factors, such as road geometry, traffic operations (volume and speed), weather conditions, and driver/vehicle characteristics must also be considered together with friction data when evaluating the safety of a particular segment of roadway (AASHTO, 2007-2008). The Circular concluded the consensus among transportation officials is that discontinuities in the roadway surface may precipitate or aggravate crashes the importance of this

relationship (i.e. quantifying the effect) is not clear for many of the discontinuities and warrants further investigation.

Similar to the NCHRP Circular, a report prepared by the transportation research arm of Australia the ARRB Transport Research (McLean and Foley, 1998) reviewed the effects of road surface type and condition on safety, vehicle operation costs, operating costs and noise. With respect to pavement type, the report focused primarily on noise generation, effect on speed and vehicle operation costs. The analysis of road surface condition used skid resistance as the parameter to measure the effect on accident rates and confirmed the instinctive conclusion that wet weather crash rates increase with the decreasing skid resistance. Road roughness measured by the IRI was considered the common measurement for overall surface ride quality but was only used to assess the impact on fuel consumption, tire wear and other vehicle operation costs.

A similar report conducted by the Swedish National Road and Transport Research Institute reviewed the influence of road surface texture on various traffic characteristics, safety being one of them. It states that friction is the primary safety characteristic of the pavement surface. Although it does not statistically validate this claim, this relationship is implied throughout the manual. It even lists as one of its "Vision for the Future" the importance of an international index to quantify the safety standard represented by friction (Sandberg, 1998).

The reports listed above recognize that there is extensive literature on skid resistance and its impact on crashes/crash rates dating back to late 1920s. The following table is a synopsis of some of the pertinent literature on this subject:

Table 1. Previous Research on Surface Distress Characteristics and Safety

Year	Country	Authors	Data	Study/Method	Results/Conclusions
2004	Switzerland	L. Seiler-Scherer	1999-2002 SCRIM at 100m intervals and 1994-1998 accident data for freeway and main road segments and accident	Data was collected and classified	Based on the large data, no correlation between skid resistance and accident frequency was found. It was able to identify hot spots.
2007	U.S.	Milton, Shankar, and Mannering	Crash segment data 1990-1994 (Washington State)	Mixed logit estimation for annual accident-severity on road segments	Pavement friction number found to be significant in the possible injury function. Decrease likelihood of possible injury and increase probability of property damage.
2005	U.K.	Viner, Sinhal, and Parry	Road segments skid resistance data and crash data 1994-2000	Sites were categorized by junction classification, values of mean and 95 percentile accident risk were calculated for different band of skid resistance	Results varied for each site category and therefore, Investigatory Level thresholds for each site category were determined.
1995	U.S.	Shankar, Mannering and Barfield	61 km section of I-90 and 5 year accident data from 1988-1993	Estimation of overall nested logit model accident severity probabilities	Wet-pavement rear-end accident indicator found greater probability of possible injury relative to property damage only
2008	U.S.	Oh, Chung, Ragland and Chan	400 miles of freeways in California and collision data 2000-2003	Continuous Risk Profile (CRP) approach to indentify high collision concentration locations (HCCL)	Hot spots were identified for high wet pavement collisions to be vertical sags with high vegetation. Collision rates at the sites did not exceed 99.5 percentile for dry, but did for wet conditions

Table 1 (cont.). Research on Surface Distress Characteristics and Safety

Year	Country	Authors	Data	Study/Method	Results/Conclusions
2007	Greece	Kopelias, Papdimitriou, Papandreou, and Prevedouros	533 crashes on Attica Tollway in 2004 and 2005. Geometric and operational features were added for each crash location.	Pearson's correlation analysis for each independent variable, ANOVA and linear regression with three forms of crash severity as dependent variables	Freeway characteristics (roughness included) and environmental conditions (wet-dry conditions) explain 5% to 10% of crash number. 80% attributed to driver behavior
1998	Finland	Leden, Hamalainen, and Manninen	Before and after crash on main roads in Finland from April 1 to September 30 for years 1989, 1990, 1992 and 1993	Before and after: Effect of resurfacing on safety based on (Hauer) observational study involving a treated and a comparison group	Risk of fatal and injury crash increase immediately after resurfacing by somewhat less than 7%. Friction was highly dependent on the type of resurfacing treatment.
2010	U.S.	Oh, Ragland, and Chan	Before and after data crash data	Before -and-After comparisons of experimental pavement types in California	After resurfacing with open-graded AC significantly decreased the number of wet-related collisions. No significant conclusions drawn for grooved pavement and rubberized open-graded AC.
2004	U.S.	Agent, Pigman and Green	Three years of crash data before and after years 1998 and 1999 when study resurfacing projects were completed. 120 locations	Before and After comparison	Analysis did not show a reduction in total crashes after resurfacing, but there was a reduction in crashes on wet pavement after resurfacing.

Although studies have confirmed the link between skid resistance/surface friction and wet crash rates, the research has not been able to exactly quantify the importance of this relationship nor have specific thresholds for friction values and substantial decrease in crash rates been determined. This is attributed to the dynamic nature of the pavement skid resistance characteristics, traffic operations, and driver/vehicle variables, and similar

to the nature of overall pavement conditions. That is why, like all other pavement characteristics, control of crash rates due to inadequate skid resistance must be based on regular assessments of the friction at the roadway segment level (AASHTO, 2007-2008).

2.1b Previous Research on Pavement Ride Quality and Safety

"Smoothness matters" (APA, 2009) is the title of an asphalt industry article promoting the fuel conservation benefits of asphalt pavement, and "to most people, a smooth road is a good road", starts the executive summary for the State of Virginia's 2004 Interstate Annual Report (Reid and Clark, 2004). This report goes on to say that ride quality is the primary attribute measured by the traveling public, and this sentiment is reflected in a number of national studies. Accordingly, roughness thresholds for the national highway system are established by the FHWA. Furthermore, a study conducted by Shafizadeh and Mannering to determine the acceptability of pavement roughness on the urban driving public in the state of Washington, pavement condition came in a close second to safety (21 % versus 26%, respectively) (Shafizadeh and Mannering). To determine an acceptable IRI threshold, real-world driving scenarios were used, and the drivers were asked their opinion on the condition/roughness of the pavement. The findings of the report validated the FHWA guidelines, and the conclusion was to keep that threshold in place.

"Rough Roads Ahead: Fix Them Now or Pay for Them Later" is the title of a 2009 publication from the American Association of State Highway Officials (AASHTO) that paints a grim picture of the present condition of the nation's roads and punctuates the

urgency of taking immediate action (AASHTO, 2009). While safety is implied, the primary focus is on the tangible costs to the motorist of this contagious condition that has impacted the majority of the states in this country. Some interesting statistics on the current condition of our roads as reported by the states using the IRI rating index follow.

13 percent of the 47,000 miles of interstate in the U.S. are in poor condition, and only 50% of America's main roads are in good condition. Rural roads are in much better condition than urban roads, and major urban centers have the worse roads – more than 60%. On the top of the list of strategies for saving America's highways in this report is using the best materials throughout the life span of the highway and matching these materials to the ambient environmental and traffic conditions. In addition, the authors stress the importance of new materials research and constant pavement condition monitoring to achieve longevity and make the best use of the limited maintenance funds.

Table 2. Previous Research on Ride Quality and Safety

Year	Country	Authors	Data	Study/Method	Results/Conclusions
2004	Norway	Elvik and Vaa	n/a	n/a	Improving unevenness increased crashes by 10%. Drivers may compensate for unevenness by slowing down
1983	Israel	Craus, Livneh, and Ishai	Accident, pavement condition rating (PR),	Normalized accident crash rates obtained by regression	Upgrading the state of the pavement does not necessarily decrease accident rates. Road safety should be viewed not generally, but by identifying hot spots
1981	U.S.	Burns	Accident history/data, ADT, ride quality/roughness	Before and after analysis of a road segment concrete pavement grinding	Roughness and ride quality improvement after grinding reduced in total wet-pavement and lost-load accidents by 40% and 51%, respectively.
2002	S. Africa	Bester	Segment, terrain, ride quality (PSI), traffic and accident data	Correlation matrix, direct comparison, and regression analysis	Bad riding quality can lead to an increase in the accident rate especially in more rugged terrain
1997	Jordan	Al-Masaeid	Pavement condition (IRI and PSR), ADT, geometry, accident data rural roads	Correlation matrix and regression	Single-vehicle and multiple-vehicle accident rate found to have high correlation with pavement condition. S-V accident rate reduced with increase in roughness (possibly due to decrease speeds) and M-V accident rate increased with increase in roughness.
2003	Sweden	Ihs, Velin, and Wiklund	Accident data, rut depth, and unevenness (IRI) all state roads for 1992-1998 and 2000	Linear regression and variance analysis	Both methodologies show the higher the IRI, the higher the accident risk.

Table 2 (cont.). Previous Research on Ride Quality and Safety

Year	Country	Authors	Data	Study/Method	Results/Conclusions
2008	U.S.	Chan, Huang, Yan, and Richards	2006 Accident, highway segment and pavement condition on 4 urban divided median asphalt roads in Tennessee	Negative Binomial comparing rut depth (RD), IRI and PSI pavement distress parameters.	RD model did not perform well except for accidents at night and in wet conditions. IRI and PSI always significant prediction variables in all types of accident models.
	U.S.	Titi, and Rasoulian	Louisiana pavement condition in profile index (PI). 23 sections of asphalt pavement	Regression analysis to evaluate correlation between IRI and PI	Conducted to assess possibility of getting reliable prediction of IRI from PI ratings. From PI ratings IRI smoothness criteria for asphalt pavement was developed.
2007	U.S.	Anastasopoulos, Tarko, and Mannering	337 roadway segment data and 1995-1999 accident data in Indiana	Tobit analyses	Factors relating to pavement condition and quality including IRI were found to significantly influence vehicle accident rates.
2008	Australia	Cairney and Bennett	Two-way undivided rural roads with speed limit of 100 km/h. Roughness, traffic and crash data	Pivot tables used to analyze distribution of crashes related to rutting, roughness and texture. Results graphed and correlation co-efficient calculated	Relationship between roughness and crashes followed a polynomial function and the fit was excellent. The extreme roughness condition resulted in noticeably higher crash rates but only over a small portion of the network. This relationship requires confirmation from other studies.

2.1c Previous Research on Pavement Type and Safety

This section highlights the originality of this study because despite the many variables considered in analyzing safety, pavement type has not been largely explored. To date, the effects of weather on roadway surface condition as a risk factor have been the primary focus for traffic safety studies. These conditions (i.e., visibility and road surface friction) and their influence on the distribution of accident severity has been studied but not the relationship between type of roadway surface and crash characteristics.

Additionally, state and local governments do not consider safety as a part of the selection criteria. Instead, smoothness, fuel economy, cost and longevity are the primary determinants (APA, 2009). Only recently has there been a study on the safety performance of experimental types of pavement in California (Oh, et al., 2010).

In a study Khattak, Khattak, Hummer and Sickling used California Highway Safety Information System (HSIS) examined crash rates and modeled crash and injury frequency on urban and rural principal arterials and minor arterials to determine if there is a relationship between pavement type (concrete vs. asphalt) and roadway safety (Khattak, et al., 2007). Regarding the crash rate analysis, the researchers found few clear trends on whether or not concrete pavement had lower rates over asphalt. However, much more conclusive and defensible findings were realized upon completion of the Negative Binomial models for crash frequency and injury severity. The modeling results showed concrete pavements are associated with lower crash frequency in California, but the frequency was marginally stronger in the relationship between crashes with injuries in

some of the roadway functional classes. Asphalt pavements showed conclusively a stronger association with higher total crash frequencies, as well as injury and non-injury crash counts. For future research, the authors suggest enhancing the thoroughness of the study and establishing a stronger causal link by correlating pavement type with pavement roughness information over a longer timeframe.

Three other studies of interest used pavement type and/or condition as an indicator variable(s) for assessing the effect on crash rate/frequency. Karlaftis and Golias revisited the question of whether road geometry and traffic volumes have a relationship to the crash rates on rural roadways (Karlaftis and Golias, 2002). They developed a model to assist in the prediction of crash rates for a given rural highway road segment. In this analysis, they used pavement, friction and pavement serviceability index (PSI) as binary, continuous and qualitative independent variables, respectively. Using a non-parametric statistical methodology they concluded the PSI, friction and pavement type variables had relative importance with respect to crash rates on rural two-lane road segments of 59, 32 and 30 percent, respectively and for rural multilane segments friction, PSI and pavement type showed relative importance of 25, 21 and 11 percent respectively. Das, Abdel-Aty and Pande used skid resistance, surface type and generic pavement condition (defects vs. non-defects) rating as continuous and binary predictor variables to model crash frequency on high-speed urban multilane arterials in Florida and provide conclusions for each crash type: rear-end, angle and head-on (Das, et al., 2010). Crashes were divided into separate models developed for three roadway segment groups: midblock, signalized, and access point. Intuition prevailed in the conclusions with respect to the significance of average

daily traffic (ADT), and wet conditions (less friction) for all three roadway segment models. In daylight conditions, these two variables were found to be significant in their relationship to the frequency of rear-end crashes. Of interest is a particular interaction that reflects roadways with no defects have fewer rear-end crashes near access points at higher posted speed limits. Additionally, the study showed dry surface conditions on asphalt pavement lead to fewer angle crashes at signalized intersections, and good road conditions coupled with dry conditions decrease the instances of head-on crashes for the segments near access points.

In a study conducted by Park and Saccomanno the authors reviewed the relationship various factors, one of which is paved and non-paved roadways, have on the frequency of crashes at highway-railroad at-grade crossings in Canada (Park and Saccomanno, 2005). They developed non-parametric models to try and establish statistical relationships between vehicle wrecks and various countermeasures. Considering the improvement in the percentage of expected collision reduction at railroad crossings they concluded that while a reduction in crashes after paving a roadway surface cannot be estimated at Class 1 and Class 4 crossings, one can speculate that a 32% reduction will be realized at all types of crossings after the roadway is paved.

In one study using Kansas work zone crash data, Li and Bai used the concept of crash severity index (CSI) to evaluate the safety risk factors associated with work zones on injurious and fatal crashes (Li and Bai, 2008). By examining 29 work-zone crash variables, a select group of risk factors were identified using Chi-squared and Cochran-

Mantel-Haenszel statistics. Based on these risk factors, CSI models were developed using logistic regression. Surface type (concrete and blacktop) as identified as a potential risk factor and was included in the more comprehensive models, but it was excluded from the simplified model due to the large p-value. Both concrete and blacktop were found to be potential risk factors in example conditions with high CSIs. The authors suggest future research using larger datasets to improve the accuracy of using CSIs to predict fatal crashes.

Lastly, in a Federal Highway Administration (FHWA) report FHWA-SA-96-068 the authors Hibbs and Larson investigated tire pavement noise and safety performance of various portland cement concrete (PCC) surface textures. In that study they state the purpose of surface texture is to reduce the number of severity of wet weather accidents and ask the question "Is there a safety advantage to either asphalt concrete (AC) and PCC pavements?" To answer the question they review the advantage of the skid resistance and longevity for various AC and PCC pavements but only analyze/compare crash data (wet weather accidents) for various concrete surface textures and do not compare this data to AC pavement types. Conversely, the asphalt industry through the Asphalt Institute has a publication Highway Safety with Asphalt Pavement, 1996 promoting the safety aspect of asphalt concrete. Advantages included are better visibility of pavement markings against the dark background of asphalt pavement, asphalt shoulders constructed with a lighter colored and coarser cover provides a distinct safety feature, ease of construction, no joints to distract drivers and snow and ice melt faster on asphalt (Hibbs

and Larson, 1996). Again, no real comparison between asphalt and concrete pavements using accepted statistical analyses methodology is provided.

This study will attempt to bring pavement type into the realm of traffic safety and fill a gap in the body of knowledge in this area. Given a crash occurs this research will evaluate the distribution of accidents that result in specific crash outcomes. Determining a relationship between pavement type and crash characteristics is important because it plays a critical role in the physics of a crash especially rear-end type collisions. Planners and engineers will be able to incorporate this newly found knowledge into developing more effective transportation management plans specifically for construction and maintenance work zones but with the potential to expand outside these zones.

2.1d. Previous Research on Safety and Pavement Maintenance

As highlighted above, in this literature review the studies suggest there are relationships between the pavement condition and roadway crashes. The studies have shown that skid resistance, roughness and distress, properties of the pavement mix, pavement markings and visibility can attribute to an increase in the probability of various types of crashes. As stated in a recent dissertation, pavement maintenance and rehabilitation is one of the most critical and costly forms of infrastructure asset management. Preserving the pavements in an appropriate manner extends their service life and, most importantly, improves motorists' safety and satisfaction and saves public tax dollars (Kim, et al., 2006).

Most state highway agencies now have wet accident reduction programs and believe the key factor to success is the willingness of the state to give improving pavement conditions top priority (Mahone and Sherwood, 1996). Even when studies have shown the positive safety effect of preventive maintenance, in some cases reduction factors as high as 54 % (Erwin and Tighe, 2008) and long-term improvement in pavement performance (Smith and Tighe, 2004), still 21 out of 48 state highway departments do not have specific design guidelines that address pavement skid behavior. Instead, they choose a passive approach by monitoring pavement conditions and reacting when needed (Jayawickrama, et al.).

In an effort to incorporate road safety into pavement management, one study conducted by Noyce, et al explored the relationship between asphalt mix designs, skid friction and roadway safety (Noyce, et al, 2007). Crash and pavement friction collected over a 10-year period was used to find a relationship between pavement skid resistance (friction) and crash frequency, particularly in wet weather conditions at six study sites in Wisconsin. By evaluating the trends over the 10-year period and conducting and regression analyses, the study concluded there was very little evidence of a relationship between crash frequency and skid resistance, but some results indicate that the number of wet pavement crashes over the life of the pavement. However, these results were not statistically supported. Nonetheless, the authors concluded the relationship appeared to behave inversely proportional meaning more crashes occurred at low friction numbers (FNs) which is an important indication that skid resistance may indeed be a factor affecting we weather crashes.

Additionally, while it was not possible to determine a critical skid friction threshold for triggering pavement maintenance, it was clear that the pavement surfaces with friction values less than 35 are less safe. An annual testing program to monitor the skid friction was recommended as part of an effective asphalt pavement asset management program, and friction values less than 35 should be reviewed for future rehabilitation or reconstruction.

Although the FHWA requires states to provide annual condition assessment reports with safety implied as one of the reasons, few have integrated safety directly into their pavement management and maintenance programs. This fact was identified in a study conducted by Tighe, Li, Falls and Hass (Tighe, et al). After reviewing pertinent literature, the authors conclude that while pavement conditions may only account for low percentage of roadway accidents, in absolute terms it does represent a substantial number of crashes and therefore should be considered in determining pavement surface treatment strategies. Road safety should not be considered a separate area; rather, it should be incorporated into pavement management programs simultaneously. The functional improvements in safety should be directly linked to the alternative maintenance treatment proposed.

Having said this, is it due to the current lack of funding, and is this a long-term trend? The report prepared by AASHTO reference above indicates this to be true. Before "Planners said this is what we want it to look like. Now let's figure out how to pay for it", today "Now we work in the reverse. We say here is how much money we have, and let

us decide what we want to do with that. That approach doesn't produce the best decision" (AASHTO, 2009). Additionally, a study conducted in the state of Washington noted that there is a discrepancy between the pavement preservation funding for both asphalt and concrete with the aging concrete pavement lagging significantly (Mahoney, et al, 2010). The authors suggest developing economic pavement performance indicators to monitor how efficiently the maintenance dollars are being spent.

It is clear the focus is on the dollars, or lack of, and safety is not an integral part of pavement management strategies. Encouraging, though, is the current emphasis on pavement preservation in the U.S. what we can learn from the international community, particularly the United Kingdom and Australia. In a recent report, Larson states the need for an integrated database that links crash data to detailed geometrics and to pavement condition databases that include pavement surface characteristics (Larson). Work zone approaches are identified as a critical hot spot when comparing the friction demand over the service life of the pavement. It lists the implementation of Road Safety Audits as one of the nation's most promising proactive techniques for including safety into pavement management preservation programs.

2.1e Previous Research on HLM and Roadway Safety Studies

In the literature, multilevel models (MLM) are also referred to as hierarchical linear models (HLM) and HLM is what will be used throughout this dissertation. In highway safety, arguably the crash data have a nested or hierarchical structure with several levels of hierarchy. Additionally, there is a strong correlation that exists between crashes that

occur under the same kind of human, environment/roadway, and traffic conditions. HLM is a type of regression model formulated to account for variation between crashes within each "nest" or "cluster" where the pattern of clustering is known. It is specifically design to capture correlations among these groups of data leading to better and unbiased parameter coefficient estimates and standard errors.

This methodology is relatively new to the field of transportation safety research and up until now has been limited to the clustered nature of educational and social sciences studies. It has only been used in a small number of studies to analyze the apparent nested relationships in traffic safety. One of these studies conducted by Eckhardt and Thomas (Eckhardt and Thomas, 2005) used HLM to explain the spatial occurrence of road accidents on the southern periphery of Brussels (Belgium). In the paper thier goal was to show that multilevel analysis can be used to better understand the spatial aspects of traffic safety. Specifically, they wanted to show how much influence the characteristics of the environment and roadway geometry (space) have on the location of where accidents occur in an urban setting.

Focusing on the spatial aspects of accidents (i.e., the location) this study developed a two level model with the first level being the hectometer. The explanatory variables at this level described the roadway segment's physical nature. Not related to the roadway itself, but part of the global environment of the accident, the municipality was chosen as the second level of analysis. At this level, the independent variables pertained to the socio-economic characteristics of the area where the accident occurred. The dependent

variables were (1) accident location (i.e., is it located in a black zone), (2) total number of accidents at each location, and (3) total number of accidents divided by the average daily traffic volume (adding an exposure variable to measure the risk). This study concluded HLM has potential when used to model roadway crashes but in spatial data analyses it is not easy to define the hierarchical levels. Additionally, the study found the road accidents are significantly influenced by the road geometry and the characteristics of the environment which lead to the conclusion that drivers are unable to adapt to changes in the road conditions to avoid accidents.

Aguero-Valverde and Jovanis conducted another spatial analysis of roadway safety using Hierarchical models. Estimates of county-level crash frequency using injury and fatal crash data for Pennsylvania for 1996-2000 were determined to compare Full Bayes Hierarchical and traditional Negative Binomial analysis methods. The independent variables included socio-economics, weather conditions, roadway geometrics and traffic volume (exposure). This study concluded that the Full Bayes models were effective in measuring the existence of spatial correlation in crash data and provides a mechanism to quantify, and reduce the effect of, this correlation (Aguero-Valverde and Jovannis, 2006). They also concluded that spatial correlation may be more important in road segment and intersection-level crash models where this correlation is even more predominant.

Severity of crashes was analyzed in two studies using multilevel models. In one study conducted by Lenguerrand, Martin and Laumon's application of the Monte-Carlo method on observed and simulated French road crash data between 1996 and 2000 was used to

compare multilevel logistic model (MLM), Generalized Estimating Equation models (GEE) and logistic models (LM) (Lenguerrand and Laumon, 2006). Severity is a discrete, non-normally distributed binary variable and requires a non-liner model. A three-level hierarchical structure MLM3 using crash level, car level and occupant level and a two-level structure MLM2 along with GEE and LM were compared.

In a study conducted by Jones and Jorgensen, the authors analyzed predictors of severity outcome amongst over 16,000 fatally and seriously injured casualties involved in accidents between 1985 and 1996 in Norway. The data were analyzed using a three-level hierarchical structure considering the characteristic of the casualties at level 1, accidents information at level 2 and within which municipalities as level 3. Again, the aim was to model the odds of survival for each casualty so logistic regression models were fitted due to the binary nature of the dependent variables (Jones and Jørgensen, 2003).

Benefits of modeling the hierarchical nature of accident data were found in both studies and concluded that it is possible to formulate multilevel models that are both technically viable and highly relevant (Lenguerrand and Laumon, 2006). When compared to GEE and LM, MLM was found to be the most efficient model while both the GEE and LM underestimated parameters and confidence intervals.

Finally, a study reviewed applying a hierarchical binomial logistic model-to-model crash type outcome probability at rural intersections. Kim, et al. documented a small study of 548 vehicle crashes collected from 91 two-lane rural intersections in the state of Georgia

in a paper (Kim, et al., 2007). The data came from 38 counties in the state for a period of 2 years and are postulated as hierarchical and analyzed using two levels: crash-level characteristics and intersection-level characteristics. The dependent variables are dichotomous and captured each type of accident; angle, head-on, rear-end, side-swipe same direction and side-swipe opposite direction. To determine if multi-level modeling techniques have advantages over traditional methods, all of the crash models were fitted using non-multilevel models as well.

The findings of this study clearly recognize the hierarchical structure of crashes at rural intersections in Georgia and the limitations of traditional modeling techniques with respect to the correlation of crash data characteristics within the nested clusters can be overcome through the use of multilevel modeling.

CHAPTER THREE

3.0. DATA SOURCES AND AGGREGATION

3.1. Overview

The Commonwealth of Virginia manages the nation's third largest road network with approximately 57,900 miles of roadways. Given a crash occurs, this study will measure the risk/probability of an individual being in a crash with a certain outcome due to the perceived hazardous pavement conditions and ride quality within an empirical setting using historical crash and roadway condition data from Virginia. This chapter describes the data sources, collection and aggregation, the empirical setting, the pertinent variables studied and the selected methodologies applied to this research.

There are a number of various driving and environmental conditions inherent in the different functional class of roadways and many contribute to the safety of the roadway (e.g. average speed, lane width, access points, etc.). This study elected to limit the data to interstate routes to increase the control of a number of potentially correlated predictor variables. In addition, this study will be aggregating pavement condition data with crash information for these two roadway classifications for each of the three main pavement types: asphalt concrete (ACP), continuously reinforced concrete (CRCP), and jointed reinforced concrete (JRCP).

VDOT has nine Districts and pavement condition/ride quality data are collected for each District. There are 1,945 miles of interstate in Virginia; the ones included in this study are

I-81, I-95, I-64, I-264, and I-395. The study's routes are shown in Figure 2 (VDOT, 2009). I-81 is a south to north route that traverses the western part of the state, I-95 is a south to north route traversing the central part, I-64 is an east to west route through central Virginia, I-264 travels through the cities of Portsmouth, Norfolk and Virginia Beach in the Hampton Roads area, and I-395 is an extension of I-95 in Northern Virginia.

