Incorporating Wet Pavement Friction into Traffic Safety Analysis

November 2010

John N. Ivan
Nalini Ravishanker
Eric Jackson
Brien Aronov
Sizhen Guo

JHR 10-324 Project 07-5

This research was sponsored by the Joint Highway Research Advisory Council (JHRAC) of the University of Connecticut and the Connecticut Department of Transportation and was performed through the Connecticut Transportation Institute of the University of Connecticut.

The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the University of Connecticut or the Connecticut Department of Transportation. This report does not constitute a standard, specification, or regulation.

Technical Report Documentation Page

1 Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JHR 10-324	N/A	

4. Title and Subtitle	5. Report Date November 2010
Incorporating Wet Pavement Friction into Traffic Safety Analysis	6. Performing Organization Code N/A
7. Authors John N. Ivan Nalini Ravishanker Eric Jackson Sizhen Guo Brien Aronov	8. Performing Organization Report No. JHR 10-324
9. Performing Organization Name and Address University of Connecticut Connecticut Transportation Institute Storrs, CT 06269	10. Work Unit No (TRAIS) N/A 11. Contract or Grant No. N/A
12. Sponsoring Agency name and Address Connecticut Department of Transportation 280 West Street Rocky Hill, CT 06067-0207	13. Type of Report and Period Covered Final 14. Sponsoring Agency Code N/A

15. Supplementary Notes

This study was conducted under the Connecticut Cooperative Transportation Research Program (CCTRP, http://www.cti.uconn.edu/chwrp/index.php).

16. Abstract

This report focuses on analysis of statistical association between wet pavement friction and crash incidence, controlling for pertinent road characteristics such as shoulder width, speed limit, curves, intersections, driveways and area type. The objective of this project was to answer the following questions: (1) Is wet pavement friction significant for explaining variation in crash history among similar locations on the road network? (2) Is this association more relevant at locations with high expected braking frequency, such as sharp curves and intersections?

The analysis data were combined from "found" data locations at which ConnDOT has previously measured the wet pavement friction, and "random" data locations at which the friction was measured specifically for this project. Including the random data locations was necessary to avoid bias in the found data due to those locations having been selected due to having experienced one or more crashes or the segments having been placed on the SLOSSS. Road characteristics, including shoulder width, speed limit, grade, curvature, presence of driveways and intersections and three years of crash count data (moderate severity to fatal crashes only) were collected and incorporated into the data set. Negative binominal regression was used to estimate models with coefficients for the main factors and interactions.

The results show an association between wet pavement friction and crash frequency. The locations where improving the wet pavement friction will most reduce crashes include sections with non-isolated curves on undivided roads and sections with driveways or mild curves on divided roads. The finding of an increase in crash frequency with increases in friction in urban areas, especially on divided roads, suggests reevaluating decisions to improve friction in those locations.

17. Key Words:	18. Distribution Statement
Wet pavement friction, Accidents, Safety, Crashes, Found data, Negative binomial regression	No restrictions. This document is available to the public through the National Technical Information Service Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	52	N/A

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: volumes greater than 1000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

Table of Contents

List of Figures

List of Tables

1 Introduction

In 2006, there were in total 71,724 motor vehicle collisions in Connecticut, of which 293 were fatal and 27,366 involved injuries. For injury collisions, "following too closely", "fail to grant right of way", "driver lost control" and "speed too fast for conditions" were the four most common contributing factors, causing 30.73%, 20.71%, 13.09% and 10.66% of the crashes involving injuries. For fatal collision, “influenced by alcohol/drugs”, “driver lost control”, “speed too fast for conditions” and “fail to grant right of way” were the four most common contributing factors, causing 30.38%, 23.55%, 11.60% and 11.26% of the crashes involving fatalities. These contributing factors obviously point to some sort of aberrant driver behavior; however, insufficient pavement friction can often be a determining factor for whether or not this behavior results in a collision, especially under wet conditions. For example, when a driver is traveling on a road at a high rate of speed, pavement friction is a critical component of the braking distance required for the driver to stop suddenly in order to avoid colliding with an object on the road, or to slow down for an unexpected, sharp curve. Thus, the same driver behavior exhibited on a road with excellent pavement friction under wet conditions is probably less likely to result in a collision, leading to a fatality or serious injury. This suggests that wet pavement friction is a critical element of road safety, especially in the vicinity of sharp curves, or other locations where drivers need to frequently brake, such as intersections and/or driveways. While drivers may follow other vehicles too closely on straight-aways without intersections or driveways, there is much less of a need to brake suddenly in such road segments. There are thus two questions of interest here:

(i) Is wet pavement friction a significant factor for explaining variation in crash history among similar locations on the road network?

(ii) Is this factor more relevant at locations with high expected braking frequency, such as sharp curves and intersections/driveways?

There has been some effort into incorporating pavement friction as a predictor of crash risk and in the identification of hazardous road locations. A recent study at the University of Wisconsin–Madison (Noyce et al., 2007) investigated the relationship between skid resistance and traffic safety, but only focused on the relationship between pavement material and skid resistance and their effect on safety. The study did not consider the interaction between this relationship and other road characteristics that might affect the need for braking. Our study describes a framework to determine to what degree wet pavement friction is associated with the number of collisions, and to determine the types of locations on the road network at which the association is strongest. Many state highway agencies, including Connecticut Department of Transportation (ConnDOT), measure wet pavement friction at locations where collisions have occurred, or which have been flagged as potentially high collision locations. These measurements help to determine whether or not the wet pavement friction in these areas meets ConnDOT standards, and to correct the situation if necessary, or to confirm whether or not it may have been a factor contributing to the collisions.

A statistical analysis of the association between wet pavement friction and road safety experience, controlling for pertinent roadway characteristics, enables highway agencies to better identify road locations where the wet pavement friction should be tested and improved in order to most effectively reduce the incidence of fatal and serious injury collisions. For describing the road safety experience, we focus on serious injury and fatal collisions rather than property damage or minor injury collisions, as the reporting rate of the former is more reliable, and these collisions result in the greatest cost to society.

In order to control for their effects, we considered road characteristics that fall into two categories: those that define conditions under which wet pavement friction is more likely to be a contributing factor and others that are known to have some association with road safety. The characteristics that we included under each category are:

Test characteristics: degree of horizontal curvature, rate of change of vertical curvature, number of intersections and driveways;
Control characteristics: pavement width, area type (rural, suburban, urban), and speed limit.

Study sites are characterized by these variables and the data are analyzed statistically to determine whether or not the wet pavement friction is associated with the rate of fatal and serious injury collisions. Whether or not this association is also dependent on the combination of test characteristics expected to exacerbate the presence of low wet pavement friction with regard to safety is examined. The results tell us the extent to which there is a possibility of reducing traffic injuries and fatalities by improving wet pavement friction, and where the greatest benefit is likely to be achieved.

ConnDOT Division of Transportation Research measures wet pavement friction on roads specified by other offices within ConnDOT, Traffic Engineering in particular. Because the specified road sections are generally selected on the basis of membership on the SLOSSS (State List of Surveillance Study Sites) or because of having recently experienced a crash, this list can be considered *neither representative nor random*, thereby possibly introducing selection bias (or selection effect). This may have the effect of distorting statistical analysis and results due to the non-probability based sampling scheme, which may include samples that preferentially include or exclude certain kinds of results. Consequently, measures of statistical significance such as p-values of hypothesis tests may appear stronger than they really are, leading to incorrect inference and decision making. We refer to such sites as “found sites” and refer to data from such sites as “found data”. Arguing that restriction solely to such sites will not provide adequate inference, we augment this “found” data with information observed for “random” data locations. Statistical analysis is then carried out on data from “combined” sites.

This study uses the Generalized Linear Models (GLIMs) for modeling crash counts as a function of explanatory variables including pavement friction. GLIMs include Poisson regression, over-dispersed Poisson regression, and Negative Binominal regression, all of which are more suitable for modeling crash *counts* modeling than simple linear regression. We specifically use these models to investigate the relationship between crash counts and pavement friction, adjusting for main and interaction effects of other relevant factors, such as road characteristics (such as horizontal curves, driveways, intersections, and area type). The interactions between road characteristics and different crash types are also estimated and will enable us to study potential differences in the effect of pavement friction on crashes specifically at locations with high braking frequency.

Section 2 provides a description of pavement friction and cites previous literature on studying the effect of pavement friction on highway safety. Section 3 provides a description of the study design, i.e., a description of the "found data", the sample size determination that leads to collection of data at "random" locations in order to avoid selection bias in the data analysis, and the determination of the "combined" data. Section 4 describes the details of the statistical data analysis using log-linear model fits to the "combined" data. Section 5 provides an interpretation of results and a detailed summary.

2 Background

2.1 *Pavement Friction*

Early research on pavement surface conditions suggested skid resistance by two categories of texture: microtexture and macrotexture (Csathy *et al.* 1968, Henry 2000). Microtexture describes roughness on the surface of the individual coarse aggregates in the mix. Achieving an adequate microtexture involves selecting aggregate that starts with a rough microtexture and is able to resist the polishing effect of vehicle traffic. Macrotexture is used to describe the pavement roughness due to the particular arrangement of aggregate particles along the surface. The magnitude of this component will depend on several factors. The macrotexture of a freshly laid pavement depends on the mix of aggregate and construction method used to place the surface layer. Over time, the microtexture of a pavement can change due to a combination of traffic loading and environmental impacts (increased temperatures and freeze-thaw mechanics). The dynamic nature of the pavement, and thus skid resistance, is why periodic testing is required to ensure surface friction is adequate and safe.

Adequate friction on a paved surface is critical to the braking needs of a vehicle. Without adequate friction between the pavement and the tire vehicles are unable to stop in a reasonable manor, thus leading to crashes. The friction of a pavement is often measured in terms of a "Skid Number". Due to the large variation in pavement materials, aggregate properties, vehicle speed, tire wear and tire types used on the roadway, standard test procedures were developed. To control for tire type and tire wear, a standard test tire must be used and the wear on that tire must be within a set standard established in AASHTO T 242. Typically pavement friction is not an issue under dry conditions. However, when a pavement becomes wet is when pavement friction becomes a much larger problem. Therefore, the majority of friction testing involves determining the wet pavement friction. Friction testing equipment has been designed to spray a metered amount of water onto the roadway immediately in front of the test tire before the breaks are applied. Figure 2-1 depicts the friction vs. tire slip curve generated from this test procedure. The friction number is obtained from coefficient of friction values obtained after the tire has been fully locked. Anti-lock brakes prevent the vehicle from exceeding the peak coefficient friction portion of the curve. In future research, peak coefficient friction could be a more realistic value to use when calculating friction number. However, there are still a significant number of vehicles on the road without anti-lock brakes.

Pavement friction data are collected using a friction testing machine (Figure 2-2). The tow vehicle (loaded with water) pulls a trailer and maintains a speed of 40 mi/h. Once the operator starts a test, one of the tires (either the smooth or the ribbed) on the trailer is locked using the breaks on the trailer. The smooth tire is more sensitive to pavement macro-texture, while the ribbed tire is more sensitive to micro-texture changes in the pavement. The amount of force needed to lock the tire and start the skid is used to calculate the skid number. If the test is performed at any other speed than 40 mi/h, there is a speed correction factor that has to be applied to the resulting skid number. Therefore, skid number is reported at a standard speed of 40 mi/h (FN_{40R}). Skid number is calculated using equation (1) and corrected to FN_{40R} using equation (2).

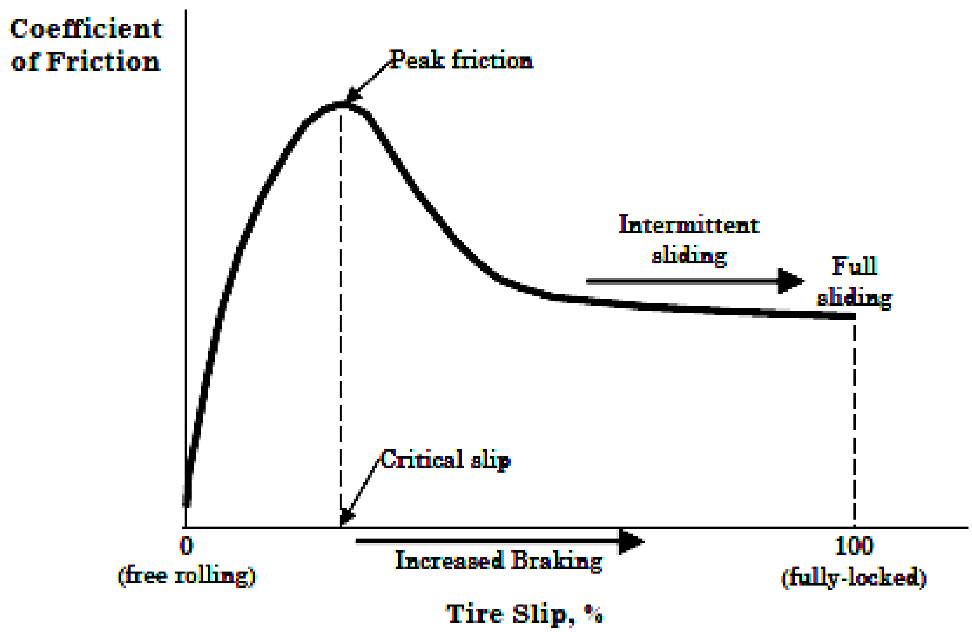

Figure 2-1. Coefficient of Friction V. Tire Slip (Henry 2000).

