SAFETY ISSUES RELATED TO

TWO-WAY LEFT-TURN LANES

Prepared for:

Florida Department of Transportation

January 2004

SAFETY ISSUES RELATED TO TWO-WAY LEFT-TURN LANES

Written by

Juan C. Pernia
Research Associate

Jian John Lu, Ph.D., P.E.,
Associate Professor

and

Haolei Peng
Research Assistant

Transportation Program
Department of Civil and Environment Engineering
University of South Florida
4202 E. Fowler Ave., ENB 118,
Tampa, FL 33620
Phone: (813) 974-5817
Fax: (813) 974-2957

Presented to
Florida Department of Transportation

A Project Sponsored by
Florida Department of Transportation

January 2004

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Safety Issues Related to Two-Way Left-Turn Lanes		5. Report Date January. 2004 6. Performing Organization Code
7. Author(s) Juan Pernia, Jian Johm Lu, and Haolei Peng		8. Performing Organization Report No.
9. Performing Organization Name and Address University of South Florida Department of Civil and Environmental Engineering 4202 E. Fowler Avenue, ENB 118 Tampa, FL 33620		10. Work Unit No. (TRAIS) 11. Contract or Grant No. BC353-40
12. Sponsoring Agency Name and Address Florida Department of Transportation 605 Suwannee Street Tallahassee, FL 32399-0450		13. Type of Report and Period Covered Final, Technical Report 2002-2004 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

In the past, the two-way left-turn lane (TWLTL) median treatment was frequently used in Florida to inexpensively improve motor traffic flow. A procedure to identify the factors that are influential in the safety experience of TWLTL sections was developed during this research. This procedure would also allow the identification of groups of TWLTL locations that present existing and future safety concerns and need further analysis to determine possible improvements. For this purpose, a three-year crash history database, from 1996 to 1998, including traffic crashes from totally 1688 TWLTL sections all over Florida was used in this research. Statistical models were used to determine the relationship between number of crashes per mile per year and several factors such as traffic volume, access density, posted speed and number of lanes. During the analysis, distribution fitting for the Poisson, Negative Binomial and Lognormal distributions was performed for the crash data. Then, a Negative Binomial regression model was developed to estimate the number of crashes per mile per year for the TWLTL sections. The regression parameters were estimated by using the maximum likelihood method. The goodness-of-fit for the developed model was evaluated based on Pearson's R-square and likelihood ratio index. In regard to the methodology to identify locations that need further study to determine if improvements are necessary, the procedure consisted of several steps. The steps included: plotting crash data to determine the distribution of the actual data for six different groups of locations based on posted speed and number of lanes, distribution fitting of crash data to determine the statistical distribution that better represents the actual data; determining percentile values for the average numbers of crashes from distribution fitting of the original crash database; estimating the average number of crashes per mile per year for sections with the same characteristics using the parameters of the crash predictive model; estimating critical values for the variables considered in the research with the percentile value of the average number of crashes and the curves plotted for number of crashes from the models; calculation of tables of critical average AADT values based on the values of access density number of lanes and posted speed for each one of the six groups; and the generation of a list of locations identified as critical according to the selected percentile value and the critical average AADT based on access density, posted speed, and number of lanes.

17. Key Word	18. Distribution Statement
two-way left-turn lanes, median treatment, safety issues, access density, posted speed, AADT, modeling, driveways	No restriction This report is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	125	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Disclaimer

The opinions, findings, and conclusions expressed in this publication are those of the author(s) and not necessarily represent those of the Florida Department of Transportation.

This report is prepared in cooperation with the State of Florida Department of Transportation.

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES vi

ABSTRACT vii

ACKNOWLEDGMENT ix

CHAPTER 1. INTRODUCTION 1
- 1.1 Background 1
- 1.2 Statistical Modeling 4
  - 1.2.1 Poisson Regression Model 4
  - 1.2.2 Negative Binomial Model 6
- 1.3 Research Statement 8
- 1.4 Research Purposes and Objectives 10
- 1.5 Outline of the Report 10

CHAPTER 2. LITERATURE REVIEW 12
- 2.1 Characteristics of TWLTL 12
- 2.2 Existing Guideline for TWLTL 13
- 2.3 Simple Linear Regression vs. Generalized Linear Regression 15
- 2.4 Poisson Regression vs. Negative Binomial Regression 17
- 2.5 Regression Models for TWLTLs 18

CHAPTER 3. RESEARCH METHODOLOGY 19
- 3.1 Crash Frequency and Crash Rates 19
  - 3.1.1 Crash Frequency 19
  - 3.1.2 Crash Rates 20
- 3.2 Statistical Distribution of Crashes 20
  - 3.2.1 Distribution Fitting 21
  - 3.2.2 The Chi-Square Test 22
- 3.3 Crash Prediction Models 23
  - 3.3.1 Statistical Prediction Modeling Procedure 23
  - 3.3.2 Test of Over-Dispersion 25
  - 3.3.3 Evaluation of Goodness-of-fit of Models 27
  - 3.3.4 Application of Crash Prediction Models 28
- 3.4 Methodology Developed to Identify Critical TWLTL Sections 28
- 3.5 Summary 29

CHAPTER 4. DATA COLLECTION 31
4.1 TWLTL Section Definition 31
4.2 Segment Length 31
4.3 Crash Database 33
4.4 Location of study sites 33
4.5 Identification of Crashes 34
4.6 Time Frame 35
4.7 Access Density 36
4.8 Setting-up the Crash Database 37
4.8.1 Extracting the Original Database 37
4.8.2 Sorting the Data Set 38
4.8.3 Converting the Crash-based Database to Section-based Database 38
4.9 Summary 40

CHAPTER 5. DATA ANALYSIS AND MODELING 41
5.1 Distribution Fitting 41
5.2 Statistical Modeling for Average Number of Crashes per Mile per Year 50
5.2.1 Statistical Analysis of the Variables 51
5.2.1.1 Traffic Volume (AADT) 51
5.2.1.2 Access Density 52
5.2.1.3 Number of Lanes 53
5.2.1.4 Posted Speed 54
5.2.2 The Model for Average Number of Crashes 54
5.2.3 Goodness-of fit of the Model 58
5.3 Linear Regression for Crash Rates 60

CHAPTER 6. MODELS FOR TWLTL SECTION RANKING AND CASE STUDY 64
6.1 Overview 64
6.2 Plotting Actual Crash Data 65
6.3 Distribution Fitting 66
6.4 Percentile Values of Crashes 68
6.5 Critical AADT Values 70
6.6 Tables for Critical AADT Values 72
6.7 List of Critical TWLTL Sections 73
6.8 Case Study 73
6.8.1 Setting up the Threshold 74
6.8.2 Critical AADT Values 74
6.8.3 Critical TWLTL Sections 75
6.9 Other Considerations 76
6.9.1 Number of Crashes 77

6.9.2 Assessing the Road 77
6.9.3 Assessing Future Development 79
6.10 Software for the Developed Methodology 79

CHAPTER 7 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS 80
7.1 Summary 80
7.2 Conclusions 81
7.3 Recommendations 83

APPENDICES 84

LIST OF TABLES

Table 5.1 Mean and Variance of Average Number of Crashes per Mile per Year 43

Table 5.2 Chi-square Test for Poisson Distribution Fitted for Average Number of Crashes per Mile per Year 44

Table 5.3 Chi-square Test for Negative Binomial Distribution Fitted for Average Number of Crashes per Mile per Year 46

Table 5.4 Chi-square Test for Log-normal Distribution Fitted for Average Number of Crashes per Mile per Year 47

Table 5.5 Descriptive Statistics for Average Number of Crashes per Mile per Year 51

Table 5.6 Descriptive Statistics for AADT 52

Table 5.7 Descriptive Statistics for Access Density 53

Table 5.8 Descriptive Statistics for Number of Lanes 53

Table 5.9 Descriptive Statistics for Posted Speed 54

Table 5.10 Estimated Parameters of the Negative Binomial Regression Model 56

Table 5.11 Explanations of Contents of the Results 56

Table 5.12 Criteria for Assessing the Goodness-of-Fit 59

Table 5.13 Statistics for the Model with All Crashes 61

Table 5.14 Estimated Parameters for the Model with All Crashes 61

Table 5.15 Statistics for the Models considering Speed Levels 62

Table 5.16 Estimated Parameters for the Models considering Speed Levels 63

Table 6.1 Description of Categories for Analysis 66

Table 6.2 Chi-square Test for Poisson, Negative Binomial and Lognormal Distribution Fitting 67

Table 6.3 Percentile Values for Distributions of Average Number of Crashes 70

Table 6.4 Final Percentile Values for Average Number of Crashes Based on Linear Regression 70

Table 6.5 Critical AADT for 85$^{\%}$ for Different Densities and Section Characteristics 72

Table 6.6 Average Number of Crashes for Case Study 74

Table 6.7 Critical AADT for the 85th Percentile Value 75

Table 6.8 Identified TWLTL Sections for District 7 76

Table 6.9 Evaluation of TWLTL Sections and the Road as a Whole 78

LIST OF FIGURES

Figure 1.1 Two-way-Left-turn Lane Layout 2

Figure 1.2 Conflicting Movements Related to Two-way Left-turn Lanes 2

Figure 4.1 Roadway Section Numbering System 34

Figure 5.1 Distribution of Average Number of Crashes per Mile per Year 42

Figure 5.2 Poisson Distribution of Average Number of Crashes per Mile per Year 49

Figure 5.3 Negative Binomial Distribution of Average Number of Crashes per Mile per Year 49

Figure 5.4 Log-normal Distribution of Average Number of Crashes per Mile per Year 50

Figure 5.5 Model Curve for Higher Speed Category for Four-lane Sections 56

Figure 6.1 Frequency Distribution for a Two-lane Road in the Higher Speed Category 66

Figure 6.2 The 50% and 85% Percentile Value of the Average Number of Crashes for Higher Speed and Four-lane Sections 69

Figure 6.3 Critical Value for AADT according to 50th Percentile Value for a Four-lane Section in the Higher Speed Category 71

Figure 6.4 Critical Value for AADT according to 85th Percentile Value for a Four-lane Section in the Higher Speed Category 71

ABSTRACT

In the past, the two-way left-turn lane (TWLTL) median treatment was frequently used in Florida to inexpensively improve motor traffic flow. A procedure to identify the factors that are influential in the safety experience of TWLTL sections was developed during this research. This procedure would also allow the identification of groups of TWLTL locations that present existing and future safety concerns and need further analysis to determine possible improvements. For this purpose, a three-year crash history database, from 1996 to 1998, including traffic crashes from totally 1688 TWLTL sections all over Florida was used in this research. Statistical models were used to determine the relationship between number of crashes per mile per year and several factors such as traffic volume, access density, posted speed and number of lanes. During the analysis, distribution fitting for the Poisson, Negative Binomial and Lognormal distributions was performed for the crash data. Then, a Negative Binomial regression model was developed to estimate the number of crashes per mile per year for the TWLTL sections. The regression parameters were estimated by using the maximum likelihood method. The goodness-of-fit for the developed model was evaluated based on Pearson's R-square and likelihood ratio index.

In regard to the methodology to identify locations that need further study to determine if improvements are necessary, the procedure consisted of several steps. The steps included: plotting crash data to determine the distribution of the actual data for six different groups of locations based on posted speed and number of lanes, distribution fitting of crash data to determine the statistical distribution that better represents the actual data; determining percentile values for the average numbers of crashes from distribution fitting of the original crash database; estimating the average number of crashes per mile per year for sections with the same characteristics using the parameters of the crash predictive model; estimating critical values for the variables considered in the research with the percentile value of the average number of crashes and the curves plotted for number of crashes from

the models; calculation of tables of critical average AADT values based on the values of access density number of lanes and posted speed for each one of the six groups; and the generation of a list of locations identified as critical according to the selected percentile value and the critical average AADT based on access density, posted speed, and number of lanes.