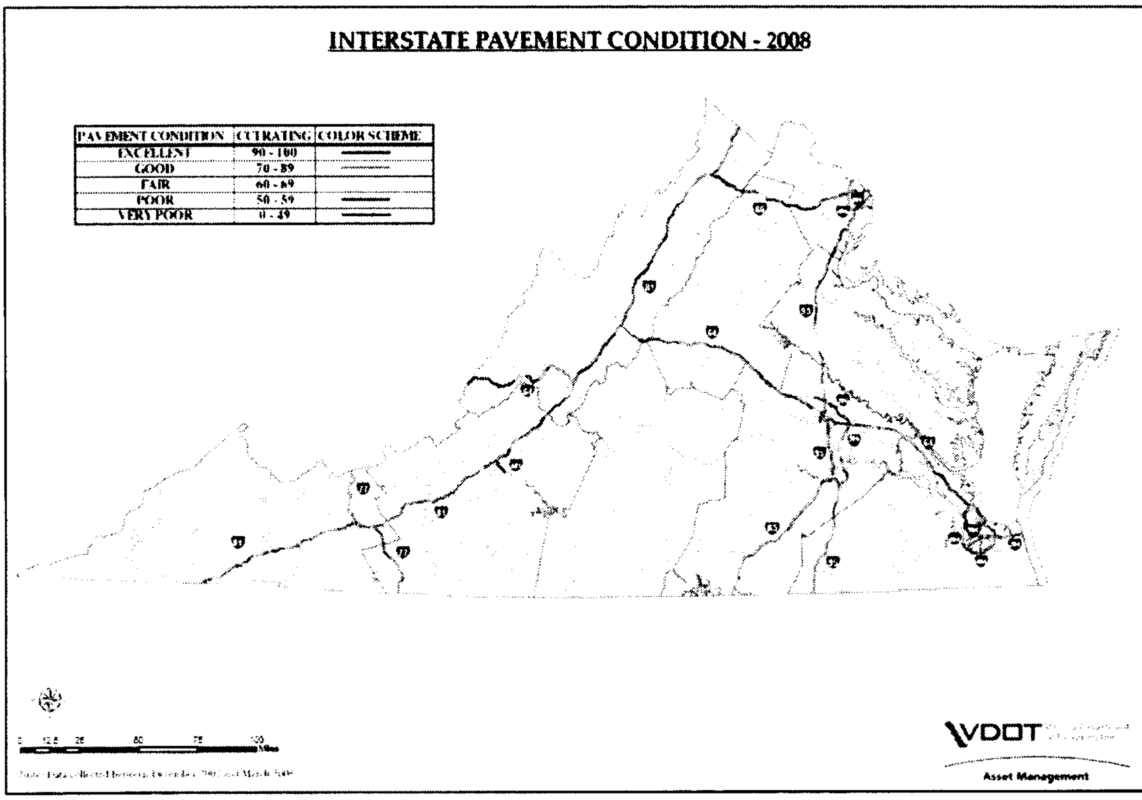


Figure 2. Virginia's major interstate routes

Crash data are collected at the crash site (i.e. point location). For ease of computation and analysis and to increase control of the independent variables, it is important to determine the best method for aggregating/segmenting the data. The method for compiling/aggregating the data for this study is further explained in the methodology section below.

3.2. Data Sources

Various unique databases will be developed for this study by combining roadway inventory, pavement performance, historical crash, environmental, and socio-economic information on roadway segments. The following lists the sources used to collect data for this research:

- The Maintenance Division of VDOT collects pavement condition data such as distress, rutting, faulting and pavement ride quality data roughness (IRI) for the interstate highway network on annual basis and non-interstate bi-annually. VDOT collects this data using digital imagery and automated crack detection methodology as part of the Inventory and Condition Assessment Survey (ICAS). Included in this dataset is roadway profile data: IRI left wheel path, right wheel path and average, the severity of transverse and longitudinal cracking, traverse and longitudinal reflective cracking, alligator cracking, and bleeding, along with pothole count, area of delamination and patching, average rut depth, pavement type, and critical condition index. VDOT first began collecting this data in 2007 in addition to the pavement performance data collected; various pertinent roadway inventory data are collected.
- Accident-level crash data was generated from the Virginia's crash database which is a collection of data for actual police crash reports. In order to aggregate this data with the pavement condition information, the raw crash data file consisted of statewide crash information from the VDOT's Traffic Engineering Division. It provides detailed information on each individual crash but is limited to only those crashes responded to and reported by a State Police officer. This database also includes roadway inventory, environmental and travel exposure data at the time and scene of

the crash. Across the country, similar databases are used to perform problem identifications and to support the development and evaluation of potential countermeasures.

- Roadway inventory, environmental and travel exposure data (i.e., traffic volumes) are collected from both the pavement performance maintenance and the crash databases.
- Socio-economic data was collected from the American Community Survey (ACS). The ACS is a household survey conducted by the U.S. Census Bureau that currently has an annual sample size of about 3.5 million addresses. ACS estimates provide communities with the current information they need to plan investments and services. Information from the survey generates estimates that help determine how more than $400 billion in federal and state funds are distributed annually. Each year the survey produces data that cover the periods of 1-year, 3-year, and 5-year estimates for geographic areas in the United States and Puerto Rico, ranging from neighborhoods to Congressional districts to the entire nation.

3.2a. Pavement Condition Data: Roadway Inventory Evaluated

The surveys were conducted in the rightmost traffic lane, usually designated lane one (1) in the VDOT pavement inventory, while the tabulations, graphs, and discussions below were extended to a lane mile basis. For example, a one-mile long pavement section with three lanes in the direction of rating would be reported as three lane miles. Using the method described above, approximately 5,250 lane miles on interstates are accounted for in 2007 and 2008 surveys. This inventory is broken down into VDOT districts and then counties within the districts.

3.2b. Pavement Condition Evaluation Criteria

Table 3 provides a scale for evaluation for the pavement surface distress condition. This index, developed by VDOT, is the result of experience with previous windshield surveys and reflects earlier action of the VDOT Pavement Management Engineering Team (PMET). The PMET action was a decision that pavements with a condition index of less than 60, referred to as the deficient pavements, would be evaluated further for possible higher types of maintenance and rehabilitation.

In order to make informed pavement management decisions in Virginia, pavement distress data are used to determine pavement condition indices. For flexible/asphalt surfaced pavements the load related distress index (LDR) and non-load related distress index (NDR) are deduct based indices. For pavements with no discernible load related distresses, the LDR and NDR are 100. From this value, points are deducted for each distress that is load related based on severity and/or frequency. These factors are weighted to those distresses that are more detrimental to the performance of the pavement. The critical condition index (CCI) is defined as the lower of the two values of the LDR and NDR.

For rigid pavement, similar deduct values are used to determine the pavement condition index. The slab distress rating (SDR) is used instead for JCP pavements and the Concrete Punchout Rating (CPR) and the Concrete Distress Rating (CDR) are used for CRCP pavements. However, the same concept of CCI and the same scale in Table 3 apply to the latter two pavement types as well: SDR is directly equivalent to CCI for JCP pavements,

and the lower of CDR and CPR is equivalent to CCI for CRCP pavements. In general, VDOT determined pavements rating less than 60 by either index are considered deficient, i.e., they need some kind of attention, more specifically, some heavier type of maintenance/rehabilitation actions.

Table 3. Pavement Condition Definition

Pavement Condition	Index Scale (CCI)
Excellent	90 and above
Good	70-89
Fair	60-69
Poor	50-59
Very Poor	49 and below

3.2c. Pavement Condition Data Collection Overview

There are different guidelines in determining the severity of pavement distresses. The VDOT classifies surface distresses for Asphalt (ACP), Continuously Reinforced Concrete (CRCP), and Jointed Reinforced Concrete Pavement (JRCP) distresses, as shown in Table 4. Asphalt is classified as flexible/composite pavement and includes asphalt/bituminous overlays over jointed concrete (BOJ) and asphalt/bituminous overlays over continuously reinforced concrete (BOC).

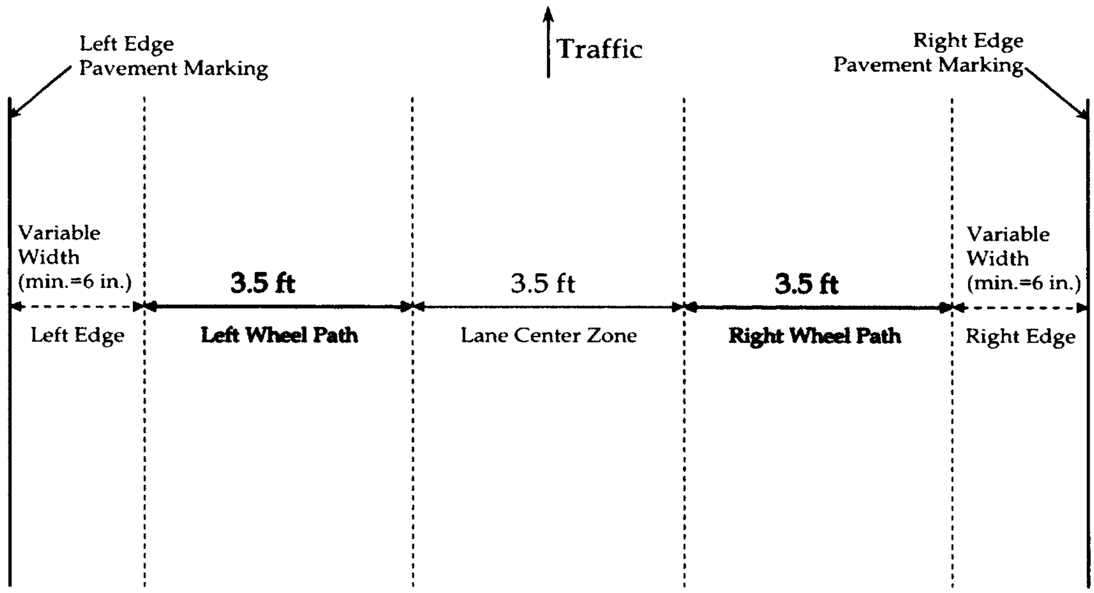
Table 4. Pavement Condition Distress Variables (PDVs)

Distress Category	Composite/Flexible Pavement	Concrete Pavements
	Distress Type	*Distress Type*
Cracking	Alligator Cracks Transverse Cracks Longitudinal Cracks Longitudinal Lane Joint Cracks Reflective Transverse Cracks Longitudinal Reflective Cracks	Clustering Cracks Punchout/Spalled Cracks Transverse Cracks Longitudinal Cracks
Patching and Potholes	Patching Potholes	PCC Patch Asphalt Patch
Surface Defects	Delamination Bleeding	
Surface Deformation	Average Deeper Rutting	
Joint Deficiencies		Transverse Joint Spalling Longitudinal Jt. Spalling Joint Fault Severity
Miscellaneous		Corner Breaks Blowups

As mentioned, VDOT maintains the third largest public road network in this country, covering a total of about 57,900 miles consisting of around 1,945 miles of interstate highways, 10,405 miles of primary highways and 45,550 miles of secondary roads. The surface distress data are collected and analyzed on a lane mile basis on all of the Interstate and Primary pavements (VDOT, 2009).

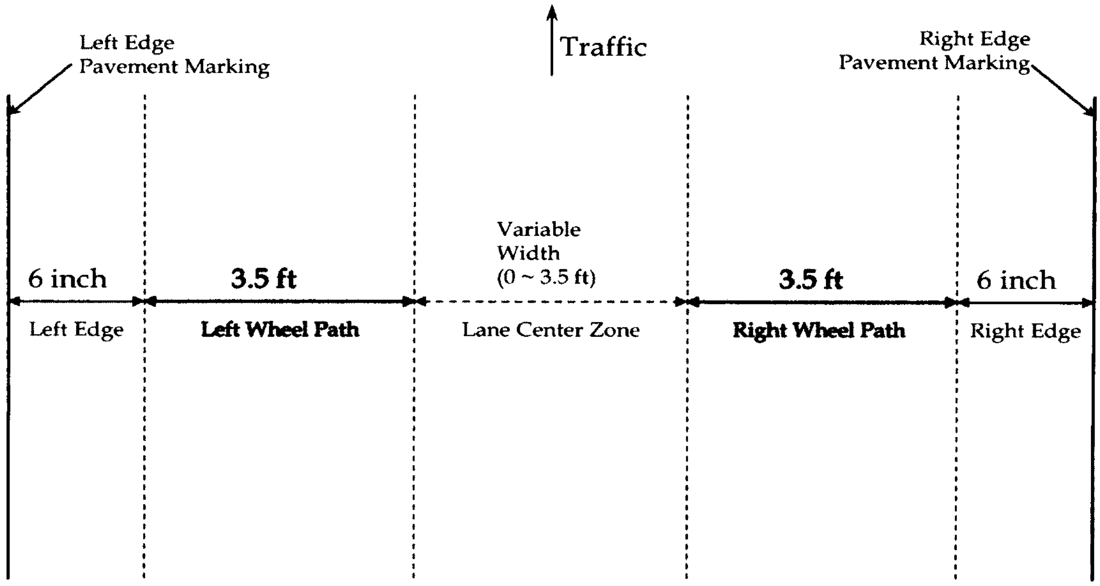
Vehicles equipped with continuous digital imaging cameras and automated crack detection technology are used to collect, reduce and process the pavement condition data

digitally captured for each road segment. These special cameras shoot downward to digitally capture pavement images for crack detection. Additionally, a camera is used for a forward perspective view, and two cameras (one left and one right) are used for the collection of sideward or right-of-way images. This specialized equipment is attached to vans which allows the data to be collected at highway speeds as they drive along the pavement. As these vans move along the roadway, specialized automated distress detection software reduces the downward and sideways digital images and quantifies the visual pavement distresses. Roughness and rutting data are also captured simultaneously with the sensors mounted on the vehicles (VDOT, 2009). This automated system uses route and milepost points for identifying pavement condition locations and summarized in one-tenth (0.10) mile roadway segments. The four figures below show the size and location of the wheel paths of the tracing vehicle (VDOT, 2007).



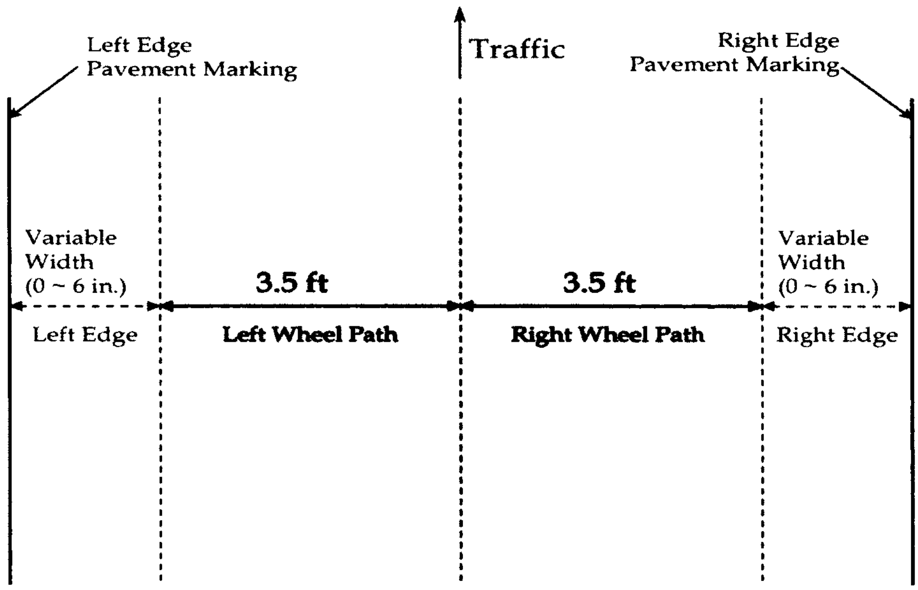
(1) Lane Width 12 ft or higher

Figure 3. Lane width 12 ft. or higher



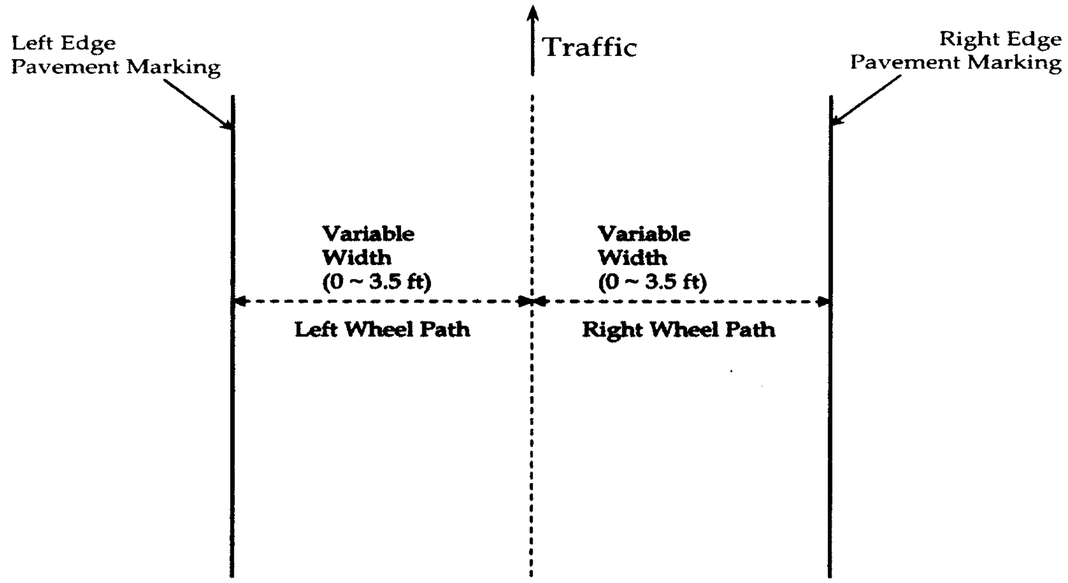
(2) 8.0 ft ≤ Lane Width < 11.5 ft

Figure 4. Lane width greater than or equal to 8 ft. but less than 11.5 ft.



(3) 7.0 ft ≤ Lane Width < 8.0 ft

Figure 5. Lane width greater than or equal to 7 ft. but less than 8 ft.



(4) Lane Width < 7.0 ft

Figure 6. Lane width less than 7 ft.

For evaluating distresses in the pavement prism, the digital imagines are required to produce a full 14-foot wide view of the pavement surface and have sufficient resolution to identify cracking of 1/8th inch wide in the downward perspective and 1/4th inch wide in the forward perspective. Using these detailed images, 100% of the pavement segment between the pavement stripes is evaluated for each applicable distress type. Distress types are rated/ranked by severity level which are clearly defined as levels 1, 2 and 3 and quantified in terms of linear feet, square feet or other quantity as applicable.

3.2d. PDVs Collected for Pavement Condition (VDOT, 2007)

The data elements collected are listed in Table 4 (above) for the three main pavement types: asphalt concrete pavement (ACP), continuously reinforced concrete pavement (CRCP), and jointed concrete pavement (JRCP). Detailed distress data in terms of extents and severities are collected and summarized for each 0.1 mile as well as for each homogeneous section.

As listed above in Table 4, the following are the six types of cracking used to rate asphalt or flexible/composite pavement which includes asphalt concrete overlays: transverse, longitudinal, longitudinal lane joint, reflective transverse and longitudinal, and alligator. Reflective transverse and longitudinal cracks are used only in rating asphalt concrete overlays. In general, cracks are rated as "open" if they are greater than ¼ inch wide and deemed "close" if less than or equal to ¼ inch wide. Transverse cracks run perpendicular to the roadway centerline are random, and the presence of these cracks reduces the NDR index value. Transverse cracks are rated at two severity levels with Severity Level 1

being cracks that are sealed and the sealant is in good condition and Severity Level 2, unsealed cracks. Longitudinal cracks run predominately parallel to the roadway centerline and for rating purposes; only those cracks outside the limits of the wheel paths are counted as longitudinal. The ones found in the wheel path are counted as alligator cracks as explained below. Longitudinal cracks reduce the NDR index value and are classified in two severity levels similar to transverse cracks. Longitudinal lane joint cracks are those typically found near the edges of the lanes but is only classified as a distress crack when the crack is severe enough to hold water. The presence of this type of crack reduces the NDR index value.

Reflective cracks, both transverse and longitudinal, are used to rate BOJ and BOC pavements occur in asphalt overlays over jointed concrete pavements primarily at the location of the joints. Transverse reflective cracks are rated over bituminous overlays where the underlying concrete pavement is jointed perpendicular to the centerline of the roadway. Longitudinal reflective cracks are rated for asphalt overlays (BOC) over continuously reinforced concrete pavement (CRCP) primarily located along the straight longitudinal joints parallel to the centerline of the roadway that break through the asphalt. Both types of reflective cracks have three levels of severity. Level one is a closed but unsealed crack with a width less than ¼-inch. Level two cracks are open with a width equal to or greater than ¼-inch but less than ¾-inches or if level one or level two cracks are within one foot spacing. And level three cracks have a width greater than or equal to ¾-inch or have width more than ¼-inch and are deteriorated for a width greater than six inches or two adjacent level two and/or level three are within one foot of each other.

Fatigue, alligator cracks, occur in areas that are subjected to repetitive heavy wheel loads and consequently are found only in wheel paths. They usually begin as longitudinal cracks and over time under heavy loads start to branch out into groups of cracks that connect forming a pattern that resembles the skin of an alligator. They are ranked in three severity levels with the two highest levels showing signs of the asphalt spalling.

Patching, potholes, delamination, bleeding and rutting are the other pavement distresses that VDOT captures in order to evaluate bituminous concrete pavement. Patching is areas where the pavement overlay has been repaired by removing and replacing small sections or where asphalt material has been placed to repair a crack. There are no severity levels and the presence of patching will reduce the NDR if it occurs outside the wheel path and reduces the LDR if they are found in the wheel path. Potholes are holes in the pavement usually extending into more than one layer. There are no severity levels, and their presence reduces the LDR index. Delamination reduces the LDR index, has no severity levels and occurs primarily in the wheel path where there is a loss of adhesion between the surface and underlying layers. Bleeding has two levels of severity and can occur randomly throughout the pavement width but primarily in the wheel path. The surface appears shiny or reflective, and it is due to excessive liquid asphalt in the bituminous concrete mix. Under warm temperatures, the surface can feel tacky. The presence of bleeding reduces the NDR index. The presence of rutting reduces the LDR index and is primarily found in the wheel path where the pavement is subjected to heavy loads. Rut depths are collected using laser sensors where a minimum of twelve points are required to calculate the depth. These calculations are made using two methods as

defined in ASTM – "straight edge" and "wire". Table 5 summarizes the pavement condition distress variable for ACP.

Table 5. PDVs for Asphalt Pavement

Distress Variable (ACP)	Pavement Type	Severity /Units	Variable Type
Transverse Cracking Severity	BIT/BOJ/BOC	1, 2	Categorical
Longitudinal Cracking Severity	BIT/BOJ/BOC	1, 2	Categorical
Longitudinal Lane Joint Severity	BIT/BOJ/BOC	1, 2	Categorical
Reflective Transverse Cracking Severity	BOJ/BOC	1, 2 ,3	Categorical
Reflective Longitudinal Cracking Severity	BOJ/BOC	1, 2 ,3	Categorical
Alligator Cracking Severity	BIT/BOJ/BOC	1, 2 ,3	Categorical
Patching Area - wheel path	BIT/BOJ/BOC	S.F.	Continuous
Potholes Count	BIT/BOJ/BOC	Count	Continuous
Delamination Area	BIT/BOJ/BOC	S.F.	Continuous
Bleeding Severity	BIT/BOJ/BOC	1, 2	Categorical
Average Depth Rut	BIT/BOJ/BOC	Inch	Continuous
Critical Condition Index (CCI)	ALL	Index	Continuous

For this study, the PDVs for jointed concrete pavement (JRCP) are corner breaks, spalling of transverse and longitudinal joints, transverse and longitudinal cracking, PCC and asphalt patching, blowups and joint fault severity. Corner-breaks have two severity levels and are characterized by a cracked portion of the slab where the transverse joint intersects the longitudinal joint which makes approximately 45-degree angle with the direction of travel. Spalling at the joints is not ranked in severity; rather, these areas are counted. They are defined as breaking or chipping at the edge or within one foot of the joint and the count is number per slab. Transverse and Longitudinal cracks are defined on concrete pavements similar to those found on bituminous concrete pavements.

Patching is also defined and rated similar to asphalt pavement where a portion greater than one square foot has been removed and replaced with like material or with hot mix asphalt. Concrete patches have three levels of severity, and asphalt patches are counted per slab. Localized upward movement of the concrete surface at transverse joints, or at random cracks, are called blowups. They are not ranked by severity but may greatly impair the ride quality of the pavement. They are counted by number-per-slab and often cause shattering of the concrete in, or near, the area. The joint fault severity distress variable is defined by the condition of the joint seal. It is considered to be damaged when the seal allows water and other incompressible substances to infiltrate the joint. They are rated by either being fully sealed or unsealed. Table 6 summarizes the PDVs for JRCP.

Table 6. PDVs for Jointed Reinforced Concrete Pavements

Distress Variable (JRCP)	**Pavement Type**	**Severity/ Units**	**Variable Type**
Transverse Cracking Severity	JRCP	1, 2	Categorical
Longitudinal Cracking Severity	JRCP	1, 2	Categorical
PCC Patch Severity	JRCP	1, 2 ,3	Categorical
Asphalt Patch	JRCP	# of Slabs	Continuous
Number of Transverse Joints	JRCP	Count	Continuous
Transverse Joint Spalled	JRCP	# of Slabs	Continuous
Longitudinal Joint Spalled	JRCP	# of Slabs	Continuous
Corner Breaks Severity	JRCP	1, 2	Categorical
Blowups	JRCP	# of Slabs	Continuous
Joint Fault Severity	JRCP	1, 2 ,3	Categorical
Index Value Slab Distress Rating (SDR)	JRCP	Index	Continuous
Index Value Slab Faulting Index (SFI)	JRCP	Index	Continuous
Critical Condition Index (CCI)	ALL	Index	Continuous

Continuously reinforced concrete pavement (CRCP) is rated by VDOT using the following PDVs: traverse, longitudinal, clustering, punchouts and spalled "Y" cracks, patches and longitudinal joint spalling. Traverse and longitudinal cracks are ranked in three levels of severity and are collected and defined similar to asphalt and jointed concrete pavement. Clustered cracking has two levels of severity and is defined as a group of transverse (3 or more) with an average spacing of two feet or less. Punchout and spalled cracks are counted in number of occurrences and square feet. A punchout is a section of the concrete slab that has broken into two or more pieces and is often comprised of two closely spaced transverse cracks, one short longitudinal located at the edge of the pavement or a longitudinal joint. When one transverse crack begins within another transverse crack and radiates to the edge of the pavement it forms a "Y" crack. They are counted and not rated by severity. Concrete and Asphalt Patching of CRCP is defined and rated similar to JCP. Spalling of longitudinal joints is defined as the breaking or chipping at the slab edges and within one foot of a longitudinal joint. These areas may be filled temporarily with hot mix asphalt concrete. There are no levels of severity for longitudinal joint spalling and they are recorded/counted in linear feet. Table 7 summarizes the PDVs for CRCP.