Figure 2-2. ConnDOT Friction Tester.

$$FN_s = 100 \times \mu = 100 \times (F/W) \qquad (2\text{-}1)$$

where FN_s : Friction number at the measured speed s
μ : Coefficient of friction during break lockup.
F: Tractive horizontal force applied to the tire, lb.
W : Vertical load applied to the tire, lb.

and

$$FN_{40R} = FN_s - 0.5(40 - s) \qquad (2\text{-}2)$$

where FN_{40R} = Ribbed tire friction number at 40 mi/h.

2.2 *Previous Research*

There has been considerable recent research conducted on pavement friction and its effect on traffic safety. It is well documented that a pavement with a high fiction number can be a significant factor in reducing the likelihood of a crash (Noyce *et al.* 2007). Wallman *et al.* (2001) reported several earlier studies on this issue in a literature review. An early study conducted by Giles *et al.* (1964) investigated the correlation between skid number and skidding crashes. They measured skid number on two grouped sites: crash sites and random sites. Crash sites were the locations where skidding crashes occur frequently in wet weather. Random sites were selected randomly. It was found that the difference between the mean skid number for random sites and the mean skid number for crash sites was 15. The study shows that a skid-related crash is more likely to occur on a pavement with a lower skid number. Schulze *et al.* (1976) conducted a regression analysis between the friction number and crashes that occur on wet pavements in Germany. They found that the percentage of wet road crashes increases with decreasing friction. To incorporate additional factors into the analysis, Griffin (1984) modeled wet weather crashes by using multiple linear regression analysis. Several variables were used as surrogates for vehicle demand for friction, which include average daily traffic (ADT), access (ACC), skid number at 40 mi/h (FN), proportion of time wet (TW), mean traffic speed (VM), standard deviation of the speed distribution (V), lanes of traffic (LN). They showed that about 58 percent of the variance of wet crash rate (WCR) on high speed roads could be explained by the following equation:

$$\begin{aligned} WCR &= -21.7 + 0.0009 \times ADT + 2.34 \times ACC - 0.4 \times SN \\ &+ 286 \times TW + 1.32 \times LN \end{aligned} \tag{2-3}$$

and that about 46 percent of the variance of (WCR) on low speed roads could be explained by the following equation:

$$\begin{aligned} WCR &= -0.75 + 0.001 \times ADT - 0.053 \times VM + 0.54 \times V \\ &+ 0.69 \times ACC - 0.025 \times SN \end{aligned} \tag{2-4}$$

Unfortunately, no additional relevant information for the equations was provided in the reference. However, their findings suggested that increasing friction helps to reduce wet crash rate.

Al-Mansour (2006) collected a massive amount of friction measurements using a Mu-meter covering more of the major highway network in the Kingdom of Saudi Arabia. A regression model was used to model the crash density based on skid resistance number skid number and the best model was found to be $AD = aSN^b$. It was concluded that a decreasing skid number leads to an increase in crash density. In another study, Caliendo *et al.* (2007) used Poisson, Negative Binominal and Multinominal regression models to analyze crash and other factors, including length, curvature, annual average daily traffic, sight distance, side friction coefficient, longitudinal slop and the presence of junctions. It is found that a wet pavement significantly increase the number of crashes.

As stated above, pavement friction is important for traffic safety, but it is not very easy to identify the effect of poor friction on crash occurrence (Wallman *et al.* 2001). Drivers adjust their behavior depending on many factors, such as sharp horizontal curves, the surrounding environment and their perception of road friction. Henry (2000) plotted the ratio of wet-dry crash versus skid number based on the fiction measurement from sites in Kentucky, but no significant trends were observed. Ohio DOT used the results of the analysis of real ODOT crash, friction, texture and roughness data to investigate the relationship between crash data and these surface-related variables (Larson *et al.* 2008). However, no strong correlation between even one of the surface-related variables and crash was discovered. They pointed out that the poor statistical relationship indicated that there are other types of factors outside of surface characteristics that might have significant influence on crashes. A recent study at the University of Wisconsin-Madison explored the relationship between asphalt mix design, skid friction and roadway safety and found that there is no relationship between crash frequency and pavement skid friction (Noyce *et al.* 2007). In addition, these studies only focused on the relationship between skid resistance and the effect on safety. There is no consideration of any interaction between this relationship and other road characteristics that might affect the need for braking.

3 Study Design

3.1 *Found Data*

Skid resistance of the pavement is the opposing force developed at the tire-pavement contact area. It plays an important role in maintaining vehicle control and reducing the stopping distance in braking emergency situations. Skid number is measured as a critical parameter to evaluate the performance of pavement. This value is generally measured by testing the force required to hold a tire at a full skid on a wet pavement surface (Cairney 1997). FN_{40R} describes the friction number measured at 40 mi/h with a ribbed tire. Representative values for friction and associated recommendations are depicted in Table 3-1.

Table 3-1. Example Skid Numbers and Prescribed Response (Jayawickrama *et al.*, 1996)

Skid Number	Comments
< 30	Take measures to correct
≥ 30	Acceptable for low volume roads
31 – 34	Monitor pavement frequently
≥ 35	Acceptable for heavily traveled roads

As noted in the introduction, ConnDOT Division of Transportation Research measures wet pavement friction on roads specified by other offices within ConnDOT, Traffic Engineering in particular. Because the specified road sections are generally selected on the basis of membership on the SLOSSS (State List of Surveillance Study Sites) or because of having recently experienced a crash, this list can be considered *neither representative nor random*, thereby possibly introducing selection bias (or selection effect). This may have the effect of distorting statistical analysis and results due to the non-probability based sampling scheme, which may include samples that preferentially include or exclude certain kinds of results. Consequently, measures of statistical significance such as *p*-values of hypothesis tests may appear stronger than they really are, leading to incorrect inference and decision making. We refer to such sites as “found sites” and refer to data from such sites as “found data”. Arguing that restriction solely to such sites will not provide adequate inference, we and augment this “found data” with information observed for “random” data locations. A similar approach was discussed in Overton *et al.* (1993). They augmented a regional probability sample for stream surveys of US EPA’s National Surface Water Survey (NSWS) with data from “found sites” contained in WATSTORE, a vast hydrological and water quality database of USGS, and pointed out that this combination of data from a non-probability based method (“found data”) with observations from a true probability based statistical sample (“random” data), enables better retention of the estimation properties inherent in a probability sample. This is discussed in Sections 3.2 through 3.4.

3.1.1 Road Characteristics

ConnDOT provides various tools for data collection, including photolog, highway geometry program, and other archived data. The Automatic Road Analyzer (ARAN) photologging van is equipped with high definition forward facing, side facing and downward facing video cameras. The cameras on the van are configured to take a picture every 10 m to generate a detailed photolog of **all** state roads on a **yearly** basis. Furthermore, the ARAN van contains gyroscopes that use pitch and yaw ("pitch" refers to the vertical angle of incline in the direction of travel; "yaw" refers to the amount of change in the horizontal direction as a vehicle travels), in order to collect detailed road geometry such as curvature and grade.

Divided and Undivided Roads:

We created separate datasets for Divided roads and Undivided roads, and carried out separate statistical analysis for each. A Divided road is divided down the middle by a barrier that separates traffic going in different directions. An Undivided road has no physical barrier in the middle. Divided and Undivided roads have different characteristics in terms of traffic flow, road facilities and crash occurrence. Divided roads generally have heavier traffic and wider shoulder width than Undivided roads. In addition, crash types also differ significantly on Undivided and Divided roads. For example, head-on crashes are common on Undivided roads, however, seldom occur on Divided roads. For these reasons, it would be inappropriate to analyze data for Undivided roads and Divided roads together using the same model.

The **horizontal curvature data** are collected using gyroscope heading to obtain the relative direction in which ARAN is pointed and a Distance Measurement Instrument (DMI) sensor collects the distance traveled. Then, using the heading (in degrees) and distance in meters the radius of curvature can be calculated.

The **vertical curve or grade data** are collected by the ARAN using the pitch and roll gyro sensors. The pitch, jointly with the DMI, measures the longitudinal gradient and the roll gyro measures the transverse slope of the road. The grade data are reported in percent (%) of rise over run. Grade and curvature data were collected every 10 m along the entire length of the test road. According to the manufacturer, the ARAN system is capable of providing grade and curvature data at "rod and level" accuracy and meets the Federal Highway Administrations (FHWA) regulations for curve classification (Roadware 2007). The photolog was used to assure the consistency of road characteristics over the years during which the study was performed. It was also utilized to determine the area type surrounding the road segments under investigation. The digital highway geometry program served as the source for radius values of road lines, based on which we determine if a curve exists.

Connecticut DOT (ConnDOT) Photolog was used to assure the consistency of all road characteristics over the years during which the study was performed. It was also utilized to observe the number of access points (intersection and driveway) and determine the area type surrounding the road segments under investigation. Photolog is a roadway viewing system which is updated annually. Each state-maintained highway in Connecticut could be viewed with Photolog, which consist of images of the roadway taken every 0.01 km. The photolog also consists of a set of corresponding highway geometric data. These data are all collected by Automatic Road Analyzer (ARAN).

The ConnDOT Automatic Road Analyzer (ARAN) photolog van is equipped with high definition forward facing, side facing (now replaced with wide angle forward only) and downward facing video cameras. The cameras on the van are configured to take a picture once every 10 m in order to generate a detailed photolog of all state roads on an annual basis. Furthermore, the ARAN van contains gyroscopes that use pitch and yaw to collect detailed road geometry such as **curvature and grade**. The horizontal curvature data are collected using gyroscope heading to obtain the relative direction in which ARAN is pointed and a Distance Measurement Instrument (DMI) sensor collects the distance traveled. Then, using the heading (in degrees) and distance in meters the radius of curvature can be calculated. The vertical curve or grade data are collected by the ARAN using the pitch and roll gyro sensors. The pitch, jointly with the DMI, measures the longitudinal gradient and the roll gyro measures the transverse slope of the road. The grade data are reported in percent of rise over run. Grade and curvature data were collected every 10 m along the entire length of the test route. According to the manufacturer, the ARAN system is capable of providing grade and curvature data at “rod and level” accuracy and meets the Federal Highway Administrations (FHWA) regulations for curve classification (Roadware 2007).

The radii of horizontal curves were collected for each segment by observing the photolog. The minimum value of radiuses was selected to present horizontal curvature characteristic for a segment. A curve in our study was defined as a location where high brake frequency is expected. Thus, if the expected operation speed through the curve is lower than the posted speed limit in the vicinity of the curve, the curve is considered a significant curve in our study. A literature review on the relationship between operation speed and radius of curve was conducted. Emmanuel *et al.* (1998) plotted the results based observation data on a highway in mountainous area, as shown in Figure 3-1.

Ritchie *et al.* (1968) recorded lateral acceleration and speed for fifty subjects during normal road driving. He found that lateral acceleration was constant below 9 m/s (32 km/h), so that speed V was related to curve radius R by

$$Latacc = V^2/R \tag{3-1}$$

At speeds above 9 m/s, drivers were found to choose a lower speed than that which would yield constant lateral acceleration, such that there was a linear relationship between lateral acceleration and speed.

Emmerson *et al.* (1969) proposed an exponential model to describe the relationship between speed of vehicles and radius of curve.

$$V = 74\left(1-e^{-0.017R}\right) \tag{3-2}$$

The maximum speed he focused is 74km/hour, *i.e.* 46.25 mi/h. Some calculations are shown in Table 3-2. Schurr *et al* (2002) summarized more equations for operating speeds on horizontal curves, which are shown in Table 3-3. According to this analysis, the definition of a curve selected for this study is shown in Table 3-4.