ACNOWLEGMENT

This project was sponsored by the Florida Department of Transportation. This support is greatly appreciated. The authors would like to express especial thanks to David Blodgett from the Office of Policy Planning; Gary Sokolow from the Systems Planning Office; and Patrick Brady from the Safety Office for their technical support and guidance.

Thanks are also given to different people in the staff of Office of Planning, Systems Planning Office and Safety Office of the Florida Department of Transportation for their support in accessing data and other resources in order to perform the research.

CHAPTER 1. INTRODUCTION

1.1 Background

From the 1950s through the 1970s, many arterials, collector roads and streets were constructed with either two lanes or four lanes and without turn lanes or medians. In these types of roads, all lanes served both the through traffic and turning traffic. Congestion delays and crashes increase with the growth of turning traffic volume, which is caused in many cases by unmanaged development and access along the roadway. The types of crashes associated with turning vehicles include rear-end crashes and turning crashes. To help solve some of these problems, traffic engineers and highway designers have used two-way left-turn lanes (TWLTL) as a median treatment on the roadway to separate left-turning traffic from through traffic.

A two-way left-turn lane is a lane in the center of a road that is dedicated to left turn movements for both directions of traffic. This type of median treatment provides separate storage lanes for left-turn vehicles along the roadway. It is commonly used to address the safety and operational problems on roadways caused by conflicts between through- and mid-block left-turn traffic. In other words, TWLTLs are intended to provide refuge from through traffic to left-turning vehicles waiting for a safe gap in the traffic flow. Figure 1.1 shows the basic concept of TWLTL.

When using TWLTL, several conflicting movements related to vehicles will occur. These conflicts include (1) motorists trying to cross the arterial from a driveway to a driveway or from a street to a street; (2) making a left turn off the arterial to a driveway or side street; (3) using the left-turn lane to pass stopped vehicles in the main through lanes; (4) allowing uncontrolled U-turns across two through lanes; (5) making a left turn from a side street or driveway onto the arterial; (6) accelerating in TWLTL to merge right; and (7) head-on accidents in the TWLTL. All of these conflicts are potential traffic crashes. Figure 1.2 shows these conflicting movements.

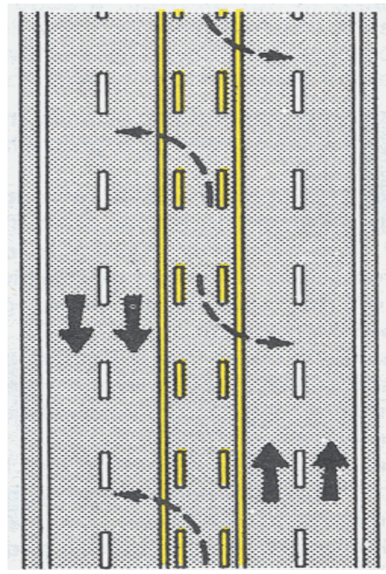


Figure 1.1 Two-way-Left-turn Lane Layout

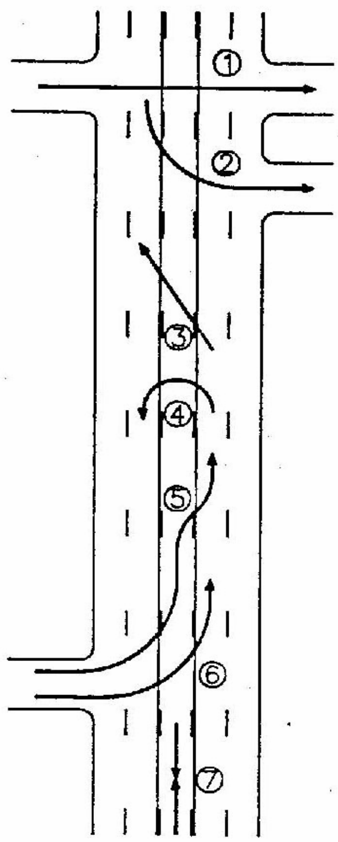


Figure 1.2 Conflicting Movements Related to Two-way Left-turn Lanes

These conflicting movements are more evident with high traffic volumes on the roadway. Previous studies have indicated that TWLTLs should generally not be used in situations where the through traffic volume is substantial [Harwood et. al. (1985), Bonneson et. al. (1997)]. If the volume in the roadway is very high, a TWLTL road might become ineffective since left-turning vehicle might decelerate or even stop in the inside through lane, creating delay to through traffic and a loss of capacity and efficiency to the road. Furthermore, heavy volumes on multiple through lanes may prevent a left-turning vehicle from finding a safe acceptable gap for an extended period of time. Moreover, if many left-turning vehicles queue up behind a vehicle in the TWLTL, its driver may feel under pressure to accept an unsafe gap, which may result in an increase in crashes or near crashes.

Consequently, traffic volume of the roadway is a very significant factor that influences safety at TWLTLs locations. In addition, other factors seem to affect the operational and safety performance of TWLTLs. The book “A Policy on Geometric Design of Highways and Streets”, published by the American Association of State Highway and Transportation Officials [AASHTO (2001)], provides a few specific suggestions about the use of a TWLTL, which includes: “[TWLTL] works well where the speed on the arterial highway is relatively low (25 to 45 mph) and there are no heavy concentrations of left-turning traffic,” and “[TWLTL] should be used only in an urban setting … where there are no more than two through lanes in each direction”. Azzeh et al. (1975) presented results of a comparative analysis related to safety aspects of a raised median and TWLTL, in a report prepared for the Federal Highway Administration (FHWA). In the report, the authors indicated that when driveway density was high, a raised median was safer than a TWLTL. Therefore, the significance of factors such as traffic volume, posted speed, access density and others should be taken into consideration when analyzing the safety of a TWLTL location

For this reason, this research was conducted to determine the relationship between variables such as traffic volume, posted speed, access density and number of lanes to crash occurrence at TWLTLs locations in Florida. Variables were obtained from FDOT databases or from video logs of state roadways provided by FDOT. This relationship was estimated using statistical models. Furthermore, several tables and curves for different values of the variables considered in the research were determined from the results of this relationship in order to identify locations where safety is a concern. This group of locations would need further analysis to determine possible improvements.

1.2 Statistical Modeling

1.2.1 Poisson Regression Model

Many types of regression models have been used to develop crash prediction models in the past 30 years. However, conventional regression models are proved to be inappropriate by many studies [Jovanis et al. (1985), Hauer et al. (1988), Saccomanno et al. (1988), Miauo et al. (1993)]. Meanwhile, recent researches show that Poisson regression model possesses the most desirable statistical properties when describing vehicle crash events that are random, discrete, nonnegative and typically sporadic [Miauo et al. (1993), Bauer et. al. (1996), Joshua et. al. (1990)].

Consider a set of n locations with given characteristics (e.g. medium traffic volume, 50 mph posted speed, in business area,). Associated with each location i^{th} is a set of parameters, X_{i1}, X_{i2}, …, X_{iq}, which describe the safety-related characteristics of this location, such as traffic volume, number of lanes, posted speed, etc.. Let the average number of crashes occurring at the i^{th} location during a specific time interval (e.g. crashes/per year or crashes/three-year) be denoted by Y_i, where, i= 1, 2, …, n. Then, denotes the actual observation of Y_i during the same time interval by y_i, where y_i = 0, 1, 2, …. and i = 1, 2, …, n. The objective of a statistical model is to provide a relationship between a function of the expected number of crashes, $E(Y_i) = \mu_i$, at the i^{th} location, and the q parameters at this location, X_{i1}, X_{i2}, …, X_{iq}. This relationship can be formulated through a general linear form:

$$g(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq} \tag{1-1}$$

where, the regression coefficients, β_0, β_1, β_2, … , β_q are to be estimated from the data. The estimation procedure to be adopted is dependent on the assumption made about the distribution of Y_i. The assumption underlying Poisson regression is that the number of crashes, Y_i, follows a Poisson distribution with mean μ_i. The probability that a location defined by a set of explanatory variables, X_{i1}, X_{i2}, …, X_{iq}, experiences y_i crashes during a fixed time interval can be expressed as:

$$P(Y_i = y_i, \mu_i) = \frac{\mu_i^{y_i} \times e^{-\mu_i}}{y_i!} \tag{1-2}$$

where,

Y_i – discrete and random variable representing the number of crashes occurring at i^{th} location during a period of time;

y_i – actual or observed number of crash at i^{th} location during a period of time;

μ_i – expected number of crashes, the dependent variable corresponding to a set of predictor variables.

The natural logarithm link function is adopted in Poisson regression models.

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq} \tag{1-3}$$

From mathematics perspective, it is not always clear in practice what link should be employed, and very often the data are analyzed by comparing several alternative choices. The reason to choose the natural logarithm link function here is taking into account the non-negative feature of crash count data. As stated by McCullagh and Nelder (1989), although canonical link may be found to be adequate over the range of the data, it is often dubious and logically unsatisfactory for extrapolation. By using a natural logarithm link, ln (μ_i) rather than μ_i obeys the linear model. This construction ensures that μ_i remains positive for all combination of independent variables and parameters. In addition, recent crash prediction studies also show that the natural logarithm link function is a reasonable choice. The Poisson probability function has only one parameter, mean, μ_i, and the variance, σ^2, equals the mean of the distribution. This inherent limitation of Poisson model is uncovered to be the major shortcoming of applying Poisson regression to crash prediction study.

Under the assumption of Poisson distribution, the regression coefficients, β_0, β_1, β_2, … , β_q, are estimated by the maximum likelihood method. The likelihood function is the product of the individual probability density function:

$$L(\mu) = \prod_{i=1}^{n} \frac{\mu_i^{y_i} \times e^{-\mu_i}}{y_i!} \tag{1-4}$$

This is a function of the parameter, μ_i, and through them, the parameters, β_0, β_1, β_2, … , β_q, are estimated by maximizing the likelihood, or more usually, by maximizing the logarithm of the likelihood. Because the logarithm is a strictly monotone transformation, the values that maximize L will also maximize log-L, which can be written as,

$$LL(\mu) = \sum_{i=1}^{n} [y_i \ln(\mu_i) - \mu_i - \ln(y_i!)] \tag{1-5}$$

The actual maximization procedure always requires an iterative calculation.

1.2.2 Negative Binomial Model

Regarding the types of models used for crash frequency studies, the Poisson regression model has been shown to be more appropriate than conventional linear regression models. However, the inability of the Poisson model to handle over-dispersed data is a major concern with regard to studying crash frequencies. This inability results from the major limitation of the Poisson regression model, which requires the variance of the dependent variable to be equal to its mean. Literature shows that most crash count data are likely to be over-dispersed, which means that the variance will likely be significantly greater than the mean [Shankar et al. (1995)]. When the mean and the variance of the data are not approximately equal, the variances of the estimated Poisson model coefficients tend to be underestimated and the coefficients themselves are biased.

This limitation can be readily overcome by using the negative binomial regression model, which assumes that the crash frequencies are distributed by negative binomial distribution. The negative binomial regression model is an extension of Poisson regression model and arises from Poisson regression model by adding an extra and independently distributed error term ε. For mathematics convenience the error term, exp (ε), is usually assumed to follow a gamma distribution with mean 1 and variance α. The resulting joint probability function, which is called negative binomial probability function, can be expressed as:

$$P\left(Y_i = y_i; \mu_i, \alpha\right) = \frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right) y_i} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} \qquad (1\text{-}6)$$

where,

Y_i – discrete, random variable representing the number of crashes occurring at i^{th} location during a period of time;

y_i – actual or observed number of crash at i^{th} location during a period of time;

μ_i - expected number of crashes, the dependent variable corresponding to a set of predictor variables;

α - dispersion parameter.