Table 7. PDVs for Continuously Reinforced Concrete Pavement

Distress Variable (CRCP)	PType	Severity/ Units	Variable Type
Transverse Cracking Severity	CRCP	1, 2 ,3	Continuous
Longitudinal Cracking Severity	CRCP	1, 2 ,3	Categorical
Clustered Cracking Severity	CRCP	1, 2	Categorical
Longitudinal Joint Spalling	CRCP	Feet	Continuous
Punchouts and Spalled Ycracks	CRCP	Count	Continuous
PCC Patch Severity	CRCP	1, 2 ,3	Categorical
Asphalt Patching	CRCP	S.F.	Continuous
Index Values Concrete Distress Rating (CDR)	CRCP	Index	Continuous
Index Values Concrete Punchout Rating (CPR)	CRCP	Index	Continuous
Critical Condition Index (CCI)	ALL	Index	Continuous

3.2e. Roughness Data Collected for Ride Quality

As mentioned above, along with the pavement condition data roughness and rutting data are simultaneously using sensors mounted on the data collection vehicle. From this data, ride quality information is determined and reported in IRI format for the left and right wheel path and from this information the average IRI is calculated all three types of pavement.

3.2f. Ride Quality Evaluation Criteria

Ride quality is defined by the public as a road that has minimal bumps/roughness and provides a smooth ride for an extended length. Pavement roughness is an expression of roadway irregularities that adversely affect the ride quality of a vehicle and thus the user. Ride quality is expressed in terms of International Roughness Index (IRI), measured in inches/mile.

Table 8 (VDOT, 2009) contains two IRI scales used for evaluation of the pavement ride quality survey: one set for Interstate and Primary highways, and the other for Secondary roads. It needs to be pointed out that ranges of IRI values corresponding to qualitative descriptors of ride quality were built upon similar categories promulgated by FHWA and incorporated consensuses from VDOT pavement experts regarding what thresholds were considered appropriate to represent acceptable roughness levels on Virginia highways. Pavements with poor and very poor ride quality are said to have deficient ride quality.

Table 8. Pavement Ride Quality Definition

Pavement Quality Category	**IRI Rating (inch/mile)**	
	Interstate and Primary	**Secondary**
Excellent	< 60	< 95
Good	60 to 99	95 to 169
Fair	100 to 139	170 to 219
Poor	140 to 199	220 to 279
Very Poor	≥ 200	≥ 280

On interstate and primaries, the data are collected on the entire network, but on the secondary pavements the data are collected on a 20 to 25% sampling basis. (VDOT, 2009)

3.2g. Pavement Data Collection Quality Assurance

To ensure the final pavement condition data are meaningful, VDOT conducts an independent QA process. For data collection, the QA process began with evaluation of control sections comprised of ACP, CRCP and JCP for interstate, primary and secondary systems. Field evaluations are conducted on three control sections, and image evaluations were completed on 19 control sections distributed over the system and pavement types. The control sections were used to calibrate the pavement distress rating process and to establish the precision and bias values for the roughness and rutting measurements (VDOT, 2009).

The precision term as specified in ASTM E177 is used for QA checks for the rutting and roughness comparison, and the data-collection vehicle is considered acceptable if it is capable of collecting rutting and roughness data within a specific precision range.

For the production ratings, nineteen batches of data including interstate, primary and secondary system ACP, JCP and CRCP pavements, are delivered to and reviewed by the independent data verification and validation (IV&V) contractor. Five percent of the data delivered in each batch were randomly chosen for QA and rated independently by the IV&V contractor. A batch is considered to have passed the QA checks when the CCI index values from the production data lies within 10 points of the CCI values from the IV&V ratings for 90% of the pavement length. In addition to the random 5% QA checks, a "high-level" data review consisted of reasonableness, and a completeness check was conducted for each delivery table (VDOT, 2009).

3.3 Data Processing and Aggregation

To create the unique study specific datasets that combine crash data with pavement ride quality and condition data (PDVs) at the site/point of the crash and intervals upstream of the crash, the following steps were conducted:

Crash data:

- VDOT compiles crash data by year. For this study, the baseline data was collected for the years 2007 and 2008.
- Querying both crash data sets determined the number of crashes on each interstate route.
- A table was created showing the number of crash cases for each interstate to determine how crashes were distributed throughout the routes. From this information, specific interstate routes were selected for the study.
- As mentioned above, five (5) of the major interstate routes were selected for this study, and they are: I-64, I-264, I-95, I-81 and I-395.
- Once the routes were selected, crash data for each interstate route was queried from the baseline data set for the years 2007 and 2008 and sorted by direction and milepost.
- Ten (10) crash datasets were created for each of the study's five (5) interstate routes for each year in both directions. This was done for years 2007 and 2008 for 20 datasets.

Table 9. List of crash datasets

Direction/Route(Total Crashes)			
East	West	North	South
I-64 (9342)	I-64 (9063)		
I-264 (1480)	I-264 (1844)		
		I-95 (3560)	I-95 (4219)
		I-81 (2037)	I-81 (2111)
		I-395 (366)	I-395 (347)

Pavement Condition Data:

- VDOT is broken into nine (9) districts with District 9 being Northern Virginia (NOVA). Each District is responsible for the condition of the pavement within its geographical boundaries and therefore the pavement condition data are provided by District.
- Annual pavement condition data are collected for each of the nine Districts by type of pavement (ACP, JRCP, and CRCP) and facility classification (interstate, primaries and secondaries).
- VDOT creates Excel spreadsheets for each type of pavement, facility classification by district. This serves as the baseline pavement condition data for this study.
- As mentioned above, this study is limited to five of the major interstates in the state. From the beginning to the end of each route, the counties that the routes traveled through each district were determined by direction. A list of these counties was developed along the direction of each route (i.e. east, west, or north, south).

- From the district-wide baseline pavement condition dataset, pavement condition data was queried for the counties listed in the step above for each of the study's routes. This provided pavement condition data by route for each county along the route.
- This county data was compiled to provide complete pavement condition data per direction per type of pavement (ACP, CRCP, and JRCP) for each of the study's interstate highways.
- Route I-64 contains all three pavement types; ACP, CRCP and JRCP. Route I-264 contains ACP and JRCP, and the remaining routes, I-95, I-81 and I-395, have only ACP. For the two routes, I-264 and I-64, that have different types of pavement two and three separate pavement type datasets were compiled, respectively.
- The result was 16 pavement condition datasets for each year, 2007 and 2008, for 32 datasets. There is one final pavement condition dataset per direction for each of the study's interstate roadways.

Table 10. List of pavement condition datasets

Pavement Type	**Direction / Routes (Total Crashes)**			
	East	**West**	**North**	**South**
ACP	I-64 (3759)	I-64 (3786)		
	I-264 (773)	I-264 (757)		
			I-95 (3560)	I-95 (4219)
			I-81 (2037)	I-81 (2111)
			I-395 (366)	I-395 (347)
JRCP	I-64 (2730)	I-64 (2903)		
	I-264 (707)	I-264 (1087)		
CRCP	I-64 (2853)	I-64 (2374)		

Determining the pavement condition at the point of the crash:

- Pavement condition is collected directionally along the route based on the direction of the route for each county in the district. If collecting from a north/south route the data are collected starting south and moving north. Data are collected every 1/10th of a mile. For this example, I-81N data was collected starting at the south end at the Virginia state line in Washington County.
- The milepost of the route where the crash occurred is collected in the crash data.
- However, pavement condition data are collected and start at 0.0 mile for each county in each district. For this example in Washington County, the pavement condition data for I-81N start at the southern end of the county, and the data point is 0.0. This is also the beginning milepost for I-81N in Virginia. When crossing into the next county to the north, the pavement condition data point goes back to 0.0. This is not the milepost for the route. However, knowing the milepost at the southern point of I-81N in this county, the pavement condition location data point can be translated to a milepost.
- Once this calculation is completed, both the crash data and the pavement condition data have mileposts as a spatial reference.
- As mentioned above, there are 10 crash data sets and 16 pavement condition datasets for each year. Therefore, 6 additional crash datasets had to be created for the JRCP and CRCP pavement types of I-64 and I-264.
- Using the milepost as a common reference the crash data was correlated for the JRCP and CRCP pavement types of I-264 and CRCP of I-64. Three additional crash data sets were created for these pavement types for each direction resulting in 6 additional

crash datasets. These were 16 total crash datasets for both study years totaling 32 crash datasets.

- The crash dataset was combined with the pavement condition dataset with crash data in one sub-worksheet and pavement condition data in another sub-worksheet. This resulted in 32 datasets, 16 for each study year.
- Using the VLOOKUP command in the Excel software the milepost for the crash was used as the lookup value. This was correlated to the same milepost for the pavement condition dataset, and the pavement condition data for this specific milepost (crash scene) was extracted from the appropriate pavement condition dataset and inserted into the crash data set. This is an automated process, and it was repeated for each route.
- In order to meet objective two of this research and evaluate in more detail the spatial component of the pavement condition the condition, of the pavement at specific intervals upstream of the crash was determined. The intervals upstream of the crash used were 0.10, 0.15 and 0.20 miles and the following process was used.
- The process is similar to that described above but before using the VLOOKUP command a point of interest location upstream of the crash was calculated using one of the specified intervals (0.10, 0.15, or 0.20). This was done by simple subtracting one of the upstream intervals from the crash site milepost.
- For example, an additional column was added, and using a simple formula, values were calculated by subtracting 0.1 from the values in the actual milepost of the crash site column.

- These newly calculated values (0.1miles upstream of crash) were used as the lookup value, and the pavement condition data for a specific milepost upstream of the crash were extracted from the appropriate pavement condition dataset and inserted into the crash data set.
- The process was repeated for each interval and each route and then combined into one dataset which included crash and pavement condition information at each interval upstream of the crash.
- At this point, there remained 32 datasets, each containing five (5) sub-worksheets for pavement condition data, and crash data combined with pavement condition data for locations at the site of the crash and 0.10, 0.15 and 0.20 miles upstream of the crash site.
- Final data aggregation and compiling for modeling occurred.
- To ensure the integrity of both the crash and pavement condition data during the aggregation process the datasets remained close to their original form with the routes broken out by directions and study year.
- However, to serve the objectives of the study and for modeling purposes, these individual datasets were combined by year and direction to create one dataset for each route for both direction and both study years.
- The final step was to turn the pavement condition predictor variables into binary variables. Using the IF statement in Excel, if the predictor variable was zero the number 0 was entered; if not, the number 1 was entered. Therefore, if the distress variable was present a 1 was entered and if not 0.

- Table 11 lists the number of files and crashes in both directions for the two study years by route.
- These datasets were further combined for type of pavement, and Table 11 lists the final datasets and the number of crashes by pavement types. This resulted in a total of three (3) primary datasets, one for each type of pavement ACP, CRCP and JRCP.
- To evaluate the effects of pavement type on crash outcomes common, pavement distress variables were determined, and datasets with these predictor variables were created for each pavement type ACP, CRCP and JRCP.

Table 11. Total number of Crashes by route

Route	Total Number of Crashes		
	ACP (63%)	CRCP (15%)	JRCP (22%)
I-64	7472	5190	5577
I-264	1517		1772
I-81	4122		
I-95	7740		
I-395	704		

On road-use level, five roadway physical characteristic, three environmental are used along with the PDVs are used for each of the three pavement types; ACP, JRCP and CRCP. For ACP there are twenty-one pavement distress variables for each of the four critical locations for a total of ninety-six predictor variables. For JRCP there are nineteen PDVs for a total of seventy-six predictor variables, and for CRCP there are eighteen PDVs for a total of seventy-two predictors. Descriptions of the common dependent and explanatory variables used in the analysis and the explanatory PDVs used in the

evaluation of the influence of PDVs on the outcomes of crashes, given a crash occurs, are provided in the four tables below.

Table 12. Common explanatory variables used in the analysis

	Variable	Description	
Dependent Variables			
	INJURY	Crash had an injury	1 if injury, 0 if no injuries
	REAREND	Crash type was rear-end	1 if rear-end, 0 if other
Explanatory Variables			
Road Use	SPEEDLIMIT	Posted speed limit of route	Mph
Physical Characteristics of the Road	LANECOUNT	Number of lanes	Count
	ALIGN	Alignment of the road	1= Straight, 0= all other
	WZ	Work zone present	1= Yes, 0= No
	SURFWIDTH	Width of roadway	Feet
	SHLDRWIDTH	Width of shoulder	Feet
Environmental	WEATHER	Weather conditions during crash	1= Clear, 0= Not clear
	SURFCOND	Condition of the riding surface	1= Dry, 0= Not dry
	LIGHTING	Light condition	1= Day, 0= Night

Table 13. Explanatory variables used in the analysis for ACP

Variables	Description	
IRI	Internal Roughness Index Average	Value
TC1	Transverse Cracking Severity 1	Presence=1, absence=0
TC2	Transverse Cracking Severity 2	Presence=1, absence=0
LC1	Longitudinal Cracking Severity 1	Presence=1, absence=0
LC2	Longitudinal Cracking Severity 2	Presence=1, absence=0
LLJ1	Longitudinal Lane Joint Severity 1	Presence=1, absence=0
LLJ2	Longitudinal Lane Joint Severity 2	Presence=1, absence=0
RTC1	Reflective Transverse Cracking Severity 1	Presence=1, absence=0
RTC2	Reflective Transverse Cracking Severity 2	Presence=1, absence=0
RTC3	Reflective Transverse Cracking Severity 3	Presence=1, absence=0
RLC1	Reflective Longitudinal Cracking Severity 1	Presence=1, absence=0
RLC2	Reflective Longitudinal Cracking Severity 2	Presence=1, absence=0
RLC3	Reflective Longitudinal Cracking Severity 3	Presence=1, absence=0
AC1	Alligator Cracking Severity 1	Presence=1, absence=0
AC2	Alligator Cracking Severity 2	Presence=1, absence=0
AC3	Alligator Cracking Severity 3	Presence=1, absence=0
PATCH	Patching Area - wheel path	Presence=1, absence=0
POT	Potholes Count	Presence=1, absence=0
DELAM	Delaminations Area	Presence=1, absence=0
BLEED1	Bleeding Severity 1	Presence=1, absence=0
BLEED2	Bleeding Severity 2	Presence=1, absence=0
RUT SE	Average Deeper Rut (Straight-edge)	Value
RUT WM	Average Deeper Rut (Wire method)	Value
CCI	Critical Condition Index	Value

Table 14. Explanatory variables used in the analysis CRCP

Variables	Description	
IRI	Internal Roughness Index Average	Value
TCS1	Transverse Cracking Severity 1	Presence=1, absence=0
TCS2	Transverse Cracking Severity 2	Presence=1, absence=0
TCS3	Transverse Cracking Severity 3	Presence=1, absence=0
TC TOT	Transverse Cracking Total Number	Presence=1, absence=0
TC SPACING	Transverse Crack Average Spacing	Presence=1, absence=0
LC1	Longitudinal Cracking Severity 1	Presence=1, absence=0
LC2	Longitudinal Cracking Severity 2	Presence=1, absence=0
LC3	Longitudinal Cracking Severity 3	Presence=1, absence=0
CCS1	Clustered Cracking Severity 1	Presence=1, absence=0
CCS2	Clustered Cracking Severity 2	Presence=1, absence=0
LJS	Longitudinal Joint Spalling	Presence=1, absence=0
PUNCHOUT	Punchouts and Spalled Ycracks	Presence=1, absence=0
PCCP1	PCC Patch Severity 1	Presence=1, absence=0
PCCP2	PCC Patch Severity 2	Presence=1, absence=0
PCCP3	PCC Patch Severity 3	Presence=1, absence=0
ASP PATCH	Asphalt Patching	Presence=1, absence=0
CDR	Index Values (Concrete Distress Rating)	Value
CPR	Index Values (Concrete Distress Rating)	Value
CCI	Critical Condition Index	Value

Table 15. Explanatory variables used in the analysis JRCP

Variable	Description	
IRI	Internal Roughness Index Average	Value
TCS1	Transverse Cracking Severity 1	Presence=1, absence=0
TCS2	Transverse Cracking Severity 2	Presence=1, absence=0
LCS1	Longitudinal Cracking Severity 1	Presence=1, absence=0
LCS2	Longitudinal Cracking Severity 2	Presence=1, absence=0
PCCP1	PCC Patch Severity 1	Presence=1, absence=0
PCCP2	PCC Patch Severity 2	Presence=1, absence=0
PCCP3	PCC Patch Severity 3	Presence=1, absence=0
ASP PATCH	Asphalt Patch	Presence=1, absence=0
NUMBER T_JTS	Number of Transverse Joints	Count
TJS	Transverse Joint Spalled	Presence=1, absence=0
LJS	Longitudinal Joint Spalled	Presence=1, absence=0
CBS1	Corner Breaks Severity 1	Presence=1, absence=0
CBS2	Corner Breaks Severity 2	Presence=1, absence=0
BLOWUPS	Blowups	Presence=1, absence=0
JFS1	Joint Fault Severity 1	Presence=1, absence=0
JFS2	Joint Fault Severity 2	Presence=1, absence=0
JFS3	Joint Fault Severity 3	Presence=1, absence=0
CCI	Critical Condition Index	Value

CHAPTER FOUR

4.0 EMPIRICAL SETTING AND MODELING METHODOLOGY

4.1 Empirical Setting

The modeling process began by reviewing the objectives of the study and the descriptive statistics to determine trends in the data, so the appropriate models can be used to make accurate inference from the sample data to the population. The primary purpose of this research is to determine if pavement condition, type and ride quality based on various PDVs are related to crash outcomes. As mentioned, this study uses the detailed pavement condition data collected annually by VDOT to assess overall pavement condition for each of the nine (9) Districts. Each of these PDVs are described/listed above in the previous chapter and used in this pavement condition assessment. It was shown in the literature review of previous transportation safety studies that the condition of the pavement is associated with both the frequency and severity of crashes. Therefore, because each of the PDVs listed affects the conditions of the pavement, it is the primary hypothesis of this study that each of these PDVs could possibly play a role in this relationship.

In the data aggregation procedure described in the previous chapter the pavement condition data and the crash data collected at the site of the crash were combined by correlating a milepost to the pavement condition data collect every 10^{th} of a mile. By determing a common milepost the spatial aspects of the data are confirmed. However, crash data are collected at, or near, the time of the crash, and the pavement condition data

are collected annually. This annual data is used for future planning purposes as funding allows for pavement maintenance and not for scheduling immediate or regular maintenance activities. While there may be some areas that need and receive immediate attention, the majority of the interstate segments in this research do not fall into this category. Therefore, this study makes a reasonable assumption that the annual pavement condition data as tabulated is indicative of the actual pavement condition at the time of the crash. In addition, only a few PDVs would be critical enough to require immediate remedial actions (e.g., potholes), thus nullifying this assumption. For those interstate roadway segments that could possibly require more frequent maintenance, the results and conclusions may not be as robust. This will be noted when interpreting the results in the following chapters.

Like most traffic safety problems, this study uses the traditional three-dimensional approach which includes exposure, crash risk and injury consequence. In this case, exposure is the condition and ride quality of the pavement along with socio-economic environment where the crash occurs. Risk is defined as the relationship between the crash and exposure. However, in order to be able to compare and rank road safety problems it is also necessary to find out the magnitude and character of the activities that generate the problems – the exposure (OECD, 1997). Therefore, the empirical relationship (magnitude and significance) between the crash consequences and these exposure elements shall be used to predict an individual's risk of having a crash with a certain outcome. This study is limited to the major interstate routes in Virginia where accurate and robust data are available for the years 2007 and 2008.

4.2 Modeling Methodology Overview

This section explains and outlines the development of the statistical models used to obtain the goals of this research. Presented in this chapter, and subsequent chapters, is general information on traffic safety modeling techniques, brief description/reasons for applying multilevel models, and a more detailed description of the models developed to meet the objectives of this research. So that we might discover un-bias associations between the crash outcome response variables and the many pavement distress predictor variables included in the datasets, it is important to note that this research makes a primary hypothesis but is an exploratory test that makes no a-priori assumptions.

As mentioned, traffic safety has promulgated an enormous amount of research. To date, the factors shown to have a direct relationship to vehicle crashes have been grouped into three primary categories: human, environmental/roadway and traffic. From a general quality of life issue it is critical that research focus on ways to better understand the detailed dynamics of the crash so countermeasures can be identified to reduce the number and/or severity of accidents. Unfortunately, capturing the amount of data necessary to comprehensively account for each of these primary categories is extremely time consuming and costly and, therefore, in most cases, is not available. In fact, only recently has there been a movement to initiate studies to focus on the fundamental questions relating to the casualty of crashes – the casual mechanisms – based on naturalistic driving information.

As a result, in order to handle the spatial and temporal elements associated with crashes and ensure adequate data are available for the estimation of statistical models, researchers have framed their approach to study the casual mechanisms occurring in a specific area over a specified time frame (Lord and Mannering, 2010). To date, two general types of models are used: models to predict the frequency/totals of crashes (predictive model) and ones to estimate the significance of the factors related to the severity or type of crashes (consequential model i.e., based on the post-crash condition given a crash occurred). While crash frequency modeling offers many benefits, it only provides weighted average effects of various factors on crashes (Kim, et al., 2007). It is limited in its ability to quantify the significance of the relationships between specific crash types/severity and various roadway characteristics. With this being the primary objective of this study, this research will use crash-level data to develop statistical models focused on predicting which pavement condition factors increase/decrease the probability of a certain crash outcome. Additionally, and because different crash types require different safety countermeasures (Kim, et al., 2006), this will provide unique information to traffic safety engineers allowing them to target countermeasures more appropriately.

This research is predicated on the premise that condition, type of pavement and ride quality of the pavement surface have an effect on the outcome of a crash and uses crash data combined with pavement condition data. The crash outcomes (dependent variables) are binary variables that are discrete, non-normally distributed. Therefore, to estimate the significance of the independent variables (pavement condition and ride quality) on discrete binary outcomes, non-linear conditional consequence models shall be applied. In

these models, a complex matrix of explanatory variables using only information about the crash are analyzed to determine which factor(s) may influence the probability of having one type of crash.

The most widely used consequence model used for this application is the logistic regression model where the odds-ratios are calculated in order to test whether one phenomenon has an effect on another. In other words, it estimates the probability an outcome (dependent variable) will occur with certain characteristics (independent variable(s)). However, it does not make exact predictions, only probabilities within a specified confidence level. For example, if the speed(s) of the vehicle(s) are known from the crash report, the model will estimate by what amount the odds of having a rear-end crash are increased or decreased due to going too fast. However, it cannot predict how many rear-end crashes will occur from speeding (or other factors), and care has to be taken to ensure it does not over simplify the complex nature of these events resulting in conclusions that could be potentially misinterpreted (Jones and Jorgensen, 2001).

Additionally, it is important to thoroughly analyze each crash type in order to identify specific sites that may be at risk because different crash types at different locations require different safety precautions. There is a-priori reason to believe that roadway, environmental and traffic variables are associated with different crash types (Kim, et al., 2007). This research, for the reasons explained below, will explore in more detail the effects of the location of the crash by exploiting the hierarchical nature of crash to analyze the land use/socioeconomic factors that may impact the number and

consequences of crashes. This will provide unique and valuable insights so that effective countermeasures can be identified. This objective is a common one to most research in traffic safety, but seeking to model crashes for various types of outcomes can lead to potentially serious statistical problems because there is an obvious correlation between injury severities and collision type (Lord and Mannering, 2010). To account for these cross-model correlations and accomplish the third objective of this study, multilevel modeling will be employed.

In the literature, multilevel models are referred to as hierarchical linear models (HLM) and that is what will be used throughout this dissertation. In highway safety, arguably the crash data have a nested or hierarchical structure with several levels of hierarchy. Additionally, there is a strong correlation that exists between crashes that occur under the same kind of human, environment/roadway, and traffic conditions. HLM is a type of regression model formulated to account for variation between crashes within each "nest" or "cluster" where the pattern of clustering is known. It is specifically designed to capture correlations among these groups of data leading to better and unbiased parameter coefficient estimates and standard errors. Further discussion of this technique and a procedure for applying it are presented in Chapter Six.

In order to thoroughly analyze the outcome of a crash and test the main hypotheses of this research, three regression-like consequence empirical models are applied because of their appropriateness in investigating and quantifying the relationship between two or more variables. The research objectives are:

1. Given a crash occurs, evaluate the relationship between pavement condition, type of pavement and ride quality on crash outcomes by estimating the strength and significance of various pavement condition distress parameters along with the type of pavement and ride quality as expressed by IRI.
2. To explore the spatial aspect of this relationship, models will be developed to evaluate crash outcomes at the crash site and specific intervals upstream. These models will be used to evaluate where within the footprint of the crash is the most critical location with respect to pavement condition and/or ride quality.
3. In conjunction with goals one and two, uniquely model the hierarchical nature of a crash to determine which regional socio-economic factors are related to traffic safety.

For the first and second objectives this study applies binary logistic modeling for non-linear dichotomous and non-dichotomous dependent variables using maximum-likelihood estimation (MLE). The dependent variables are turned into logits (natural log of the odds of the event occurring) and MLE applied, subsequently. This methodology is widely used to determine the relationship between a single dependent variable and several independent, or predictor, variables.

4.3 Conceptual Structure for Modeling

Figure 7 below presents the modeling structure that will be used to test the study's first and second hypotheses: the condition and/or ride quality of the pavement has an association with certain crash outcomes, and the condition and/or ride quality of the pavement may not be critical at the crash site but at some location upstream. Figure 8

presents the conceptual structure for the third hypothesis: the socio-economic characteristics of the county where the crash occurred have a relationship to the crash outcome.

These diagrams illustrate the causal hypothesized relationship between the predictor and response variables. The double straight arrows indicate a possible correlated causal relationship between two or more predictor variables, and the single straight arrows indicate the causal association between a predictor variable and outcome variable. Potential causal relationships between variables may consist of direct and indirect effects (Lleras, 2005). In this study, for example, a rear-end crash is assumed to be dependent directly on a pavement distress, roadway or environmental predictor variable. Then, given a rear-end crash occurs, the crash could be a crash with or with-out an injury. Continuing with this example as illustrated, a rear-end crash indirectly affects a crash through its relationship with an injurious crash. For illustration purposes, these two figures show the path to a crash with an injury by this approach, but one should understand there are direct relationships between other indicators. For example, a crash with an injury could be directly associated with a particular PDV or roadway characteristic. Socio-economic predictor variables are included in Figure 8 to illustrate the hierarchical modeling structure with correlation arrows connecting the level-2 county variables with the level-1 crash variables.