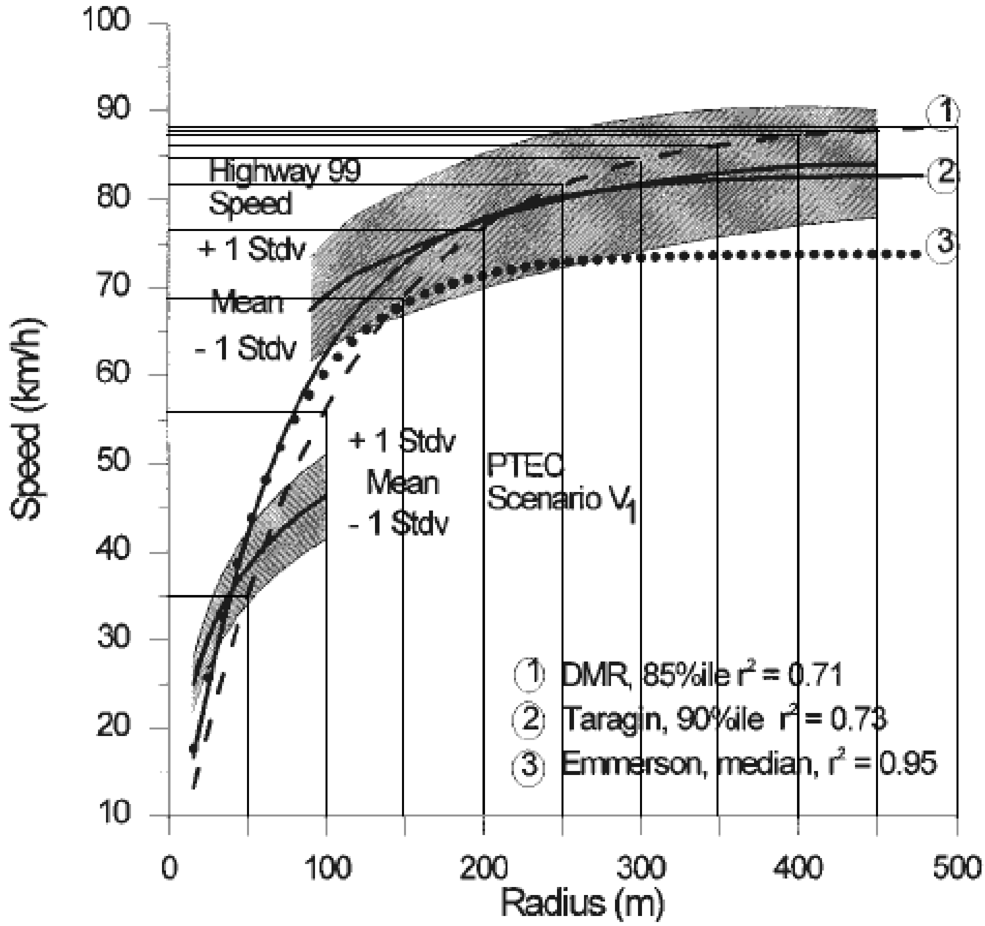

Figure 3-1. Speed vs. Radius (Emmanuel *et al.* 1998).

Table 3-2. Calculations Using Emmerson (1969) Formula.

Radius (meter)	Speed (mi/h)
50	26.48
100	37.80
150	42.64
200	44.71
250	45.59
300	45.97
350	46.13
400	46.20
450	46.23
500	46.24
550	46.25
600	46.25

Table 3-3. Equations for Operating Speeds on Horizontal Curves (Schurr *et al.* 2002).

Author	Eq. No.	Equation	R^2	Sample Size	Location	Year
Glennon et al.	5	$V_{85} = 103.96 - \frac{4524.94}{R}$	0.84	56 curves	Florida, Ohio, Illinois, Texas	1985
Lamm and Choueiri	6	$V_{85} = 94.39 - \frac{3189.94}{R}$	0.79	261 curves	New York	1986
Islam and Seneviratne	7	$V_{85} = 103.03 - \frac{4208.76}{R} - \frac{36597.92}{R^2}$	0.98	8 curves	Utah	1994
Ottesen	8	$V_{85} = 103.64 - \frac{3400.73}{R}$	0.80	138 curves	New York, Pennsylvania, Oregon, Texas, Washington	1993
Krammes et al.	9	$V_{85} = 102.44 - \frac{2741.81}{R} + 0.012L - 0.10I$	0.82			
Voight	10	$V_{85} = 99.61 - \frac{2951.37}{R} + 0.014L - 0.13I + 71.82e$	0.84			1996
Fitzpatrick et al.	11	$V_{85} = 102.10 - \frac{3077.13}{R}$ $(-9\% \leq G < -4\%)^a$	0.58	21 curves	Minnesota, New York, Pennsylvania, Oregon, Washington, Texas	1996
	12	$V_{85} = 105.98 - \frac{3709.90}{R}$ $(-4\% \leq G < 0\%)^a$	0.76	25 curves		
	13	$V_{85} = 104.82 - \frac{3574.51}{R}$ $(0\% \leq G < 4\%)^a$	0.76	25 curves		
	14	$V_{85} = 96.61 - \frac{2752.19}{R}$ $(4\% \leq G < 9\%)^a$	0.53	23 curves		
	15	$V_{85} = 105.32 - \frac{3438.19^b}{R}$	0.92	25 curves		
	16	$V_{85} = 103.24 - \frac{3576.51^c}{R}$	0.74	22 curves		
		d	n/a	13 curves		

Where:
V_{85} = 85[th] percentile speed on a curve (km/h)
L = Length (m)
I = Deflection Angle (deg)
R = Radius of curvature (m)
e = Superelevation (m/m)
G = Grade (%)
K = Rate of vertical curvature

a Horizontal curve on grade.
b Horizontal curve combined with sag vertical curve.
c Horizontal curve combined with limited sight distance crest vertical curve (i.e., $K \leq 43$ m/%). In addition, check the speeds predicted from Fitzpatrick's Equation 11 or 12 (for the downgrade) and Equation 13 and 14 (for the upgrade) and use the lowest speed. This will ensure that the speed predicted along the combined curve will not be better than if just the horizontal curve was present (i.e., that the inclusion of a limited sight distance crest vertical curve results in a higher speed).
d Horizontal curve combined with a non-limited sight distance crest vertical curve. Use the lowest speed predicted from Fitzpatrick's Equation 11 and 12 (for the downgrade) and Equations 13 and 14 (for the upgrade).

Table 3-4. Definition of Curve.

Speed limit (mi/h)	Radius (m)
25	< 50
30	< 100
35	< 100
40	< 150
45	< 250
50	< 350
55	< 500
60	< 1000
65	< 2000

Vertical curvatures (grade) were also collected from the ConnDOT ARAN photolog. The largest absolute value of grade for each given road segment was used. Data on other road characteristics, including speed limit, shoulder width, driveways, intersections and area type were collected from spreadsheets or databases available from ConnDOT. Table 3-5 lists the variables that are potential predictors in the log-linear statistical models for counts. The ConnDOT Photolog system (including the road geometry archive) served as the primary source for these additional data. We discuss these predictors in more detail in Section 3.3.

3.1.2 Crash Data

One of the response variables is *crash count***,** which denotes the number of crashes occurring on a given segment of road for three years prior to the pavement surface friction testing date. Crash counts were extracted from databases provided by the Accident Records section of ConnDOT and were merged with the database consisting of the predictor variables for the "found sites". We only used collisions where the police report lists at least one individual involved sustaining an injury code of K, A or B, for "killed", "life-threatening injury", or "visible injury". This is because the rate of non-reporting of lower severity collisions is not consistent in all jurisdictions.

ConnDOT crash data record collision type, pavement surface condition and other relevant information for each crash. The collision types in the data have 17 categories as shown in Table 3-6. Four types of crash were collected for three years prior to the friction test dates for each predetermined road segment, which are as follows:

- *Total Crashes*: The sum of all crashes
- *Wet Crashes:* The crashes that occur on wet road surface
- *Type 1 Crashes (Segment Related Crashes)*: Sideswipe-Opposite Directions, Head-on, Fixed Object, Moving Object
- *Type 2 Crashes (Intersection Related Crashes)*: Turning-Same Direction, Turning-Intersecting Paths, Sideswipe-Same Direction, Angle, Rear-end, Pedestrian.

Type 1 crashes are expected to be associated with the presence of a curve, while Type 2 crashes may be associated with the presence of an intersection or driveway.

Table 3-5. Description of the Predictor variables

Variable	Definition	Values
Volume	The average value of ADTs for three years prior to pavement surface friction testing date. The ADTs are estimates of vehicles passing through the defined section of highway on an average day in a certain year. For divided roads, 50% directional distribution is used.	> 0 continuous
Speed	Speed limits divided into several categories. Speed limit is the maximum speed legally permitted on a given stretch of road.	25-30 35-40 45-50 >55
***Curve* (Classification I)**	Presence of horizontal curve on a given segment of road.	Yes, No
***Curve* (Classification (II)**	Presence and severity of horizontal curve on a given segment of road.	None Mild Severe
***Curve* (Classification III)**	Presence of horizontal curve on a given segment of road and whether or not it is isolated.	None Isolated Non-isolated
Shoulder Width	The sum of left side and right side shoulder widths of a given road segment. The values are divided into several categories. For undivided roads it is the sum of the shoulder on both sides of the road; for divided, it is the sum of the shoulder on the side of the road and between the left edge of each travel way and the median.	< 4 ft 4-11 ft 12-19 ft > 19 ft
Mean FN_{40R}	The average skid numbers at 40 mile/hour, which were measured by ConnDOT locked-wheel skid trailer with a standard ribbed tire.	> 0 continuous
Driveway	A binary variable indicating the presence of a driveway on a given segment of road. A driveway is a private road that provides vehicular access from a property to the study road segment.	0 = no 1 = yes
Intersection	A binary variable indicating the presence of an intersection on a given segment of road. An intersection is a road junction where two or more public roads either meet or cross at grade.	0 = no 1 = yes
Area Type	The surrounding area type of a given segment of road	urban or rural
Grade	A measure of the road's incline or slope. The amount of grade indicates how much the road is inclined from the horizontal.	Continuous

Table 3-6. Taxonomy of Crash Types

ConnDOT Collision Type Code	Description of Collision Type	Assignment to Collision Type Numbers for this Project
1	Turning - Same Direction	Type 2
2	Turning - Opposite Direction	Type 2
3	Turning - Intersecting Paths	Type 2
4	Sideswipe - Same Direction	Type 2
5	Sideswipe - Opposite Directions	Type 1
6	Miscellaneous Non-Collision	
7	Overturn	
8	Angle	Type 2
9	Rear-end	Type 2
10	Head-on	Type 1
11	Backing	
12	Parking	
13	Pedestrian	Type 2
14	Jackknife	
15	Fixed Object	Type 1
16	Moving Object	Type 1
17	Unknown	

3.2 *Sample Size Determination and Selection of "Random" Locations*

The data from "found sites" were augmented with data from "random" sites with similar characteristics on explanatory variables, and inference was made on the "combined" data using log-linear models. For sample size determination in the Poisson log-linear regression model setup, we used the approach in Signorini (1991). We also extended this approach for the Negative Binomial log-linear regression. The technical details are described in Appendix I.

In our situation, let Y_i be the crash count at the *i*th "found site", with mean $\lambda_i = EY_i$. The log-linear model under a Poisson sampling distribution assumes that given the Volume and $MeanFN_{40R}$ at site *i*, Y_i follows an overdispersed Poisson (or a Negative Binomial) distribution with mean λ_i, and

$$\ln(\lambda_i) = \beta_0 + \beta_1 MeanFN40_i + \beta_2 \ln(Volume_i) \tag{3-3}$$

Since we assumed here that all segments were 0.5 mile long, we did not have to include ln*(SL)* as an offset (note that we do include ln*(SL)* as an offset in the analysis discussed in Section 4, where the road segments have different lengths). If other explanatory variables, such as road characteristics are available, they may be included in the model formulation on the right hand side of (5) as well. At the start of our analysis, we only had reliable information on $MeanFN_{40R}$ and Volume, and hence, we only used this information for the sample size determination, using the approach in Signorini (1991) (also see Shieh 2001).

Clearly, incorporation of such other explanatory variables "will not reduce the sample size" determined by using only Volume and $MeanFN_{40R}$, and therefore cannot result in bias due to undersampling at "random" sites. Information on other covariates, such as Speed, Curve, Intersection, Driveway, etc. were constructed later, and have been used in fitting the log-linear models for crash counts at the "combined" ("found" and "random") locations, as described in Section 4. Also, we only used *severe crashes* as the response variable in Equation ($\ln(\lambda_i) = \beta_0 + \beta_1 MeanFN40_i + \beta_2 \ln(Volume_i)$ (3-3). Models involving "Total Crashes" as response variable should show higher power in the Wald tests (McCullagh and Nelder 1989) discussed below.

We determined the sample size required to carry out a Wald test of the hypothesis H_0: $\beta_1=0$**,** where β_1 denotes the coefficient of $MeanFN_{40R}$ in Equation ($\ln(\lambda_i) = \beta_0 + \beta_1 MeanFN40_i + \beta_2 \ln(Volume_i)$ (3-3)**.** Under H_0, wet friction has no effect on the expected crash count. We would expect that the estimated value of β_1 is negative, indicating that as the $MeanFN_{40R}$ value increases, the expected average crash count decreases (on the log scale). A power calculation must be done at a value of β_1 which is consistent with the alternate hypothesis H_1: $\beta_1 \neq 0$. The power of the Wald test is based on the independent data, and on the "nuisance" parameters, *i.e.*, parameters other than β_1. In order to construct Table 3-7 and Table 3-8, the idea is to estimate the power via simulations as follows:

Step 1. For a specific value of the parameter β_1 which is chosen consistent with the alternative hypothesis, we fit the log-linear model (overdispersed Poisson, or Negative Binomial), constraining β_1 coefficient to be this specified value (implemented by treating $MeanFN_{40R}$ as another offset in the model).

Step 2. Assuming that the road characteristics would be (roughly) the same for the "found sites" and "random" sites, we use the joint empirical distribution of the data to randomly select the characteristics for the number of observations desired.

Step 3. We simulate the number of crashes based on the model fit in Step 1 and the data created in Step 2.

Step 4. We fit the appropriate model to the data set simulated in Step 3, but no longer "knowing/constraining" the value of the β_1 coefficient.

Step 5. At a selected level of significance, say 0.05, we check whether the model fit in Step 4 rejects the null hypothesis H_0: $\beta_1=0$; count a rejection of H_0 as a success, and a non-rejection as a failure.