Note that the mean and variance of the negative binomial distribution of crash data can be expressed as:

$$E\left(Y_i\right) = \mu_i \qquad (1\text{-}7)$$

$$Var(Y_i) = E(Y_i) * [1 + \alpha E(Y_i)] = \mu_i + \alpha\mu_i^2 \qquad (1\text{-}8)$$

The second term on the right hand of equation (1-8), $\alpha\mu^2$, arises from the combination of the Poisson distribution with the gamma distribution assumption, and relaxes the constraints of Poisson distribution. Actually, as α goes to zero, the negative binomial regression yields the Poisson regression.

Like Poisson regression model, the relationship between the expected value of dependent variable and the corresponding q parameters, X_{i1}, X_{i2}, …, X_{iq}, is still taken to be:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq} \qquad (1\text{-}9)$$

The model coefficients, β_0, β_1, β_2, … , β_q, and the extra parameter, dispersion parameter α, are estimated by maximum likelihood method [Lawless (1987)]. The likelihood function is:

$$L(\mu_i, \alpha) = \prod_{i=1}^{n} \frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right) y_i!} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\alpha^{-1}} \qquad (1\text{-}10)$$

The log-likelihood function is:

$$LL(\mu_i; \alpha = \sum_{i=1}^{n} \left[\ln\left[\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1}) y_i!}\right] + y_i \ln\left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right) + \alpha^{-1} \ln\left(\frac{1}{1 + \alpha\mu_i}\right) \right] \qquad (1\text{-}11)$$

All the estimation procedure of coefficients was done using SPSS.

1.3 Research Statement

State Departments of Transportation and other responsible agencies lack of proper guidelines on the correct application of different median treatments. Often in many situations, only one factor might be taken into consideration in the decision making procedure without due regard to other influential factors. The major reason behind such circumstances is the lack of information on other related factors. More detailed research is therefore needed in this area to identify the factors that are influential in the safety experience of TWLTLs.

In the past, two-way left-turn lanes were frequently used in Florida to inexpensively improve motor traffic flow. While this may have been an appropriate application on low speed, low volume roadways with relatively few conflicts, the safety of this application is questionable at roadways with higher volumes and speeds. The Florida Department of Transportation has a policy of generally not using two-way left turn lanes on sections with design speeds greater than 40 mph. This research is needed to determine if traffic volumes and/or access density should also be limiting factors.

In this regard, this report describes a procedure developed to identify a group of TWLTLs locations that present safety concerns and need further analysis to determine possible improvements. Several factors that are influential in the safety experience of TWLTLs were also determined during this research. This research focused in the relationship between average number of crashes per mile per year and several factors such as traffic volume, access density, posted speed and number of lanes. FDOT crash database and RCI database were used to obtain information regarding number of crashes, traffic volume, posted speed and number of lanes. Video logs of all the state roads in Florida provided by FDOT were used to determine the access density of all TWLTLs locations considered in this research. Statistical modeling was applied to develop models to estimate the relationship between crash frequencies and the factors considered. Graphs and tables were developed based on the modeling results. Using these graphs and tables, the TWLTLS locations that might need improvements could be determined. Further analysis of this group of locations is needed since the procedure serves as a screening tool. A software that facilitates the use of the methodology developed on this research was also prepared.

1.4 Research Purposes and Objectives

The primary purpose of this research was to identify TWLTLs locations that present safety concerns and might need improvements based on factors that are influential in the safety experience of TWLTL sections. The specific objectives of the study were:

1) to review the existing literature and any ongoing projects related to TWLTLs;
2) to obtain information for traffic volume, access density, posted speed and others from the proper source;
3) to determine access density from video logs and verify that all locations considered were TWLTLs,
4) to conduct a detailed crash data analysis to establish factors influencing crashes and the best fitted distribution;
5) to develop statistical models for crash expectancy as a function of roadway traffic volume, access density, and posted speed at TWLTLs;
6) to recommend the criteria concerning the selection of TWLTLs that present safety concerns for possible improvements;

1.5 Outline of the Report

This report on the safety issues related to TWLTLs consists of seven chapters and appendices. Chapter 1 provides an overview of the research project with some backgrounds in this subject area. Chapter 2 presents brief summaries of previous studies related to TWLTLs. Chapter 3 describes the methodology employed to obtain the required information for the different factors considered and to determine the access density from video logs. It also describes the methodology used to develop the statistical models for crash expectancy as function of several factors. This chapter also presents the recommended procedure to select TWLTLs locations for possible improvements. Chapter 4 presents the data performing process, which was obtained from the FDOT crash database and other data sources. Chapter 5 presents analysis results and findings of the study including distribution fitting and the models developed. Chapter 6 describes in detail the developed methodology to identify critical TWLTL sections that might need improvements and a case study where the methodology is applied. Chapter 7 provides the summary, conclusions and recommendations of this study. The appendices contain figures and tables that complement the results and findings of the different sections of the project.

CHAPTER 2. LITERATURE REVIEW

2.1 Characteristics of TWLTL

Two-way-left-turn lanes and raised median are two common median treatments for roadways. The business sector and motoring public seems to prefer the TWLTL operation to raised median designs that restrict mid-block left turns. In 1978, a research of TWLTLs by Nemeth at the Ohio State University listed the advantages and disadvantages of TWLTLs over raised medians as follows:

Advantages of TWLTL over raised medians:

- Removal of left-turning vehicles from through traffic while still providing maximum left-turning access
- Reduction of delay to left-turning vehicles
- Direct access to adjoining property
- Flexibility in roadway use, as for a detour lane, a path for emergency vehicles, refuge for disabled vehicles

Disadvantages of TWLTL when comparing with raised medians:

- No refuge area for pedestrians crossing wide arterials
- Unsafe operation where sight distance is inadequate (such as where a TWLTL goes over a steep hill)
- Visibility problem of painted median (on rainy nights)
- More traffic conflict points, especially at driveways
- Possible misuse as a passing lane or even a travel lane
- Burden of instructing public in proper use. Some motorists do not know that the solid yellow line prohibits passing

Harwood and St. John (1985) also indicated the characteristics and appropriate use of TWLTLs and raised medians. In regard to TWLTLs, the authors indicated that these locations decrease travel time for drivers who wish to turn left and reduce delay to left-turning vehicles when comparing with locations where median openings are not provided. Raised medians reduce operational flexibility, such as allowing for emergency vehicle operation, lane closures, and work zones. The authors also mention that TWLTLs do not provide any refuge area for pedestrians, and that the inappropriate use of TWLTLs by drivers may cause many vehicular conflicts. Harwood and St. John also indicated that TWLTL should be used when there are low to moderate volumes of through traffic.

2.2 Existing Guideline for TWLTL

In the past decade, there have been many studies regarding median treatment selection. Many of this research compare TWLTLs with other median treatments. The main focus of the studies refers to the operational and safety effects of TWLTLs and other median treatments, and where each type of median could be appropriately used. Bonneson and McCoy (1997) indicated that accidents are more frequent on street segments with higher traffic demands, driveway densities, or public street densities. Also, they noted that accidents are more frequent where land use is business or office as opposed to residential or industrial.

FHWA conducted a study [Azzeh et al. (1975)] regarding accident-rate of TWLTLs and raised medians for a four-lane highway. They measured the accident rate reduction of these two types of median from a previously undivided roadway. The relative safety of the two medians remained constant for all average daily traffic (ADT) levels studied (less than 5,000 vehicles per day, 5,000 to 15,000 vpd, and more than 15,000 vpd). From the comparison of the results, TWLTLs resulted to be safer in areas with several concentrated sources of traffic and fewer than 60 commercial low-volume driveways per mile for all ADT levels. The report also indicated that for areas with no high-volume driveways and a large number of low-volume driveways, raised medians are safer. Furthermore, the report points out that a TWLTL should be considered to improve undivided roadways when there were frequent rear-end conflicts caused by left-turning vehicles and on moderate to high volume highways that have few cross streets and many driveways.

A research conducted by Parker (1983) on a four-lane road presents expected-value tables and a set of general guidelines. These expected- value tables indicated that TWLTLs have a lower number of accidents per mile when the roadway have ADTs ranging from 10,000 to 30,000 vpd, fewer than 30 driveways per mile and fewer than 5 streets per mile. The general guidelines for using different types of median presented in the research indicated that if the stopping sight distance is less than the AASHTO standards, a TWLTL should not be used. Additionally, the report indicated that TWLTLs should not be used when access is required on only one side of the street.

Harwood (1986) indicated that even though accidents per million vehicles miles (MVM) remained constant with different ADTs, accidents per mile per year preclude the use of TWLTL where ADT was high. He also specified that frequent driveways must be accompanied by few signals per mile and few approaches per mile. The author also addressed the need to accommodate pedestrians by suggesting that TWLTL should be used instead of raised medians when the number of driveways per mile exceeds 45.

Conversely, the study of development of policies and guidelines governing median selection conducted by Georgia Tech [Parsonson, (1990)] indicated that driveways per mile were not found to be significant for either raised median or TWLTL. Furthermore, the report specified that ADT is quite significant. Their results showed that with ADT of 10,000 vpd, TWLTL are safer except when the number of approaches per mile is low. With ADT of 30,000 vpd, raised medians are safer except with seven or more approaches per mile and two or fewer signals per mile. In the research, a TWLTL median would not be recommended for a four-lane road on the basis of accident per MVM, even though, the Georgia Tech results agree with other guidelines based on the number of accidents per mile per year.

The green book from AASHTO (2001), A Policy on Geometric Design of Highways and Street, does not present a comparative analysis of medians and TWLTLs. However, it made a few specific comments about the use of a TWLTL, "[TWLTL] works well where the speed on the arterial highway is relatively low (25 to 45 mph) and there is no heavy concentration of left-turning traffic", and "[TWLTL] should be used only in an urban setting where operating speeds are relatively low and where there are no more than two through lanes in each direction".

2.3 Simple Linear Regression vs. Generalized Linear Regression

Researchers have attempted several statistical approaches when relating traffic safety measures (e.g. crash frequencies, severity-weighted crash frequencies, crash rates) to traffic related explanatory variables. Among them, simple linear regression and generalized linear regression are the two most commonly used statistical techniques to develop crash prediction models. Simple linear regression is the traditional approach to develop crash prediction models. In the classical linear model, the dependent variable (e.g. crash frequency) is expressed as a linear combination of explanatory parameters with or without interactions, under the assumption that the dependent variable is normally distributed. Unlike conventional simple linear regression, generalized linear models, such as Poisson regression, negative binomial regression and lognormal regression, are based on alternative distributions. Poisson regression is appropriate for dependent variables that have a Poisson distribution, as crash counts often do. Negative binomial regression assumes the negative binomial distribution, and lognormal regression assumes the lognormal distribution. For each of these models, the dependent variable can be crash frequencies, or similar safety measures mentioned above.

Even though simple linear regression has generated many useful findings, studies show that this approach suffers some undesirable statistical properties, for example, the poor explanatory ability of the variation in crash data. In the study performed by King et al.

(1975), the author indicated that a linear regression model, even one with many independent variables, would not furnish an adequate model of crash experience associated with a given type of intersection control, and suggested to explore some more complex, probably non-linear regression model.

Joshua and Garber (1990) studied the relationship between crash involvements of trucks and associated traffic and geometric variables using both linear and Poisson regression models. The authors concluded that the multiple linear regression models did not adequately describe that relationship, but that the Poisson models did.

Miaou and Lum (1993) completed a study to evaluate the statistical properties of two conventional linear regression models and two Poisson regression models. The four types of models considered were: (1) an additive linear regression model; (2) a multiplicative linear regression model; (3) a multiplicative Poisson regression with exponential function; and (4) a multiplicative Poisson regression with non-exponential rate function. The authors concluded that of the four models tested, Poisson regression models outperformed linear regression models. Furthermore, the Poisson regression model with exponential rate function was the favored model.