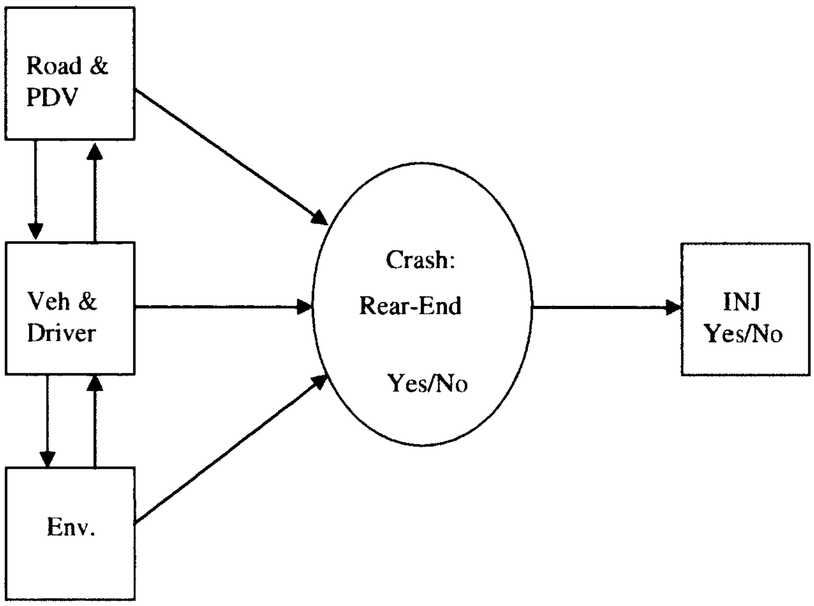


Figure 7. Conceptual Structure for crashes and injuries.

Acronyms:

PDV = Pavement Distress Variables

Veh & Driver = Vehicle and Driver Variables

Road = Roadway Variables

Env. = Environmental Variables

Rear-End = Rear-End Crash

INJ = Crash with Injury

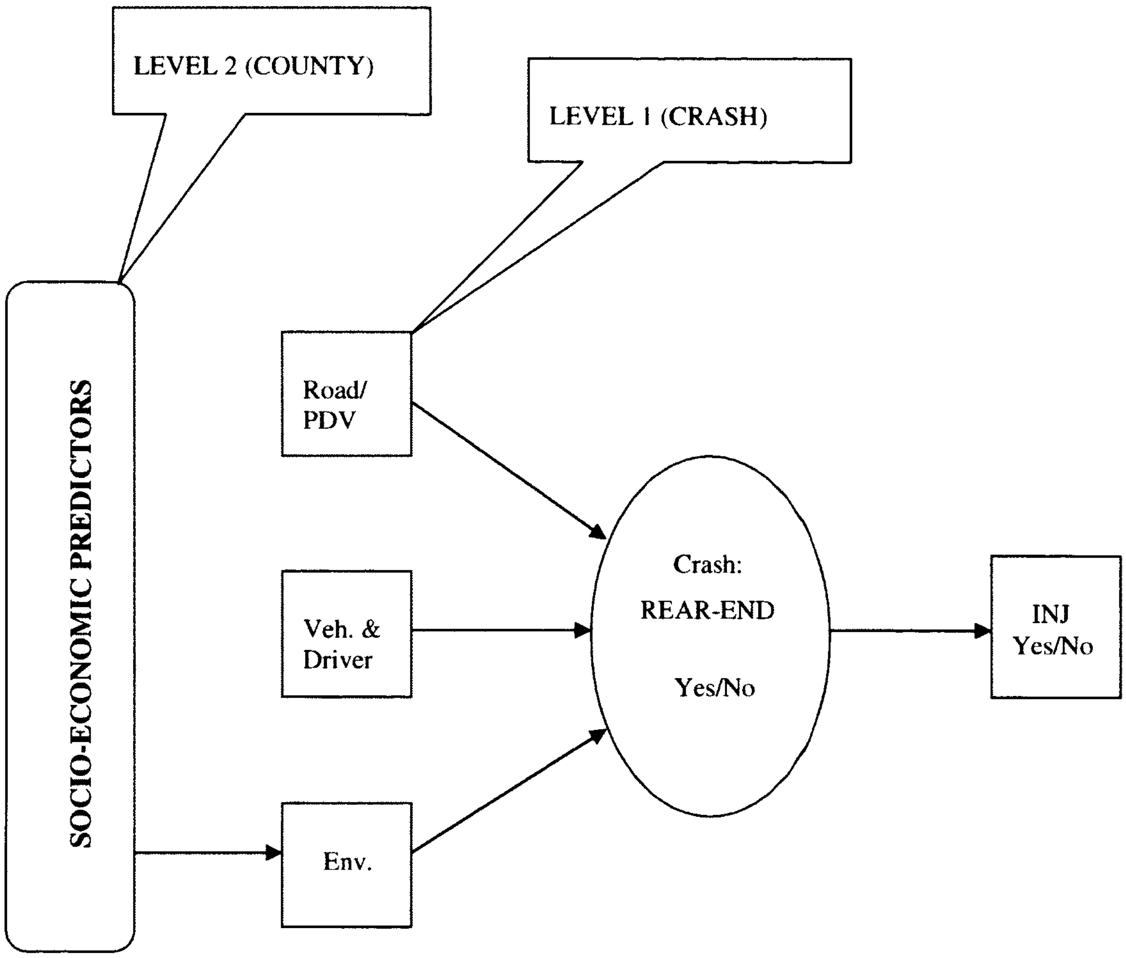


Figure 8. Conceptual Structure capturing levels of associated factors.

4.4 Logistic Regression Modeling Methodology

This study focuses on the severity of the outcome of a crash, given a crash occurs, by chosing the binary event (yes or no) for a crash with an injury. Additionally, this study determines if there is a relationship to pavement condition, type and ride quality on the specific type of crash rear-end. Current statistics show, when compared to other crash types like head-on and sideswipes, rear-end crashes are argueably the most common type of accident - particularly on high-speed freeways. It has been shown that a large percentage of rear-end crashes are caused, or contributed to, by inattentive drivers, or the inability to cope with defects in the pavement (Pant and Panta, 2009). Reducing the risk of rear-end crashes while driving on high-speed, high-volume, interstates where travel speeds vary requires all the senses. In the current fast-paced environment, impatience, multi-tasking, and inattentiveness is more common than ever.

In addition, traveling at high speeds increases the risk of a rear-end crash because it drastically decreases the vehicle's stopping distance and the driver's ability to brake on time. Historically, this type of crash, on these types of roadways, has the potential to cause multi-vehicle crashes and crashes with costly damages and injuries. With rear-end crashes being common and having the potential to cause major consequences, they are currently the subject/concern for many DOT agencies. Choosing this type of crash adds to the timeliness attribute of this research.

The models were developed to further exploit the casual mechanisms of a crash and examine the type and severity of a crash. The above binary dependent variables for crash

severity and type with discrete outcomes were chosen and logistic regression chosen as the modeling methodology. This technique is a maximum-likelihood estimation method that has been used for many crash consequence studies and has been recognized for many years in its ability to provide significant results in the field of traffic safety. It is relatively free from any restriction and allows the prediction of a discrete outcome from independent variables that may be continuous, discrete, or a mix (Nowakowska, 2010).

This multivariate technique was chosen for its ability to assess the affect of pavement condition and/or ride quality on the type and severity of crashes while controlling for other variables, accounting for interdependencies, and allowing examination of interactions. The goal is to predict the probability of the response variables using the most parsimonious model. The pavement condition and ride quality variables at the site of the crash are the primary predictors, along with the characteristics of the environment and roadway, given a crash occurs. As mentioned above, the logistic regression procedure predicts the logit of the binary dependent variable and the log odds of the dichotomous event occurring.

The output from the statistical model estimates the parameters (b coefficients) which are used to forecast the logit of the dependent variable. The logistic prediction equation is:

$$Y = Logit = z = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \ldots\ldots \gamma_k X_k$$

where: γ_0 = the constant

$\gamma_{1,} \gamma_{2,} \ldots \gamma_k$ = unstandardized logistic regression coefficients

X = independent variables, also called predictors

$\mathrm{Exp}(\gamma)$ = the odds ratio of a predictor variable

$\mathrm{Exp}(z)$ = the odds that the event (e.g. the particular crash outcome).

There are two ways to enter dichotomous variables: as covariates or categorical factors. The results are interpreted depending on how they are coded. For this research, they are entered as dichotomous covariates. The first step in determining a relationship using the binary logit model is to assign a value of 1 or 0 to the dependent variable. For example, modeling crash type with an INJURY = 1 and crash with NO INJURY = 0 the binary logistic regression estimates the probability of the "1" value (predicted), using the "0" value as the reference. Once the model predicts the logit, calculating the Exp (z) determines the odds the dependent variable equals the "1" value rather than the reference level "0" value.

The "γ" is also called the parameter estimate and calculating Exp (γ) gives the odds ratio for the independent variable. Exp (γ) is the factor by which the independent variable increases or decreases the binary outcome variable (logit). It is the estimate of the change in the log odds of the event occurring (predicted value 1, rather than the reference 0). For continuous independent variables, it is the change in the logit resulting from a unit change in the independent variable. The sign of the "γ" parameter determines whether the independent variables increase or decrease the odds that the event occurs.

In this research, the dependent variables INJURY and REAR-END are binary and coded 1 for the event occurring and 0 for the event not occurring. For example, given a crash

occurs, if there is even a single injury, the dependent variable INJURY is coded a 1 and if there are no injuries, it is coded a 0. The 1 is the highest number; therefore, the model uses an injury occurring as the predicted outcome and the model results will be interpreted as the odds of an injury occurring. The primary objective of this research is to understand the relationship between pavement condition, type and ride distress variables on specific crash outcomes that or related to the severity of the crash. Therefore, the model will estimate the probability a crash outcome due to the presence of a PDV. Given a crash occurs, if the odds ratio of a PDV has a positive sign it means that the presence of that particular PDV increases the probability of the dependent crash outcome and vice versa if the sign in negative. In addition and in accordance with the conceptual structure, the dependent variable REAR-END was included as and independent variable in all INJURY models. This was done to estimate the likelihood of having a rear-end crash that resulted in an injury. Therefore, given a rear-end crash, do the odds of having an injury increase or decrease?

As presented in the literature review, there have only been a few studies conducted evaluating the condition of pavement or type of pavement as it relates to traffic safety. Because this research is uniquely exploring this relationship at the micro-level, using detailed pavement condition data there is no previous research to guide in trying to determine which pavement condition, type and ride quality variables to include in the models. Therefore, for exploratory purposes a stepwise logistic regression approach was used which determines automatically which variables to use in the model or drop. The forward stepwise method was chosen starting with the base constant –only model and

entering independent variables one at a time until the variables not in the model have a statistical significance greater than 0.05. It uses the likelihood ratio test (chi-square difference) restimating the -2LL as target variables are entered or removed.

As mentioned, this study limits the data to interstate routes to increase the control of a number of potentially correlated predictor variables. In addition, this study will be aggregating pavement condition data with crash information for these two roadway classifications for each of the three main pavement types: asphalt concrete (ACP) (including bituminous concrete overlays over concrete pavement (BOJ) and (BOC)), continuously reinforced concrete (CRCP), and jointed reinforced concrete (JRCP).

To evaluate the effect of pavement type on crash outcomes a dataset was created using all the crash data and PDVs for each of the three pavement types: ACP, CRCP and JRCP. The objective is to estimate the relationship between the type of pavement and roadway safety based on how it affects the study's two crash outcomes: crashes with injuries and rear-end crashes. For this analysis CRCP and JRCP were considered to be one type of pavement: concrete. The dummy predictor variable PTYPE was included in the dataset along with the crash data and PDVs. For this predictor variable (PTYPE) a "0" was given to all crash cases occurring on asphalt pavement and a "1" for crashes on concrete pavement. The variable PTYPE was entered into two binary logistic regresson models as a dichotomous covariate with the asphalt pavement being the reference. The dependent variables are INJURY and REAR-END with REAR-END included as an independent variable in the INJURY model.

The second objective of this research is to gain a better understanding of the spatial component of the pavement condition as it relates to crash outcomes. Therefore, the pavement condition was modeled at the site of the crash and specific intervals upstream of the crash. As described above, specific dataset combining pavement condition and ride quality data and crash data were developed to determine where in the footprint of the crash is the significance of the pavement condition optimal to the outcome of the crash. This was accomplished by developing crash prediction models using the same logistic regression modeling techniques as described above with the exception that only pavement distress variables were included as predictors. The locations of interest within the footprint of the crash are at the site of the crash and 0.10, 0.15, and 0.20 miles upstream of the crash. As previously mentioned, the pavement type for the majority of the study's interstate roadways is asphalt and subsequently yields the largest sample of crashes. Therefore, this effort will be limited to asphalt pavement.

Again, the dependent variables INJURY, and REAR-END are binary and coded 1 for the event occurring and 0 for the event not occurring. For example, in the event of a crash, if there is even a single injury, the dependent variable INJURY is coded a 1, and if there are no injuries, it is coded a 0. The 1 is the highest number; therefore, the model uses an injury occurring as the predicted outcome and the model results will be interpreted as the odds of an injury occurring. The logit, or log odds determined from the regression coefficients "γ " will estimate if the corresponding independent variable increases or decreases the odds of the injury occurring. A total of eight (8) models were developed; one for each of the two dependent variables INJURY, and REAR-END at each of the

four locations within the footprint of the crash. As mentioned, this research objective is limited to the interstates with asphalt pavements and the descriptive statistics are the same as presented for objective one.

To determine the critical location for assessing the importance of pavement condition and ride quality on traffic safety each model was reviewed for best fit and significance. The Hosmer and Lemeshow (H-L test) chi-square test of goodness of fit, the pseudo R squares and the classification table percent correct were used to evaluate the model's fit. The H-L test is recommended for overall fit of a binary logistic regression model (statnotes NC State) but it is important to also review the pseudo R squared values and percent complete from the classification table. Each of these indicators is presented in the model summaries presented in Chapter Five.

CHAPTER FIVE

5.0 DATA ANALYSIS AND RESULTS FOR LOGISTIC MODELING

5.1 Raw Data Analysis

As mentioned, previous literature suggests, and it is the primary hypothesis of this study, that pavement condition has either a direct or an indirect association with the safety of the roadway. In order to study this relationship further, unique datasets were created that combine PDVs with crash, environmental and roadway characteristics. These datasets were created at the crash site and specific intervals upstream of the site in order to evaluate the spatial component of this phenomenon. This data will serve as the basis for the evaluation of this relationship; therefore, it is important to examine the descriptive statistics of this raw data for bias potential.

5.2 Descriptive Statistics for Binary and Multinomial Logistic Models

The following tables present the summary statistics used in the binary and multinomial logistic models for research objectives one and two. The tables are presented in order according to pavement type. They provide the descriptive statistics for the dependent variables and the roadway characteristics predictor variables at the crash site and predictor PDVs at each of the four (4) critical locations: at the crash site and 0.10, 0.15 and 0.20 miles upstream of the crash. Tables 16-20 list the descriptive statistics for ACP. Tables 21-25 list the summary statistics for JRCP, and tables 26-30 present the statistics for CRCP.

5.2a Crash, Roadway and Environmental Data

Crash data for the two (2) year period of 2007 and 2008 was collected from the police reports filed by Virginia State Troopers. This study focuses on determining the significance of the relationship pavement conditions have on these crashes, so this two year period coincides with the same period for collecting the pavement performance data. This sets the temporal boundaries for this study. In addition, the crash data was reduced to include only the crashes on interstate routes where data are available and provided a wide range of scenarios that will improve the sample. This helps define the spatial boundaries of the study. Cross section and roadway alignment variables along with environmental, pavement type and other related predictor variables pertaining to the crash are included in the crash database.

In order to develop policies and traffic safety programs that improve the safety and operation of the nation's roadways, stakeholders need high-quality data to use a basis for statistical analysis and ultimate decision-making. Therefore, it is important to record complete, accurate and timely crash information so the potential measurement errors in the variables used in statistical studies can be reduced. Although for the most part all states and localities endeavor to collect high-quality crash data there are many inconsistencies in the way they collect it. In Virginia, one difference affects this study. Virginia does not use the KABCO scale for measuring injuries which in past studies has been proven reliable when compared to the actual injury diagnosed in a follow-up with a physician. Instead, they use a progressive 1 to 4 severity scale with 1 being a crash where there is a fatality before report is made; 2 for crashes with visible signs of injury,

such as a bleeding wound or distorted member or the injured had to be carried from the scene; 3 for other visible injury, such as bruises, abrasions, swelling, limping, etc; and 4 for no visible injury but complaint of pain or momentary unconsciousness. This procedure leads to inconsistency between the injury report at the crash and the actual injury. This should be considered when interpreting the results of this study.

To ensure the integrity of the crash data, quality control checks were made on the data and crashes with outliers were removed. For example, some crash descriptors had the number 999 listed, in some cases the surface width was zero, and for some single vehicle crashes the damage amount was well over $100,000.00. In cases like this, because accurate data was unavailable, the entire case was removed from the database. However, there were a number of cases where the speed was coded as zero. Investigating this further determined the speed was coded this way because it was probably unknown at the time of the crash. Because this is not a primary predictor variable, these cases were left in the dataset. Further reductions of the crash data are expected while conducting the research to ensure only relevant crashes are included in the analysis. Thus, the final sample of crashes shall include the necessary information to meet our study objectives.

From the descriptive statistics for two dependent variables on each of the three types of pavement in the sample, follow national trends with crashes with injuries being the fewest at approximately 30% followed by rear-end crashes ranging from approximately 40% to 60%. One would expect to find roughly the same number of rear-end crashes as crashes with injuries because one would expect rear-end crashes to result in at least one

injury. However, it is noteworthy that in this dataset there is a wide range of difference depending on the type of pavement. On asphalt pavement there is a 10% difference, and on concrete pavement the difference almost doubles from 32% to 59%. One explanation may be that in rear-end crashes it is common for the injury to be actually felt/realized days after the accident occurs and is therefore not reported on or added to the crash report.

In addition, the road use, physical characteristics and environmental variables collected from the crash data sample at the crash site are reasonable for interstate highways in Virginia for these years. The mean speed ranges between 55-60 miles per hour; the alignment of the roadways is predominately straight; there are a reasonable number of crashes in work zones at 4%; and the width of the pavement and shoulder averages are within design standards. In addition, the majority of crashes happen in the daytime when the volume of traffic is highest, on predominately-dry surfaces and on clear/non-rainy days. After the data was reduced from the quality control checks, there remained 21,541 crash cases on ACP, 5,190 on CRCP and 7,349 on JRCP.

Overall, using crash data of this type is an accepted methodology for studying the relationships of crash characteristics to independent crash variables given a crash has occurred on a specific roadway segment. This data set is similar to the crash data used in a number of the studies listed in the literature review and this sample size allows estimation of a population in statistically significant models.

5.2b Pavement Condition/Distress Data

As mentioned in the introduction, there are numerous factors affecting general road safety, and research in roadway safety is categorized into three main categories, vehicle, driver and environment each with their own set of sub-variables. The condition of the pavement is a general roadway condition factor and falls within the environmental section of the roadway safety matrix. As mentioned in the introduction, the driver's ability to collect information and carry out intended maneuvers is greatly compromised due to the vibrations encountered on rough roadways (TRB, 2009). Studies have shown that roughness affects safety in many ways particularly the ability of the driver to steer and brake which can significantly affect the overall controllability of the vehicle (Burns, 1981).

In theory, each of the predictor PDVs described in detail in section 3.2d above could compromise the smoothness of the roadway and thus have an effect on its safety. This is the primary hypothesis of the research. Additionally, the effect of inconsistency traveling from a smooth section of pavement to a rough section has the potential to hinder the driver's ability to control a vehicle, and at high speeds this effect is compounded. Therefore, the presence of one or more of the pavement distress variables (at various severity levels) presented in this study may have a bearing on roadway safety.

The PDVs were described in detail above, and they were combined with the crash data at the crash site and three specific intervals upstream of the site. As previously mentioned, there are ninety-six predictor PDVs for ACP, seventy-six for JRCP, and seventy-two for

CRCP. In the tables below, the PDVs at the crash site are simply designated with the acronym listed in the description tables above, but for those located upstream of the crash site an "_2" is added after the acronym for those located 0.01 mile from the crash site, an "_3" for those located 0.15 miles upstream, and an "_4" for those 0.20 miles upstream.

These factors are originally collected and coded as continuous variables, but for this study the majority were converted to categorical/binary predictor variables to determine if the presence of the distress factor alone has a relationship to the response variables (crash outcomes). The ride quality as expressed in IRI and the overall pavement condition as expressed in CCI remain continuous variables along with the distress variable rutting (RUT SE and RUT WM) for ACP, the variable number of transverse joints for JRCP, and the two components CDR and CPR that make up the CCI for CRCP. For each of the predictor PDVs, descriptive statistics were computed and examined. The minimum, maximum, and mean computed for the continuous variables and minimum, maximum and percent for the categorical variables. The percent value presented is the percent of all crashes where a particular distress variable is present. For example, 34.21% of all crashes had the TC1 pavement condition at the site of the crash.

The distress variable alligator cracking AC has the highest presence percentage on ACP followed by transverse cracking TC/RTC and longitudinal cracking LC/RLC. For JRCP the distress variables with the highest presence percentages are spalled longitudinal joints LJS with 61%, spalled transverse joints TJS at 60%, asphalt patch ASP PATCH at 43% and corner breaks CBS and patching PCCP at 35%. The distress variables with the

highest presence percentages on CRCP are transverse cracking TCS at 88%, clustered cracking CCS at 64% and asphalt patching ASP PATCH at 49%. These distress variables, especially the ones with relatively high percentages, are likely to be statistically significant in the relationship between pavement condition and the response crash outcome variables.

Table 16. Descriptives Statistics for D.V.s and Roadway I.V.s (ACP)

	Variables	**N**	**Minimum**	**Maximum**	**No.**	**%**
Dependent Variables						
	INJURY	21541	0.00	1.00	6777.00	32%
	REAREND	21541	0.00	1.00	8883.00	41%
Road Use					Mean	Std. Dev.
	SPEEDLIMIT	21541	0.00	65.00	59.18	7.33
Physical Characteristics of the Road					No.	%
	ALIGN	21541	0.00	1.00	15080.00	70%
	WZ	21541	0.00	1.00	956.00	4%
					Mean	Std. Dev.
	SURFACEWIDTH	21541	12.00	86.00	33.20	9.20
	SHOULDERWIDTH	21541	0.00	11.00	5.09	0.82
Environmental					No.	%
	WEATHER	21541	0.00	1.00	17066.00	79%
	SURFCOND	21541	0.00	1.00	16459.00	76%
	LIGHTING	21541	0.00	1.00	13800.00	64%

Table 17. Descriptives for PDVs at the Crash Site (ACP)

Variables	N	Minimum	Maximum	% / Mean
IRI	21541	27.00	435.00	100.42
TC1	21541	0.00	1.00	34.21%
TC2	21541	0.00	1.00	22.05%
LC1	21541	0.00	1.00	30.15%
LC2	21541	0.00	1.00	24.92%
LLJ1	21541	0.00	1.00	8.19%
LLJ2	21541	0.00	1.00	0.25%
RTC1	21541	0.00	1.00	31.97%
RTC2	21541	0.00	1.00	18.05%
RTC3	21541	0.00	1.00	13.82%
RLC1	21541	0.00	1.00	29.91%
RLC2	21541	0.00	1.00	13.10%
RLC3	21541	0.00	1.00	3.33%
AC1	21541	0.00	1.00	54.62%
AC2	21541	0.00	1.00	37.99%
AC3	21541	0.00	1.00	8.00%
PATCH	21541	0.00	1.00	13.70%
POT	21541	0.00	1.00	0.17%
DELAM	21541	0.00	1.00	1.05%
BLEED1	21541	0.00	1.00	0.71%
BLEED2	21541	0.00	1.00	0.02%
RUT SE	21541	0.00	0.69	0.13
RUT WM	21541	0.00	0.71	0.20
CCI	21541	0.00	100.00	76.59

Table 18. Descriptives for PDVs 0.10 miles upstream of Crash Site (ACP)

Variables	N	Minimum	Maximum	% / Mean
IRI_2	21541	27.00	435.00	100.09
TC1_2	21541	0.00	1.00	34.21%
TC2_2	21541	0.00	1.00	22.01%
LC1_2	21541	0.00	1.00	30.05%
LC2_2	21541	0.00	1.00	24.62%
LLJ1_2	21541	0.00	1.00	8.31%
LLJ2_2	21541	0.00	1.00	0.25%
RTC1_2	21541	0.00	1.00	31.99%
RTC2_2	21541	0.00	1.00	18.35%
RTC3_2	21541	0.00	1.00	13.82%
RLC1_2	21541	0.00	1.00	29.93%
RLC2_2	21541	0.00	1.00	13.18%
RLC3_2	21541	0.00	1.00	3.54%
AC1_2	21541	0.00	1.00	54.99%
AC2_2	21541	0.00	1.00	38.02%
AC3_2	21541	0.00	1.00	8.19%
PATCH_2	21541	0.00	1.00	13.83%
POT_2	21541	0.00	1.00	0.12%
DELAM_2	21541	0.00	1.00	0.90%
BLEED1_2	21541	0.00	1.00	0.71%
BLEED2_2	21541	0.00	1.00	0.02%
RUT SE_2	21541	0.00	0.75	0.14
RUT WM_2	21541	0.00	0.75	0.20
CCI_2	21541	0.00	100.00	76.54

Table 19. Descriptives for PDVs at 0.15 mile upstream of Crash Site (ACP)

Variables	N	Minimum	Maximum	% / Mean
IRI_3	21541	29.00	435.00	100.03
TC1_3	21541	0.00	1.00	34.49%
TC2_3	21541	0.00	1.00	22.05%
LC1_3	21541	0.00	1.00	30.11%
LC2_3	21541	0.00	1.00	24.77%
LLJ1_3	21541	0.00	1.00	8.32%
LLJ2_3	21541	0.00	1.00	0.23%
RTC1_3	21541	0.00	1.00	32.19%
RTC2_3	21541	0.00	1.00	18.07%
RTC3_3	21541	0.00	1.00	13.65%
RLC1_3	21541	0.00	1.00	30.30%
RLC2_3	21541	0.00	1.00	13.03%
RLC3_3	21541	0.00	1.00	3.56%
AC1_3	21541	0.00	1.00	55.24%
AC2_3	21541	0.00	1.00	38.16%
AC3_3	21541	0.00	1.00	8.19%
PATCH_3	21541	0.00	1.00	13.75%
POT_3	21541	0.00	1.00	0.16%
DELAM_3	21541	0.00	1.00	1.07%
BLEED1_3	21541	0.00	1.00	0.70%
BLEED2_3	21541	0.00	1.00	0.03%
RUT SE_3	21541	0.00	0.69	0.14
RUT WM_3	21541	0.00	0.71	0.20
CCI_3	21541	0.00	100.00	76.52