Step 6. Repeat Steps 2-5 for each of the chosen N values shown in Table 3-7 and Table 3-8. The power is calculated as the number of rejections in Step 5 divided by N. Note that N denotes the number of ½ mile segments (for a total of N/2 miles).

In Table 3-7 and Table 3-8, power calculations are reported for several selected sample sizes N, under five scenarios, each consistent with the alternate hypothesis H_1. Power calculations are based on M=10,000 sets of simulated data, based on parameter estimates from data for the "found sites" in each case. The Percentages in the column headings indicate the percentage of crashes of "baseline" for a 1 standard deviation (about 8 units) increase in $MeanFN_{40R}$. As expected, the power increases as we move away further from

H_0. From Table 3-7, we see for example that, using the overdispersed Poisson model, and a sample of size N=100 of one-half mile segments (*i.e.*, a total mile-length of 50 miles), a test for detecting a 70% decrease of $MeanFN_{40R}$ of crashes will yield a power of 0.5146. To summarize, these power calculations are based on only Volume, Segment length and $MeanFN_{40R}$ (and in that sense, they are approximations to the numbers we might get when we include other road characteristics as covariates into the log-linear model).

Table 3-7. Power tables for the Over-dispersed Poisson model ("Found data")

N	90%	80%	70%	60%	50%
100	0.1416	0.2907	0.5146	0.742	0.8744
150	0.1476	0.3494	0.652	0.8706	0.9579
200	0.1662	0.4285	0.7557	0.9353	0.9851
250	0.1760	0.4908	0.8325	0.9699	0.9944
300	0.1963	0.5565	0.8775	0.9845	0.9978
350	0.2066	0.6072	0.923	0.9924	0.9992
400	0.2263	0.6537	0.9521	0.9968	0.9996
460	0.2442	0.7209	0.9678	0.9992	0.9998
500	0.2612	0.7481	0.9800	0.9994	1.0000
540	0.2702	0.7730	0.9839	0.9997	1.0000
720	0.3373	0.8810	0.9971	1.0000	1.0000

N: number of sample locations/sites of ½ mile segments selected for data collection.
Percentage: Percent of crashes of baseline for 1 standard deviation increase in $MeanFN_{40R}$ (see Appendix I).

Table 3-8. Power tables for the Negative Binomial model ("Found data")

N	90%	80%	70%	60%	50%
100	0.0942	0.1716	0.3170	0.5250	0.7363
150	0.0994	0.2141	0.4154	0.6874	0.8869
200	0.1032	0.2566	0.5248	0.8054	0.9533
250	0.1156	0.3024	0.6197	0.8807	0.9808
300	0.1216	0.3533	0.6994	0.9290	0.9936
350	0.1331	0.3901	0.7640	0.9570	0.9980
400	0.1400	0.4351	0.8110	0.9757	0.9993
460	0.1593	0.4975	0.8653	0.9904	0.9999
500	0.1675	0.5338	0.8902	0.9921	0.9999
540	0.1712	0.5560	0.9119	0.9954	1.0000
720	0.2130	0.6821	0.9688	0.9993	1.0000

N: number of sample locations/sites of ½ mile segments selected for data collection.
Percentage: Percent of crashes of baseline for 1 standard deviation increase in $MeanFN_{40R}$ (see Appendix I).

3.3 *Collection of Other Road Characteristics*

Having determined the optimal sample size of $N = 300$ (corresponding to 150 miles), we obtained other characteristics for the road locations at the "found" and "random" sites. For each location in the "found" data set, we randomly selected from the rest of the Connecticut State highway network another highway location with a similar combination of control and test characteristics based on probabilistic methods. Any variables not used for matching (e.g., Volume), but rather to be used as covariates, was gathered from the videolog or other archival databases. These measurements and the road characteristics data were collated into a single database indexed by road location and the matching factors.

Speed and shoulder width match almost perfectly between the "found" and "random" data. One reason they do not exactly match is that some roads whose pavements were replaced within 3 years prior to friction test date were deleted. The variables **Driveway** and **Intersection** also nearly match between the "found" and "random" sites. There are more **curves** in the "found" data than the "random" data, most likely because ConnDOT chose those roads with curves intentionally, since they are more crash prone.

The $\mathbf{MeanFN_{40R}}$ in the "found" data is about 10 units smaller than in the "random" data; again the reason may be that ConnDOT mainly focused on roads with suspected low pavement friction. The **wet pavement friction** for the "random" sites were collected by the Division of Research using extra runs with the measurement equipment during the summer of 2008 (it is not possible to make these measurements in freezing weather). These friction data, along with that from the measurements at the "found" sites, was augmented with road characteristics and traffic volume data to build the analysis database.

Curve Definitions:

a) In the initial stage of our project, we treated *Curve* as a binary variable indicating whether there is a presence of horizontal curve on a given segment of road. The definition of curve is based on the curve radius criterion we set, and we first incorporated the road characteristic *Curve* as a binary variable: (Yes, No).

b) Later in the project, we included an alternate and finer definition of *Curve* based on the classification: (No, Mild, Moderate and Severe). Since the data in the *Curve* category Severe was too sparse, we combined Moderate and Severe into a single category, Severe.

c) We included another alternate and finer definition of *Curve* classification: (No, Isolated, and Non-isolated). This classification is based on the following definitions: *Isolated*: (a) simple curve (only contains one curve), or, (b) there are long tangents (> 180 m) between curves; and *Non-isolated:* (a) compound curve (two simple curves joined together or with a short tangent and curving in the same direction), or, (b) reverse curve (two simple curves joined together or with a short tangent, but curving in opposite directions).

3.4 *Combined Data*

The “found” data and “random” data were merged into a single database, which we refer to as the “combined” data, in which each record contains date, location, Volume, $MeanFN_{40R}$, Speed Limit, Shoulder width, Curve, Slope, Intersection and Driveway, as well as Area type. Separate datasets for Divided roads and Undivided roads were created in the “found” data and the “random” data, and separate statistical analysis was carried out for each case.

We noticed that on some roads, pavements had been replaced less than three years before the friction test date. In that case, a portion of the three year crash data period would have been observed on a different payment, which would introduce error into the analysis. To correct this, a dataset indicating the age of the pavement was consulted, and any locations that were repaved less than three years before the friction observation were removed from the data set. After this censoring, there were 424 segments in the “combined” data for Divided roads (for a total of 99.75 miles) and 704 segments on Undivided roads (with 119.8 miles total road length). A summary of the continuous predictors at “found” sites, “random” sites and “combined” sites are shown in Table 3-9, while a similar summary for categorical predictors is shown in Table 3-10. The number of crashes by collision type and road type is given in Table 3-11.

Speed limit and Shoulder Width match *almost* perfectly between the “found” sites and “random” sites because these variables were the criteria for identifying “random” sites. One reason they do not match exactly is that some roads whose pavements were replaced within 3 years prior to friction test date were deleted after the process of selecting “random” sites (as explained earlier). The variables Driveway and Intersection also match nearly perfectly between the “found” and “random” sites. There are more curves in the “found” data than in the “random” data, most likely because ConnDOT intentionally chose testing sites with curves. The $MeanFN_{40R}$ in the “found” data is about 10 units smaller than in the “random” data; again the reason may be that ConnDOT mainly focused on roads with suspected low pavement friction.

Table 3-9. Characteristics of Continuous Predictors at “Found”, “Random” and “Combined” Sites.

	Divided			Undivided		
	Found	**Random**	**Combined**	**Found**	**Random**	**Combined**
	Volume					
Mean	24,984.24	29,889.16	27,575.52	8319.97	14,153.53	12,540.00
Std.Dev.	22,561.72	16,418.32	19,687.35	5657.28	7886.05	7784.75
N	200	224	424	195.00	510.00	705.00
Min	4604.00	4674.00	4604.00	1227.00	350.00	350.00
25th	10,622.75	14,947.41	10,900.00	5130.00	8086.12	6262.00
50th	165,05.00	33,345.66	21,658.00	6507.00	14,297.26	11,913.97
75th	29,376.00	381,14.93	36,856.85	10,439.45	19,651.85	17,858.36
Max	87,735.00	67,853.00	87,735.00	33,162.00	31,206.00	33,162.00
	FN_{40R}					
Mean	37.20	46.26	41.95	33.93	48.28	44.31
Std.Dev.	10.86	7.22	10.19	9.04	8.18	10.59
N	200	224	424	195	510	705
Min	19.60	32.93	19.60	16.60	31.80	16.60
25th	25.45	40.56	36.05	25.73	72.05	37.05
50th	36.35	45.81	42.90	34.35	47.93	44.88
75th	45.43	50.80	49.45	39.75	54.23	51.84
Max	61.40	69.85	69.85	61.25	72.05	72.05
	Grade (%)					
Mean	2.39	2.4	2.4	4.64	2.7	3.24
Std.Dev.	2.11	1.84	1.75	2.81	2.29	2.59
N	200	224	424	195	510	705
Min	0.17	0	0	0.02	0	0
25th	1.1	0.85	8.66	2.37	0.77	13.29
50th	2.11	1.87	2.02	4.31	2.05	2.63
75th	3.09	3.54	3.26	6.87	4.17	4.91
Max	8.66	7.84	8.66	13.29	11.54	13.29

Table 3-10. Characteristics of Categorical Predictors at "Found", "Random" and "Combined" Sites.

		Divided			Undivided		
Variable	Level	Found (%)	Random (%)	Combined (%)	Found (%)	Random (%)	Combined (%)
Speed	25-30	0	1.43	0.71	18.55	19.08	18.90
	35-40	1.48	6.07	3.80	32.34	34.46	33.73
	45-50	12.45	13.37	12.90	48.9	43.02	45.04
	55-65	86.07	79.13	82.59	0.21	3.44	2.33
Shoulder	<4	1.36	12.53	6.93	19.42	13.01	15.22
	4-11	25.94	21.4	23.68	65.67	55.53	59
	12-19	66.30	57.68	62	13.43	26.17	21.79
	>19	6.40	8.39	7.39	1.48	5.29	3.99
Curve Class. 1	Yes	47.03	11.00	29.12	19.00	3.90	15.8
	No	52.97	89.00	70.88	81.00	96.10	84.2
Curve Class. 2	Mild	*	*	8.25	*	*	3.83
	Severe	*	*	12.75	*	*	3.40
	No	*	*	79.00	*	*	92.77
Curve Class. 3	Isolated	*	*	17.92	*	*	5.53
	Non-isolated	*	*	3.53	*	*	1.70
	No	*	*	78.55	*	*	92.77
Driveway	Yes	0	11.40	5.68	87.80	93.40	91.48
	No	100	88.60	94.32	12.20	6.60	8.52
Intersection	Yes	0	13.98	6.97	59.65	77.46	71.33
	No	100	86.02	93.03	40.35	22.54	28.67
Area Type	Urban	11.43	39.40	25.94	9.83	42.68	31.39
	Rural	88.57	60.60	74.06	90.17	57.32	68.61

* Curve classifications 2 and 3 were not calculated until after the found and random data had been merged, so the distributions in the sub-datasets are not available.

Table 3-11. Number and Percent of Crashes by Collision Type on Divided and Undivided Road Segments in the Database.

Collision Type	Divided Roads		Undivided Roads	
	Number	Percent	Number	Percent
Type 1 (Segment Related)	564	31.35%	474	17.00%
Type 2 (Intersection Related)	1224	68.04%	2303	82.57%
Wet Pavement	496	27.57%	668	23.95%
Total Crashes*	1799	100.00%	2789	100.00%

*Note that the total crashes include some not classified as one of the three sub-types.

4 Statistical Data Analysis

4.1 *Methodology and Model Selection*

The "combined" data set (with data from "found" sites and "random" sites) was analyzed to determine whether or not wet pavement friction is statistically associated with collision incidence and also to determine whether this association is dependent upon other road characteristics. The Poisson, overdispersed Poisson, and Negative Binominal generalized linear models (GLIMs) were developed for the estimation and prediction of the counts of different types of crashes as the response variable: **Total Crashes, Wet Pavement Crashes, Type 1 Crashes, and Type 2 Crashes**. Some details about these distributions and GLIM models are given in Appendix II. We use the log link, and model $\ln(\lambda_i)$ as a function of explanatory variables. Here, $\lambda_i = E(Y_i)$, where Y_i denote the crash count (response variable). The R and SAS statistical packages were used as the platforms for the statistical analysis.

For each location/segment *i*, let SL_i denote the segment length, and $Volume_i$ denote the volume of that segment. Both $\ln(SL_i)$ and $\ln(Prop_i)$ will enter the GLIM as offsets, *i.e.*, they have regression coefficients set equal to 1. The variable *Prop* corresponds to "proportion", and was introduced for the following reason. Friction tests were conducted on "random sites" during July and August, 2008. However, the latest crash database available at that time from ConnDOT only provided crash counts until June, 2008 (first half year of 2008). Therefore, we defined *Prop* as the proportion of days for which we have valid crash counts. For example, if a friction test is conducted at location *i* on July 25th, there would be 25 days for which data on crash counts were missing; $Prop_i$ is calculated as

$$\Pr op_i = \frac{365 \times 3 - 25}{365 \times 3} = 0.977 \qquad (4\text{-}1)$$

The values of $Prop_i$ in the data for "found" sites are all equal to 1, since no crash counts were missing. For "random" sites, we use Equation (4-1) to compute $Prop_i$ and include ln ($Prop_i$) as another offset in the GLIM. This method allows us to regain missing data in a reasonable way. Because the missed data only accounted for a small portion of the data, there should not be much impact on the results from the statistical analysis.