Bauer and Harwood (1996) summarized several reasons indicating why conventional linear regression models are inappropriate for modeling crash frequencies or crash rates. The first reason indicates that traffic crashes are random and discrete events that are sporadic in nature. Secondly, crash frequencies for particular intersections or relatively small roadway sections are typically very small integers even if several years of crash data are obtained for those intersections and roadway sections. In fact, it is not uncommon for a substantial proportion of the sites in a crash study to have experienced no crashes at all during the study period. Small integer counts, often zero or close to zero, do not typically follow a normal distribution. Finally, crash frequencies and crash rates are necessarily non-negative, and traditional linear regression models could predict negative values for them.

2.4 Poisson Regression vs. Negative Binomial Regression

According to previous research, generalized linear regression definitely is a more adequate crash prediction approach than simple linear regression. Poisson regression models and negative binomial regression models are the generalized linear regression models that are being used widely. For the Poisson regression model, one important basic assumption is that the mean and the variance of the error distribution are equal. This feature simplifies the probability function, which only has one parameter. On the other hand, this advantage turns out to be the major disadvantage of Poisson regression models when applied to modeling crash data, which exhibits extra variation. If the variance of the crash frequencies exceeds the mean, then the data are over dispersed. When over dispersion exists in the data and Poisson regression models are used, the variances of the estimated model coefficients tend to be underestimated, which means the significance of the models will be overstated.

In their study, Miaou and Lum (1993) suggested the use of a more general probability distribution such as the negative binomial distribution to overcome the over-dispersion problem. In the follow-up study, Miaou (1994) recommended that the Poisson regression model should be used as the initial step to establish the relationship between the dependent variable and independent variables. Then, if over dispersion exists and is found to be moderate or high, both the negative binomial regression models and zero-inflated Poisson regression models can be explored.

J. Nicholson (1985) analyzed the considerable variation in the variability of crash counts. His results revealed that the pattern of crash occurrence at many locations was either too regular or too irregular to be well described by the Poisson process. Thus, the procedure for analyzing temporal variations in crash occurrences at particular locations should take into account the variations in the variability of crash counts. Based on the variance/mean ratio, the Binomial (variance/mean < 1.0) and Negative Binomial distribution (variance/mean ratio > 1.0) were complements to the Poisson distribution.

2.5 Regression Models for TWLTLs

Previous researches have developed statistical models for predicting the expected annual accident frequency for an arterial street. The independent variables for these models include the following factors: traffic volume or ADT, road length or segment length, driveway density, signalized intersection density, unsignalized public street approach density, median treatment type, number of arterial traffic lanes, and adjacent land use.

Bonneson and McCoy (1997) used statistical models to identify common trends related to median types. However, the large number of independent variables in each model prevented an exhaustive analysis. A typical combination of variables was established and used to compute the accident frequency predicted by each model for a range of daily traffic volumes.

Several studies used statistical models to analyze the performance of different median treatments. Harwood (1986) used accident data from California and Michigan to develop a procedure for estimating accidents for roads with different design features, which included a TWLTL and a raised median. For the estimation of delay he used a simulation model. Harwood's procedures utilize tabular data and graphs. Parker (1983) used data collected in Virginia to develop equations for estimating accidents and left-turn delay for roads with a TWLTL and a non-traversable median, respectively. Squires and Parsonson (1989) used data from urban areas in Georgia to develop equations for estimating accidents for roads with a raised median and a TWLTL, respectively. They did not analyze delay. Parsonson (1990) performed further analysis of accident data from Georgia, and he arrived at the conclusion that a raised median is safer than a TWLTL under all traffic conditions when used with either four-lane or six-lane roads

CHAPTER 3. RESEARCH METHODOLOGY

This chapter presents a detailed description of the methodology used in this research to analyze the data, namely, basic concepts of crash frequency and crash rate, distribution fitting, and statistical modeling, and how they were applied under the given situation. The basic idea of the methodology developed to identify critical TWLTL sections is also presented on this section.

3.1 Crash Frequency and Crash Rates

This section presents the definitions of crash frequency and crash rates based upon the Manual of Transportation Engineering Studies [Robertson et. al. {1998)]. *C*rash frequencies and crash rates are calculated in this study.

3.1.1 Crash Frequency

Crash frequency is the actual number of reported crashes that has occurred at a certain location, which could either be a roadway section or an intersection. By ranking the number of reported crashes, safety analysis can identify crash-prone locations. The primary virtues of using crash frequency are that it is simple and it makes intuitive sense. The number of crashes at each of the sections with TWLTLs considered in the research was obtained from the FDOT Crash Database.

The average number of crashes per mile per year, which is the arithmetic mean of number of crashes, was calculated for each TWLTLs section. In statistical inference, the mean is generally the most efficient estimator of the central tendency of the population characteristics being studied. The average number of crashes for section i was defined as:

$$N_i = \frac{n_i}{Y \cdot L} \tag{3-1}$$

where,

N_i = average number of crashes per mile per year for section i,

n_i = number of crashes at section i for the time frame considered,

Y = time frame for the analysis (years),

L = length of section i (miles).

3.1.2 Crash Rates

Crash-prone sections are also identified by crash rates, which are defined as crashes per million vehicle miles traveled for segments of crashes and per million entering vehicles for spots. Crash rates are used as a criterion for identifying high crash locations and statistical tests could be applied to determine whether the crash rate “is significantly different” compared with a predetermined crash rate for segments or locations with similar characteristics. Crash rates for all crashes were calculated. The following equation is used to calculate crash rates at sections:

$$\text{CRS} = \frac{1{,}000{,}000 \times A}{365 \times T \times V \times L} \tag{3-2}$$

where,

CRS = Crash Rate for the Segment

A = number of police reported crashes,

T = time frame of the analysis (years),

V = average annual daily traffic (AADT, vpd),

L = length of the segment (miles).

The main concern over the use of crash rates is to reduce the influence of traffic volume on the result and thus to improve the accuracy of the safety analysis. Traffic volume, or the average annual daily traffic volume (AADT), is a variable that has previously been suggested as possibly being able to affect crash rates, although its exact effect on crash rate is not yet well understood. It is believed that the crash frequency tends to increase as the through way traffic volume (or AADT) goes up even though there are many other factors affecting the situation. In this study, the corresponding AADT for each site was obtained according to the date of the crash. In other words, the time period for the volume data matches the time period of the crash data being analyzed.

3.2 Statistical Distribution of Crashes

This section presents a description of the distribution fitting process performed to determine the statistical distribution that best fits the crash data and the Chi-Square test procedure to select the best choice of distribution.

3.2.1 Distribution Fitting

The average number of crashes and crash rates per year were calculated for each segment with the use of SPSS. Details of this procedure to process data will be explained in the next chapter. The estimated values are then plotted into histograms, where the independent variable (x-axis) is the average number of crashes per section and the dependent variable (y-axis) is the number of sections. Poisson, Negative Binomial and Lognormal distributions are used to fit the frequency of crash data for high and low speed section using the observed mean and variance. Subsequently, the Chi-Square goodness-of-fit test was used to test the hypothesis whether the number of crashes follows a particular probability distribution. The following paragraphs present a brief introduction to Poisson, Negative Binomial and Lognormal distribution.

The definition of Poisson distribution is: if the mean number of counts (λ) in the interval is greater than zero ($\lambda>0$), the random variable X that equals the number of counts in the interval has a Poisson distribution with parameter λ, and the probability mass function of X is

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x=0,1,2,\ldots.. \qquad (3\text{-}10)$$

where,

λ -- observed mean value of the crash frequency

In regard to the negative binomial distribution, the probability function of X is:

$$f(x) = \binom{x-1}{r-1} p^{r}(1-p)^{x-r}, \qquad x = r, r+1, \ldots\ldots \qquad (3\text{-}11)$$

where

r, p – two parameters calculated from observed mean and variance.

The mean and variance of this distribution of crash counts can be expressed in terms of parameters p and r as follows:

$$\text{Mean} = E(Y) = r/p \qquad (3\text{-}12)$$

$$\text{Variance} = Var(Y) = r(1-p)/p^{2} \qquad (3\text{-}13)$$

The Log-normal distribution is the continuous probability distribution of a random variable whose logarithm follows the normal distribution. The random variable x has the range space of $R_x=\{x:0<x<\delta\}$ and y=ln x, is normally distributed with two parameters, mean μ_y and variance σ_y^2. The density function of x, say f(x), is defined as:

$$f(x)=\frac{1}{x\sigma_\gamma\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\ln x-\mu_\gamma}{\sigma_\gamma}\right]^2} \tag{3-14}$$

The mean *E(x)* and the variance *V(x)* of the log-normal distribution are:

$$E(x)=e^{\mu_\gamma+\frac{1}{2}\sigma_\gamma^2} \tag{3-15}$$

$$V(x)=e^{2\mu_\gamma+\sigma_\gamma^2}\left(e^{\sigma_\gamma^2}-1\right) \tag{3-16}$$

3.3.2 The Chi-Square Test

The Chi-Square goodness-of-fit test is used to test the hypothesis whether the number of crashes follows a particular probability distribution. The test procedure requires a set of randomly chosen samples of size n from X, whose probability density function is unknown. These n observations are then plotted into a frequency histogram of k class interviews.

O_i represents the observed frequency in the i^{th} class interval. The expected frequency in the i^{th} class interval denoted E_i could be calculated from the hypothesized probability distribution. The test statistic is,

$$\chi_0^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i} \tag{3-17}$$

where

O – observed frequency in the class interval i,

E – expected frequency in the class interval i.

It can be shown that, if the population follows the hypothesized distribution, χ_0^2 has, approximately a Chi-square distribution with *k-p-*1 degrees of freedom, where *p* represents the number of parameters of the hypothesized distribution estimated by sample statistics. This approximation improves as n increases. If the calculated value of the test statistic $\chi_0^2 > \chi_{\alpha,k-p-1}^2$, the hypothesis that the distribution of the population is the hypothesized distribution would be rejected. $\alpha = 0.05$.

3.3 Crash Prediction Models

Statistical crash prediction models can estimate the average number of section related crashes as well as the corresponding variances at a TWLTL section in terms of all crashes. The models considered in this research use crash rate or crash frequency as the dependent variable together with various site characteristics for a large number of sites over a period of time. The modeling approach finds a relationship between crash frequency (or crash rate), traffic characteristics (such as volume and speed), and roadway geometry (such as segment length and number of lanes).

The regression models adopted in this study are based on observed crash frequency distributions. Two general types of statistical regression models have been considered to apply to the crash data: (1) conventional linear regression models; and (2) generalized linear models, including Poisson regression model and negative binomial regression model.

As mentioned before, many previous researches in this field show that conventional linear regression models are incapable to model the traffic crash data, which are non-negative, random, discrete and sporadic in nature. As alternatives, generalized linear models were explored and adopted in recent crash studies due to their advantages over conventional linear regression models.

3.3.1 Statistical Prediction Modeling Procedure

The crash modeling consists of seven major tasks: (1) to collect and reduce the crash data; (2) to analyze the crash data to determine the safety measure that was adopted as dependent variables in the modeling, and find appropriate probability functions to describe the random variation of crash frequencies; (3) to select and analyze the predictor variables; (4) to determine an appropriate functional form and parameterization, $f(.;\beta)$, to describe the effects of predictor variables on expected crash frequencies; (5) to estimate the regression parameters β in $f(.;\beta)$ using appropriate statistical algorithm based on crash data and probability assumptions; (6) to assess the quality of developed models, and make sure that the models make good engineering sense in addition to fulfilling statistical goodness-of-fit criteria; and (7) to apply the developed models, and convert the modeling results to tables or graphs for use. These tasks are briefly presented in the following paragraphs.

The modeling database was built by 1688 sections with TWLTL treatment collected from the state of Florida. This crash database was created from the Florida crash database maintained by FDOT, which consists of all crashes occurred on state roadways from 1996 to 1998. The 1688 sections included in the modeling database contained safety related characteristics and crash counts occurred within the TWLTL sections. The process of generating the modeling database will be presented in detail later.