Table 20. Descriptives for PDVs at 0.20 mile upstream of Crash Site (ACP)

Variables	N	Minimum	Maximum	% / Mean
IRI_4	21541	28.00	435.00	99.92
TC1_4	21541	0.00	1.00	34.42%
TC2_4	21541	0.00	1.00	22.06%
LC1_4	21541	0.00	1.00	30.21%
LC2_4	21541	0.00	1.00	24.82%
LLJ1_4	21541	0.00	1.00	8.29%
LLJ2_4	21541	0.00	1.00	0.22%
RTC1_4	21541	0.00	1.00	32.19%
RTC2_4	21541	0.00	1.00	18.02%
RTC3_4	21541	0.00	1.00	13.72%
RLC1_4	21541	0.00	1.00	30.15%
RLC2_4	21541	0.00	1.00	12.93%
RLC3_4	21541	0.00	1.00	3.47%
AC1_4	21541	0.00	1.00	55.27%
AC2_4	21541	0.00	1.00	37.90%
AC3_4	21541	0.00	1.00	8.24%
PATCH_4	21541	0.00	1.00	13.58%
POT_4	21541	0.00	1.00	0.16%
DELAM_4	21541	0.00	1.00	1.07%
BLEED1_4	21541	0.00	1.00	0.78%
BLEED2_4	21541	0.00	1.00	0.04%
RUT SE_4	21541	0.00	0.69	0.14
RUT WM_4	21541	0.00	0.71	0.20
CCI_4	21541	0.00	100.00	76.48

Table 21. Descriptive Statistics D.V.s and Roadway I.V.s (JRCP)

	Variables	**N**	**Minimum**	**Maximum**	**No.**	**%**
Dependent Variables						
	INJURY	7349	0.00	1.00	2276.00	31%
	REAREND	7349	0.00	1.00	4082.00	56%
Road Use					Mean	Std. Dev.
	SPEEDLIMIT	7349	0.00	65.00	56.26	6.93
Physical Characteristics of the Road					No.	%
	ALIGN	7349	0.00	1.00	5627.00	77%
	WZ	7349	0.00	1.00	0.00	0%
					Mean	Std. Dev.
	SURFACEWIDTH	7349	22.00	72.00	35.68	11.07
	SHOULDERWIDTH	7349	0.00	8.00	4.97	0.57
Environmental					No.	%
	WEATHER	7349	0.00	1.00	6091.00	83%
	SURFCOND	7349	0.00	1.00	5959.00	81%
	LIGHTING	7349	0.00	1.00	5040.00	69%

Table 22. Descriptive Statistics PDVs at the Crash Site (JRCP)

Variable	N	Minimum	Maximum	% / Mean
IRI	7349	42.00	381.00	151.33
TCS1	7349	0.00	1.00	34.20%
TCS2	7349	0.00	1.00	20.52%
LCS1	7349	0.00	1.00	27.77%
LCS2	7349	0.00	1.00	6.01%
PCCP1	7349	0.00	1.00	34.41%
PCCP2	7349	0.00	1.00	21.13%
PCCP3	7349	0.00	1.00	9.38%
ASP PATCH	7349	0.00	1.00	44.32%
NUMBER T_JTS	7349	0.00	56.00	12.25
TJS	7349	0.00	1.00	60.73%
LJS	7349	0.00	1.00	37.77%
CBS1	7349	0.00	1.00	14.55%
CBS2	7349	0.00	1.00	10.41%
BLOWUPS	7349	0.00	1.00	0.76%
JFS1	7349	0.00	1.00	36.78%
JFS2	7349	0.00	1.00	3.93%
JFS3	7349	0.00	0.00	0.00%
CCI	7349	0.00	100.00	67.97

Table 23. Descriptive Statistics PDVs at 0.10 mile upstream (JRCP)

Variable	N	Minimum	Maximum	% / Mean
IRI_2	7349	50.00	381.00	152.84
TCS1_2	7349	0.00	1.00	34.25%
TCS2_2	7349	0.00	1.00	21.45%
LCS1_2	7349	0.00	1.00	27.38%
LCS2_2	7349	0.00	1.00	5.59%
PCCP1_2	7349	0.00	1.00	33.98%
PCCP2_2	7349	0.00	1.00	22.08%
PCCP3_2	7349	0.00	1.00	9.44%
ASP PATCH_2	7349	0.00	1.00	43.00%
NUMBER T_JTS_2	7349	0.00	56.00	12.19
TJS_2	7349	0.00	1.00	61.23%
LJS_2	7349	0.00	1.00	60.82%
CBS1_2	7349	0.00	1.00	37.03%
CBS2_2	7349	0.00	1.00	22.06%
BLOWUPS_2	7349	0.00	1.00	13.76%
JFS1_2	7349	0.00	1.00	11.04%
JFS2_2	7349	0.00	1.00	0.60%
JFS3_2	7349	0.00	1.00	31.60%
CCI_2	7349	0.00	100.00	61.01

Table 24. Descriptive Statistics PDVs at 0.15 miles upstream (JRCP)

Variable	N	Minimum	Maximum	% / Mean
IRI_3	7349	42.00	381.00	152.80
TCS1_3	7349	0.00	1.00	33.96%
TCS2_3	7349	0.00	1.00	20.86%
LCS1_3	7349	0.00	1.00	27.16%
LCS2_3	7349	0.00	1.00	5.58%
PCCP1_3	7349	0.00	1.00	34.09%
PCCP2_3	7349	0.00	1.00	21.92%
PCCP3_3	7349	0.00	1.00	9.46%
ASP PATCH_3	7349	0.00	1.00	43.16%
NUMBER T_JTS_3	7349	0.00	56.00	12.17
TJS_3	7349	0.00	1.00	60.69%
LJS_3	7349	0.00	1.00	60.20%
CBS1_3	7349	0.00	1.00	35.37%
CBS2_3	7349	0.00	1.00	21.35%
BLOWUPS_3	7349	0.00	1.00	13.25%
JFS1_3	7349	0.00	1.00	11.44%
JFS2_3	7349	0.00	1.00	1.05%
JFS3_3	7349	0.00	1.00	32.29%
CCI-3	7349	0.00	100.00	60.99

Table 25. Descriptive Statistics PDVs at 0.20 mile upstream (JRCP)

Variable	N	Minimum	Maximum	% / Mean
IRI_4	7349	42.00	381.00	152.41
TCS1_4	7349	0.00	1.00	33.95%
TCS2_4	7349	0.00	1.00	20.63%
LCS1_4	7349	0.00	1.00	27.54%
LCS2_4	7349	0.00	1.00	5.17%
PCCP1_4	7349	0.00	1.00	35.27%
PCCP2_4	7349	0.00	1.00	21.88%
PCCP3_4	7349	0.00	1.00	9.77%
ASP PATCH_4	7349	0.00	1.00	43.07%
NUMBER T_JTS_4	7349	0.00	56.00	12.15
TJS_4	7349	0.00	1.00	60.12%
LJS_4	7349	0.00	1.00	60.85%
CBS1_4	7349	0.00	1.00	35.26%
CBS2_4	7349	0.00	1.00	21.69%
BLOWUPS_4	7349	0.00	1.00	13.77%
JFS1_4	7349	0.00	1.00	11.39%
JFS2_4	7349	0.00	1.00	0.98%
JFS3_4	7349	0.00	1.00	32.71%
CCI_4	7349	0.00	100.00	61.03

Table 26. Descriptive Statistics D.V.s and Roadway I.V.s (CRCP)

	Variables	N	Minimum	Maximum	No.	%
Dependent Variables						
	INJURY	5190	0.00	1.00	1560.00	30%
	REAREND	5190	0.00	1.00	3038.00	59%
Road Use					Mean	Std. Dev.
	SPEEDLIMIT	5190	0.00	65.00	57.25	6.17
Physical Characteristics of the Road					No.	%
	ALIGN	5190	0.00	1.00	4035.00	78%
	WZ	5190	0.00	1.00	218.00	4%
					Mean	Std. Dev.
	SURFACEWIDTH	5190	24.00	60.00	33.60	10.33
	SHOULDERWIDTH	5190	0.00	10.00	4.97	0.37
Environmental					No.	%
	WEATHER	5190	0.00	1.00	4431.00	85%
	SURFCOND	5190	0.00	1.00	4337.00	84%
	LIGHTING	5190	0.00	1.00	3661.00	71%

Table 27. Descriptive Statistics PDVs at the Crash Site (CRCP)

Variables	N	Minimum	Maximum	% / Mean
IRI	5190	59.00	265.00	112.32
TCS1	5190	0.00	1.00	87.69%
TCS2	5190	0.00	1.00	16.36%
TCS3	5190	0.00	1.00	0.81%
LC1	5190	0.00	1.00	16.30%
LC2	5190	0.00	1.00	9.81%
LC3	5190	0.00	1.00	1.43%
CCS1	5190	0.00	1.00	64.24%
CCS2	5190	0.00	1.00	1.71%
LJS	5190	0.00	1.00	13.43%
PUNCHOUT	5190	0.00	1.00	26.13%
PCCP1	5190	0.00	1.00	25.18%
PCCP2	5190	0.00	1.00	9.27%
PCCP3	5190	0.00	1.00	0.02%
ASP PATCH	5190	0.00	1.00	49.29%
CDR	5190	26.00	100.00	94.31
CPR	5190	0.00	100.00	68.31
CCI	5190	0.00	100.00	66.90

Table 28. Descriptive Statistics PDVs at 0.10 mile upstream (CRCP)

Variables	N	Minimum	Maximum	% / Mean
IRI_2	5190	67.00	246.00	112.56
TCS1_2	5190	0.00	1.00	88.25%
TCS2_2	5190	0.00	1.00	16.11%
TCS3_2	5190	0.00	1.00	0.71%
LC1_2	5190	0.00	1.00	15.63%
LC2_2	5190	0.00	1.00	9.75%
LC3_2	5190	0.00	1.00	1.46%
CCS1_2	5190	0.00	1.00	64.28%
CCS2_2	5190	0.00	1.00	1.25%
LJS_2	5190	0.00	1.00	13.47%
PUNCHOUT_2	5190	0.00	1.00	25.39%
PCCP1_2	5190	0.00	1.00	25.57%
PCCP2_2	5190	0.00	1.00	8.98%
PCCP3_2	5190	0.00	1.00	0.02%
ASP PATCH_2	5190	0.00	1.00	50.23%
CDR_2	5190	26.00	100.00	94.38
CPR_2	5190	0.00	100.00	68.63
CCI_2	5190	0.00	100.00	67.27

Table 29. Descriptive Statistics PDVs at 0.15 mile upstream (CRCP)

Variables	N	Minimum	Maximum	% / Mean
IRI_3	5190	56.00	294.00	112.91
TCS1_3	5190	0.00	1.00	88.59%
TCS2_3	5190	0.00	1.00	16.20%
TCS3_3	5190	0.00	1.00	0.69%
LC1_3	5190	0.00	1.00	15.41%
LC2_3	5190	0.00	1.00	9.85%
LC3_3	5190	0.00	1.00	1.43%
CCS1_3	5190	0.00	1.00	64.28%
CCS2_3	5190	0.00	1.00	1.52%
LJS_3	5190	0.00	1.00	14.01%
PUNCHOUT_3	5190	0.00	1.00	25.47%
PCCP1_3	5190	0.00	1.00	25.90%
PCCP2_3	5190	0.00	1.00	8.94%
PCCP3_3	5190	0.00	1.00	0.02%
ASP PATCH_3	5190	0.00	1.00	49.61%
CDR_3	5190	9.00	100.00	94.32
CPR_3	5190	0.00	100.00	68.56
CCI_3	5190	0.00	100.00	67.17

Table 30. Descriptive Statistics PDVs at 0.20 mile upstream (CRCP)

Variables	N	Minimum	Maximum	% / Mean
IRI_4	5190	56.00	294.00	112.99
TCS1_4	5190	0.00	1.00	88.84%
TCS2_4	5190	0.00	1.00	16.28%
TCS3_4	5190	0.00	1.00	0.60%
LC1_4	5190	0.00	1.00	15.74%
LC2_4	5190	0.00	1.00	9.92%
LC3_4	5190	0.00	1.00	1.37%
CCS1_4	5190	0.00	1.00	64.24%
CCS2_4	5190	0.00	1.00	1.43%
LJS_4	5190	0.00	1.00	14.24%
PUNCHOUT_4	5190	0.00	1.00	26.05%
PCCP1_4	5190	0.00	1.00	25.82%
PCCP2_4	5190	0.00	1.00	9.31%
PCCP3_4	5190	0.00	0.00	0.00%
ASP PATCH_4	5190	0.00	1.00	49.77%
CDR_4	5190	9.00	100.00	94.19
CPR_4	5190	0.00	100.00	68.83
CCI_4	5190	0.00	100.00	67.27

5.3 Crash Rate Analysis

A crash rate analysis was conducted, and the results are presented in Table 31. The formula for calculating the rate "R" is expressed in the standard nomenclature of crashes per million vehicle miles traveled (MVMT).

$$R = \frac{A * 1{,}000{,}000}{L * V * 730}$$

where:

C = No. of crashes along the roadway segment for two year period 2007-2008

L = Length of roadway segment

V = Annual Average Daily Traffic Volume (2007-2008)

The analysis shows the highest crash rates on asphalt pavement on I-64 and I-264. These segments are relatively short in length and have a large number of crashes, especially I-264 in the urban area of Hampton Roads. For I-64, the AADT was averaged over the entire length of the asphalt paved segment but ranges from 9200 to 147000. In accordance with crash versus traffic volume theory, there are considerably more crashes on the higher volume segments. Determining the crash rate for each segment per the AADT is not relevent to the purpose of this research, and the analysis presented provides an overall general assessment of crash rates on the study's roadway segments. Noteworthy is the fact the crash rate for JRCP and CRCP is relatively equal ranging from 0.55 to 0.59 and will be considered when interpreting the results from the statistical models.

Table 31. Crash Rate Analysis

Route	Ptype	AADT	Length (mi.)	Crashes	R
I-64	ACP	38800	93.00	7472	2.84
I-64	JRCP	111800	115.00	5577	0.59
I-64	CRCP	130600	92.00	5190	0.59
I-264	ACP	63250	6.81	1517	4.82
I-264	JRCP	187600	23.35	1772	0.55
I-81	ACP	40800	325.00	4122	0.43
I-95	ACP	103100	179.00	7740	0.57
I-395	ACP	179200	13.00	704	0.41

5.4 Results of Logistic Regression Modeling: Objective One

Objective: Given a crash occurs, evaluate the relationship between pavement condition, type of pavement and ride quality on crash outcomes by estimating the strength and significance of various pavement condition distress parameters along with the type of pavement and ride quality as expressed by IRI.

The assumptions for logistic regression are much different than for OLS regression. Instead of a normal distribution, logistic regression follows a binary or Bernoulli distribution, and there is no linear relationship between the dependent variables and the predictors. This study meets these assumptions along with the following: the dependent variables (crash outcomes REAR-END and INJURY) are dichotomous, each crash case is mutually exclusive, and each model uses a large sample group which is needed because coefficient estimates are calculated using the maximum likelihood method.

The model output provides the results of model fit tests with their associated chi-squared, *p* significance levels, and degrees of freedom. Additionally, the model provides pseudo-

R^2 values listed as Cox & Snell R Square, Nagelkerke R Square, and McFadden's R Square. The ominbus tests of model parameters is provided and tests if the model with the predictor variables included is significantly different than the baseline model with only the intercept within a confidence interval of 95%. A significant finding indicates the data adequately fits the model and at least one predictor variable has a significant association with the outcome variable.

In traditional OLS regression, the R^2 value attempts to determine the amount of variance explained in the model, but in logistic regression, the variance of the predictor variables depends on the frequency distribution. As an indicator of variance these pseudo-R^2 should be used in conjunction with the results from the classification table that calculate the correct and incorrect estimates. The classification table assesses the predictive correctness of the logistic regression. If the resulting percent predicted correctly value shown in the model summary is 100%, the model is perfect. For each of the models, the histogram of predicted probabilities or the plot of observed groups and predicted probabilities was plotted as an alternative way of assessing correct or incorrect predictions and discussed for each model. The ideal configuration of the plot is a U-shape which indicates a good distribution of the predictions, but a non-U-shape configuration with wide distribution of predicted probabilities is acceptable.

For interpretation purposes, for each of the dependent variables INJURY, and REAREND having a crash with an injury, or being a rear-end crash is the predictor variable and having a crash with no injuries, or not being a rear-end is the reference

variable. As mentioned above, the independent variables were entered as dichotomous covariates, and the model provides both the logistic regression coefficient (B) and the odds ratio (Exp (B)). The odds ratio is the factor by which the predictor variable increases or decreases the log odds of the outcome variable. Each of the significant independent variables is interpreted in the following manner. If the sign of the parameter estimate "B" of a significant independent variable is positive, the presence of the independent variable (e.g. PDV) is associated with an increase in the chance the predictor variable (e.g. crash with injury) occurs, and vice versa for significant variables with parameter estimates with a negative sign. The amount of the increase or decrease is equal to the odds ratio of the parameter Exp (B).

Table 32 lists the modeling results for the dependent variable INJURY for ACP all routes. In the model for INJURY the results of the omnibus tests of model coefficients presented in the model summary indicates, the whole model is significant. However, from the McFadden's R-squared results the model is weak in predicting the variance in the independents. Additionally, from the classification table, a large percentage of observed data are predicted correctly, but when broken down, only a small percentage of the observed INJURY data are predicted as INJURIES. This is shown graphically in the model output by the tight normal distribution of the plot of observed groups versus predicted probabilities. The predicted probabilities are normally distributed, skewed to the left, close to the cut point and not well distributed.

For this dependent variable, there are a relatively small number of significant pavement distress explanatory variables ascertained in the model associated with an increase in the chance of an injury given a crash occurs. For INJURY there are four pavement condition distress variables associated with the odds of a crash having an injury that are significant in the model. They are LC2 (longitudinal cracking severity 1), RUTSE (rutting measured by straight-edge), LLJ2 (longitudinal lane joint severity 2) and AC2 (alligator cracking severity 2), but only LLJ2 and RUTSE show substantial association in the odds a crash will have an injury. Their respective odds ratios are 1.878 and 2.360 with RUTSE being a continuous variable. As mentioned above, the variable REAR-END was included in all INJURY models to determine if the odds of having an injury increase or decrease with rear-end crashes, and from the results of this INJURY model REAR-END as an independent variable is not statistically significant.

Noteworthy for the dependent variable INJURY, ride quality, while significant in the model, it has very little influence on whether a crash results in an injury. Based on these results, one can conclude that pavement condition distress and ride quality variables have little to no impact on whether a crash results in an injury. However, one should note that the distress variable RUTSE is very substantial in increasing the odds a crash results in an injury. In theory and as explained above, this condition, if severe enough, could significantly impair a driver's ability to control a vehicle on a high speed interstate route and very possibly result in a crash with an injury.

Table 33 lists the modeling results for the dependent variable REAREND. The model summary indicates the results of the omnibus tests of model coefficients presented in the model summary indicates the whole model is significant, and from the McFadden's R-squared results the models are slightly better in predicting the variance in the independents. Additionally, from the classification table, a good percentage of observed data are predicted correctly. When broken down, each of the models have a good "hit rate" for predicting the observed REAREND crashes. Graphically, the plot of observed groups versus predicted probabilities is much better distributed across the X-axis of predicted probabilities and not grouped closer to the cut point than the INJURY model, which indicates a better model fit. Also, there are few errors.

Unlike the results for injuries, there a number of PDVs that are significantly associated with rear-end crashes. For rear-end crashes, it is worthwhile to note that as pavement condition distress variables increase in severity so does the odds of having a rear-end crash. This can be seen in the TC (transverse cracking) and AC (alligator cracking). At the crash site, odds ratio for TC increases from 1.117 to 1.222 as the severity goes from level 1 to 2, and as the severity of AC increases from a 2 to a 3, the odds ratio increases from 0.700 to 1.232. Additionally, RLC2 (reflective cracking severity 2), RTC1 (reflective transverse cracking severity 1) and POT (potholes) increase the odds of a rear-end crash by 1.285, 1.218 and 2.035, respectively. Therefore, the results for certain distress variables indicate that as the severity of the pavement distress variable increases (i.e., the condition of the pavement worsens) it is likely to increase the association between the chance of a rear-end crash and these diminishing pavement conditions.

Table 32. Results for dependent variable INJURY all ACP routes

Variables in the Equation	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
SURFACEWIDTH	0.004	0.002	6.411	1.000	0.011	1.004
ALIGN	-0.089	0.032	7.467	1.000	0.006	0.915
WEATHER	0.180	0.037	23.318	1.000	0.000	1.197
IRI	-0.001	0.000	8.526	1.000	0.004	0.999
LC2	0.094	0.036	6.852	1.000	0.009	1.099
RTC3	-0.135	0.046	8.548	1.000	0.003	0.874
RUTSE	0.859	0.226	14.482	1.000	0.000	2.360
PATCH_3	-0.084	0.044	3.656	1.000	0.056	0.919
LLJ2_4	0.630	0.294	4.597	1.000	0.032	1.878
AC2_4	0.102	0.032	9.878	1.000	0.002	1.107
Constant	-1.064	0.071	225.149	1.000	0.000	0.345
Model Summary		Chi-square	df	Sig.		
n	21541					
-2 Log likelihood	26735.300					
Cox & Snell R Square	0.004					
Nagelkerke R Square	0.006					
McFadden's R Square	0.003					
Omnibus tests of model coefficients		93.549	10	0.000		
% Predicted correctly	68.500					

Table 33. Results for dependent variable REAREND all ACP routes

Variables in the Equation	B	S.E.	Wald	df	Sig.	Exp(B)
SURFACEWIDTH	0.012	0.002	50.839	1.000	0.000	1.013
SPEEDLIMIT	-0.033	0.002	215.491	1.000	0.000	0.967
ALIGN	0.283	0.034	71.159	1.000	0.000	1.327
WEATHER	0.275	0.086	10.087	1.000	0.001	1.316
SURFCOND	0.528	0.082	41.407	1.000	0.000	1.695
LIGHTING	0.746	0.032	551.252	1.000	0.000	2.109
WZ	0.201	0.072	7.814	1.000	0.005	1.222
IRI	0.004	0.000	198.552	1.000	0.000	1.004
TC1	0.110	0.055	4.015	1.000	0.045	1.117
TC2	0.201	0.062	10.640	1.000	0.001	1.222
LC1	-0.279	0.052	29.170	1.000	0.000	0.756
LC2	0.202	0.065	9.749	1.000	0.002	1.224
LLJ1	-0.438	0.060	52.548	1.000	0.000	0.646
RLC1	0.180	0.045	16.240	1.000	0.000	1.198
AC2	-0.357	0.049	52.780	1.000	0.000	0.700
AC3	0.209	0.067	9.686	1.000	0.002	1.232
PATCH	0.268	0.049	29.663	1.000	0.000	1.307
POT	-0.995	0.386	6.655	1.000	0.010	0.370
CCI	-0.003	0.001	9.836	1.000	0.002	0.997
RLC3_2	-0.237	0.103	5.337	1.000	0.021	0.789
TC2_3	0.137	0.061	5.097	1.000	0.024	1.147
LC2_3	0.136	0.062	4.835	1.000	0.028	1.145
RLC2_3	0.251	0.062	16.296	1.000	0.000	1.285
AC1_3	-0.116	0.035	11.087	1.000	0.001	0.891
AC2_3	-0.171	0.052	10.818	1.000	0.001	0.843
RTC1_4	0.197	0.043	21.052	1.000	0.000	1.218
AC3_4	0.206	0.065	9.902	1.000	0.002	1.228
POT_4	0.710	0.368	3.732	1.000	0.053	2.035
Constant	-0.332	0.207	2.568	1.000	0.109	0.717
Model Summary		Chi-square	df	Sig.		
n	21541					
-2 Log likelihood	26393.873					
Cox & Snell R Square	0.122					
Nagelkerke R Square	0.164					
McFadden's R Square	0.074					
Omnibus tests of model coefficients		2803.307	28	0.000		
% Predicted correctly	66.500					

In the rear-end crash model the environmental condition variable LIGHTING at the crash was significant in increasing the odds of the crash outcome. Therefore, to further explore the relationship between crash outcomes and pavement condition/ride quality and to help in model building and refining the model specification, the crashes were broken into daytime vs. nighttime crashes. For rear-end crashes, the results are presented in Table 34. Again, longitudinal and longitudinal joint cracking along with alligator cracking have a positive relationship and are associated with increasing the odds of a rear-end crash. Additionally, rutting, patching and potholes are also associated with an increase in the odds of a rear-end crash and are ranked as follows: rutting and potholes are associated with an increase in the odds in daytime rear-end crashes by 2.779 and 3.379, respectively and patching increases the odds of a rear-end crash at night by 1.313.