Corresponding to each response variable, *viz*., Total Crashes, Wet Crashes, Type 1 crashes and Type 2 crashes, we model $\ln(\lambda_i)$ as

$$\ln(\lambda_i) = \ln(SL_i) + \ln(\Pr op_i) + x_i{}'\beta \qquad (4\text{-}2)$$

where the last term on the right side of Equation (4-2) is a condensed way to represent a linear function of explanatory variables and regression coefficients. The following predictors (explanatory variables) were input into the model for $\ln(\lambda_i)$: MeanFN_{40R}, ln(Volume), Shoulder Width (4 levels), Speed Limit (4 levels), Intersection (binary), Driveway (binary), Area Type (binary), Curve (binary, or new classification 1 with 3 levels, or new classification 1 with 3 levels), Slope (continuous-valued). The fitted models also include interactions of these variables. In particular, we are interested in first order and second order interactions between MeanFN_{40R} and variables such as Curve, Driveway, Intersections, Area Type and Slope. We then selected significant predictors and in particular, studied the role of MeanFN_{40R} and its interactions with other variables, on the expected mean crash.

These GLIM models were fit using R (or SAS) and variable selection (via Wald tests and deviance difference tests) was used to find the most parsimonious model(s) with significant regression coefficients. In each case, the crash count was the response variable, *SL* and *Prop* were offsets, and the road characteristics described above were the predictors.

A series of models were developed for each crash type with a combination of predictors as described above. In each case, the first model was a model that included all the main effects and the interactions between MeanFN_{40R} and the road characteristics. The second model was the first model with one interaction term removed, one that is least significant. Similarly, interactions were removed in successive stages until only significant interactions remain (many models will be generated in succession). Then, we began removing non-significant main effects which did not appear in the remaining interactions. Interactions were reviewed again to see that they were still significant. If no longer significant, the interaction was removed, and a new model was developed. The final model displayed after this successive variable selection only contained significant factors. The best model was then selected from these adequate models, based on Akaike's information criterion (AIC). Note that AIC is defined as

$$AIC = -2\ln\left[L\left(\hat{\theta}\right)\right] + 2K \tag{4-3}$$

where $L(\hat{\theta})$ is the value of the maximized likelihood for a model with K freely estimated regression coefficients. AIC may be used to compare nested and non-nested models. The model yielding the smallest value of AIC among the candidate models is selected as the "best model" and to be the closest to the "true" model. This process of selecting as set of adequate models and then selecting the best model was repeated for each crash type, using all the road characteristics.

4.2 *Results and Discussion*

Table 4-1 through Table 4-6 provide a summary of the coefficients and fit statistics for the best performing (lowest AIC) models, organized by type of road and crash type. Table 4-1 and Table 4-2 show the models for divided and undivided roads with the original (binary) curve classification, respectively, for all four crash types: total crashes, type 1 (segment-related) crashes, type 2 (intersection-related) crashes and wet pavement crashes. Table 4-3 and Table 4-4 show the models for the second curve classification scheme, "no curve, mild curve, or severe curve", for divided and undivided roads, respectively, for all four crash types, and Table 4-5 and Table 4-6 show the best models for the third curve classification scheme, "no curve, isolated curve and non-isolated curve", for divided and undivided roads, respectively, for all four crash types. These results are discussed in more detail in the following paragraphs.

Table 4-1. Models with Lowest AIC for Curve Classification 1 (Binary) on Divided Roads by Crash Type.

Parameter		Total Crashes		Crash Type 1		Crash Type 2		Wet Crashes	
		Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
(Intercept)		-7.0338	< 0.0001	-5.7444	0.0008	-12.6559	< 0.0001	-10.5951	0.0002
log(Volume)		1.0523	< 0.0001	0.8426	< 0.0001	1.6931	< 0.0001	1.3101	< 0.0001
sw>19								0.0696	0.9182
sw4-11								-1.2887	0.0361
sw12-19								-0.4232	0.4439
speed35-40						-0.9094	0.3612	2.0299	0.1357
speed45-50						-1.8358	0.0510	1.9817	0.1647
speed55-65						-2.2176	0.0240	0.1498	0.9155
intersectionyes		0.9959	0.0273						
MeanSN40		-0.0306	0.0184	-0.0249	0.0245	-0.0209	0.1772	-0.0462	0.0217
curveyes								-2.4723	0.0813
intersectionyes						1.2268	0.0379		
drivewayyes						6.4296	0.0427		
areatypeUrban		-0.9404	0.3641			-1.7110	0.1914	-3.7864	0.0101
Slope1									
MeanSN40:curveyes								0.0676	0.0391
MeanSN40:drivewayyes						-0.1621	0.0315		
MeanSN40:areatypeUrban		0.0544	0.0323			0.0814	0.0082	0.1044	0.0021
curveyes:areatypeUrban								1.9402	0.0163
	Min	-1.6844		-1.4991		-1.6155		-1.5585	
Deviance Residuals	1Q	-1.0788		-0.9743		-0.9172		-0.7584	
	Median	-0.6229		-0.6055		-0.5317		-0.5141	
	3Q	0.1250		0.0032		-0.1058		-0.1835	
	Max	2.2140		2.4862		2.9364		2.2006	
Dispersion Parameter		0.2729		0.2976		0.2482		0.2579	
AIC		1585.30		1087.90		1230.60		812.12	

Note: * = significant at 0.1 level; ** = significant at 0.05 level; df = degrees of freedom

Table 4-2. Models with Lowest AIC for Curve Classification 1 (Binary) on Undivided Roads by Crash Type.

Parameter		Total Crashes		Crash Type 1		Crash Type 2		Wet Crashes	
		Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Intercept		-5.3837	< 0.0001	0.6904	0.4614	-7.3075	< 0.0001	-5.6689	< 0.0001
Log(Volume)		0.9255	< 0.0001	0.1833	0.0525	1.1280	< 0.0001	0.7376	< 0.0001
sw4-11		-0.3223	0.0145			-0.6010	0.0156	-0.6116	0.0575
sw12-19		-0.4296	0.0056			-0.3549	0.0171	-0.3059	0.0668
sw>19		-0.6093	0.0068			-0.3487	0.0439	-0.3764	0.0601
speed35-40		-0.3470	0.0034	-0.4650	0.0055	-0.2842	0.0341	-0.0481	0.7517
speed45-50		-0.9177	< 0.0001	-1.1852	< 0.0001	-0.5784	0.0006	-0.8160	0.0001
speed55-65		-0.2485	0.3606	-1.4634	0.0033	0.1094	0.7176	-0.1937	0.5874
Mean SN40		0.0082	0.0854	-0.0130	0.0459	0.0027	0.6127	0.0206	0.1733
curveyes		1.1831	0.0666	1.1119	< 0.0001	-1.7884	0.0220	2.3900	0.0018
intersectionyes						0.1226	0.2766		
drivewayyes						-0.4209	0.0042	1.2561	0.0860
areatypeUrban		0.7923	< 0.0001			1.1631	< 0.0001	0.8938	< 0.0001
MeanSN40:curveyes		-0.0246	0.1217					-0.0433	0.0283
MeanSN40:drivewayyes								-0.0361	0.0252
curveyes:intersectionyes						1.1494	0.0423		
curveyes:drivewayyes						2.3896	0.0458		
MeanSN40:curveyes:drivewayyes						-0.0467	0.0678		
	Min	-2.3778		-1.9075		-2.1830		-2.0073	
Deviance Residuals	1Q	-1.0026		-0.7938		-0.9822		-0.8829	
	Median	-0.4769		-0.4959		-0.5285		-0.4932	
	3Q	0.2091		-0.0561		0.1019		0.1279	
	Max	4.2414		4.0278		4.6723		3.4465	
Dispersion Parameter		1.1330		1.0274		0.9504		1.1102	
AIC		3000.90		1294.40		2717.70		1634.60	

Note: * = significant at 0.1 level; ** = significant at 0.05 level; df = degrees of freedom

Table 4-3. Models with Lowest AIC for Curve Classification 2 (none, mild or severe) on Divided Roads by Crash Type.

Parameter		Total Crashes		Crash Type 1		Crash Type 2		Wet Crashes	
		Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
(Intercept)		-0.8420	0.7809			-12.5458	< 0.0001	-11.4542	< 0.0001
log(Volume)		1.0474	< 0.0001			1.5457	< 0.0001	1.4916	< 0.0001
speed35-40								1.5396	0.2449
speed45-50								1.1470	0.3816
speed55-65								0.1914	0.8848
intersectionyes		0.9913	0.0270			1.5454	0.0015		
MeanSN40		-0.1845	0.0140			-0.0239	0.1283	-0.0351	0.0441
areatypeUrban		-1.3895	0.1930			-1.1471	0.3196	-3.3120	0.0186
newcurve1no		-6.0033	0.0253			-0.5131	0.5470	-2.1236	0.0154
newcurve1severe		-9.2712	0.0062			-1.7475	0.1103	-2.6914	0.0148
Slope1		-3.0638	0.0068			-0.6995	0.0573	-1.0341	0.0091
MeanSN40:areatypeUrban		0.0644	0.0141			0.0749	0.0079	0.1096	0.0009
MeanSN40:newcurve1no		0.1524	0.0504						
MeanSN40:newcurve1severe		0.2269	0.0167						
MeanSN40:Slope1		0.0699	0.0410						
newcurve1no:Slope1		3.0844	0.0083			0.6690	0.0744	1.1015	0.0065
newcurve1severe:Slope1		3.5048	0.0053			0.9830	0.0212	1.4596	0.0011
MeanSN40:newcurve1no:Slope1		-0.0710	0.0416						
MeanSN40:newcurve1severe:Slope1		-0.0791	0.0319						
Deviance Residuals	Min	-1.6945				-1.6561		-1.4372	
	1Q	-1.0890				-0.8946		-0.7798	
	Median	-0.6228				-0.5295		-0.4898	
	3Q	0.1450				-0.0630		-0.1738	
	Max	2.3483				2.9560		2.3059	
Dispersion Parameter		0.2853				0.2538		0.2624	
AIC		1596.50				1229.60		812.18	

Note: * = significant at 0.1 level; ** = significant at 0.05 level; df = degrees of freedom; the algorithm for Crash Type 1 did not converge.

Table 4-4. Models with Lowest AIC for Curve Classification 2 (none, mild or severe) on Undivided Roads by Crash Type.

Parameter		Total Crashes		Crash Type 1		Crash Type 2		Wet Crashes	
		Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
(Intercept)		-2.4043	0.0353	4.9082	0.0020	-7.4701	< 0.0001	-2.8981	0.0676
log(Volume)		0.9325	< 0.0001	0.1588	0.1311	1.1478	< 0.0001	0.7382	< 0.0001
sw>19		-0.6438	0.0044			-0.6317	0.0109	-0.6087	0.0587
sw4-11		-0.3406	0.0099			-0.3852	0.0096	-0.2968	0.0776
sw12-19		-0.4376	0.0048			-0.3905	0.0238	-0.3662	0.0683
speed35-40		-0.3594	0.0025	-0.4994	0.0043	-0.2805	0.0365	-0.0471	0.7573
speed45-50		-0.8753	< 0.0001	-1.1340	< 0.0001	-0.5698	0.0007	-0.8041	0.0001
speed55-65		-0.1871	0.5050	-1.5020	0.0036	0.1590	0.6043	-0.2369	0.5131
MeanSN40		-0.0730	0.0045	-0.0797	0.0039	-0.0417	0.1467	-0.0383	0.2372
newcurve1no		-3.0318	0.0014	-4.1223	0.0010	-0.0022	0.9986	-2.7593	0.0093
newcurve1severe		-2.6563	0.0758	-3.4527	0.0780	-1.2826	0.4242	-0.5020	0.7302
intersectionyes						1.8948	0.0120		
drivewayyes						-0.3792	0.0085	1.2235	0.1005
areatypeUrban		-12.1113	0.0893	1.2783	0.1684	1.1510	< 0.0001	0.8997	< 0.0001
Slope1				-0.1529	0.0880				
MeanSN40:newcurve1no		0.0807	0.0022	0.0695	0.0146	0.0438	0.1334	0.0580	0.0471
MeanSN40:newcurve1severe		0.0797	0.0304	0.0715	0.0740	0.0687	0.0953	0.0220	0.5691
MeanSN40:drivewayyes								-0.0351	0.0324
MeanSN40:areatypeUrban		0.2814	0.0488						
newcurve1no:intersectionyes						-1.7763	0.0196		
newcurve1severe:intersectionyes						-2.1667	0.0526		
newcurve1no:areatypeUrban		12.8873	0.0711	-1.2642	0.1793				
newcurve1severe:areatypeUrban		13.6842	0.0686	-2.1498	0.0901				
newcurve1no:Slope1				0.1529	0.0882				
newcurve1severe:Slope1				0.2918	0.0268				
MeanSN40:newcurve1no:areatypeUrban		-0.2804	0.0502						
MeanSN40:newcurve1severe:areatypeUrban		-0.3236	0.0436						
	Min	-2.3941		-1.9004		-2.1880		-2.0100	
Deviance Residuals	1Q	-0.9995		-0.7842		-0.9843		-0.8810	
	Median	-0.4619		-0.4959		-0.5319		-0.4929	
	3Q	0.2322		-0.0557		0.1012		0.1229	
	Max	4.2570		4.1076		4.6898		3.4487	
Dispersion Parameter		1.1554		1.1402		0.9597		1.1119	
AIC		3003.80		1299.40		2720.60		1637.70	

Note: * = significant at 0.1 level; ** = significant at 0.05 level; df = degrees of freedom

Table 4-5. Models with Lowest AIC for Curve Classification 3 (none, isolated, non-isolated) on Divided Roads by Crash Type.