Based on data analyses, the average number of crashes per mile per year, and crash rate were adopted as the safety measures. For these safety measures, models were developed. Another important issue was to determine which roadway characteristics should be used as predictor variables in the model. The variables should be easy to obtain by FDOT traffic engineers when applying the models. Totally four roadway characteristics including AADT of section with TWLTL, access density in the section, posted speed in the section and number of lanes on the section were included in the model as predictor variables.

Based on crash frequency distributions and previous studies Poisson regression and negative binomial regression were chosen to estimate the model parameters. Generally, Poisson regressions can be used to build the relationships between crash frequencies and a set of predictor variables under the assumption that crash frequencies are Poisson distributed. However, Poisson regression has a limitation requiring the variance of the data to be equal to the mean. This restraint can be overcome by negative binomial regressions assuming crash frequencies are negative binomial distributed. Thus, for the model, Poisson regression was used as an initial step in the modeling process, with a negative binomial regression being applied if over-dispersion was founded existed in the crash data. For both Poisson and negative binomial regressions, the regression parameters were estimated by maximum likelihood method with GENMOD procedure in SPSS. Once the models were developed, two methods were applied to test the goodness-of-fit of the models: Pearson's R-square, and likelihood ratio index.

With the developed model, the expected number of crashes at TWLTL sections was estimated. The calculated results were tabulated in order and presented in graphs to furnish a simple and clear overview of the impacts of the TWLTL treatment on crashes on roadways with different characteristics.

Regarding crash rates, Poisson regression model and Negative Binomial regressions models did not give good results. Therefore, a relationship between this safety measure and the predictor variables was determined using the conventional linear regression. Even though, this method is not the best alternative, this relationship will give an overview of how several variables considered influence crash rates.

3.3.2 Test of Over-Dispersion

Firstly, Poisson regression was performed during the modeling process. After that, the crash data were tested for over-dispersion related to Poisson regression. If extra-Poisson variation is proved to be significant, the Poisson distribution assumption is violated; then the negative binomial regression model would be a more appropriate choice.

To test the over-dispersion of data, the mean deviance and Pearson's χ^2 ratio were used due to two reasons. First, these methods are widely used [Bauer et al. (1996), McCullagh and Nelder (1983)]; secondly, these methods are adopted by SPSS software. Let L_s denote the maximum likelihood estimated from the saturated model that has as many parameters as observations, making each fitted value equal to the observed value, and let L_β denote the likelihood estimated by the current model. For Poisson regression model, log-likelihood can be expressed as,

$$\log(L_\beta) = \sum_{i=1}^{n} [-\mu_i + y_i \ln \mu_i - \ln(y_i!)] \quad (3\text{-}3)$$

$$\log(L_s) = \sum_{i=1}^{n} [-y_i + y_i \ln y_i - \ln(y_i!)] \quad (3\text{-}4)$$

where,

y_i - actual or observed number of crash at i^{th} section during a period of time;

μ_i - expected number of crashes, the dependent variable corresponding to a set of predictor variables.

The deviance, or G^2, is defined as minus twice the logarithm of the ratio of likelihood of the current model to the saturated model [Nelder et. al. (1972), Agresti (1990), Greene (1997)], and for Poisson regression, can be expressed as,

$$G^2 = 2\sum_{i=1}^{n}(y_i \ln \frac{y_i}{\mu_i}) \tag{3-5}$$

The deviance has an asymptotic distribution that is Chi-squared with degree of freedom equal to n-p, where n is the sample size and p is the number of parameters estimated. By forming the ratio of the deviance to its residual degree of freedom, n-p, an estimate of the scale constant $G^2/(n-p)$, called the mean deviance, can be found. For the Poisson regression, this scale constant should theoretically be equal to one. Values substantially in excess of one reflect over-dispersion of the data. The acceptable range for the mean deviance, $G^2/(n-p)$, is from 0.8 to 1.2.

Similar to the mean deviance statistic, the Pearson's χ^2 ratio statistic is also used to test the over-dispersion of crash data. The over-dispersion index can be calculated as,

$$\sigma_d = \frac{Pearson's\,\chi^2}{n-p} \tag{3-6}$$

where, n is the number of observations and p is the number of parameters used in the model. Pearson's χ^2 can be calculated by,

$$Pearson\chi^2 = \sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{Var(Y_i)} \tag{3-7}$$

where, for Poisson regression, $Var(Y_i)=\mu_i$. The value of σ_d tends to be one. If $\sigma_d>1.0$, then the data have greater dispersion than is explained by the Poisson distribution and a further analysis with a negative binomial error structure is required.

3.3.3 Evaluation of Goodness-of-fit of Models

So far there is no commonly acceptable measure that can give an absolute assessment of goodness-of-fit for generalized linear models. Therefore, several measures are selected and calculated, and jointly will give a relatively accurate evaluation of the models. First, deviance, as stated previously, is defined as minus twice the logarithm of the ratio of the maximum likelihood under current model and the maximum likelihood under saturated model. Thus, deviance describes lack of fit, greater deviance indicates poorer fit [Agresti (1990)]. Secondly, according to McCullagh and Nelder (1983), the Pearson's χ^2 is asymptotic to the χ^2 distribution with n-p-1 degrees of freedom for large sample sizes and exact for normally distributed error structures. Therefore, for a model, similar to deviance, the greater the Pearson's χ^2 , the poorer the fit. However, this statistic is not well defined in terms of minimum sample size when applied to non-normal distributions. Therefore, it should not be used as an absolute measure of model significance.

In traditional least square regression, the coefficient of determination, R^2, is frequently used to assess the goodness-of-fit of a model. It represents the proportion of variation in the data that is explained by the model. However, it was shown that R^2 is not an appropriate measure to assess the goodness-of-fit of crash prediction models due to their non-normal and nonlinear nature [Miaou et al. (1985)]. As a variation, a measure based on the standardized residuals, Pearson's R^2, can be calculated for each model to give some indication of the goodness-of-fit,

$$R_p^2 = 1 - \frac{\sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i}}{\sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{\bar{y}}} \tag{3-8}$$

where,

R_p^2 – Pearson's R-square statistic;

y_i – observed number of crash at i^{th} section during a time period;

μ_i – estimated number of crashes during a time period;

$\bar{y}$ – average crash counts at all sections of interest.

In addition, as the counterpart of R^2 in nonlinear regression, a measure of overall statistical fit, the likelihood ratio index can be computed as,

$$\rho^2 = 1 - \frac{L(\beta)}{L(0)} \tag{3-9}$$

where,

$L(\beta)$ – Log-likelihood at convergence;

$L(0)$ – restricted log-likelihood (all parameters are set to zero except for the intercept).

The value of 0.200 is quite satisfactory considering the variance in the data, and values tend to be generally lower than typical R^2 values [Ben-Akiva and Lerman (1985), Poch and Mannering (1996)].

3.3.4 Application of Crash Prediction Models

Once the parameters of crash predictive models were estimated, the average number of crashes per mile per year can be estimated by replacing the regression parameters, β_0, β_1, β_2, … , β_q, with the estimated values, and the variables X_{i1}, X_{i2}, …, X_{iq}, with the corresponding values of the characteristics of the TWLTL sections. However, the estimated average number of crashes per mile per year will only provide a statistic of the safety measure either for an infinite number of sections with the same characteristics or a section in an infinite time period with every characteristic unchanged. The estimated number of crashes obtained from the models will be plotted in different curves. From these curves, values for the different variables could be determined for the desired percentile values. These percentile values would be obtained from crash data. TWLTLs sections that are located above the values for the different variables based on the percentile value selected would be identified as locations that need further study in order to determine if improvements are needed.

3.4 Methodology Developed to Identify Critical TWLTL Sections

The methodology developed to identify locations that need further study to determine if improvements are necessary consists of several steps. The first step of the methodology includes plotting the crash data to determine the distribution of the actual data for six different groups of locations. The crash data was divided into six different groups based on posted speed and number of lanes in order to facilitate the application of the methodology. The second step consists of the distribution fitting of the crash data. Distribution fitting is used to determine the best fitted distribution for the crash data in order to decide which statistical distribution better represents the actual data. The third step refers to the determination of percentile values for the data according to distribution fitting. A desired percentile value for the number of crashes is set as a threshold. The average numbers of crashes for the percentile values are determined from distribution fitting of the original crash database. The fourth step is the estimation of the critical AADT based on percentile values and modeling results. The average number of crashes per mile per year could be determined for sections with the same characteristics once the parameters of the crash predictive model are estimated. These estimated number of crashes obtained from the model is plotted in different curves. With the percentile value of the average number of crashes and the curves plotted from the models, the critical values for the variables considered in the research could be obtain. The fifth step is the calculation of tables of the critical average AADT values for each one of the six groups considered. These AADT are based on the values of the other variables considered: access density number of lanes and posted speed. The sixth step is the generation of a list of locations identified as critical according to the selected percentile value and the critical average AADT based on access density, posted speed, and number of lanes. This methodology will be explained in detail on Chapter 6.

3.5 Summary

The research methodology used in the project was presented in this chapter. An overview of some basic definitions such as crash frequency and crash rates were presented. Distribution fitting were described in order to determine the number of crashes for a desired percentile from the crash data. Once, the statistical properties of crash frequencies were explored to determine the best regression model to use, the modeling procedure was presented with a detailed description of each one of its steps. The steps included: use of Poisson regression and negative binomial regression, test of over-dispersion, evaluation of the goodness-of-fit of developed models, and applications of modeling results. Finally, an overview of the methodology developed to identify critical TWLTL sections is presented. This methodology will be presented in detail in Chapter 6 along with a case study.

CHAPTER 4. DATA COLLECTION

The purpose of this chapter is to describe the process of the site selection and data collection effort in this research. This chapter addresses criteria to define an arterial segment length, FDOT crash database, the FDOT coding system for identifying roadway sections, and the procedures for gathering relevant crash data and creating a specific crash database for the research.

4.1 TWLTL Section Definition

This research study was intended to concentrate on TWLTL sections. The original list of TWLTL sections was provided by FDOT in EXCEL format. This list contained 3535 sections with information regarding their roadway ID number, beginning and ending milepost, and other characteristics. The characteristics included in the list were street name, number of lanes, posted speed, lane widths and median type. Number of lanes range from 2 to 7 lanes. Posted speed range from 25 to 60 mph. Segment length could be determined from the beginning and ending milepost. Median type basically referred to TWLTL median treatment. This list did not include any information regarding traffic volume and access density (number of driveways per mile).

4.2 Segment Length

Crash rates (crashes per million vehicle miles traveled) were computed by dividing the number of crashes that have occurred on a section per year by the segment length and the average daily traffic volume. The average number of crashes was also determined per year and per mile by dividing the number of crashes on a section by the time frame and the segment length. The segment length affects the magnitude of the crash rate and the average number of crashes, and extremely short segment lengths may affect the conclusions

Resende and Benekohal (1997) analyzed the influence of segment lengths on crash rates and showed that it influenced the geometric variables used in crash predictions for rural interstate highways and rural two-lane highways. To get reliable crash models, they indicated that the rates should be computed based on segment lengths of 0.5 miles or longer. However, Resende and Benekohal's study involved only rural highways, while the study sites of this project are located in urban and rural areas. Thus, the suggested value, 0.5 miles or longer was not considered as appropriate for used as the criteria for limiting the section length in this study. Resende and Benekohal did agree that their study didn't imply that short sections should be deleted from the data sets. It showed that crash rates computed from too short and too long sections could lead to misrepresentation of real conditions. Attention must be paid in collecting and investigating the most standardized number of sections as possible, so groups of similar section lengths could be created.