Table 34. Results for REAREND all ACP routes: Day vs. Night Crashes

Lighting Day Crashes					Lighting Night Crashes				
Variables in the Equation	**B**	**S.E.**	**Sig.**	**Exp(B)**	**Variables in the Equation**	**B**	**S.E.**	**Sig.**	**Exp(B)**
IRI	0.006	0.000	0.000	1.006	IRI	0.005	0.001	0.000	1.005
TC2	0.258	0.072	0.000	1.295	TC2	-0.436	0.100	0.000	0.647
LC1	-0.322	0.056	0.000	0.725	LC1				
LC2	0.295	0.075	0.000	1.343	LC2	-0.272	0.101	0.007	0.762
LLJ1	-0.523	0.074	0.000	0.593	LLJ1	0.620	0.106	0.000	1.859
RTC1	-0.154	0.060	0.010	0.857	RTC1	-0.334	0.072	0.000	0.716
RLC1	0.276	0.058	0.000	1.318					
RLC3					RLC3	-0.418	0.161	0.009	0.658
AC1					AC1	0.183	0.058	0.002	1.201
AC2	-0.406	0.057	0.000	1.501	AC2	0.362	0.079	0.000	1.436
AC3	0.252	0.082	0.002	0.777					
PATCH	0.342	0.060	0.000	0.711	PATCH	-0.193	0.084	0.022	0.825
PATCH_2					PATCH_2	0.272	0.082	0.001	1.313
POT	-1.218	0.469	0.009	0.296					
RUTSE	-1.119	0.356	0.002	0.327					
RUTWM	1.022	0.326	0.002	2.779					
CCI	-0.004	0.001	0.002	0.996					
TC2_2	0.202	0.058	0.001	1.223					
RLC1_2	0.121	0.054	0.025	1.129					
AC3_2	-0.249	0.107	0.020	0.779					
IRI_3					IRI_3	0.002	0.001	0.002	1.002
RTC3_3					RTC3_3	-0.287	0.084	0.001	0.751
AC1_3	-0.132	0.041	0.001	0.876					
RUTWM_3	-0.581	0.287	0.043	0.559					
IRI_4	0.001	0.000	0.001	1.001					
TC2_4					TC2_4	-0.196	0.071	0.006	0.822
RLC2_4					RLC2_4	-0.242	0.087	0.005	0.785
RTC1_4	0.275	0.056	0.000	1.316					
AC3_4	0.399	0.107	0.000	1.490					
POT_4	1.106	0.531	0.037	3.022					
Constant	-0.466	0.172	0.007	0.628					

Table 34 (cont.). Results for REAREND all ACP routes: Day vs. Night

Model Summary		Chi-square	df	Sig.	Model Summary		Chi-square	df	Sig.
n	13811				n	7745			
-2 Log likelihood	18222.32				-2 Log likelihood	9061.74			
Cox & Snell R Square	0.063				Cox & Snell R Square	0.041			
Nagelkerke R Square	0.084				Nagelkerke R Square	0.058			
McFadden's R Square	0.020				McFadden's R Square	0.020			
Hosmer and Lemeshow		48.63	8	0.000	Hosmer and Lemeshow		15.87	8	0.044
% Predicted correctly	60.80				% Predicted correctly	70.30			

Tables 35 and 37 list the modeling results for the dependent variable INJURY for JRCP and CRCP all routes, respectively. Similar to the ACP models, in the model for INJURY the results of the omnibus tests of model coefficients presented in the model summary indicate the whole model is significant, but the McFadden's R-squared results indicate that the model is weak in predicting the variance in the independents. Additionally, from the classification table, a large percentage of observed data are predicted correctly, but when broken down, only a small percentage of the observed INJURY data are predicted as INJURIES. This is the same for both the JRCP and CRCP models and puts into question the conclusiveness of the results for the INJURY models. Nonetheless, the few distress variables that are statistically significant shall be explained with this reservation.

Again, for the dependent variable INJURY for both JRCP and CRCP pavements there are a relatively small number of statistically significant pavement distress explanatory variables ascertained in the model associated with an increase in the chance of an injury

given a crash occurs and interpreted with caution for the reasons mentioned above. The pavement condition distress variable on JRCP pavements associated with the odds of a crash having an injury and significant in the model is JFS2 (joint fault severity 2) with an odds ratio of 2.030. The other pavement distress variable on CRCP pavements with an increased association with the odds of having a crash with injuries are TCS1 and 2 (transverse cracking severity 1 and 2) and have odds ratios of 1.490 and 1.343, respectively. Based on these results, one can conclude that pavement condition distress and ride quality variables have little to no impact on whether a crash results in an injury on JRCP or CRCP.

Noteworthy for both of these models, when the variable REAR-END is introduced as an independent variable the odds of having an injury, given a rear-end crash occurs, decrease by a factor of 0.712 and 0.656 for crashes that occur on JRCP and CRCP, respectively.

Table 36 lists the modeling results for the dependent variable REAREND for JRCP pavement. The results of the omnibus tests of model coefficients presented in the model summary indicate the whole model is significant. From the McFadden's R-squared results, we see that the model is slightly better in predicting the variance in the independents. Additionally, from the classification table, a good percentage of observed data are predicted correctly, and when broken down, the model has a good "hit rate" for predicting the observed REAREND crashes.

Unlike the results for injuries, there a number of PDVs that are statistically significantly associated with rear-end crashes on JRCP pavements. On JRCP the pavement distress predictors that have an association with an increase in the odds of having a rear-end crash are TJS (transverse joint spalled) with odds ratios 1.225, CBS1 and CBS2 (corner breaks severity 1 and 2) with odds ratios of 1.261 and 1.151, respectively, JFS1, JFS2 and JFS3 (joint fault severity 1, 2 and 3) with odds ratios of 1.213, 1.803 and 1.227, respectively and LJS (longitudinal joint spalled) with an odds ratio of 1.125. It is worthwhile to note that as pavement condition distress variable JFS increase in severity from 1 to 2, the odds ratio for a rear-end crash increases substantially. Therefore, the results for this distress variable indicate as the severity of the pavement distress variable increases from severity level 1 to 2 (i.e., the condition of the pavement worsens) it is likely to increase the association between the chance of a rear-end crash and these diminishing pavement conditions. However, at severity level 3 the association returns to level 1 odds which may be explained by the fact that there are fewer areas where this more severely distressed condition is present. Lastly, noteworthy is the fact there are no PDVs associated with an increase in the odds of having a rear-end crash on CRCP pavements.

Table 35. Results for dependent variable INJURY all JRCP routes

Variables in the Equation	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
LANECOUNT	0.121	0.028	18.911	1	0.000	1.129
BLOWUPS	-0.800	0.350	5.219	1	0.022	0.450
SURFCOND	0.165	0.067	6.160	1	0.013	1.180
REAREND	-0.340	0.052	43.147	1	0.000	0.712
CBS1_2	-0.249	0.054	21.426	1	0.000	0.780
CBS2_4	-0.231	0.064	13.109	1	0.000	0.794
JFS2_4	0.708	0.258	7.548	1	0.006	2.030
Constant	-0.978	0.109	80.505	1	0.000	0.376
Model Summary		Chi-square	df	Sig.		
n	7349					
-2 Log likelihood	8987.69					
Cox & Snell R Square	0.015					
Nagelkerke R Square	0.021					
McFadden's R Square	0.007					
Hosmer and Lemeshow		4.038	8	0.854		
% Predicted correctly	69.10					

Table 36. Results for dependent variable REAREND all JRCP routes

Variables in the Equation	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
SURFACEWIDTH	-0.017	0.002	49.508	1.000	0.000	0.983
SHOULDERWIDTH	-0.179	0.047	14.388	1.000	0.000	0.836
SPEEDLIMIT	-0.019	0.004	21.993	1.000	0.000	0.981
PCCP2	-0.390	0.067	33.548	1.000	0.000	0.677
NUMBER_JTS	0.027	0.006	23.651	1.000	0.000	1.027
TJS	0.203	0.062	10.856	1.000	0.001	1.225
CBS1	-0.266	0.080	11.037	1.000	0.001	0.766
ALIGN	0.279	0.060	21.935	1.000	0.000	1.322
SURFCOND	0.738	0.064	132.947	1.000	0.000	2.092
LIGHTING	0.802	0.053	225.836	1.000	0.000	2.230
CBS1_2	0.232	0.104	4.940	1.000	0.026	1.261
CBS2_2	0.141	0.065	4.643	1.000	0.031	1.151
PCCP1_3	-0.202	0.059	11.830	1.000	0.001	0.817
NUMBERT_JTS_3	-0.015	0.005	7.052	1.000	0.008	0.986
CBS1_3	-0.359	0.106	11.478	1.000	0.001	0.698
JFS3_3	0.204	0.056	13.338	1.000	0.000	1.227
IRI_4	-0.001	0.001	5.116	1.000	0.024	0.999
LCS1_4	-0.186	0.060	9.766	1.000	0.002	0.830
LJS_4	0.118	0.055	4.527	1.000	0.033	1.125
JFS1_4	0.193	0.090	4.598	1.000	0.032	1.213
JFS2_4	0.589	0.260	5.139	1.000	0.023	1.803
Constant	1.481	0.381	15.082	1.000	0.000	4.399
Model Summary		Chi-square	df	Sig.		
n	7349					
-2 Log likelihood	9364.64					
Cox & Snell R Square	0.095					
Nagelkerke R Square	0.127					
McFadden's R Square	0.042					
Omnibus tests of model coefficients		732.660	21	0.000		
% Predicted correctly	64.10					

Table 37. Results for dependent variable INJURY all CRCP routes

Variables in the Equation	B	S.E.	Wald	df	Sig.	Exp(B)
TCS1	0.399	0.149	7.186	1	0.007	1.490
TCS2	0.295	0.089	10.952	1	0.001	1.343
CCS1	-0.396	0.098	16.361	1	0.000	0.673
PCCP2	-0.337	0.116	8.398	1	0.004	0.714
CCI	-0.008	0.002	13.227	1	0.000	0.992
SURFCOND	0.177	0.084	4.390	1	0.036	1.193
REAREND	-0.422	0.062	45.694	1	0.000	0.656
ASPPATCH_3	-0.192	0.063	9.261	1	0.002	0.825
Constant	-0.221	0.174	1.608	1	0.205	0.802
Model Summary		Chi-square	df	Sig.		
n	5190					
-2 Log likelihood	6259.08					
Cox & Snell R Square	0.017					
Nagelkerke R Square	0.023					
McFadden's R Square	0.007					
Hosmer and Lemeshow		11.617	8	0.169		
% Predicted correctly	69.90					

As mentioned in the previous section, to evaluate the effects of pavement type on crash outcomes a dataset was created using all the crash data and PDVs for each of the three pavement types: ACP, CRCP and JRCP. The objective is to estimate the relationship between the type of pavement and roadway safety based on how it affects the study's two crash outcomes: crashes with injuries and rear-end crashes. For this analysis CRCP and JRCP were considered to be one type of pavement: concrete. The dummy predictor variable PTYPE was included in the dataset along with the crash data and PDVs. For this predictor variable (PTYPE) a "0" was given to all crash cases occurring on asphalt pavement and a "1" for crashes on concrete pavement. The variable PTYPE was entered into two binary logistic regresson models as a dichotomous covariate with the asphalt

pavement being the reference. The dependent variables are INJURY and REAR-END, and again, with REAR-END included as an independent variable in the INJURY model.

Presented in Tables 38 and 39 are the results from the REAR-END and INJURY models. The PDVs associated with an increase in the odds of having a rear-end crash or a crash with an injury are similar to those presented above for the individual pavement type models, IRI, TC2, LC2, and RLC1, AC3, PATCH, RUTWM and CCS2. With respect to PTYPE, the type of pavement is not statistically significant in the INJURY model and is significant in the REAR-END model. Therefore, given a crash occurs, the type of pavement has a statistically significant relationship with rear-end crashes, and the odds of having a rear-end crash is increased by a factor of 1.910 on concrete pavement. The type of pavement does not have a statistically significant association with crashes with injuries, given a crash occurs.

Table 38. Results for the dependent variable REAR-END w/ PTYPE

Variables in the Equation	B	S.E.	Wald	df	Sig.	Exp(B)
SURFACEWIDTH	-0.158	0.006	688.077	1.000	0.000	0.853
SHOULDERWIDTH	-0.047	0.016	8.212	1.000	0.004	0.954
LANECOUNT	1.857	0.070	701.459	1.000	0.000	6.405
SPEEDLIMIT	-0.024	0.002	166.677	1.000	0.000	0.976
PTYPE	0.647	0.057	130.911	1.000	0.000	1.910
ALIGN	0.275	0.027	101.875	1.000	0.000	1.317
WEATHER	0.272	0.071	14.837	1.000	0.000	1.313
SURFCOND	0.517	0.067	58.852	1.000	0.000	1.676
LIGHTING	0.776	0.025	943.539	1.000	0.000	2.172
WZ	0.244	0.064	14.427	1.000	0.000	1.276
IRI	0.002	0.000	66.060	1.000	0.000	1.002
TC2	0.268	0.056	22.837	1.000	0.000	1.308
LC1	-0.232	0.034	47.570	1.000	0.000	0.793
LC2	0.178	0.049	13.440	1.000	0.000	1.195
LLJ1	-0.400	0.061	42.468	1.000	0.000	0.670
RLC1	0.092	0.040	5.229	1.000	0.022	1.096
AC2	-0.296	0.044	44.516	1.000	0.000	0.744
AC3	0.198	0.062	10.329	1.000	0.001	1.219
PATCH	0.308	0.048	41.877	1.000	0.000	1.360
POT	-0.946	0.383	6.106	1.000	0.013	0.388
RUTWM	0.803	0.222	13.040	1.000	0.000	2.233
CCI	-0.002	0.001	8.390	1.000	0.004	0.998
TCS2	-0.129	0.050	6.644	1.000	0.010	0.879
LC3	-0.665	0.247	7.262	1.000	0.007	0.514
CCS2	0.686	0.251	7.458	1.000	0.006	1.986
LJS	-0.181	0.043	17.737	1.000	0.000	0.834
PCCP2	-0.512	0.055	85.791	1.000	0.000	0.599
Constant	0.005	0.173	0.001	1.000	0.978	1.005
Model Summary		Chi-square	df	Sig.		
n						
-2 Log likelihood	41861.02					
Cox & Snell R Square	0.143					
Nagelkerke R Square	0.191					
McFadden's R Square	0.085					
Hosmer and Lemeshow		52.777	8	0.000		
% Predicted correctly						

Table 39. Results for the dependent variable INJURY w/ PTYPE

Variables in the Equation	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
SURFACEWIDTH	0.007	0.001	34.002	1.000	0.000	1.007
ALIGN	-0.066	0.027	6.251	1.000	0.012	0.936
WEATHER	0.176	0.031	32.397	1.000	0.000	1.192
LIGHTING	-0.057	0.025	5.103	1.000	0.024	0.945
REAREND	-0.153	0.024	39.030	1.000	0.000	0.858
LLJ1	0.149	0.052	8.167	1.000	0.004	1.161
RTC3	-0.168	0.044	14.529	1.000	0.000	0.846
AC2	0.114	0.029	15.464	1.000	0.000	1.120
TCS2	0.108	0.048	5.146	1.000	0.023	1.114
LJS	-0.132	0.041	10.147	1.000	0.001	0.877
Constant	-1.034	0.053	376.808	1.000	0.000	0.356
Model Summary		Chi-square	df	Sig.		
n						
-2 Log likelihood	42116.31					
Cox & Snell R Square	0.005					
Nagelkerke R Square	0.007					
McFadden's R Square	0.004					
Hosmer and Lemeshow		14.775	8	0.064		
% Predicted correctly						

5.5 Results of Logistic Regression Modeling: Objective Two

Objective: To explore the spatial aspect of this relationship, models will be developed to evaluate crash outcomes at the crash site and specific intervals upstream. These models will be used to evaluate where within the footprint of the crash is the most critical location with respect to pavement condition and/or ride quality.

Traditionally, crashes with injuries have been studied to determine the severity of the crash. Therefore, the results of this analysis will be presented in two categories using the two dependent variables INJURY for severity and REAREND for crash type. The model estimated for INJURY, for all asphalt routes and all crash types, resulted in pavement condition and ride quality parameters consistent with those of Objective One. For these dependent variables, there are a relatively small number of pavement distress explanatory variables that are significant in the model and increase the odds of an injury, given a crash occurs. The results from the INJURY model that the overall best fitting model is not at the crash site, but approximately 0.15 upstream and the pavement distress parameter RUTSE show the highest odds ratio of 2.282 approximately 0.20 miles upstream. Noteworthy is the fact this parameter was also identified in the results for Objective One above as having a significant relationship to crashes with injuries.

Generally, the results from these models are consistent with those presented in Objective One of this study in that the pavement condition distress and ride quality variables have little to no impact on whether a crash results in an injury, but the objective was to estimate the critical location within the footprint of the crash for assessing the importance

of pavement condition and ride quality on the severity of crashes, given a crash occurs. Based on the results from the models, this location is upstream of the crash site approximately 0.15 to 0.20 miles. Additionally, one should again note that in theory, the distress variable RUTSE has the potential to substantially increasing the odds that a crash results in an injury. This condition, if severe enough, could significantly impair a driver's ability to control a vehicle, very possibly resulting in a crash with an injury. This effect is also likely to originate somewhere upstream of the final crash site; therefore, these results are consistent with crash theory.

For the crash type REAREND, the results from the model indicate the critical location is approximately at the crash site and 0.15 to 0.20 miles upstream of the crash site, respectively. The majority of the pavement distress parameters (Exp (B)) estimated in the model that increase the odds of having a rear-end crash are highest at the crash site. The model summary also shows the data best fits this model. Therefore, based on the results of these models, the critical location for assessing the importance of pavement condition on rear-end crashes is at the site of the crash. Again, this is consistent with crash theory in that rear-end crashes are usually sudden, and the effecting factors are usually located in close vicinity to the crash site. The results are presented in Tables 40 and 41.

Table 40. Results for dependent variable INJURY for all ACP Routes

Distress Variable	Exp(B)			
	At site	Miles Upstream		
		0.100	0.150	0.200
IRI	0.999			
LC2	1.171			
LLJ2				1.781
RTC3	0.861	0.891	0.846	0.882
AC2		1.110	1.086	1.151
RLC1	1.085			
RLC3			1.211	
PATCH				0.901
RUTSE	2.219	1.922	1.932	2.282
Constant	0.423	0.410	0.413	0.401
Model Summaries				
N	21541			
-2 Log likelihood	26782.07	26804.70	26799.72	26786.98
Cox & Snell R Square	0.002	0.001	0.001	0.002
Nagelkerke R Square	0.003	0.002	0.002	0.003
McFadden's R Square	0.000	0.000	0.001	0.001
Omnibus tests of model coefficients	0.000	0.000	0.000	0.000
% Predicted Correctly	68.50	68.50	68.50	68.50

Table 41. Results for dependent variable REAREND for all ACP Routes

Distress Variable	Exp(B)			
	At site	Miles Upstream		
		0.100	0.150	0.200
IRI	1.006	1.003	1.003	1.003
TC2	1.431	1.320	1.299	1.271
LC1	0.706	0.776	0.799	0.809
LC2	1.402	1.301	1.318	1.236
LLJ1	0.585	0.745	0.721	0.751
RTC1_2		1.115	1.188	1.185
RTC3	1.134			
RLC1	1.277	1.242	1.153	1.183
RLC2_2		1.212	1.243	1.177
AC2	0.665	0.826	0.779	0.816
AC3	1.298		1.191	1.237
PATCH	1.306			
POT	0.357			
BLEED1_2		1.412		
RUTSE	0.419	0.596		
RUTWM	1.803		0.583	0.571
CCI	0.997			
Constant	0.479	0.518	0.538	0.539
Model Summaries				
N	21541			
-2 Log likelihood	28095.28	28828.10	28835.82	28847.29
Cox & Snell R Square	0.050	0.017	0.017	0.016
Nagelkerke R Square	0.067	0.023	0.022	0.022
McFadden's R Square	0.015	0.008	0.008	0.008
Omnibus tests of model coefficients	0.000	0.000	0.000	0.000
% Predicted Correctly	61.60	58.20	58.60	58.30

CHAPTER SIX

6.0 HIERARCHICAL GENERALIZED LINEAR MODELING (HGLM)

6.1 HGLM: General Introduction

To further explore the relationship between pavement distress and conditions factors and meet objective three of this research, crash prediction models using HGLM shall be developed to estimate the odds of having an injury crash or rear-end crash. This methodology is relatively new to the field of traffic safety, and applying this technique to the hierarchical nature of crash data greatly enhances the overall contribution of this research.

Inherent in all statistical models is the inability to perfectly predict the value of the dependent variable and, therefore, to try to compensate, models are built with two parts: a fixed or deterministic part where the variation in the dependent variable can be predicted by the independent variable and the random or stochastic part where the variance cannot be predicted. It is in the independence assumption of the random part that rarely holds true and is not realistic in traffic safety research. In this assumption the error "e" is assumed to not be correlated to the independent variable X, i.e. all observed autocorrelations of the errors are zero (Dupont and Martensen (Eds), 2007). For example, it is reasonable to assume there are commonalities among crashes that occur within the same county because these crashes may share characteristics that were not recorded in the crash-level police report. The socio-economic environment that may shape the behavior of individual drivers will differ from county to county, and it is

important to identify all the factors that may contribute to vehicle crashes. These factors may also be related to the amount of funds allocated to ensure that proper pavement condition is maintained. Traditionally, the hierarchical structure of crash data has been somewhat ignored and not thoroughly considered when developing models to quantify the casualty factors related to various crash outcomes, and, in doing so, studies fail to adequately relate injury severity and/or type of crash with the primary contributing factors.

HGLM is mainly predicated on the idea that the characteristics of a group have some influence over the characteristics of an individual. Hierarchy means consisting of *units* grouped at different *levels*. For example, the individual exposure characteristics of the crash site may be the level-1 units in a two-level hierarchy and the level-2 units might be the socio-economic characteristics of the locality where the crash occurred. These groupings tend to become differentiated, and this differentiation implies that the group and its members both influence and are influenced by being in the group (Goldstein, 1999). Observations within a group are arguably often more similar that would be predicted on a pooled-data basis. Therefore, to ignore this relationship render some of the traditional statistical techniques for studying data relationship invalid (Goldstein, 1999). Specifically, by ignoring the hierarchical structure of the random part of the equation generally causes bias in the random error "*e*" value leading to either over- or underestimated standard error terms and subsequently bias the parameter estimate values of the independent variable leading to increase in Type 1 errors in the conclusions.

Generally, HGLM is a complex form of ordinary least squares (OLS) regression that is used to analyze variance in the outcome variable when the predictor variables are at varying hierarchical levels (Woltman et al., 2012). The output results from the HGLM are a set of regression coefficients interpreted somewhat like OLS with the difference being that HGLM has additional estimates of the variances and co-variances that reflect the hierarchical structure. In order to further explain HGLM modeling one must recognize that it is an extension of the traditional regression model.

As previously mentioned, crashes by nature show a hierarchical structure because drivers/passengers are in cars, cars are in crashes and crashes happen at specific locations all of which happens randomly. To account for this randomness, statistical models have traditionally been used and are comprised of a deterministic (fixed) part and random part (error term). The expected value of the dependent variable is based on the values of the independent variables. In a traditional one-level linear regression model:

$$y_i = \beta_0 + \beta x_i + e_i$$

where y is the outcome variable and β_0 is the intercept, β is the slope or regression coefficient that explains the X to Y relationship, and x_i is the explanatory variable making up the fixed part and the e_i is the random term or residual. The e_i term is the difference between the actual data point and the predicted data point forecasted by the regression equation and is assumed due only to random error. HGLM have additional error terms to reflect the complex pattern of variation introduced by the hierarchical data structure (Roberts, 2004).

To simultaneously analyze the group relationship the subscript *"j"* is added and gives the following:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}.$$

In this model *"j"* refers to the *level-2* unit, and *"i"* refers to the *level-1* unit. Again, β_{0j} is the intercept and β_{1j} is regression coefficient associated with the explanatory variable X_{ij}, and r_{ij} is the Level-1 random error. This is still a single level-1 model, but it allows for the county/clustering relationship to be analyzed.

The most basic concept to HGLM lies within the explanation of fixed and random coefficients and is best explained by a simple example related to this study. Suppose the driver's outcome of a crash can be predicted by the individual driver's socio-economic status (SES). In OLS, one random error term explains the variance between the actual outcome of crash and the predicted outcome of crash given the driver's SES. However, it does not take into consideration driver's differences in crash outcomes within the counties where the crashes occur. In other words, and as explained above, there may be a group effect that is not accounted for in simple linear regression.

In HGLM, the so-called "fixed effect" is the coefficient for the slope of the SES variable and is the average of the effect of this variable over all the counties sampled. This fixed effect coefficient of SES also has a random error term "u_j" which allows the effect of the driver's SES to fluctuate from county to county. By adding this random term for the driver's SES allows for the possibility that the individual drivers within similar counties may be more alike than individual drivers in non-similar counties. In particular,

estimates of the variance and covariance between counties in their slopes and intercepts will allow a comparison of counties with different/similar characteristics. In essence, in a two-level model the intercept (β_{0j}) and the slope (β_{1j}) that describe the difference/similarities of the counties become outcome variables in the level-2 model and, depending on the research question, are either fixed or allowed to vary across the counties.

As mentioned above, HGLM is a step-wise optimization procedure which allows for various combinations of fixed and random slopes and/or intercepts. The main objective of HGLM is to uncover variables that are able to reasonably describe the level of variation of independent variables on the dependent variable (Roberts, 2004). Referencing Bryk and Raudenbush, there are five submodels used to estimate these cross group effects, and each one has a specific purpose. The primary objective here is to determine if there is a relationship between the socio-economic traits of a county (level-2) and the outcomes of a crash given a crash occurs. Therefore, the submodel used for this application will be a random-intercept model and include level-2 predictors as well as level-1 predictors. This is an extension of the random-effects ANCOVA model and provides for a level-2 covariate, W_j, while also controlling for the effect of a level-1 covariate, X_{ij}, and the random effects of the level-2 units, u_{0j} (Bryk and Raundenbush). In this model, the intercept β_{0j} is allowed to vary across the counties to try to explain the differences among the socio-economic characteristics of the different counties on crash outcomes.

The level-1, level-2 and combined models are as follows assuming the normally distributed error terms:

$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$ (level-1 model)

$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}$ (level-2 model)

$\beta_{1j} = \gamma_{10}$ (level-2 model).

Substituting the level-2 models into the level-1 model yields the following combined model:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_{0j} + r_{ij} \quad \text{(combined model)}$$

where $\mu_{0j} \sim N(0,\tau_{00})$ and $r_{ij} \sim N(0,\sigma^2)$.

The definition of the hierarchical levels in the HGLM models for this study is dependent on the availability of data. As mentioned above, because this pseudo segment data that contain pavement condition attributes as well as the details of the crash, the lowest aggregated level of analysis is the segment on which the crash occurs. The response variables at this level will be the distress variables associated with the condition and ride quality of the pavement.

6.2 HGLM: Description of Data and Hierarchical Structure

For this analysis, the same crash and pavement distress/condition predictor variables used for objectives one and two will be included in the dataset. This dataset will then be supplemented with county-based socio-economic data. Based on empirical data the models will be developed to estimate, under certain conditions, the odds of a certain

type/severity of crash, given a crash occurs (categorical outcome- ordinal and dichotomous) based upon identifiable and measureable socio-economic explanatory variables specific to the county in which the crash occurs. As previously mentioned the pavement type for the majority of the study's interstate roadways is asphalt and subsequently yields the largest sample of crashes. Therefore, this effort will be limited to asphalt pavement.