Parameter		Total Crashes		Crash Type 1		Crash Type 2		Wet crashes	
		Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
(Intercept)		-3.1508	0.1911	-1.3250	0.1852	-12.6999	< 0.0001	-11.0800	0.0001
log(Volume)		0.9410	< 0.0001	5.2580	< 0.0001	1.4655	< 0.0001	1.2496	< 0.0001
MeanSN40		-0.1181	0.0194	-2.3210	0.0203	-0.0274	0.4082	-0.0317	0.2276
areatypeUrban		-0.7093	0.5019	-0.8470	0.3968	1.9898	< 0.0001	-1.8419	0.2305
newcurve2no		-2.8004	0.1814	-2.1100	0.0349	0.9456	0.4533	0.3032	0.6787
newcurve2nonisolated		-2.7360	0.6112	-0.3560	0.7221	9.5029	0.0077	1.5815	0.4304
intersectionyes						1.6436	0.0008		
Slope1		-1.2828	0.0554	-2.9320	0.0034	-0.6867	0.0248	-1.2642	0.0378
sw>19								-0.0873	0.9011
sw4-11								-1.2098	0.0542
sw12-19								-0.5502	0.3509
speed35-40								2.0295	0.1418
speed45-50								2.1516	0.1331
speed55-65								0.5157	0.7141
MeanSN40:areatypeUrban		0.0576	0.0250	2.3190	0.0204			0.1264	0.0002
areatypeUrban:newcurve2no				-1.9350	0.0530			-2.7144	0.0015
areatypeUrban:newcurve2nonisolated				0.7410	0.4587			0.4824	0.8349
MeanSN40:newcurve2no		0.0897	0.1000	2.1850	0.0289	-0.0095	0.7584		
MeanSN40:newcurve2nonisolated		0.1485	0.3357	0.7670	0.4429	-0.2287	0.0200		
MeanSN40:Slope1		0.0359	0.0304	2.9920	0.0028	0.0152	0.0291	0.0320	0.0113
newcurve2no:Slope1		1.2214	0.0959	3.4940	0.0005			1.5575	0.0087
newcurve2nonisolated:Slope1		2.9443	0.0686	2.2710	0.0231			2.3683	0.0579
MeanSN40:newcurve2no:Slope1		-0.0350	0.0509	-3.6590	0.0003			-0.0391	0.0008
MeanSN40:newcurve2nonisolated:Slope1		-0.1068	0.0473	-2.1700	0.0300			-0.0797	0.0982
	Min	-1.7071		-1.5153		-1.6861		-1.3859	
Deviance Residuals	1Q	-1.1057		-0.9358		-0.8805		-0.7524	
	Median	-0.6275		-0.6122		-0.5450		-0.5024	
	3Q	0.1825		0.0452		-0.0775		-0.1694	
	Max	2.7405		2.6892		2.8920		2.1446	
Dispersion Parameter		0.2783		0.3598		0.2475		0.2854	
AIC		1599.60		1088.90		1230.00		817.00	

Note: * = significant at 0.1 level; ** = significant at 0.05 level; df = degrees of freedom

Table 4-6. Models with Lowest AIC for Curve Classification 3 (none, isolated, non-isolated) on Undivided Roads by Crash Type

Parameter		Total Crashes		Crash Type 1		Crash Type 2		Wet Crashes	
		Coefficient	P-Value	Coefficient	P-Value	Coefficient	P-value	Coefficient	P-Value
(Intercept)		-5.2070	< 0.0001	1.8447	0.0401	-8.3240	< 0.0001	-2.9436	0.0339
log(Volume)		0.9291	< 0.0001	0.1824	0.0535	1.1490	< 0.0001	0.7428	< 0.0001
sw>19		-0.6225	0.0057			-0.6374	0.0106	-0.6023	0.0614
sw4-11		-0.3279	0.0125			-0.3747	0.0117	-0.2964	0.0757
sw12-19		-0.4467	0.0038			-0.3775	0.0288	-0.3762	0.0597
speed35-40		-0.3737	0.0015	-0.4736	0.0049	-0.3116	0.0200	-0.0574	0.7059
speed45-50		-0.9676	< 0.0001	-1.1916	< 0.0001	-0.6119	0.0003	-0.8273	0.0001
speed55-65		-0.2920	0.2714	-1.4502	0.0037	0.0340	0.9089	-0.1840	0.6056
MeanSN40		0.0051	0.2592	-0.0128	0.0513	0.0007	0.8941	-0.0348	0.1678
newcurve2no				-1.1527	< 0.0001	0.9490	0.0354	-2.8744	0.0008
newcurve2nonisolated				-0.1665	0.6945	-28.3000	1.0000	-2.8384	0.1529
intersectionyes						1.0390	0.0507		
drivewayyes						-0.3937	0.0063	1.3780	0.0612
areatypeUrban		0.7526	< 0.0001			1.1360	< 0.0001	0.8894	< 0.0001
newcurve2no:intersectionyes						-0.9210	0.0898		
newcurve2nonisolated:intersectionyes						27.8500	1.0000		
MeanSN40:newcurve2no								0.0576	0.0144
MeanSN40:newcurve2nonisolated								0.0700	0.1440
MeanSN40:drivewayyes								-0.0386	0.0169
	Min	-2.3805		-1.9248		-2.1880		-2.0162	
Deviance Residuals	1Q	-2.3805		-0.7952		-0.9861		-0.8788	
	Median	-0.4702		-0.4953		-0.5338		-0.4902	
	3Q	0.1968		-0.0818		0.1331		0.1344	
	Max	4.2562		4.0307		4.7052		3.4521	
Dispersion Parameter		1.1235		1.0275		0.9488		1.1247	
AIC		3001.10		1296.20		2720.50		1636.70	

Note: * = significant at 0.1 level; ** = significant at 0.05 level; df = degrees of freedom

Because many of these models include interactions between the coefficient on friction and the various other explanatory variables, the actual association between friction and crashes is not readily apparent by reviewing these model results, especially when the coefficients on the main and interaction effects have different signs. To help with this interpretation, Table 4-7 through Table 4-9 show the total coefficient on the wet pavement friction measure (FN_{40R}) for both divided and undivided roads with the original, second and third curve classifications, respectively, for each crash type and for different road scenarios defined by the area type (urban or rural), presence of curves, and presence of driveways. Then, Table 4-10 through Table 4-12 reports the exponent of these coefficients; these numbers are more useful as they indicate the multiplicative factor by which the expected crash count is predicted to change for a unit increase in the wet pavement friction (FN_{40R}) measurement for each road scenario and crash type. Multiplicative factors greater than 1.00 indicate an increase in crashes, and factors less than 1.00 indicate a decrease in crashes; obviously the desired result is a factor less than 1.00. For example, looking at Table 4-10, for total crashes on a section of a divided road in an urban area with a curve and a driveway, the multiplicative factor is 1.0241, meaning that if the wet pavement friction on such a road section were to increase by 1 unit, the total crash count on that section is expected to also increase by 2.41 percent. In contrast, the factor for the same type of road section in a rural area is 0.9698, indicating a decrease in the crash count by 3.02 percent. Note that many of the factors repeat from one row to the next; this is because many of the interactions between the corresponding road characteristic and the friction coefficient were not significant. The rest of this section summarizes the notable findings with respect to the project objectives.

Table 4-10 gives the multiplicative factors for the original (binary) curve classification for divided and undivided road sections for all four crash types. Many of the interactions between the friction coefficient and the other road characteristics were not significant, so that there is not much variation in the table for some crash types. For example, the coefficient on friction varies on divided roads only for urban and rural roads; the results show that the total crashes are expected to increase by 2.41 percent in urban areas if the friction improves, but decrease by 3.02 percent in rural areas. On undivided roads, the only factor that changes the friction effect is the presence of a curve: on sections where a curve is present, the crash count decreases by 1.62 percent for each unit increase in the wet pavement friction; where there is no curve it increases by 0.83 percent. Regarding the situations where the total crash count is expected to increase (*e.g.*, divided roads in urban areas and sections without curves in rural areas), it is important to remember that higher road friction facilitates not only braking, but also accelerating. In these situations, having higher pavement friction may facilitate drivers choosing to drive too fast for conditions. For type 1 (segment) crashes, greater pavement friction reduces crashes on divided roads by 2.46 percent and on undivided roads by 1.30 percent under all conditions. Higher friction reduces type 2 crashes in the vicinity of driveways under all conditions on divided roads – as much as 16.72 percent in rural areas, but on undivided roads, only on curves, and only as much as 4.31 percent. Wet pavement crashes are increased by improved friction under all conditions on divided roads with the exception of rural road sections without curves or driveways; on undivided roads they decrease everywhere except locations without curves or driveways.

Table 4-7. Total Coefficient on Friction (FN_{40R}) by Road Condition Scenario: Divided and Undivided Roads with Curve Classification 1 (Binary).

Setting	Curve	Driveway	Total Crashes	Type 1 Crashes	Type 2 Crashes	Wet Crashes
Divided Roads						
Urban	Yes	Yes	0.0238	(0.0249)	(0.1016)	0.1259
		No	0.0238	(0.0249)	0.0605	0.1259
	No	Yes	0.0238	(0.0249)	(0.1016)	0.0582
		No	0.0238	(0.0249)	0.0605	0.0582
Rural	Yes	Yes	(0.0306)	(0.0249)	(0.1830)	0.0214
		No	(0.0306)	(0.0249)	(0.0209)	0.0214
	No	Yes	(0.0306)	(0.0249)	(0.1830)	(0.0462)
		No	(0.0306)	(0.0249)	(0.0209)	(0.0462)
Undivided Roads						
Urban or Rural	Yes	Yes	(0.0163)	(0.0130)	(0.0440)	(0.0588)
		No	(0.0163)	(0.0130)	0.0027	(0.0227)
	No	Yes	0.0082	(0.0130)	0.0027	(0.0155)
		No	0.0082	(0.0130)	0.0027	0.0206

Note: Negative coefficients are shown in ().

Table 4-11 shows the multiplicative factors for divided and undivided roads with Curve Classification 2: no curve, mild curve and severe curve. Note that the algorithm did not converge for estimating the model for Type 1 crashes. As for the original curve classification (binary), friction reduces crashes more in rural areas than in urban areas, especially in urban areas. On divided roads, type 2 and wet pavement crashes are reduced only in rural areas, by 2.46 and 3.45 percent, respectively, while they are increased in urban areas, by 5.24 and 7.73 percent, respectively. Total crashes on divided roads vary most by the degree of the curve, justifying this further categorization. Specifically, crashes reduce with friction the most on mild curves, by 11.32 and 16.85 percent in urban and rural areas, respectively. Crashes increase with friction in urban for sections with no curves or with severe curves, by 3.28 and 11.26 percent, respectively. In rural areas sections without curves show a decrease in crashes with friction of 3.16 percent and sections with severe curves show an increase of 4.33 percent. This is a somewhat counter-intuitive result; perhaps the increased friction is sufficient to assist with braking to avoid crashes on mild curves but on severe curves the increased friction in the vicinity encourages drivers to travel faster than it is safe to do so.

Table 4-8. Total Coefficient on Friction (FN_{40R}) by Road Condition Scenario: Divided and Undivided Roads with Curve Classification 2 (Mild/Severe).

Setting	Curve	Driveway	Total Crashes	Type 1 Crashes	Type 2 Crashes	Wet Crashes
Divided Roads						
Urban	No	Yes	0.0323	*	0.0511	0.0745
		No	0.0323	*	0.0511	0.0745
	Mild	Yes	(0.1201)	*	0.0511	0.0745
		No	(0.1201)	*	0.0511	0.0745
	Severe	Yes	0.1067	*	0.0511	0.0745
		No	0.1067	*	0.0511	0.0745
Rural	No	Yes	(0.0321)	*	(0.0239)	(0.0351)
		No	(0.0321)	*	(0.0239)	(0.0351)
	Mild	Yes	(0.1845)	*	(0.0239)	(0.0351)
		No	(0.1845)	*	(0.0239)	(0.0351)
	Severe	Yes	0.0424	*	(0.0239)	(0.0351)
		No	0.0424	*	(0.0239)	(0.0351)
Undivided Roads						
Urban	No	Yes	0.0087	(0.0102)	0.0021	(0.0154)
		No	0.0087	(0.0102)	0.0021	0.0197
	Mild	Yes	0.2084	(0.0797)	(0.0417)	(0.0734)
		No	0.2084	(0.0797)	(0.0417)	(0.0383)
	Severe	Yes	(0.0355)	(0.0083)	0.0270	(0.0514)
		No	(0.0355)	(0.0083)	0.0270	(0.0163)
Rural	No	Yes	0.0077	(0.0102)	0.0021	(0.0154)
		No	0.0077	(0.0102)	0.0021	0.0197
	Mild	Yes	(0.0730)	(0.0797)	(0.0417)	(0.0734)
		No	(0.0730)	(0.0797)	(0.0417)	(0.0383)
	Severe	Yes	0.0067	(0.0083)	0.0270	(0.0514)
		No	0.0067	(0.0083)	0.0270	(0.0163)

Note: Negative coefficients are shown in ().