Criteria for establishing segment length were found in several other papers studying median treatments for urban arterial streets, although none of them provided a detailed discussion on how the criteria were identified [Bonneson et. al. (1997 and 1998)]. Council and Stewart (2000) chose a section length of 0.07 miles as the minimum length for which reported crash locations could be considered reliable for merging with the roadway inventory database. In highway safety improvement programs, it was presented that all highway segments used to calculate the safety ratio were 0.101 miles to 3 miles (HSIP). Bonneson and McCoy (1997) constructed a safety model to predict the expected annual frequency for a quarter-mile segment of arterial streets, that is, the signal spacing is 1,000 feet or more. Bonneson and McCoy presented the concept that the evaluation of the operational and safety effects of the three alternative median treatments was limited to their "mid-signal" performance. Their field observation of 71 arterials with a raised-curb median indicated that 93 percent had at least one median opening and that U-turn activity at these openings was negligible. In this context, the primary effects of a median treatment were assumed to occur outside of the functional boundary of any signalized intersection located on the arterial.

This project defined a segment as a street section with a median treatment of TWLTL, having signalized and unsignalized access points (driveways or side streets) along its length. Segments with length from 0.06 miles were selected.

4.3 Crash Database

FDOT has a very large crash database that is updated yearly. The database includes crashes gathered from the Department of Highway Safety and Motor Vehicles (DHSMV). The crashes included in the database are those that occurred in state roads and with a high amount of property damage, an injury, or a fatality. Crashes with high property damage are those with an estimated property damage of $500.00 or more. Crashes with minor property damage are not included in this database. The exclusion of crashes reported in short forms in the database may affect the estimated number of crashes in the sense that not all crashes occurred at TWLTLs are considered, and specific type of crashes, such as rear-end crashes, may be under reported because many of these crashes have low property damage. For each crash, there are several variables used to describe the site and time of the crash, geometric conditions, traffic control, and drivers and pedestrian’s characteristics.

4.4 Location of study sites

Florida consists of seven FDOT districts and they are numbered as District One through District Seven. Another district, District Eight or the Turnpike District, is responsible for the operations, maintenance, and construction of the toll roads in Florida. The sample sites for this study were selected from all FDOT districts.

4.5 Identification of Crashes

The state roads and the crash locations in FDOT’s crash database are recorded by using a code that utilizes FDOT section number and milepost. The main purpose of this system is to make the crash information consistent with the Street Line Diagram, which is a technical resource normally used by state road planners and traffic engineers. To perform the crash data analysis using the FDOT crash database, the definitions of section number and milepost must be clearly defined. Every roadway section on the state highway system in Florida has been identified with an eight-digit code by FDOT. This code is called “section” number, which uniquely defines that roadway. As described in Figure 4.1, the section number consists of a county number, a section number, and a subsection number. The first two digits correspond to the county number, the next three numbers are the actual section numbers for the roadway, and the last three numbers are known as the subsection number. Usually, a subsection is “000”. If a roadway is reconfigured, the subsection number may be recorded as “001”.

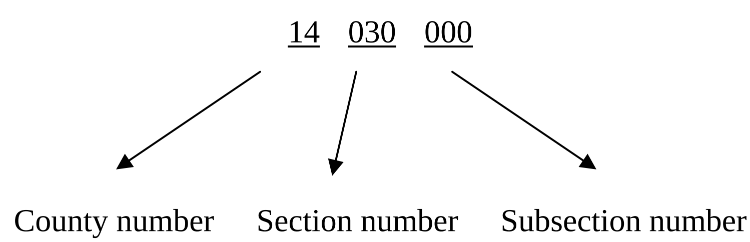


Figure 4.1 Roadway Section Numbering System

Milepost is used to describe those interacting points on the roadway, such as intersections, crossing interstates, driveways, and key commercial developments. Most roadways are labeled either from south to north or from west to east by FDOT. Thus, milepost zero is normally labeled at the southernmost or westernmost terminus of the road within that county. Mileposts are rounded to the third digit after the decimal points because 0.001 miles equals 5.28 feet. Thus, any point within the right-of-way of a roadway will be accurately identified using the numbering and milepost systems.

The roadway numbering and milepost systems are also applied in the FDOT crash database to identify the exact site location of a crash. There are totally five variables used to convey the information in the database: DISTID, COUNTYID, SECID, SUBSECID, and MILEPOST. The first four variables are used to identify on which road a crash occurred. MILEPOST is used to locate the exact position where the crash occurred. If the crash vehicle ran off the road, MILEPOST records the milepost of the point on the roadway that is nearest to the crash site. The subsection number is only used when a roadway is reconfigured (a one way pair is constructed which used to be a four-lane roadway).

Based on the numbering and milepost systems, all crashes occurred within a TWLTL section can be identified by searching the FDOT database by the crash ID number and milepost range according to the ID number and milepost of the section. Each section has a beginning and ending milepost, which allows the determination of crashes on the section.

4.6 Time Frame

In this study, crash data of three consecutive years, from 1996 through 1998, were used for the analysis process. It is commonly believed that three years will usually provide a sufficient number of crashes for analysis while reducing the possibility of extraneous factors influencing the crash data. Also, the *Manual of Transportation Engineering Studies* indicates that a three-year window is the most common choice [Hummer (1994)]. A three-year time frame enables analysts to collect sufficient crash counts. Whether or not enough crash counts are available is always the major concern for the analysts when conducting traffic crash studies, because traffic crash events are sporadic in nature, a large proportion of sections could experience no crash at all if the time window selected is too narrow.

Changes that have occurred at the site during the analysis period can result in changes to the crash characteristics. These include changes in the surrounding land use in addition to changes at the site itself. These changes have a higher probability of occurring, as the analysis period becomes longer. Thus, a three-year time frame represents a good compromise between the desire for larger crash sample size and the desire for time frames within which conditions were unlikely to have changed a great deal. Therefore, a three-year time frame was used in this study.

4.7 Access Density

Access density is a significant factor considered in the analysis. This variable was not available in the FDOT Crash Database. Therefore, the information regarding this specific variable was determined from video logs of all the state roads in Florida provided by FDOT in a hard drive. Obtaining the information of access density was the most time-consuming part of this research. In order to determine the number of driveways on the TWLTL, the video logs of each section were manually reviewed. Roadways on the video logs are also identified through the section number and milepost system. Therefore, video logs were viewed from the beginning and ending milepost of each TWLTL section. In order to count the number of driveways along the roadway for both sides of the road, the video logs had to be reviewed two times. One for each direction, since each side of the roadway is recorded separately in the video logs. The number of driveways for each TWLTL section was the result of adding the number of driveways in each side of the roadway considered. Finally, the access density per mile was calculated as follows:

$$\text{Access Density} = \frac{\text{number of driveways (in both directions)}}{\text{length of section}} \qquad (4\text{-}1)$$

During the procedure, many sections indicated as TWLTL in the original listed provided by FDOT did not have this median treatment. Thus, these sections were taken out of the database. Finally, 1789 sections with access density were available.

4.8 Setting-up the Crash Database

This section provides the general information about the creation of the crash database solely for the purpose of this project. The data set creation was conducted using the Florida Traffic Crash database, which was obtained from the State Data Program of the National Highway Traffic Safety Administration established under the U.S. Department of Transportation. [NHTSA (1998)].

4.8.1 Extracting the Original Database

Corresponding to each year, there is one data file consisting of all crashes occurred on state roads during that year. For each crash, twelve record types containing specific information related to the crash are included. Each record type contains approximately 20 variables. All these files, stored in ASCII format, have the same database structure. A SAS (Statistical Analysis System) program was written and used in order to change the ASCII format to SAS format. SAS program uses Structured Query Language (SQL) to gather all of crash data needed for the files.

Several variables were selected for the original database for the research. These variables were selected from five of the twelve record types, which included the factors that were considered having effect on the safety of TWLTLs. The record types selected were record “00” (Time and Location), record “01” (Characteristics), “09” (RCI-Features-I), record “10” (RCI-Features-II), and record “11” (RCI-Point). In order to put the variables in one file, these files with record type “00”, “01”, “09”, “10” and “11” were merged into one merged file for each year. The data of three consecutive years, from 1996 to 1998, were used for the analysis. When summarizing crash data for a merge, additional data were kept intentionally in case of the need for further study. It is considered as more economical to have extra information in a computer file than to discover later that needed information has been lost during the merge and must be rebuilt from the original recording system.

4.8.2 Sorting the Data Set

The statistical package software program SPSS was used to handle the large data sets. The original data files for each year considered in the research were merged into one file using SPSS. This file contained crash information and three variables, AADT (AADT), posted speed (POSTSPED), and number of lanes (NUMBLANE).

Additionally, other variables, district ID (DISTID), county ID (COUNTYID), section ID (SECID), subsection ID (SUBSECID), milepost (MILEPOST), accident number (ACCNUMB), and site location (SITELOC) were also included in the merged data file. These variables were used to identify the sections related to TWLTLs. In FDOT database, a certain accident number corresponds to one crash. If there is more than one vehicle involved in the crash, some characteristics variables, such as vehicle movement, may have values for each one of the vehicles involved. Thus, a crash is recorded several times since each vehicle involved in the crash has the same accident number. Therefore, in order to avoid the bias of having more crashes than the actual count, duplicate crashes were taken out from the data set.

The variable site location allows the identification of a crash that happened at an intersection or very close to it. These crashes might be influenced by the intersection and not by the TWLTL. The codes “02” and “03” of the variable site location indicate “at intersection” and “influenced by intersection”. Thus, records with these two codes for site location were taken out of the data set.

4.8.3 Converting the Crash-based Database to Section-based Database

Once the database based on crashes was all set, the next step was to convert it to section-based database. In the section-based modeling database, a record should correspond to a section. After reviewing the original list of TWLTL sections based on the video log information, approximately 1789 sections with TWLTLs for the seven districts of Florida State were considered. Each section was identified by roadway ID, beginning milepost and ending milepost. As mentioned above, the roadway ID is an eight-digit code consisting of county number, section number, and a subsection number. The breakpoints of the TWLTL on a roadway are indicated by mileposts (begin/end milepost). Crashes with mileposts within one of the ranges of begin milepost and end milepost on the list, of TWLTL sections were grouped as a section. The sections studied in this research were summarized based on the District ID, County ID, section ID subsection ID, beginning milepost and ending milepost of the list provided by FDOT.

Crash data for a section in three years could be zero, one of more crashes. This possibility of having different number of crashes also means that it could be zero, one or more crash records related to this section in the section-based database. During data processing, the total number of crashes of each section was easily determined. A problem encountered was that if there were more than one crash in the section, inconsistency of the data among the crash records could be possible. It was important to calculate or select a value for each variable. For the number of lanes, all records had the same value, which was taken for the variable in this section. For posted speed if all the records contained the same posted speed value, this was taken as the value for that variable. If the posted speed value was different for the records of crashes at the same section, the posted speed value was determine by matching values from the crash database, list of TWLTL sections given by FDOT, and video log information. For AADT, if all the records contained the same volume value, this was taken as the value for that variable. If the values were different for a section, the average AADT was calculated by averaging the AADT values of all the crashes in the section.

If there was zero crash or no crash in the section, the total number of crashes was recorded as “0”. While the values of all the variables were missing, the values of number of lanes and posted speed were obtained from the original list of TWLTL sections provided by FDOT. The information from the list from FDOT allowed to double check the variables of the database. If there were difference between them, the values from the spreadsheet by FDOT were used, which were more reliable.

The missing AADT were obtained from a computer disk of Florida Traffic Information prepared by FDOT. The FTI system contains information related to the main characteristics of roadways, including AADT. When using this program, the district number and the eight-digit road ID were input and the road was highlighted on a map of that area. By clicking on any point of the road, the AADT of that section was shown on the screen. Thus, AADT for the zero crash sections were obtained with the use of the system. In regard to the access density, this information was included in the section based database once the information for the other variables was already set for each one of the TWLTL sections.