A critical task of all road safety studies is the development of applicable databases that combine all the elements of the research, and in this case: crash data, pavement condition and ride quality data, traffic, roadway and environmental features data, and local socio-economic data. As explained above, the socio-economic data was collected for each county in the state and compiled into one dataset and the crash dataset includes the county location of the crash. For creating the datasets for the HGLM a method similar to the aggregation of the pavement condition with the crash data was used to include the county specific socio-economic data. Queries and other spreadsheet commands were applied so the select socio-economic factors could be aggregated into the pavement and crash dataset to create one unique dataset for each of the studies interstate routes.

The data consists of 21,556 vehicle crashes located within 35 counties in Virginia. The hierarchical structure of the date is shown in Figure 9. The crash-level (level-1) binary outcome variables will again be the two different crash types in which one can be potentially related to severity of the crash. They are crashes with injuries and rear-end crashes. Additionally, at the crash-level are the same roadway and environmental

predictor variables used in objective one, and only the pavement distress predictor variables at the crash "site" will be used for this research.

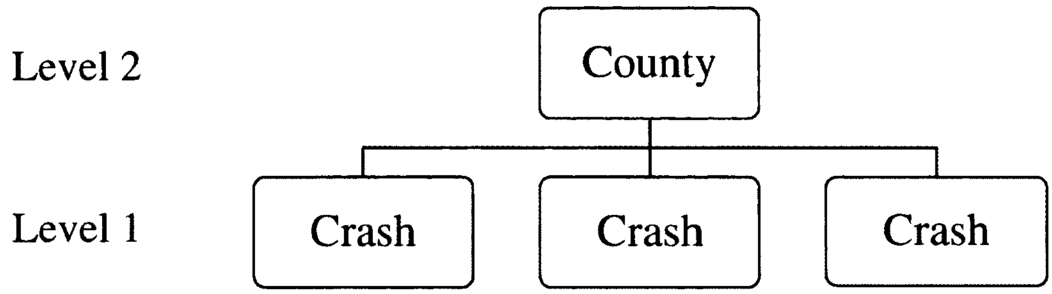


Figure 9. Hierarchical structure of the data

At the county level (level-2) the predictor variables will be related to the socio-economic characteristics of the county. There are 23 variables at the county level primarily describing the population demographics and income of the county. Descriptions of the common dependent variables along with the crash level and county level explanatory variables used in the analysis of the crash outcomes given a crash occurs are provided in Tables 42, 43, and 44 below.

Table 42. Description of crash level variables used in the HGLM

	Variable	Description	
Crash-level Dependent Variables	INJURY	Crash had and injury	1 if injury, 0 if no injuries
	REAREND	Crash type was rear-end	1 if rear-end, 0 if other
Crash-level Predictor Variables			
Physical Characteristics of the Road	ALIGN	Alignment of the road	1= Straight, 0= all other
	WZ	Work zone present	1= Yes, 0= No
Environmental	WEATHER	Weather conditions during crash	1= Clear, 0= Not clear
	SURFCOND	Condition of the riding surface	1= Dry, 0= Not dry
	LIGHTING	Light condition	1= Day, 0= Night

Table 43. Description of crash level PDVs for ACP

Variables	Description	
IRI	Internal Roughness Index Average	Value
TC1	Transverse Cracking Severity 1	Presence=1, absence=0
TC2	Transverse Cracking Severity 2	Presence=1, absence=0
LC1	Longitudinal Cracking Severity 1	Presence=1, absence=0
LC2	Longitudinal Cracking Severity 2	Presence=1, absence=0
LLJ1	Longitudinal Lane Joint Severity 1	Presence=1, absence=0
LLJ2	Longitudinal Lane Joint Severity 2	Presence=1, absence=0
RTC1	Reflective Transverse Cracking Severity 1	Presence=1, absence=0
RTC2	Reflective Transverse Cracking Severity 2	Presence=1, absence=0
RTC3	Reflective Transverse Cracking Severity 3	Presence=1, absence=0
RLC1	Reflective Longitudinal Cracking Severity 1	Presence=1, absence=0
RLC2	Reflective Longitudinal Cracking Severity 2	Presence=1, absence=0
RLC3	Reflective Longitudinal Cracking Severity 3	Presence=1, absence=0
AC1	Alligator Cracking Severity 1	Presence=1, absence=0
AC2	Alligator Cracking Severity 2	Presence=1, absence=0
AC3	Alligator Cracking Severity 3	Presence=1, absence=0
PATCH	Patching Area - wheel path	Presence=1, absence=0
POT	Potholes Count	Presence=1, absence=0
DELAM	Delaminations Area	Presence=1, absence=0
BLEED1	Bleeding Severity 1	Presence=1, absence=0
BLEED2	Bleeding Severity 2	Presence=1, absence=0
RUT SE	Average Deeper Rut (Straight-edge)	Value
RUT WM	Average Deeper Rut (Wire method)	Value
CCI	Critical Condition Index	Value

Table 44. Description of county level variables

County-level	Variables	Description
Population/Commut.	POP	Average total population
	POPWORK	Average working population
	HOUSEHLD	Average total households
	UNEMP	Average total unemployment (%)
	COMMUT	Average total number of commuters
	SOV	Average single occupant vehicle commuters
Income/Education	HHINCOME	Average household income
	PERCAPINC	Average income per capita
	NOGED	Average population with no GED
	GED	Average population with GED
	ASSDEGREE	Average population with Associates
	BACHDEGRE	Average population with Bachelors
	GRADDEGREE	Average population with Graduate
Gender	MALE	Average male population
	FEMALE	Average female population
	AVGAGE	Average age of population
	@65OLDR	Average 65yrs and older population
	M65OLDR	Average 65yrs and older male population
	F65OLDER	Average 65yrs and older female population
Race	WHITE	Average white population
	AFRAM	Average African American population
	ASIAN	Average Asian population
	OTHER	Average other race population

6.3 HGLM: Model Structure for the Research Objective

Again, wanting to provide many more insights into the nature of the casual mechanism of various crash types, this study will examine the socio-economic characteristic that may influence these crash outcomes. As in most human geography studies, it is difficult to estimate the population at risk present in the area or traversing it. There is an a priori reason that the morphology of the built-up environment could affect accident occurrence. Therefore, in the hierarchical structure of this analysis the county where the crash occurred was chosen as the second level and appropriate regional socio-economic response variables included. A very simple example of the multilevel equation for this study that considers both pavement condition and county socio-economic would be:

$$\textit{Crash outcome} = Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_{0j} + r_{ij}$$

where

X_{ij} = level-1 variable ride quality (IRI value)

W_j = level-2 variable average county population

r_{ij}= individual crash level-1 random effect

μ_{0j} = county level-2 random effect accounting for variation in socio-economics of the county.

Arguably, additional parameters may increase the fit of the model, but the results may be hard to interpret. Therefore, as the model is being stepped, all the parameters shall be tested in a chi-squared-versus-degrees-of-freedom manner or in terms of AIC and BIC to produce the best fitting and most parsimonious model.

The dependent discrete variables Y_{ij} in this part of the research are crash outcomes with binary results; therefore, discrete response models will be developed. As an example, does the IRI rating of the pavement have a positive or negative influence on whether or not a crash results in an injury? In this model, severity will be treated as a binary variable that is a discrete non-normally distributed variable that requires a non-linear discrete response model that will be modified to consider the hierarchical structure of the data. (Lenguerrand and Laumon, 2006)

For this analysis, a two level HGLM for binary data will be used with level-1 being the crash level. The factors chosen to describe this level will be PDVs that describe the condition and ride quality of the pavement at the site of the crash. Level-2 will be county/locality specific socio-economic variables. The HGLM modeling software developed by Bryk and Raudenbush will be used to analyze the models developed for this research objective. The following will reference their book and HGLM software manual to describe the multilevel binomial logistic model used in this research.

Unlike the normal multilevel regression model where the outcome variable is continuously distributed, the binary outcome model uses a binomial sampling model and logit link shown below. In the logit link function the φ_{ij} is the probability of success and η_{ij} is the log of the odds of success.

$$\eta_{ij} = \log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right)$$

using as the outcome variable in a HGLM two level combined model yields

$$\eta_{ij} = \text{logit}\ (\Theta) = \log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right) = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_{0j} + r_{ij}.$$

A predicted probability between zero and one can be calculated by

$$\varphi_{ij} = \left(\frac{1}{1+\exp\left(-\eta_{ij}\right)}\right).$$

HGLM is a stepwise model building process starting with the null or unconditional model with no predictors at any level. HGLM uses a sequential "model building" strategy in order to build the best model for reaching the research objective. The unconditional model becomes the baseline for which the subsequent models will be evaluated. In this study, the null model will determine the magnitude of the difference in the specific crash outcome across the different counties where the crashes occur. The level-1 model is

$$\eta_{ij} = \beta_{0j},$$

and the level-2 model is

$$\beta_{0j} = \gamma_{00} + \mu_{0j}, \text{ where } \mu_{0j} \sim N(0,\tau_{00}).$$

In this model, γ_{00} is the average log-odds of the crash outcome across all counties, and τ_{00} is the variance between the different counties in county log-odds of the crash outcome.

Based on the study's hypothesis a binary model via a random-intercept model will be developed to predict the effects of the county's socio-economic characteristics on crash outcomes. In this model, the intercept β_{0j} is allowed to vary across the counties (level-2) while holding the slope β_{1j} constant. The primary predictors are measured at level-2 (county), while the outcome variable is measured at level-1 (crash).

Specifically, the level-1 and level-2 models are structured as follows:

$$\eta_{ij} = \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} X_{pij} \qquad \text{(level-1)}$$

$$\beta_{0j} = \gamma_{00} + \sum_{s=1}^{S} \gamma_{0s} W_{sj} + \mu_{0j} \qquad \text{(level-2)}$$

where X_{pij} is a pavement distress or condition crash-level predictor variable and W_{sj} is socio-economic characteristic predictor variable of the county. The combined model is structured as follows where η_{ij} is the log-odds of a certain type of crash, given a crash occurs. As mentioned above, the probability of the crash can be calculated from the log-odds. The model results will provide the odds of a predictor variable that will either increase or decrease the odds that a particular crash type will occur.

$$\eta_{ij} = \gamma_{00} + \sum_{s=1}^{S} \gamma_{0s} W_{sj} + \sum_{p=1}^{P} \beta_{pj} X_{pij} + \mu_{0j}$$

6.4 HGLM: Procedure for Building the Level-2 Model

As in all modeling, choosing the predictor variables should be theory-driven, where specific hypotheses are posed about expected relationships in each of the Q+1 level-2 equations. The backward solution of entering all possible level-2 predictors and then removing the ones that are not significant may cause multi-collinearity problems (Byrk and Raudenbush). Recognizing this potential problem and because this research is primarily exploratory in nature, the predictors were divided into conceptually distinct subsets. Then the strongest of all the predictors from these sub-models were combined in the overall model (Byrk and Raudenbush). For this study, the level-2 county socio-economic predictor variables were divided into four sub-models: population/commuter, income/education, gender and race giving 23 predictor variables. The crash level (level-1) outcome variables will again be the two different crash types in which one can be

related to severity of the crash. They are crashes with injuries and rear-end crashes. Additionally, at the crash level are the same roadway and environmental predictor variables used in hypothesis tests for objective one along with the PDVs identified to be significant at the site of the crash.

The software package HLM version 6.0 was used to estimate the multi-level binomial logistic models. The first step is to prepare the data files using a statistical software package before importing the data structure into the HLM software. For this study, IBM Statistical Package for the Social Sciences (SPSS) Version 20 was used. This software is used to create two separate files for each level of the data. The uniquely aggregated data file described above containing crash, pavement condition and socio-economic data at the crash site is used as the level-1 dataset. From this dataset and using the aggregate command in SPSS the level-2 county data file was constructed. Each file should contain the outcome and predictor variables for that level, plus an identification code to link the variables between levels. In both datasets, the county where the crash occurred is given an identifier number, and this was used as the identification code to link the two levels. Once a data file has been created in this manner for each level, it is possible to import the data files into the HLM software.

To determine if the variation between the counties is significant enough to justify using multilevel modeling a model is estimated without predictor variables (unconditional or null model).

This is a simple model, and in a binary model with a logit link function the level-1 model is

$$\eta_{ij} = \beta_{0j},$$

and the level-2 model is

$$\beta_{0j} = \gamma_{00} + \mu_{0j},$$

where $\mu_{0j} \sim N(0,\tau_{00})$.

This model is estimated for each of the two binary crash outcome variables to determine the variance components of the random variables. The next step is to examine the proportion of the variance in the crash type outcome variables and the county socio-economic variables. Unlike the standard hierarchical regression model where this variance is examined by calculating the intra-class correlation (ICC) index which is the ratio of level-2 variance τ_{00} to the total variance given by the equation,

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}.$$

In multilevel binomial logistic models, a simplified approach is used. This is because the variance at the lowest level σ^2 (level-1) is determined by the proportion of successes and not available for use in the above equation. Instead, the standard logistic distribution for level-1 residual implies a variance of $\pi^2/3 = 3.29$, and the simplified equation becomes:

$$\rho = \frac{\tau_{00}}{\tau_{00} + 3.29}$$

The HGLM output for the Bernoulli case provides the level-2 variance τ_{00}, and the above equation allows for the Variance Partition Coefficient (VPC) (Dupont and Martensen, 2007, as cited in Goldstein, 2003 and Snijders and Bosker, 1999) to be calculated and examined. The results are presented along with the results for the conditional model.

Estimating a random-intercepts conditional model for each of the two binary crash outcomes to test the objective's hypotheses using pavement distress predictor variables at level-2 is the next step in the process. In this model, the intercept β_{0j} is allowed to vary randomly across the counties where the crash occurs to try to explain the differences among the socio-economic characteristics of the different counties on specific crash outcomes.

CHAPTER SEVEN

7.0 DATA ANALYSIS AND RESULTS FOR HGLM MODELS

7.1 Descriptive Analysis

The data consists of 21,556 vehicle crashes located within 35 counties in Virginia. The crash level (level-1) outcome variables will again be the two different crash types in which two can be related to severity of the crash and at the crash level are the same road-use, roadway and environmental predictor variables used in objective one along with the PDVs identified to be significant at the site of the crash. The descriptive statistics for these variables are presented in Tables 45 and 46 below, and analysis for these variables is presented above in sections 5.2a and 5.2b.

Table 45. Descriptive Statistics of crash level variables

	Variables	**N**	**Min.**	**Max.**	**No.**	**%**
Dependent Variables						
	INJURY	21556	0	1	6787	32%
	REAREND	21556	0	1	8887	41%
Physical Characteristics of the Road					No.	%
	ALIGN	21556	0	1	15092	70%
	WZ	21556	0	1	957	4%
Environmental					No.	%
	WEATHER	21556	0	1	17078	79%
	SURFCOND	21556	0	1	16471	76%
	LIGHTING	21556	0	1	13811	64%

Table 46. Descriptive statistics PDVs at crash site (ACP)

Variables	N	Minimum	Maximum	Mean / %
IRI	21556	27	435	100.42
TC1	21556	0	1	34.21%
TC2	21556	0	1	22.05%
LC1	21556	0	1	30.15%
LC2	21556	0	1	24.92%
LLJ1	21556	0	1	8.19%
LLJ2	21556	0	1	0.25%
RTC1	21556	0	1	31.97%
RTC2	21556	0	1	18.05%
RTC3	21556	0	1	13.82%
RLC1	21556	0	1	29.91%
RLC2	21556	0	1	13.10%
RLC3	21556	0	1	3.33%
AC1	21556	0	1	54.62%
AC2	21556	0	1	37.99%
AC3	21556	0	1	8.00%
PATCH	21556	0	1	13.70%
POT	21556	0	1	0.17%
DELAM	21556	0	1	1.05%
BLEED1	21556	0	1	0.71%
BLEED2	21556	0	1	0.02%
RUT SE	21556	0.00	0.69	0.13
RUT WM	21556	0.00	0.71	0.20
CCI	21556	0	100	76.59

At level-2 the predictor variables will be the socio-economic characteristics of the 35 counties where the crashes occurred. These predictors were described in the previous chapter, and Table 47 below provides the descriptive statistics. As mentioned above, the socio-economic data was collected from the 2006-2010 5-Year American Community Survey (ACS) which is a household survey conducted by the U.S. Census Bureau that currently has an annual sample size of about 3.5 million addresses and coincides with the crash and pavement condition data years. The data are found in the American Community Survey Summary File (ACSSF). This is a unique data product that includes

all the estimates and margins of error from the detailed tables and geographies that are published for the ACS. Data contained in the ACS Summary File cover demographic, social, economic, and housing subject areas. When using data from surveys of this kind, human sampling error is the primary concern of researchers. In the ACS data files, the margins of error (MOE) due to sampling are provided and based on the 90-percent confidence interval. Combining this with the accuracy of medium to large sample sizes of this study these MOEs are deemed satisfactory when considering the exploratory nature of this study. However, when conducting more detailed research in this area, the researcher may choose to use the methods recommended by the Census Bureau to calculate their own estimates of standard error due to sampling. One very noteworthy descriptive of each of the socio-economic parameters is the large range between the minimums and maximums which is indicative of long interstate highways that traverse a very diverse set of localities. For example, the parameter total population ranges from approximately 12,000 for some of the more rural counties to the urbanized counties with population over 1,000,000. This broad range of diversity is indicated in the majority of the county socio-economic characteristics which enhances the uniqueness and generalizabilty attributes of this study.

Table 47. Descriptive statistics of county level variables

County-Level Variable	N	Minimum	Maximum	Mean	Std. Deviation
POP	35	12167.00	1048554.00	125536.54	193530.09
POPWORK	35	4822.00	608225.00	69480.89	112399.71
UNEMP	35	3.20	10.90	5.96	1.95
COMMUT	35	4296.00	568600.00	64439.63	104923.51
SOV	35	3451.00	412332.00	49137.83	77169.34
HOUSEHLD	35	3385.00	381768.00	46853.31	70626.71
HHINCOME	35	44129.00	132662.00	75138.17	23172.78
PERCAPINC	35	16735.00	57724.00	29015.74	8585.37
NOGED	35	505.00	30164.00	3659.57	5293.35
GED	35	3221.00	99012.00	19400.40	21049.46
ASSDEGREE	35	218.00	37449.00	5504.94	7793.23
BACHDEGRE	35	342.00	217459.00	18997.69	38454.28
GRADDEGREE	35	133.00	194382.00	13962.57	33636.77
MALE	35	6657.00	517987.00	61970.86	95334.27
FEMALE	35	4475.00	530567.00	63565.69	98231.17
AVGAGE	35	26.00	45.30	39.31	4.40
@65OLDR	35	1460.00	98779.00	13483.03	17818.21
M65OLDR	35	655.00	44220.00	5824.57	7808.21
F65OLDER	35	805.00	54559.00	7658.46	10030.99
WHITE	35	4620.00	680404.00	88832.80	123734.57
AFRAM	35	470.00	106147.00	19812.69	31112.04
ASIAN	35	0.00	181897.00	9117.91	30953.27
OTHER	35	95.00	91296.00	7773.14	17785.77

7.2 Results of HGLM Modeling

Objective Three: In conjunction with goals one and two, uniquely model the hierarchical nature of a crash to determine which regional socio-economic factors are related to traffic safety.

Based on the study's hypothesis, a binary model via a random-intercept model was developed to predict the effects of the county's socio-economic characteristics on crash outcomes. In this model, the intercept β_{0j} is allowed to vary across the counties (level-2) while holding the slope β_{1j} constant, and the primary predictors are measured at level-2 (county), while the outcome variable is measured at level-1 (crash). From the following equation, the model results provide the odds of a predictor variable that will either increase or decrease the odds of a particular crash type will occur:

$$\eta_{ij} = \gamma_{00} + \sum_{s=1}^{S} \gamma_{0s} W_{sj} + \sum_{p=1}^{P} \beta_{pj} X_{pij} + \mu_{0j}.$$

Because this research is primarily exploratory in nature, the predictors were divided into conceptually distinct subsets, and then the strongest of all the predictors from these submodels were combined in the overall model (Byrk and Raudenbush). For this study, the level-2 county socio-economic predictor variables were divided into four submodels: population/commuter, income/education, gender and race. The crash level (level-1) outcome variables will again be the two different crash types in which two can be related to severity of the crash. They are crashes with injuries and rear-end crashes. Additionally, at the crash level are the same road-use, roadway and environmental predictor variables used in hypothesis tests for objective one along with the PDVs identified to be significant at the site of the crash.

Although technically different, the fixed part of the HGLM can be interpreted in the same manner as an ordinary logistic regression based on the coefficients (i.e., odds ratios) (Jones and Jorgenson, 2003), but one of the primary objectives was to determine if there are statistically significant variations in the crash outcomes due to the county's different socio-economic features after controlling for the fixed regressors. Using the two crash outcomes as dependent variables, two separate HGLM unconditional (empty) models were estimated and the random effects determined. As explained in the previous section, using the random effects and examining the variance proportion between level-1 and level-2 for each of the models provides meaningful information on the relevance of analyzing the clustering nature of the data. A VPC estimated close to zero means very little variation exists between the outcome variable and the county's socio-economic factors, and it therefore can be estimated by using only the level-1 crash predictor variables.

From the random effect parameters in Table 48 below the VPC for crashes with injuries is 0.010 indicating only 0.1% of the total variance comes from the variance at the level-2 county level. This is a relatively small VPC indicating little justification for estimating the outcomes from the nested structure of the data using HGLM, under these assumptions. Therefore, this dependent variable is excluded from further study using this methodology.

However, the VPC for rear-end is more significant at 0.168 meaning that 16.8% of the variability in these types of crashes can be explained by the nested structure and the

variance in the county's socio-economic factors and it would be beneficial to model the data using HGLM.

The second objective is to determine which predictor variables at each level of the nested structure are associated with an increase or decrease in the probability of a particular crash outcome. The primary interest was to determine if the socio-economic characteristics of the county where the crash occurred has an association with the study's crash outcome response variables. Since HGLM is not justified for injury crashes the study will focus on rear-end crashes. As mentioned above, for this study, the level-2 county socio-economic predictor variables were divided into four submodels: population/commuter, income/education, gender and race with 23 predictor variables modeled.

Similar to standard logistic regression, the odds ratio is used to interpret the influence of the predictor variables on the response variables. The odds ratio of estimated coefficients indicates how the odds of an event (crash) is affected by the presence of socio-economic characteristic and/or a pavement distress variable (Kim, et al. 2007). To get the likelihood of the event occurring with a unit change in the predictor variable in percentage, 1 is subtracted from the odds ratio and then multiplied by 100.

From the four submodels the results indicate only two out of the 23 socio-economic factors models are associated with rear-end crashes, and they are the average age (AVGAGE) and unemployment percentage (UNEMP). Both of these factors are

significant (after holding other predictors constant) and have a positive association in reducing the odds of these types of crashes from occurring. As the average age increases, the odds of having a rear-end crash decreases slightly by a factor of 0.926. This means that the odds of having a rear-end crash is reduced by 7.4% as the average age in the county increases. As the unemployment rate increases, the odds of having this type of crash decrease by a slightly higher percentage of 16%. One could argue that as the unemployment rate of a locality goes up, there would be less volume of traffic on the roadways making the results here consistent with other studies relating traffic volume to crashes. Also, it has been shown that the age of the driver has an association with traffic safety, and to a point, as drivers mature they are less likely to have accidents. The results are presented in Table 49.

Based on the results presented there is logical confirmation that certain socio-economic factors of the counties where crashes occurs have a statistically significant association with the likelihood of having a rear-end crash. Lastly, it is noteworthy that similar PDVs have similar consequences when using both standard logistic regression and HGLM.

Table 48. Results of HGLM for dependent variable Injury

	Crashes with Injury
Null Model	
Random effect τ_{00}	0.033
VPC	0.010

Table 49. Results of HGLM for dependent variables Rear-end

	Rear-End crashes		
Null Model			
Random effect τ_{00}	0.666		
VPC	0.168		
Conditional Model	Coefficient	*p*- value	Odds Ratio
Fixed Effects			
Intercept	-1.077		
Crash-level			
LC2	0.212	0.002	1.236
LLJ1	-0.223	0.001	0.800
RTC1	-1.130	0.016	0.880
RTC3	0.173	0.004	1.188
POT	-0.760	0.059	0.468
PATCH	0.285	0.000	1.330
AC2	-0.293	0.000	0.746
County-level			
AVGAGE	-0.077	0.018	0.926
UNEMP	-0.174	0.015	0.840
Random effects			
Level-2 variance τ	0.226		
Reliability	0.879		

CHAPTER EIGHT

8.0 FINAL CONCLUSIONS

8.1 Purpose

As mentioned in the introduction, there are numerous factors affecting general road safety and research in roadway safety is categorized into three main categories, vehicle, driver and environment each with their own set of sub-variables. The condition of the pavement is a general roadway condition factor and falls within the environmental section of the roadway safety matrix. Drivers expect smooth riding and quiet pavement and when this expectation is met, they feel safe and remain in better control of the vehicle. When this expectation changes due to rough pavement, the driver becomes anxious which affects the ability to collect information and carry out intended maneuvers. Studies have shown that roughness affects safety in many ways. In particular, the ability of the driver to steer and brake can significantly affect the overall controllability of the vehicle (Burns, 1981).

In theory, each of the predictor PDVs described in section 3.2d could compromise the smoothness of the roadway and, thus, have an effect on safety. This is the primary hypothesis of the research. Additionally, the effect of inconsistency traveling from a smooth section of pavement to a rough section has the potential to hinder the driver's ability to control a vehicle and a high speeds this effect is compounded. Therefore, the presence of one or more of the pavement distress variables (at various severity levels) presented in this study may have a bearing on roadway safety.

In general, previous studies have proven that the condition of the pavement described by certain general parameters is, within a specified confidence level, statistically significant in association with roadway safety. However, this study uniquely expanded that hypothesis by exploring this relationship from a micro-level and examined each pavement distress variable collected from annual pavement condition data. It also provided more information on the yet to be answered question, which type of pavement is safer: asphalt or concrete. In general, the goal was to determine which of these PDVs, along with pavement type, have an association with certain crash outcomes (i.e., roadway safety). In addition, the originality of this study was enhanced by examining the spatial aspect of this question. Not only was this relationship modeled at the crash site but also at specific intervals upstream of the crash site to determine where within the footprint of the crash is the critical location with respect to pavement condition. Lastly, the most unique aspect of this study was to explore the hypothesis that certain socio-economic factors of the locality where the crash occurred may have a relationship to the outcomes of a crash.

The first two questions were answered by developing a method to create unique datasets that combined crash with detailed pavement condition and pavement type data at the crash site and three intervals upstream and using proven logistic regression modeling techniques. The third question was answered by expanding the datasets to include socio-economic predictor variables at the county level and using a methodology that is somewhat unique to this area of study: hierarchical generalized logistic regression modeling.