*The algorithm did not converge for Crash Type 1 on divided roads.

Table 4-9. Total Coefficient on Friction (FN_{40R}) by Road Condition Scenario: Divided and Undivided Roads with Curve Classification 3 (Isolated/Non-Isolated).

Setting	Curve	Driveway	Total Crashes	Type 1 Crashes	Type 2 Crashes	Wet Crashes
Divided Roads						
Urban	No	Yes	0.0292	*	(0.0369)	0.0947
		No	0.0292	*	(0.0369)	0.0947
	Isolated	Yes	(0.0604)	*	(0.0274)	0.0947
		No	(0.0604)	*	(0.0274)	0.0947
	Non-Isolated	Yes	0.0880	*	(0.2561)	0.0947
		No	0.0880	*	(0.2561)	0.0947
Rural	No	Yes	(0.0284)	*	(0.0369)	(0.0317)
		No	(0.0284)	*	(0.0369)	(0.0317)
	Isolated	Yes	(0.1181)	*	(0.0274)	(0.0317)
		No	(0.1181)	*	(0.0274)	(0.0317)
	Non-Isolated	Yes	0.0304	*	(0.2561)	(0.0317)
		No	0.0304	*	(0.2561)	(0.0317)
Undivided Roads						
Urban or Rural	No	Yes	0.0051	(0.0128)	0.0007	(0.0159)
		No	0.0051	(0.0128)	0.0007	0.0227
	Isolated	Yes	0.0051	(0.0128)	0.0007	(0.0735)
		No	0.0051	(0.0128)	0.0007	(0.0348)
	Non-Isolated	Yes	0.0051	(0.0128)	0.0007	(0.0035)
		No	0.0051	(0.0128)	0.0007	0.0351

Note: Negative coefficients are shown in ().
*The algorithm did not converge for Crash Type 1 on divided roads.

Undivided roads have an entirely different pattern of results. For total crashes, only the curve severity and the setting are significant for changing the effect of friction: on sections in urban areas with mild curves the crashes increase with increased friction by 23.17 percent, while they are comparatively constant with friction for sections with no curves and actually decrease by 3.49 percent on sections with severe curves. In contrast, in rural areas total crashes decrease with friction by 7.04 percent on sections with mild curves, but are relatively constant on other sections. For type 1 and type 2 crashes, only the curve severity is significant, and crashes reduce with friction the most for mild curves, by 7.64 and 4.08, respectively. Type 2 crashes actually increase with friction on sections with no curves or with severe curves, by 0.21 and 2.74 percent, respectively. For wet pavement crashes, the crash multiplication factor varies by both curve severity and presence of driveway; crashes increase with friction by 1.99 percent on section with neither curves nor driveways and decrease in all other conditions, with the greatest decrease for sections with mild curves and driveways, where the crashes decrease with friction by 7.08 percent.

Table 4-10. Ratio of Crashes to Friction (FN_{40R}) by Road Condition Scenario: Divided and Undivided Roads with Curve Classification 1 (Binary).

Setting	Curve	Driveway	Total Crashes	Type 1 Crashes	Type 2 Crashes	Wet Crashes
Divided Roads						
Urban	Yes	Yes	1.0241	0.9754	0.9034	1.1341
		No	1.0241	0.9754	1.0623	1.1341
	No	Yes	1.0241	0.9754	0.9034	1.0599
		No	1.0241	0.9754	1.0623	1.0599
Rural	Yes	Yes	0.9698	0.9754	0.8328	1.0217
		No	0.9698	0.9754	0.9793	1.0217
	No	Yes	0.9698	0.9754	0.8328	0.9548
		No	0.9698	0.9754	0.9793	0.9548
Undivided Roads						
Urban or Rural	Yes	Yes	0.9838	0.9870	0.9569	0.9429
		No	0.9838	0.9870	1.0027	0.9775
	No	Yes	1.0083	0.9870	1.0027	0.9846
		No	1.0083	0.9870	1.0027	1.0208

Table 4-12 shows the multiplicative factors for divided and undivided roads with curve classification 3 for all crash types. Again, the algorithm did not converge in estimating the model for type 1 crashes on divided roads. The factors for total crashes show the most variation, with both the setting and the type of curve being significant, with the crashes reducing with friction more in rural than urban areas, and on section with isolated curves rather than those with no curves or with non-isolated curves. The greatest reduction in crashes per increase in friction is 11.14 percent, for sections with isolated curves in rural areas, and the greatest increase is 9.20 percent, for sections with non-isolated curves in urban areas. The factors vary only with curve type for type 2 crashes: 3.62, 2.70 and 22.60 percent for sections with no curve, isolated curves and non-isolated curves, respectively. For wet pavement crashes the factor only varies with urban and rural setting, with crashes increasing with friction by 9.93 percent in urban settings and reducing by 3.12 percent in rural areas. There is very little variation on undivided roads for total, type 1 and type 2 crashes change with friction the same for all conditions, increasing by 0.52, decreasing by 1.27 and increasing by 0.07 percent, respectively. For wet crashes on divided roads, the factors vary with both type of curve and the presence of driveway; they decrease in the presence of driveways and isolated curves. The greatest decrease in crashes with friction is 7.08 percent for sections with isolated curves and driveways, and the greatest increase is 3.58 percent for section with non-isolated curves and no driveways.

These results are interpreted in more detail in the next section.

Table 4-11. Ratio of Crashes to Friction (FN_{40R}) by Road Condition Scenario: Divided and Undivided Roads with Curve Classification 2 (Mild/Severe).

Setting	Curve	Driveway	Total Crashes	Type 1 Crashes	Type 2 Crashes	Wet Crashes
Divided Roads						
Urban	No	Yes	1.0328	*	1.0524	1.0773
		No	1.0328	*	1.0524	1.0773
	Mild	Yes	0.8868	*	1.0524	1.0773
		No	0.8868	*	1.0524	1.0773
	Severe	Yes	1.1126	*	1.0524	1.0773
		No	1.1126	*	1.0524	1.0773
Rural	No	Yes	0.9684	*	0.9764	0.9655
		No	0.9684	*	0.9764	0.9655
	Mild	Yes	0.8315	*	0.9764	0.9655
		No	0.8315	*	0.9764	0.9655
	Severe	Yes	1.0433	*	0.9764	0.9655
		No	1.0433	*	0.9764	0.9655
Undivided Roads						
Urban	No	Yes	1.0088	0.9899	1.0021	0.9847
		No	1.0088	0.9899	1.0021	1.0199
	Mild	Yes	1.2317	0.9234	0.9592	0.9292
		No	1.2317	0.9234	0.9592	0.9624
	Severe	Yes	0.9651	0.9918	1.0274	0.9499
		No	0.9651	0.9918	1.0274	0.9838
Rural	No	Yes	1.0077	0.9899	1.0021	0.9847
		No	1.0077	0.9899	1.0021	1.0199
	Mild	Yes	0.9296	0.9234	0.9592	0.9292
		No	0.9296	0.9234	0.9592	0.9624
	Severe	Yes	1.0067	0.9918	1.0274	0.9499
		No	1.0067	0.9918	1.0274	0.9838

*The algorithm did not converge for Crash Type 1 on divided roads.

Table 4-12. Ratio of Crashes to Friction (FN_{40R}) by Road Condition Scenario: Divided and Undivided Roads with Curve Classification 3 (Isolated/Non-Isolated).

Setting	Curve	Driveway	Total Crashes	Type 1 Crashes	Type 2 Crashes	Wet Crashes
Divided Roads						
Urban	No	Yes	1.0297	*	0.9638	1.0993
		No	1.0297	*	0.9638	1.0993
	Isolated	Yes	0.9414	*	0.9730	1.0993
		No	0.9414	*	0.9730	1.0993
	Non-Isolated	Yes	1.0920	*	0.7740	1.0993
		No	1.0920	*	0.7740	1.0993
Rural	No	Yes	0.9720	*	0.9638	0.9688
		No	0.9720	*	0.9638	0.9688
	Isolated	Yes	0.8886	*	0.9730	0.9688
		No	0.8886	*	0.9730	0.9688
	Non-Isolated	Yes	1.0309	*	0.7740	0.9688
		No	1.0309	*	0.7740	0.9688
Undivided Roads						
Urban or Rural	No	Yes	1.0052	0.9873	1.0007	0.9842
		No	1.0052	0.9873	1.0007	1.0230
	Isolated	Yes	1.0052	0.9873	1.0007	0.9292
		No	1.0052	0.9873	1.0007	0.9658
	Non-Isolated	Yes	1.0052	0.9873	1.0007	0.9965
		No	1.0052	0.9873	1.0007	1.0358

*The algorithm did not converge for Crash Type 1 on divided roads.

5 Summary and Conclusions

Pavement friction is an important element for transportation safety. This study focuses on analysis of statistical association between friction and different types of crashes, controlling for pertinent road characteristics, such as shoulder width, speed limit, curves, intersections, driveways and area type. In particular, the objective of this project was to answer the following questions:

- Is wet pavement friction a significant factor for explaining variation in crash history among similar locations on the road network?
- Is this factor more relevant at locations with high expected braking frequency, such as sharp curves and intersections?

The data set for analysis was combined from “found” data locations at which ConnDOT has previously measured the wet pavement friction, and “random” data locations at which the friction was measured specifically for this project. Including the random data locations was necessary to avoid bias in the found data due to those locations having been selected due to having experienced one or more crashes or the segments having been placed on the SLOSSS. Road characteristics, including shoulder width, speed limit, grade, curvature, presence of driveways and intersections and three years of crash count data (moderate severity to fatal crashes only) were collected and incorporated into the data set. Negative binominal regression was used to estimate models with coefficients for the main factors and interactions fitting the data. Table 5-1 summarizes the results qualitatively, indicating the road scenarios and crash types with the greatest expected relative reduction and increase. These results help to indicate where the greatest benefits can be achieved by correcting areas with poor road friction.

Following is a summary of potential crash reductions focused on the presence of driveways and curves, the two factors that are expected to exacerbate the safety of low friction conditions due to increased need for braking. Note that the percentage decreases and increases apply only at the predicted values and would not apply indefinitely. They are provided as an example to show where the greatest benefits can be achieved.

- **Driveways:** on **undivided** road sections where there is also a **curve**, the expected **wet pavement crash** count will decrease by 5.71 percent for each unit increase in the friction measurement. On all **divided** road sections, the expected **intersection-related crash** count will reduce by 16.72 percent for each unit increase in the friction measurement.

- **Isolated curves:** on **rural and urban divided** road sections, the expected **total crash** count will decrease by 11.14 and 6.86 percent, respectively, for each unit increase in the friction measurement.

- **Non-isolated curves:** on **undivided** road sections, the expected **intersection-related crash** count will decrease by 22.60 percent for each unit increase in the friction measurement. This is the largest expected decrease for any road scenario.

- **Mild curves on divided** road sections**:** the expected **total crash** count on **rural** and **urban** road sections will decrease by 16.85 and 11.32 percent, respectively, for each unit increase in the friction measurement.

Table 5-1. Expected Crash Reduction and Increase by Crash Type and Road Scenario.

Change in Crashes with Unit Increase in Friction	Crash Type and Road Scenario
Reduction > 10%	• Intersection crashes on rural divided roads with driveways • Total crashes on divided roads with mild curves • Total crashes on rural divided roads with isolated curves
Reduction between 5 and 10%	• Intersection crashes on urban divided roads with driveways • Wet pavement crashes on undivided roads with curves and driveways • Segment and wet pavement crashes on urban undivided roads with driveways or mild curves • Total, segment and wet pavement crashes on rural undivided roads with driveways or mild curves • Total crashes on urban divided roads with isolated curves
Increase between 5 and 10%	• Intersection crashes on urban divided roads without driveways • Wet pavement crashes on urban divided roads without curves • Intersection and wet pavement crashes on urban divided roads
Increase > 10%	• Wet pavement crashes on urban divided roads with curves • Total crashes on urban divided roads with severe curves • Total crashes on urban undivided roads with mild curves

- **Mild curves on undivided** road sections: the expected **segment-related crash** and **wet pavement crash** counts will decrease by 7.66 and 7.08 percent, respectively, for each unit increase in the friction measurement. On **rural** road sections only, the expected **total crash** count will decrease by 7.04 percent for each unit increase in the friction measurement.

Given that operating the friction measuring equipment consumes substantial resources, this suggests that ConnDOT can get the most safety benefit for the investment into the use of the equipment by using it at these types of locations. The greatest percentage reductions expected are at the types of locations listed in Table 5-2 (in rank order).

Table 5-2. Scenario and Crash Type for Top Five Expected Crash Reductions.