4.9 Summary

Data processing programs were written to retrieve data from the FDOT database. The program SPSS was chosen to conduct the database-building task once the steps choosing time frames for crash analysis, identifying intersection related crashes, and selecting variables for the database were completed. After data reduction and analyses were completed, the final modeling database was built. The database consists of one section-based data file. In the data file, totally 1789 sections were included. For each section, four variables were used to record the safety-related information. This database was set up to perform the statistical analysis.

CHAPTER 5. DATA ANALYSIS AND MODELING

This chapter presents the analysis of the data and modeling procedure in three phases. The first phase covers the distribution fitting of the actual crash data. The types of distributions considered were: Poisson distribution, Negative Binomial distribution and Lognormal distribution. The second phase presents the modeling process followed to determine a relationship between crashes per mile per year at TWLTL sections and several variables through a Negative Binomial regression model. Descriptive statistics for the variables considered are also presented in this phase. The third phase presents the linear regression modeling procedure considering crash rates as the dependable variable.

5.1 Distribution Fitting

Distribution fitting was used to determine the best fitted distribution for the crash data in order to decide which statistical model better represents the data. The dependent variable adopted for the distribution fitting was the same variable that was considered for the modeling process: average number of crashes per mile per year. The average number of crashes per mile per year was obtained considering crash data from 1996 to 1998. The distribution curve for the average number of crashes per mile per year was plotted to assess the general shape of this variable in order to provide the basis for crash distribution assumptions for modeling. Figure 5.1 shows that a large number of sections have zero or low crash experience when considering crash data from all the TWLTL sections together. Therefore, the distribution of number of crashes per mile per year seems to follow a Poisson distribution, Negative Binomial distribution or Log-normal distribution.

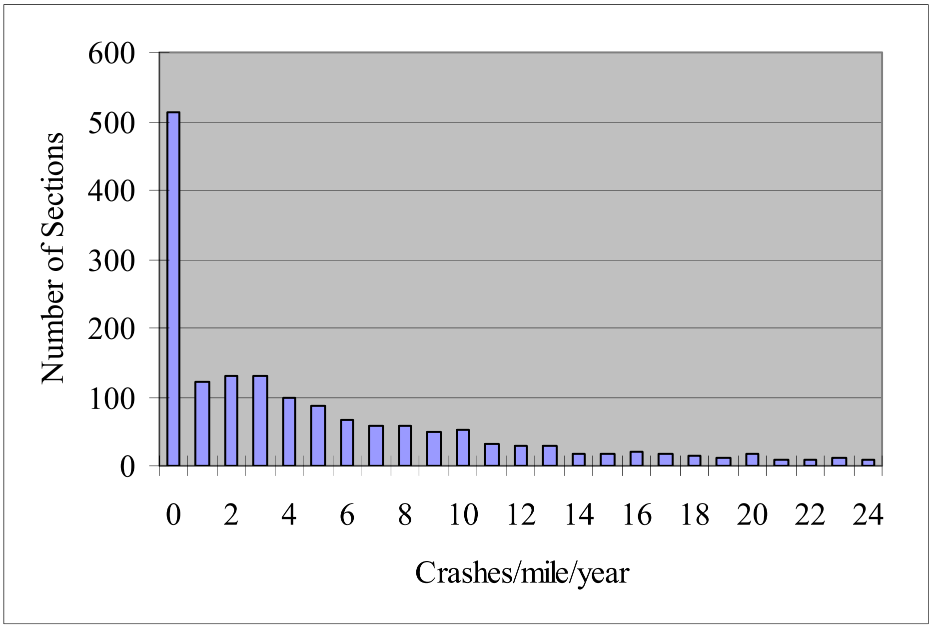


Figure 5.1 Distribution of Average Number of Crashes per Mile per Year

Based on the frequency distribution and cumulative probability for average number of crashes per mile per year, the observed mean and variance were calculated in order to perform the distribution fitting procedure. The mean or expected value of the discrete random variable x, denoted as E(x), and the variance of x, denoted as V(x), are calculated as

$$E(x) = \sum_{x} x \times f(x) \tag{5-1}$$

$$V(x) = \sum_{x} (x - E(x))^2 \times f(x) \tag{5-2}$$

Where,

$f(x)$ = the probability of each random variable x.

Table 5.1 summarizes the procedure to get the observed mean and variance for average number of crashes per mile per year when considering crash data from all the TWLTL sections together.

Table 5.1 Mean and Variance of Average Number of Crashes per Mile per Year

x	Frequency	f(x)	Cumulative percent	x*f(x)	(x-E(x))2
0	512	16.0	16.0	0.00	3.69
1	122	3.8	19.8	0.04	0.55
2	129	4.0	23.8	0.08	0.32
3	130	4.1	27.9	0.12	0.13
4	98	3.1	30.9	0.12	0.02
5	86	2.7	33.6	0.13	0.00
6	66	2.1	35.7	0.12	0.03
7	57	1.8	37.4	0.12	0.09
8	57	1.8	39.2	0.14	0.18
9	50	1.6	40.8	0.14	0.27
10	51	1.6	42.4	0.16	0.43
11	33	1.0	43.4	0.11	0.39
12	30	0.9	44.3	0.11	0.48
13	29	0.9	45.2	0.12	0.61
14	16	0.5	45.7	0.07	0.42
15	17	0.5	46.3	0.08	0.55
16	19	0.6	46.8	0.09	0.74
17	18	0.6	47.4	0.10	0.83
18	14	0.4	47.8	0.08	0.76
19	13	0.4	48.3	0.08	0.82
20	16	0.5	48.8	0.10	1.15
21	10	0.3	49.1	0.07	0.82
22	10	0.3	49.4	0.07	0.92
23	11	0.3	49.7	0.08	1.14
24	9	0.3	50.0	0.07	1.03
	1603		E(x)=2.40		V(x)=16.39

With the use of the observed mean and variance, the Poisson, Negative Binomial, and Log-normal distribution were fitted to the crash data distribution for the average number of crashes per mile per year. Table 5.2 demonstrates the Chi-Square test for Poisson distribution fitted for the variable considered. Thus:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-2.40}2.40^{x}}{x!}, \qquad x=0, 1, 2, \ldots\ldots, \qquad (5\text{-}3)$$

Where,

λ is the mean value of the observed data from Table 5.1, λ = 2.40.

Table 5.2 Chi-square Test for Poisson Distribution Fitted for Average Number of Crashes per Mile per Year

x	f(i)	f(x)-Poisson	f(i)-f(x)	$(f(i)-f(x))^2$	$(f(i)-f(x))^2/f(x)$
0	31.9	0.09072	0.22868	0.05230	0.576468
1	7.6	0.21772	-0.14162	0.02006	0.092113
2	8.0	0.26127	-0.18079	0.03269	0.125107
3	8.1	0.20901	-0.12792	0.01636	0.078284
4	6.1	0.12541	-0.06427	0.00413	0.032941
5	5.4	0.06020	-0.00655	0.00004	0.000712
6	4.1	0.02408	0.01709	0.00029	0.012136
7	3.6	0.00826	0.02730	0.00075	0.090297
8	3.6	0.00248	0.03308	0.00109	0.441889
9	3.1	0.00066	0.03053	0.00093	1.411409
10	3.2	0.00016	0.03166	0.00100	6.322554
11	2.1	0.00003	0.02055	0.00042	12.21348
12	1.9	0.00001	0.01871	0.00035	50.60148
13	1.8	0.00000	0.01809	0.00033	256.2764
14	1.0	0.00000	0.00998	0.00010	455.105
15	1.1	0.00000	0.01061	0.00011	3211.188
16	1.2	0.00000	0.01185	0.00014	26741.53
17	1.1	0.00000	0.01123	0.00013	170005.1
18	0.9	0.00000	0.00873	0.00008	771319.7
19	0.8	0.00000	0.00811	0.00007	5265110
20	1.0	0.00000	0.00998	0.00010	66462926
21	0.6	0.00000	0.00624	0.00004	2.27E+08
22	0.6	0.00000	0.00624	0.00004	2.08E+09
23	0.7	0.00000	0.00686	0.00005	2.41E+10
24	0.6	0.00000	0.00561	0.00003	1.62E+11
					χ_0^2 =1.88E+11

The Chi-square value obtained for the Poisson distribution fitted for the average number of crashes per mile per year was calculated with:

$$\chi_0^2 = \sum_{i=1}^{k} (f(i) - f(x))^2 / f(x) \tag{5-4}$$

The estimated value of χ_0^2 is 1.88E+11. This Chi-Square value is higher than the Chi-Square table value $\chi^2_{\alpha,k-p-1}$ = 35.1725 (α = 0.05, k = 25, p = 1). This Chi-square test result indicates that the hypothesis, which stated that the distribution of the average number of crashes per mile per year is the hypothesized Poisson distribution, is rejected.

Table 5.3 explains how the Chi-Square test is processed for the Negative Binomial distribution fitted for the average number of crashes per mile per year. As mentioned before, the Negative Binomial distribution has two parameters, mean E(x) and variance V(x). The probability function of x is:

$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad x = r, r+1, \ldots\ldots \tag{5-5}$$

Changing the scale in the previous equation by replacing x by x + r,

$$f(x) = \binom{x+r-1}{x} p^r (1-p)^{x}, \qquad x = 0, 1, 2, \ldots\ldots \tag{5-6}$$

In this case, from the observed mean E(x) and variance V(x), parameter *p* can be obtained from E(x)/V(x) and parameter r acquired from E(x)/(1/p-1)

In Table 5.3, *p* = 2.40/16.39 = 0.15, r = E(x)/(1/p-1)= 1.

$$f(x) = \frac{(1+x-1)!}{x!}(0.15)^1(1-0.15)^x$$

From Table 5.3, the Chi-Square calculation value estimated from the Negative Binomial distribution fitted with observed average number of crashes is χ_0^2 = 0.2578. This value is smaller than the Chi-Square table value $\chi^2_{\alpha,k-p-1}$ = 33.9244 (α = 0.05, k = 25, p=2), which indicates that the hypothesis that the distribution of the average number of crashes per mile per year is the hypothesized Negative Binomial distribution will not be rejected.

Table 5.3 Chi-square Test for Negative Binomial Distribution Fitted for Average Number of Crashes per Mile per Year

x	f(i)	f(x)-Negative Binomial	f(i)-f(x)	$(f(i)-f(x))^2$	$(f(i)-f(x))^2/f(x)$
0	31.9	0.15	17.30%	0.029919	0.2043
1	7.6	0.12	-4.89%	0.002389	0.0191
2	8.0	0.11	-2.62%	0.000687	0.0064
3	8.1	0.09	-1.00%	9.93E-05	0.0011
4	6.1	0.08	-1.66%	0.000275	0.0035
5	5.4	0.07	-1.27%	0.000161	0.0024
6	4.1	0.06	-1.55%	0.000239	0.0042
7	3.6	0.05	-1.28%	0.000163	0.0034
8	3.6	0.04	-.57%	3.25E-05	0.0008
9	3.1	0.04	-.40%	1.62E-05	0.0005
10	3.2	0.03	.18%	3.07E-06	0.0001
11	2.1	0.03	-.51%	2.57E-05	0.0010
12	1.9	0.02	-.32%	1.02E-05	0.0005
13	1.8	0.02	-.06%	3.65E-07	0.0000
14	1.0	0.02	-.60%	3.57E-05	0.0022
15	1.1	0.01	-.30%	9.1E-06	0.0007
16	1.2	0.01	.02%	5.12E-08	0.0000
17	1.1	0.01	.13%	1.7E-06	0.0002
18	0.9	0.01	.03%	6.9E-08	0.0000
19	0.8	0.01	.09%	7.73E-07	0.0001
20	1.0	0.01	.38%	1.45E-05	0.0024
21	0.6	0.01	.10%	9.41E-07	0.0002
22	0.6	0.00	.17%	3.03E-06	0.0007
23	0.7	0.00	.30%	9.14E-06	0.0024
24	0.6	0.00	.23%	5.47E-06	0.0017
					χ_0^2 =0.2578