The primary purpose of this research was to determine which of the PDVs have the potential to influence the safety of the roadway pavement and bring safety into the realm of pavement maintenance planning. Using both a large number of crash cases and pavement distress predictor variables, this study was limited to the reasonable quality of the crash and pavement condition data collected both by humans and electronically. Nonetheless, holding other factors constant, certain PDVs are, within a high level of confidence, statistically significant in the association to increase the odds of certain crash outcomes and, in particular, rear-end and multi-vehicle crashes. These initial findings, reported below, both uniquely expand existing hypotheses and bring new insights that should have a profound contribution on the body of knowledge in this area. Additionally, the unique methods for creating the study's datasets can be repeated, and the results generalized/expanded upon for use in other states.

8.2 Overview of Important Results

This research used the complex set of PDVs collected by the VDOT (which is similar to the data collected by other states) along with crash data to create a unique dataset in order to evaluate, in detail, the relationship between the type of pavement and the performance of the pavement structure and various crash outcomes. Coupling this with a proven statistical modeling methodology, this study expanded the use of already collected detailed pavement performance data by linking it to pavement/roadway safety. In addition, this dataset and procedure can be repeated for use in other states across the country.

The three pavement types considered are asphalt concrete (ACP), continuously reinforced concrete (CRCP) and jointed reinforced concrete (JRCP). ACP includes bituminous concrete over jointed reinforced concrete (BOJ), and bituminous concrete over continuously reinforced concrete (BOC). With concrete and asphalt pavement materials varying only slightly across the Mid-Atlantic, this study can be generalized for other Departments of Transportation in this region and the methodology applicable for similar studies across the country. In order to account for the randomness of the occurrence of vehicle crashes, stochastic regression models were used to determine these relationships.

For each of the three pavement types evaluated (ACP, JRCP and CRCP) in the model for INJURY crash outcomes, the results of the omnibus tests of model coefficients presented in the model summary indicate that the entire model is significant. However, as somewhat expected, the McFadden's R-squared results indicate that the model is weak in predicting the variance in the independents. This puts into question the conclusiveness of the results for the INJURY model and based on the results, one can conclude that pavement condition distress and ride quality variables have little to no impact on whether a crash results in an injury. However, with respect to ACP one should note the distress variable RUTSE is very substantial in increasing the odds a crash results in an injury. And in theory this condition, if severe enough, could significantly impair a driver's ability to control a vehicle on a high-speed interstate route and very possibly resulting in a crash with an injury.

There a number of PDVs that are significantly associated with rear-end crashes for all three-pavement types ACP, JRCP and CRCP. The omnibus tests of model coefficients presented in the model summary indicates the models are significant, and from the McFadden's R-squared results the models are slightly better in predicting the variance in the independents. And from the classification table, a good percentage of observed data are predicted correctly. When broken down, each of the models have a good "hit rate" for predicting the observed REAREND crashes.

For rear-end crashes on ACP, it is worthy to note that as PDVs increase in severity so does the odds of having a rear-end crash. Therefore, the results for certain distress variables indicate as the severity of the pavement distress variable increases (i.e., the condition of the pavement worsens) it is likely to increase the association between the chance of a rear-end crash and these diminishing pavement conditions. On CRCP pavements, noteworthy is the fact there are no PDVs associated with an increase in the odds of having rear-end crashes. On JRCP pavements there a number of PDVs that are statistically significantly associated with rear-end crashes. With respect to rear-end crashes, this overall observation is an important one and contributes to the discussion of which type of pavement is safer – concrete or asphalt.

With respect to PTYPE, the type of pavement is not statistically significant in the INJURY model and is significant in the REAR-END model. Therefore, given a crash occurs, the type of pavement has a statistically significant relationship with rear-end crashes, and the odds of having a rear-end crash is increased on concrete pavement. The

type of pavement does not have a statistically significant association with crashes with injuries, given a crash occurs.

To test the second hypothesis and to explore the spatial aspect of this relationship, models were developed to evaluate crash outcomes at the crash site and specific intervals upstream. These models were used to evaluate where within the footprint of the crash is the most critical location with respect to pavement condition and/or ride quality. It is important to note that the model estimated for INJURY, for all asphalt routes and all crash types, resulted in pavement condition and ride quality parameters consistent with those of objective one. The results from the INJURY model indicate the overall best fitting model is not at the crash site but approximately 0.15 upstream. The pavement distress parameter showing the highest odds ratio is approximately 0.20 miles upstream. Noteworthy is the fact that this parameter was also identified in the results for objective one having a significant relationship to crashes with injuries.

Generally, the results from these models are consistent with those presented in objective one of this study in that the pavement condition distress and ride quality variables have little to no impact on whether a crash results in an injury, but the objective was to estimate the critical location within the footprint of the crash for assessing the importance of pavement condition and ride quality on the severity of crashes, given a crash occurs. Based on the results from the models, this location is upstream of the crash site approximately 0.15 to 0.20 miles. Additionally, one should again note that in theory, the critical distress variable RUTSE if severe enough could significantly impair a driver's

ability to control a vehicle very possibly resulting in a crash with an injury. This effect is also likely to originate somewhere upstream of the final crash site and therefore, these results are consistent with crash theory.

For the crash type REAREND the results from the model indicate the critical location is approximately at the crash site and 0.15 to 0.20 miles upstream of the crash site, respectively. The majority of the pavement distress parameters (Exp (B)) estimated in the models that increase the odds of having a rear-end crash are highest at the crash site. The model summary also shows the data best fits this model. Therefore, based on the results of these models, the critical location for assessing the importance of pavement condition on rear-end crashes is at the site of the crash. Again, this is consistent with crash theory in that rear-end crashes are usually sudden, and the effecting factors are usually located close to the crash site.

A unique aspect of this research was to employ a relatively new modeling methodology used to fully exploit the clustering nature of road crash data. Crash, pavement condition and socioeconomic census data was combined to create unique datasets that were used to create a parsimonious comprehensive predictive model, unique datasets were developed by combining pavement condition and ride quality (IRI roughness index) along with roadway environmental and local socioeconomic data. This data was aggregated with available crash data to determine which of these parameters, solely or in combination, are important in explaining the outcomes of vehicle crashes. To analyze these unique datasets, this study used multilevel modeling techniques to analyze the hierarchical nature

of crash data and examine the socio-economic characteristic that may influence crash outcomes.

For crash outcome with injuries, only 0.1% of the total variance comes from the variance at the level-2 county level. This is a relatively small VPC indicating little justification for estimating the outcomes from the nested structure of the data using HGLM, under these assumptions. Therefore, this dependent variable was excluded from further study using this methodology.

However, the VPC for rear-end crashes are more significant at 0.168 meaning that 16.8% of the variability in these types of crashes can be explained by the nested structure and the variance in the county's socio-economic factors. It would be beneficial to model the data using HGLM. Two socio-economic factors are associated with rear-end crashes. They are the average age (AVGAGE) and unemployment percentage (UNEMP). Both of these factors are significant (after holding other predictors constant) and have a positive association in reducing the odds of this type of crash. The odds of having a rear-end crash is reduced by 7.4% as the average age in the county increases. As the unemployment rate increases, the odds of having this type of crash decrease by a slightly higher percentage of 16%. One could argue that as the unemployment rate of a locality goes up, there would be less traffic on the roadways making the results here consistent with other studies relating traffic volume to crashes. Also, it has been shown that the age of the driver is associated with traffic safety, and to a point, as drivers mature they are less likely to have accidents.

Based on the results presented there is logical confirmation that within this empirical setting certain socio-economic factors of the counties where crashes occurs have a statistically significant association with the likelihood of having a rear-end crash. Lastly, it is noteworthy that similar PDVs have similar consequences when using both standard logistic regression and HGLM.

8.3 Summary and Discussion

Based on the underlying premise that pavement condition and ride quality may directly or indirectly influence the consequences and severity of a crash, it is worthwhile to explore this relationship in detail in order to gain a better understanding of its effect on overall roadway safety. Researching this association has the potential to increase the body of knowledge and serve as the basis for bringing safety directly into the planning of pavement maintenance. This study focused on particular crash types that are related to crash severity and measured their significance to determine if these potentially safety critical parameters associated with poor pavement condition are substantial.

In spite of their limitations, statistical models are a proven tool for estimating the significance of the relationship between the various casual factors mentioned above and vehicle crashes. As shown in the Haddon matrix, there is an a-priori reason that the morphology of the built-up environment could affect accident occurrence and wanting to provide additional insights into the nature of the spatial and casual mechanism of various crash characteristics. To fully exploit the clustered nature of road crashes and develop a comprehensive predictive statistical model, unique datasets were developed by

combining pavement condition and ride quality (IRI roughness index) along with roadway environmental and local socioeconomic data. This data was aggregated with available crash data to determine which of these parameters, solely or in combination, are important in explaining the outcomes of vehicle crashes. To analyze these unique datasets, this study used multilevel modeling techniques to analyze the hierarchical nature of crash data and examine the socio-economic characteristic that may influence crash outcomes.

While this study examined the severity of the crash (given a crash occurred) by including injurious crashes, the results indicate that only a few PDVs influence these two outcomes. However, this is stated not to dismiss the potential importance of this relationship but merely as an observation that may deserve additional research. On the other hand, rear-end crashes were shown to be influenced by a number of statistically significant PDVs. As stated in the body of the study, one would expect these results on interstate highways because this is the most common type of crash on multi-lane high-speed highways where large volumes of vehicles are traveling together in the same lane. In fact, the National Highway Traffic Safety Administration (NHTSA) estimates that 30% of all crashes are rear-end crashes. Additionally, recent studies indicate these types of crashes are not only frequent but also cause a lot of property damage and injuries. This has gotten the attention of many DOTs including Virginia where the results of a traffic safety analysis conducted from 2008 to 2010 on the I-64 corridor between Newport News and Richmond, Virginia indicate 48% of all crashes were rear-end accidents (VDOT, 2012). Also, the Ohio Department of Transportation initiated a recent study with the main

objective to develop a set of strategies to reduce rear end crashes on interstate and non-interstate highways. In the abstract they recognize in order to develop a comprehensive set of strategies you have to have a good understanding likely countermeasures for various "geometric, operational, and environmental factors" (ODOT, January 2009). In this regard, it is important to not only understand these causal factors but also have a better understanding of how a driver responses to them.

Therefore, this research is timely because the driver's ability to collect information and carry out intended maneuvers is greatly compromised due to the vibrations encountered on rough roadways (TRB, 2009). Recent studies have shown that roughness affects safety in many ways. In particular, the ability of the driver to steer and brake can significantly affect the overall controllability of the vehicle (Burns, ?). In extreme cases, it is widely understood that wash- boarding surfaces and repeated cycling undulations of the surface severe enough to shake a truck or loaded vehicle so much as to lose part, or all, of its load potentially causing catastrophic multi-vehicle crashes. It has been shown that only seven to ten percent of roadway crashes can be attributed to the road and its environment, but when combined with human error, more than twenty percent of crashes can be ascribed to this combination (TRB, 2009).

With the addition of a spatial component, this research contributed to the area of transportation safety by providing results that explore in more detail the effects of the condition of pavement and location of the crash by including socioeconomic factors. The results of this study increase the body of knowledge of the spatial and causal mechanism

of pavement-related roadway crashes for use by highway agencies in developing pavement management strategies to improve traveler safety. This will assist transportation officials in developing strategies for pavement maintenance that focus on reducing the number of pavement-related crashes and in essence, by conducting this in-depth study, we have a better understanding of this important relationship and the consequences of crashes may be reduced by changes in how roads are maintained.

Additionally, this study provides unique and valuable insights into this critical relationship so effective site-specific countermeasures may be identified. It provides transportation agencies additional information on the safety of their pavements to assist them in optimizing their overall safety programs. This was the motivation and main objective behind this research.

8.4 Future Research

This micro-level study that explored the relationship between individual PDVs and crash outcomes, particularly rear-end crashes, lays the foundation for future research. The method for creating the sample data by aggregating crash, pavement condition, and county socio-economic was unique and developed specifically for this study so the spatial aspect of this relationship could be examined at this level. It is repeatable and could be used in future research.

The modeling approaches used in this research go beyond the traditional methods offering future researchers insights into the flexibility of these techniques. The

advantages of the HGLM need to be used in future research alone or in combination with traditional methods to fully exploit the nested structure of crash data. Specifically related to this study, certain PDVs and/or ride quality as expressed by IRI should be modeled as the dependent variables to determine if socio-economic factors of the locality where the crash occurred has an influence on pavement condition and/or ride quality which can now be directly related to roadway safety. This could start to answer the question do rural counties have better roads than urban counties, and if so, why?

The predictor PDVs in this study were binary exploring whether or not the presence of a certain pavement distress variable influenced this relationship. Non-binary PDVs could be used to study the relationship between roadway pavement condition and safety. The pavement condition data are collected as continuous variables. Using various traditional and multilevel regression modeling techniques future research could determine "hot spots" and safety thresholds for each of these PDVs. This information could then included in optimization models to assist state transportation agencies in answering the question, at what point does the condition of the pavement become critical? This would greatly improve pavement maintenance planning.

Lastly, modeling and simulation techniques are at the forefront of current research in the area of transportation planning and safety. These techniques could be used to simulate the behavior of drivers/vehicles under various pavement condition scenarios to further investigate on a micro-level the relationship between overall roadway safety and pavement distress variables as they relate to the general condition of the pavement.

These simulation models could expand the area of study to include primary and secondary routes by including common roadway and environmental characteristics for these types of routes.

BIBLIOGRAPHY

Agent, K.R., Pigman, J.G., and Green, E.R. *Effect of Pavement Resurfacing on Traffic Safety.* Research report Kentucky Transportation Center, University of Kentucky, Lexington, KY., February, 2004.

Aguero-Valverde, J. and Jovanis, P.P. *Spatial analysis of Fatal and Injury Crashes in Pennsylvania.* Accident Analysis and Prevention, Volume 38, Issue 3, May 2006, pp. 618-625.

Al-Masaeid, H.R. *Impact of pavement condition on rural road accidents.* Canadian Journal of Civil Engineering, Volume 24, 1997, pp. 523-531.

American Association of State Highway and Transportation Officials. *Guide for Pavement Friction.* Joint Technical Committee on Pavements, 2007-2008.

American Association of State Highway and Transportation Officials. *Rough Roads Ahead, Fix them Now or Pay for it Later.* 2009.

American Automobile Association. *2009 Traffic Safety Culture Index.* A study conducted by AAA Foundation for Traffic Safety, 2009.

American Concrete Pavement Association. *Current Perspectives on Pavement Surface Characteristics.* R&T Update Concrete Pavement Research and Technology. April, 2007.

Anastasopoulos, P.Ch., Tarko, A.P., and Mannering, F.L. *Tobit analysis of vehicle accident rates on interstate highways.* Accident Analysis and Prevention, 40, 2008, pp. 768-775.

Asphalt Pavement Alliance. *Smoothness Matters- Smooth pavements save fuel and even small changes can make a big difference.* Excerpts from an article by Howard, M., Hot Mix Asphalt Technology, Volume 14, No. 6, 2009, pp. 18-29.

Bester, C.J. *The Effect of Road Roughness on Safety.* Department of Civil Engineering, University of Stellenbosch, Matieland 7602, South Africa. Paper presented at the Transportation Research Board 2003 Annual Meeting.

Burns, J.C. *Roughness and Roadway Safety.* Transportation Research Record 836, Washington, D.C., National Academy of Science, 1981, pgs. 8-14.

Cairney, P. and Bennett, P. *Relationship between road Surface Characteristic and Crashes on Victorian Rural Roads.* 23rd ARRB Conference- Research Partnering with Practitioners, Adelaide, Australia, 2008.

Chan, C.Y., Huang, B., Yan, X., and Richards, S. *Effects of Asphalt Pavement Conditions on Traffic Accidents in Tennessee Utilizing Pavement Management System (PMS).* Southeastern Transportation Center, University of Tennessee, Knoxville, 2008.

Craus, J., Livneh, M., and Ishai, I. *Effect of Pavement and Shoulder Condition on Highway Accidents.* Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., No. 1318, 1991, pp. 51-57.

Das, A., Abdel-Aty, M., and Pande, A. *Genetic programming to investigate design parameters contributing to crash occurrence on urban arterials.* Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., No. 2147, 2010, pp. 25-32.

Dupont, E., and Martensen, H. *Multilevel modeling and time series analysis in traffic safety research- Methodology.* Deliverable D7.4 of the EU FP6 project SafetyNet, 2007.

Eckhardt, N., and Thomas, I. *Spatial Nested Scales for Road Accidents in the periphery of Brussels.* International Association of Traffic and Safety Science, Volume 29 No. 1, 2005, pp. 66-78.

Elvik, R. and Vaa, T. *The Handbook of Road Safety Measures.* Institute of Transport Economics, Oslo, Norway, 2004.

Erwin, T. and Tighe, S.L. *Safety effect of preventive maintenance.* Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., No. 2044, 2008, pp. 79-85.

Federal Highway Administration – Office of Highway Policy Information. *HPMS Reassessment 2010 Final Report.* September 2008.

Goldstein, H. *Multilevel Statistical Models.* Textbook. London: Institute of Education, Multilevel Models Project, April, 1999.

Hibbs, B.O. and Larson, R.M. *Tire Pavement Noise and Safety Performance.* Department of Transportation Federal Highway Administration, FHWA-SA-96-068, 1996.

Ihs, A., Velin, H., and Wiklund, M. *The influence of Road Surface Condition on Traffic Safety.* VTI Publication M909.

Jayawickrama, P.W., Prasanna, R., and Senadheera, S.P. *Survey of state practices to control skid resistance on hot-mix asphalt concrete pavements.* Transportation Research Record 1536.

Jones, A.P. and Jørgensen, S.H. *The use of multilevel model for the prediction of road accident outcomes.* Accident Analysis and Prevention 35, 2003, pp. 59-69.

Karlaftis, M.G., and Golias, I. *Effects of road geometry and traffic volumes on rural roadway accident rates.* Accident Analysis and Prevention 34, 2002, pp. 357-365

Khattak, A. J., Kantor, P. and Council, F. M. *Role of adverse weather in key crash types on limited-access roadways.* Transportation Research Record No. 1621, 1998, pp.10-19.

Khattak, A.J., Khattak, A.J., Hummer, J.E. and Sickling, D.L. Are Concrete Roadways safer than Asphalt Roadways? Comparison of Crash Frequency, Rates and Injury Severity using California Data. Submitted to American Concrete Pavement Association, 2007.

Kim, D-G., Lee, Y., Washington, S., and Choi, K. Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. Accident Analysis and Prevention 39, 2007, pp. 125-134.

Kopelias, P., Papadimitriou, F., Papandreou, K., and Prevedouros, P. *Urban Freeway Crash Analysis – Geometric, Operational and Weather Effects on Crash Number and Severity.* Transportation Research Record: Journal of the Transportation Research Board, No. 2015, Washington, D.C. 2007, pp 123-131.

Larson, R.M. Using friction and texture data to reduce traffic fatalities, serious injuries, and traffic delays. Applied Pavement Technology, Inc., Springfield, Virginia

Leden, L., Hamalainen, O., and Manninen, E. *The Effect of Resurfacing on Friction Speeds and Safety on Main Roads in Finland.* Accident Analysis and Prevention, Volume 30, No. 1, 1998, pp. 75-85.

Lenguerrand, E., Martin, J.L., and Laumon, B. *Modelling the hierarchical structure of road crash data – Application to severity analysis.* Accident Analysis and Prevention 38, 2006, pp. 43-53.

Li., Y. and Bai, Y. Development of crash-severity-index models for measurement of work zone risk levels. Accident Analysis and Prevention 40, 2008, pp.1724-1731.

Lleras, C. *Path Analysis.* Encyclopedia of Social Measurement, Volume 3, pgs. 25-30, 2005.

Lord, D., and Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. March 22, 2010.

Mahone, D.C., and Sherwood, C. *Virginia's wet accident reduction program: A user's manual.* Virginia Transportation Research Council, Charlottesville, Virginia, May 1996.

Mahoney, J.P., Uhlmeyer, J., Morin, P., Luhr, D., Willoughby, K., Muench, S.T., and Baker, T. *Pavement Preservation Funding and Perfromance in Washington State.*

Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., No. 2150, 2010, pp. 55-62.

McLean, J., and Foley, G. *Road surface characteristics and condition: effects on road users.* ARRB Transport Research, research report ARR 314, January 1998.

Miller, T.R., and Zaloshnja, E. *On a Crash Course, the Dangers and Health Costs of Deficient Roadways.* A study by the Pacific Institute for Research and Evaluation, May 2009.

Milton, J.C., V.N. Shankar, and F.L. Mannering. *Highway accident severities and the mixed logit model: An exploratory empirical analysis.* Accident Analysis and Prevention, 40, 2008, pp. 260-266

National Concrete Pavement Technology Center. *Strategic Plan for Improved Concrete Pavement Surface Characteristics.* Center for Transportation Research and Education, Iowa State University. Sponsored by Federal Highway Administration under Cooperative Agreement DTFH61-01-X-00042 (Project 15), 2006.

National Highway Transportation Safety Association. Traffic Safety Facts 2009 – A compilation of Motor Vehicle Cradh Data from the Fatality Analysis Reporting System and the General Estimates System. National Center for Statistics and Analysis, United States Department of Transportation, DOT-HS-811_402, Washington, D.C., 2009.

Nowakowska, M., *Logistic Models in Crash Severity Classification Based on Road Characteristics.* Transportation Research Record: Journal of the Transportation Research Board, No2148 Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 16-28.

Noyce, D.A., Bahia, H.U., Yambo, J., Chapman, J., and Bill, A. *Incorporating Road Safety into Pavement Management: Maximizing Surface Friction for Road Safety Improvements.* Midwest Regional University Transportation Center, College of Engineering, Department of Civil and Environmental Engineering, University of Wisconsin, Madison, 2007.

Oh, S., Chung, K., Ragland, D.R., and Chan, C. *Analysis of Wet Weather Related Collision Concentration Locations: Empirical Assessment of Continuous Risk Profile.* Paper presented at the Transportation Research Board Annual Meeting, 2009.

Oh, S., Ragland, D.R., and Chan, C. Safety Performance of Experimental Pavement Type in California Using Before-and-After Comparisons. Transportation Research Board Annual Meeting paper, 2010.

Organisation for Economic Co-Operation and Development. *Road Safety Principles and Models: Review of Descriptive, Predictive and Risk and Accident Consequence Models.* Research review was prepared by an OECD Scientific Expert Group, Paris, France, 1997. It is the sister report to the main publication Road safety principles and models (IRRD No. 888815).

Pant, P.D., and Panta, S. *Crash Base Rates for Freeways/Reduction Strategies for Rear End Crashes.* Volume 2: Reduction Strategies for Rear End Crashes in Ohio. Ohio Department of Transportation, 2009.

Park, Y-J., and Saccomanno, F.F. *Collision frequency analysis using tree-based stratification.* Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., No. 1908, 2005, pp. 121-129.

The PEW Center on the States and the Rockefeller Foundation. *Measuring Transportation Investments: The Road to Results.* The PEW Charitable Trusts and the Rockefeller Foundation, May 2011.

Raudenbush, S.W., and Bryk, A.S. Hierarchical Linear Modeling – Applications and Data Analysis Methods, Second Edition. Sage Publications, Inc., 2002.

Reid, R.A., and Clark, T.M. *Roughness on Virginia Roads – 2004 Annual Interstate Roughness Report.* Virginia Department of Transportation, 2004.

Roberts, J.K. *An Introductory Primer on Multilevel and Hierarchical Modeling.* Learning Diabilities: A Contemporary Journal 2 (1), 30-38, 2004.

Sandberg, U. Influence of Road Surface Texture On Traffic Characteristics Related to Environment, Economy and Safety – A State-of-the art Study Regarding Measures and Measuring Methods. Swedish National Road and Transport Research Institute, 1998.

Seiler-Scherer, L. *Is the correlation between pavement skid resistance and accident frequency significant?* Paper presented at the Swiss Transport Research Conference, March 25-26, 2004.

Shafizadeh, K. and Mannering, F. *Acceptability of Pavement Roughness on Urban Highways by Driving Public.* Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., No. 1860, Paper No. 03-4430, pp. 187-193.

Shankar, V., F. Mannering, and W. Barfield. *Statistical Analysis Accident Severity on Rural Freeways.* Accident Analysis and Prevention, Vol. 28, No. 3, 1996, pp.391-401.

Smith, J.T., and Tighe, S.L. *Assessment of overlay roughness in long-term pavement performance test sites.* Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., No. 1869, 2004, pp. 125-135.

Statnotes from North Carolina State University, Public Administration. *Logistic Regression,* 2009.

Tighe, S.L., Li, N., Cowe Falls, L., and Hass, R. *Incorporating road safety into pavement management.* Transportation Research Record 1699, Paper No. 00-0431.

Titi, H.H. and Rasoulian, M. *IRI Smootheness Criteria for Asphalt Concrete Pavements in Louisiana.* Louisana Transportation Research Center.

Transportation Research Board (TRB) of the National Academies. *Influence of Roadway Surface Discontinuities on Safety.* Transportation Research Circular E-C134, May 2009.

Vanlarr, W. and Yannis, G. *Perception of road accidents causes.* Accident Analysis and Prevention, 38, 2006, pp. 155-161.

Viner, H.E., Sinhai, R. and Parry, A.R. *Linking Road Traffic with Skid Resistance – Recent UK Developments.* Copyright TRL Limited, 2005.

Virginia Department of Transportation. *State of the Pavement – 2008.* Prepared by the Virginia Department of Transportation Maintenance Division. January, 2009.

Virginia Department of Transportation. A Guide to Evaluating Pavement Distress Through the Use of Digital Images- Version 2.42. Asset Management Division, May 2007.

Virginia Department of Transportation. Interstate 64 Peninsula Study-Purpose and Need Technical Memorandum, 2012

Woltman, H., Feldstain, A., MacKay, J.C., and Rocchi, M. *An introduction to hierarchical linear modeling.* Tutorials in Quantitative Methods for Psychology. Volume 8 (1), pg. 52-69, 2012.

VITA

Robert Alan Morgan, P.E.

Civil and Environmental Engineering Department

Old Dominion University

Norfolk, Virginia

Education:

B.S.E.T. December 1981, Old Dominion University

M.S.C.E. June 1994, Virginia Polytechnic Institute and State University

Professional Experience:

H.W. Lochner, Inc.: Bridge analyst and roadway designer, 1982-1985

Michael Baker, Inc. (formerly Maguire Associates, Inc.): Transportation Engineer to Transportation Group Manager, 1985- 2004

Virginia Department of Transportation: Area Construction Engineer, 2004- present