Rank	Road Scenario	Crash Type	Expected Reduction
1	Non-isolated curves on undivided roads	Intersection-related	22.60 %
2	Mild curves on rural divided roads	Total crashes	16.85 %
3	Driveways on divided roads	Intersection-related	16.72 %
4	Mild curves on urban divided roads	Total crashes	11.32 %
5	Isolated curves on rural divided roads	Total crashes	11.14 %

These findings confirm the hypotheses declared in the objectives: the wet pavement friction is most associated with crash frequency in the presence of road factors that tend to increase the need for unexpected braking: curves and driveways. What is interesting is that while we had expected that more severe and/or isolated curves might increase the association between wet pavement friction and crash frequency, in fact, it is on mild and non-isolated curves where the greatest reductions are observed. This may actually be more intuitive than expected. Note that on non-isolated curves it is the intersection-related crash count that is associated. On roads with many curves, there would be less sight distance and thus drivers may be less prepared to brake when a vehicle exits a driveway or minor intersection, or slows or stops to turn off the road into a driveway or intersection. Regarding mild versus severe curves, it may be that drivers pay more attention to reduce their speeds ahead of a severe curve while not adjusting their speeds sufficiently for mild curves. Then when they find a need to slow down, if the friction in the vicinity of the mild curve is insufficient, they may be unable to slow down enough to avoid a collision. Consequently, it is recommended that friction measurement activities be focused on road sections matching the scenarios listed in Table 5-2 in order to identify the locations with the greatest potential safety improvements.

Conversely, the locations with the greatest increases in crash counts related to friction are all in urban areas, suggesting that the wet pavement friction plays less of a role in improving safety in urban areas. Instead, increased friction in urban areas may actually induce drivers to choose speeds that are higher than what is safe for the land development conditions. Following are the road scenarios where increased road friction is associated with higher crash frequency:

- **Urban undivided roads:** on sections with **mild curves**, the expected **total crash** count will increase by 23.17 percent for each unit increase in the friction measurement.
- **Urban divided roads:**
    - The expected **intersection-related crash** count on sections **without driveways** will increase by 6.23 percent for each unit increase in the friction measurement.
    - The expected **total crash** count on section with **severe curves** will increase by 11.26 percent for each unit increase in the friction measurement.
    - The expected **wet pavement crash** count on sections **with and without curves** will increase by 13.41 and 5.99 percent, respectively, for each unit increase in the friction measurement.

While some of these increases are quite large (one over twenty percent), it is not suggested that actions be taken to reduce the wet pavement friction under these conditions. Rather, these findings suggest that these are locations where there are inconclusive safety benefits associated with improving the wet pavement friction, so it is recommended that the friction measuring equipment **not** be used at such locations, and choosing any road improvement that would improve the pavement friction should be traded off against other potential benefits and costs.

In conclusion, the results reported here clearly show an association between wet pavement friction and crash frequency. Due to the substantial costs of operating the friction measuring equipment, there were a limited number of observation sites available to perform the analysis. Obviously, more could be learned by including additional study locations, but these findings are generally consistent from one model to another, and show that friction is indeed related to safety, in some road locations more than others. The locations where improving the wet pavement friction will most reduce crashes include sections with non-isolated curves on undivided roads and sections with driveways or mild curves on divided roads. The really interesting finding of an increase in crash frequency with increases in friction in urban areas, especially on divided roads, suggests reevaluating decisions to improve friction in those locations.

6 References

Al-Mansour, AI. 2006. Effect of Pavement Skid Resistance on Traffic Accidents. *The Journal of Engineering Research,* 3(1):75-78.

Cairney, P. 2007. Skid Resistance and Crashes – A Review of the Literature. *Research Report No. 311*, ARRB Transport Research Ltd., Vermont South Victoria, Australia.

Caliendo, C; G. Maurizion, G; A. Parisi, A. 2007. A Crash- Prediction Model for Multilane Roads, *Accident Analysis and Prevention,* 39:657-659.

Csathy, TI; Burnett, WC; Armstrong, MD. 1968. State-of-the-Art of Skid Resistance Research. *Highway Research Board Special Report 95*, Highway Research Board, National Research Council, Washington, D.C.

Emmanuel, F; Navin, F. 1998. Automobiles on Horizontal Curves, *Transportation Research Record* 1628:50-56.

Emmerson, J. 1969. Speeds of Cars on Sharp Horizontal Curves. *Traffic Engineering and Control*, 11(3):135-137.

Giles, CG; Sabey, BE; Cardew, KHF. 1964. Development and Performance of the Portable Skid-Resistance Tester. Road Research Technical Paper No. 66, Road Research Laboratory, London.

Griffin, L. 1984. Accident Data Relationships. The Influence of Roadway Surface Discontinuities on Safety, A State-of-the-art Report. Transportation Research Board, Washington, D.C.

Henry, JJ. 2000. Evaluation of Pavement Friction Characteristics: A Synthesis of Highway Practice. NCHRP 291.Transportation Research Board, National Research Council, Washington, D.C.

Jayawickrama, PW; Prasanna, R; Senadheera, SP. 1996. Survey of State Practices to Control Skid Resistance on Hot-Mix Asphalt Concrete Pavements. *Transportation Research Record* 1536: 52–58.

Larson, RM; Todd, EH; Smith, KD; and Wolters, AS. 2008. Relationship Between Skid Resistance Numbers Measured With Ribbed and Smooth Tire and Wet Accident Locations, Project FHWA/OH-2008/11, Ohio Department of Transportation, Ohio.

McCullagh, P; Nelder, J.A. 1989. *Generalized Linear Models*. Chapman & Hall: London.

McOverton, JC; T. C. Young, TC; and W. S. Overton. 1993. Using 'found' data to augment a probability sample: Procedure and case study, Environmental Monitoring and Assessment, Volume 26, Number 1, pp. 65-83.

Noyce, DA; Bahia, HU; Yambo, J; Chapman, J; Bill, A. 2007. Incorporating Road Safety into Pavement Management: Maximizing Surface Friction for Road Safety Improvement, MRUTC 04-04, Midwest Regional University Transportation Center, University of Wisconsin- Madison.

Ritchie, ML; McCoy, WK; Welde, WL. 1968. A Study of the Relation between Forward Velocity and Lateral Acceleration in Curves during Normal Driving. *Human Factors*, 10(3):255-258.

Roadware. 2007. *Roadware ARAN brochure*, http://www.roadware.com/_lib/pdf/datasheet.icon_sheet.pdf. accessed: June 5.

Schulze, KH; Gerbaldi, A; Chavet, J. 1976. Skidding Accidents, Friction Numbers, and the Legal Aspects Involved. *Transportation Research Record* 623:1-10.

Shieh, G. 2001. Sample size calculations for logistic and Poisson regression models. *Biometrika*, 88(2):1193-1199.

Schurr, KS; McCoy, PT; Pesti, G; Huff, R. 2002. Relationship of Design, Operating, and Posted Speeds on Horizontal Curves of Rural Two-Lane Highways in Nebraska, *Transportation Research Record 1796*:60-71.
Signorini, DF. 1991. Sample Size for Poisson Regression. *Biometrika*, 78(2):446-450.
Wallman, CG; Astrom, H. 2001. Friction Measurement Methods and the Correlation between Road Friction and Traffic Safety – Literature Review. VTI Meddellande 911V, Swedish National Road and Transport Research Institute, Linkoping, Sweden.

Appendix I. Technical Details of Sample Size Determination and Selection of Random Locations

For sample size determination in the Poisson log-linear regression model setup, we used the approach in Signorini (1991) assumed count responses Y_i on N individuals subject to exposure t_i, so that $\lambda_i = E[Y_i]$ follows

$$\lambda_i = t_i \exp\left(\beta_0 + \beta^T x_i\right) \tag{I.1}$$

where β_0 is the intercept, $\boldsymbol{x_i}$ is a p–dimensional vector of covariates, $\boldsymbol{\beta^T} = (\beta_1,\ldots,\beta_p)^T$ is the corresponding p–dimensional vector of parameters, $\boldsymbol{x_i}$ and t_i are regarded as realizations of independent random variables **X** and T, where $\mathbf{X} \sim f_X(x)$ and $T \sim f_T(t)$ with mean exposure time μ_T. Let b_0 and b denote the respective MLE's of β_0 and $\boldsymbol{\beta}$, obtained by maximizing the likelihood function $L(\beta_0,\boldsymbol{\beta})$. As N increases, standard asymptotic theory states that these converge in distribution to a multivariate normal distribution, with mean $(\beta_0,\boldsymbol{\beta^T})^T = \boldsymbol{\beta}^*$ and variance– covariance matrix I^{-1}, where I is the Fisher information matrix with elements given for

$$I_{jk} = -E\left(\frac{\partial^2 \log L}{\partial\beta_j \partial\beta_k}\right) \quad (j,k = 0,\ldots,p) \tag{I.2}$$

$$I_{jk} = N\mu_T \exp(\beta_0) E\left\{X_j X_k \exp\left(\beta^T X\right)\right\} \tag{I.3}$$

The second term follows from the independence of T and X. Suppose the moment generating function of the covariates X is $m(s) = E\{\exp(S^T X)\}$. For

$i,j=1,\ldots, p$, let $m_i = \dfrac{\partial m}{\partial s_i}, m_{ij} = \dfrac{\partial^2 m}{\partial s_i \partial s_j}$, $m_0 = m_{00}$ =m and $m_{i0} = m_{0i} = m_i$ and form the

(p+1)× (p+1) matrix M = (m_{ij}).We can write

$$I(\beta_0, \beta) = N\mu_T \exp(\beta_0) M(\beta) \tag{I.4}$$

and the maximum likelihood estimate $\hat{\beta}^*$ of β^* is asymptotically, as $N \to \infty$, multivariate normal with mean β^* and covariance matrix $(N\mu_T)^{-1}\exp(\beta_0)M^{-1}(\beta)$.

Suppose β_1 is the parameter of interest, and suppose we wish to test the null hypothesis H_0: $\beta = \beta_N$ = (0, β_2, . . ., β_p) against the alternative hypothesis H_1: β_A = ($\tilde{\beta}, \beta_2, \ldots \beta_p$), at a significance level α and with power at least $1-\gamma$. Assuming N is large enough to apply the asymptotic results derived above, the asymptotic variance of $\widehat{\beta}_1$ is given by the second diagonal term of I^{-1}. Routine calculation gives

$$N\mu_T \exp(\beta_0) \geq \{z_\alpha V^{\frac{1}{2}}(\beta_N) + z_\gamma V^{\frac{1}{2}}(\beta_A)\}^2 / \tilde{\beta}^2, \tag{I.5}$$

where $V(\beta) = \{M^{-1}(\beta)\}_{22}$, the second diagonal term of M^{-1}, evaluated at β, and z_δ is the $1-\delta$ point of the standard normal distribution.

Appendix II. Generalized Linear Modeling (GLIM)

Although linear models are very versatile in many applications, there are some problems: restriction to normality, and the assumption of a linear model function relating the response to the predictors. A motivation for using GLIM is that it permits more general distributions than the normal for the response (McCullagh and Nelder, 1991). A link function is used to relate the linear model to the mean of the response variable through a suitable scale. We may think of GLIM as an extension of the usual normal linear models.

The log-linear regression model is a standard model for count response data. Poisson regression is the most often used log-linear regression model, and allows the intensity (mean) parameter of the count response to depend on regressors (or covariates) via a link function. We usually assume that we know the parametric form of the relationship and the covariates are assumed fixed. The fitting of the Poisson regression may be carried out via IRLS (iterative reweighted least squares) or maximum likelihood by standard statistical packages.

In some cases, due to clustering of events, or some contaminating influences, there is variation in the responses that does not coincide with that implied by Poisson distribution.

Suppose we fit the usual Poisson regression model to such data, the resulting fit will have the following problems. Although the parameter estimates are still approximately unbiased for the true parameters, their standard errors are smaller than they should be, so that tests give smaller p-values than truly possible from the data. In order to account for the extra variation, we include an extra overdispersion parameter into the regression model. When this parameter is one, the model corresponds to the usual Poisson model. The quasi-likelihood approach is useful for parameter estimation for the overdispersed Poisson model.

Alternately, negative binomial regression is used to estimate count models when the Poisson estimation is inappropriate due to overdispersion, which is common for crash counts. In a Poisson distribution the mean and variance are equal. When the variance is greater than the mean the distribution is said to display overdispersion, which is demonstrated in our crash data. Although the overdispersed Poisson model is an option, the Negative Binominal model which offers better performance is assumed, and the probability of observing n_i crashes is represented as

$$P(n_i) = \frac{\Gamma(n_i+\theta)}{\Gamma(n_i+1)\Gamma(\theta)}\left(\frac{\theta}{Np+\theta}\right)^{\theta}\left(\frac{Np}{Np+\theta}\right)^{n_i} \tag{II.1}$$

This formula is derived from the Poisson distribution, in which Np is the mean under a Poisson distribution, however, it cannot represent the crash frequency at different observation sites. Therefore, an error term following a Gamma distribution was incorporated to the average crash frequency because of the between–site variation in the database. After integrating on the error in Poisson distribution, the NB distribution was obtained in which θ is the inverse of the dispersion parameter in the NB distribution.