Table 5.4 explains how the Chi-Square test is processed for the Log-normal distribution fitted for the average number of crashes per mile per year. As mentioned before, the

Log-normal distribution has two parameters, mean E(x) and variance V(x). The probability function of x is:

$$f(x)=\frac{1}{x\sigma_{\gamma}\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\ln x-\mu_{\gamma}}{\sigma_{\gamma}}\right]^{2}} \tag{5-7}$$

Table 5.4 Chi-square Test for Log-normal Distribution Fitted for Average Number of Crashes per Mile per Year

X	f(i)	f(x)-Log	f(i)-f(x)	$(f(i)-f(x))^2$	$(f(i)-f(x))^2/f(x)$
0	31.9	0	0.319401	0.102017	*
1	7.6	0.338126	-0.26202	0.068654	0.203042
2	8.0	0.156943	-0.07647	0.005847	0.037258
3	8.1	0.084905	-0.00381	1.45E-05	0.000171
4	6.1	0.05099	0.010145	0.000103	0.002019
5	5.4	0.032911	0.020739	0.00043	0.013068
6	4.1	0.02239	0.018782	0.000353	0.015756
7	3.6	0.015859	0.0197	0.000388	0.024471
8	3.6	0.011596	0.023962	0.000574	0.049515
9	3.1	0.008702	0.022489	0.000506	0.058121
10	3.2	0.006673	0.025143	0.000632	0.094739
11	2.1	0.005211	0.015376	0.000236	0.045373
12	1.9	0.004133	0.014582	0.000213	0.051447
13	1.8	0.003323	0.014768	0.000218	0.065627
14	1.0	0.002704	0.007277	5.3E-05	0.019584
15	1.1	0.002224	0.008381	7.02E-05	0.031591
16	1.2	0.001846	0.010007	0.0001	0.054246
17	1.1	0.001545	0.009684	9.38E-05	0.060675
18	0.9	0.001304	0.00743	5.52E-05	0.04234
19	0.8	0.001108	0.007002	4.9E-05	0.044267
20	1.0	0.000947	0.009034	8.16E-05	0.08619
21	0.6	0.000814	0.005424	2.94E-05	0.036126
22	0.6	0.000704	0.005534	3.06E-05	0.0435
23	0.7	0.000612	0.00625	3.91E-05	0.063859
24	0.6	0.000534	0.00508	2.58E-05	0.048334
					χ_0^2 ==1.191321

Chi-Square calculation value estimated from the Log-normal distribution fitted with observed average number of crashes is $\chi_0^2 =$ 1.1918. This value is smaller than the Chi-Square table value $\chi^2_{\alpha,k-p-1}$ = 33.9244 (α = 0.05, k = 25, p=2), which indicates that the hypothesis that the distribution of the average number of crashes per mile per year is the hypothesized Log-normal distribution will not be rejected.

Finally, since the Negative Binomial and Log-normal distributions were not rejected by the Chi-square test ($\chi_0^2 < \chi^2_{\alpha,k-p-1}$), the distribution with smaller Chi-square calculation value was selected as the fitted distribution. Therefore, it could be concluded that the Negative Binomial distribution is better to fit the distribution of average number of crashes per mile per year from the Chi-square test comparison. Figures 5.2 through 5.4 present the graphs of frequency distributions, which illustrate the same outcome for distribution fitting of the average number of crashes per mile per year. Therefore, the Negative Binomial distribution was selected as the distribution to fit the average number of crashes per mile per year.

The distribution fitting equation for the selected Negative Binomial distribution is as follows:

$$f(x) = \binom{x+r-1}{x} p^r (1-p)^x, \tag{5-8}$$

where:

p = 0.15,

r = 1

Then,

$$f(x) = \frac{(1+x-1)!}{x!}(0.15)^1(1-0.15)^x \tag{5-9}$$

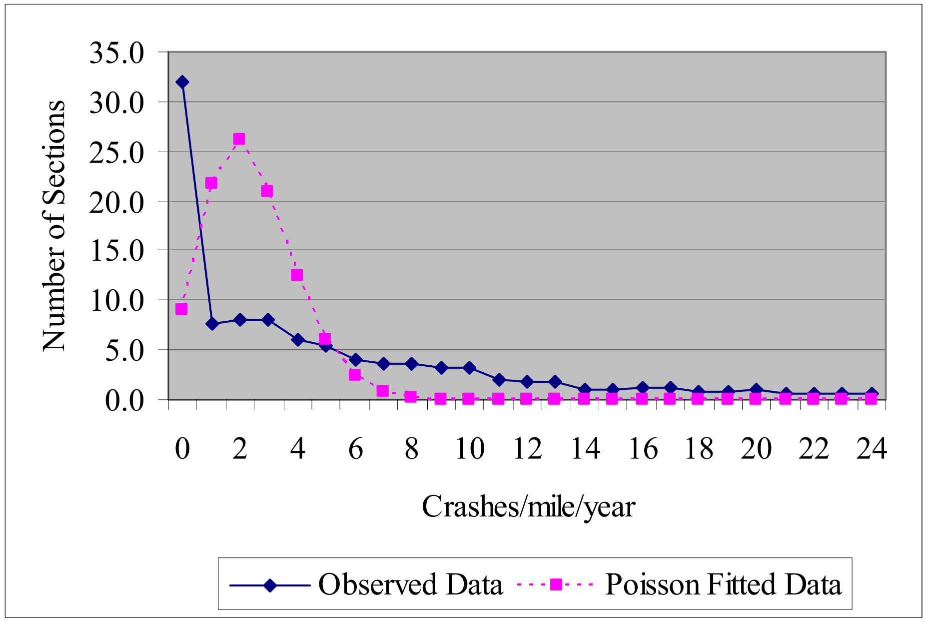
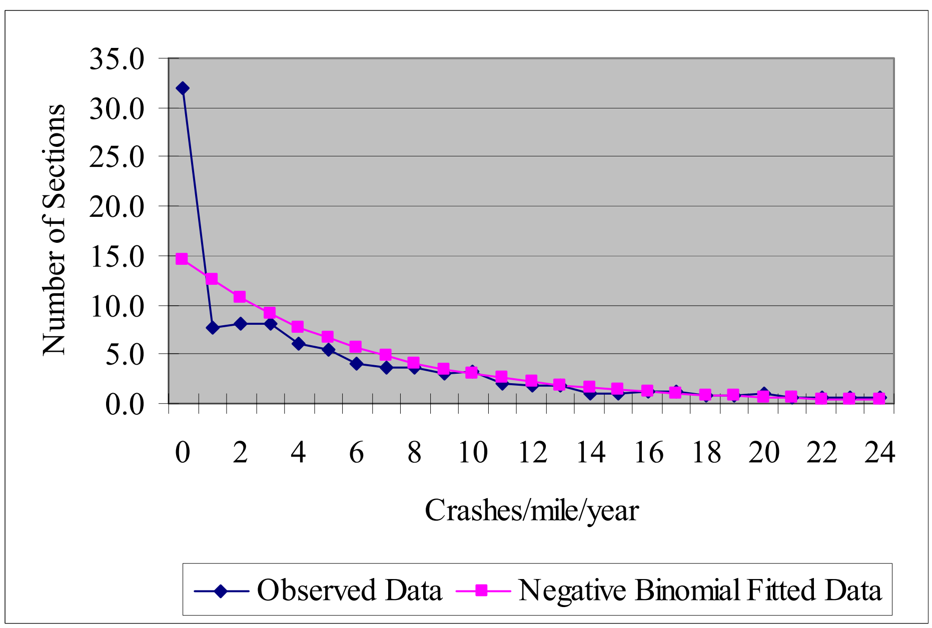


Figure 5.2 Poisson Distribution of Average Number of Crashes per Mile Per Year



\Figure 5.3 Negative Binomial Distribution of Average Number of Crashes per Mile per Year

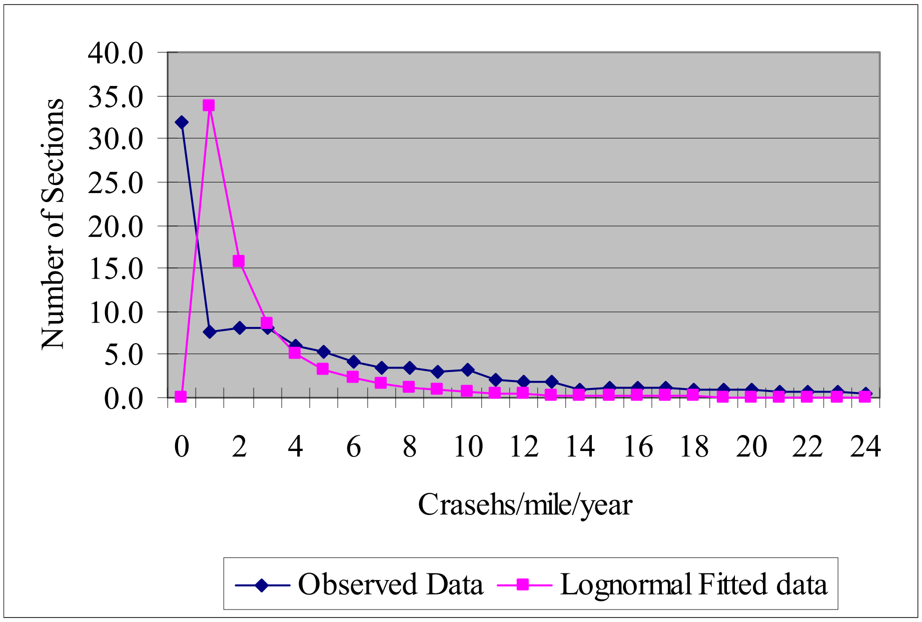


Figure 5.4 Log-normal Distribution of Average Number of Crashes per Mile per Year

5.2 Statistical Modeling for Average Number of Crashes per Mile per Year

A statistical prediction model was developed to identify the factors that are influential in the safety experience of two-way left-turn lanes. Crash data from all TWLTL sections were used together for the development of the model. The Negative Binomial regression model was developed with the average number of crashes per mile per year as the dependent variable and with four independent variables: AADT, access density, posted speed, and number of lanes. The Negative Binomial regression model format is:

$$\text{Ln(crashes/mile/year)} = \beta_0 + \beta_1 \text{ Access Density} + \beta_2 \text{ AADT} + \beta_3 \text{ Posted Speed} + \beta_4 \text{ Number of Lanes} \quad (5\text{-}10)$$

The following section presents detailed description and analyses of the dependent and independent variables as well as the detailed modeling process for average number of crashes. The parameter estimation was performed using the statistical package software program SPSS. Mean deviance and Pearson's Chi-square ratio were adopted as the criteria to test over-dispersion of the crash data. Deviance, Pearson's Chi-square, Pearson's R-square and likelihood ratio index were adopted to evaluate the goodness-of-fit of the developed model.

5.2.1 Statistical Analysis of the Variables

The dependent variable for the model was the average number of crashes per mile per year. The summary descriptive statistics for the dependent variable is presented in Table 5.5. In regard to the selection of independent variables, it was based on the available data and engineering judgment. The task was carried out through the database building process for the modeling. After the modeling database was built, totally four variables were available. The following subsections provide detailed description and analyses for the four variables used in the modeling process.

Table 5.5 Descriptive Statistics for Average Number of Crashes per Mile per Year

Statistics	Average Crashes
Mean	6.55
Standard Deviation	10.02
Minimum	0
Maximum	110

5.2.1.1 Traffic Volume (AADT)

Traffic volume is the most significant factor contributing to crash occurrence. The change of traffic volume at a TWLTL location imposes multiple effects on traffic operations and safety at the section. In this project, most AADT values were obtained from the FDOT crash database. If the AADT was not available in the crash database, the missing AADT were obtained from the Florida Traffic Information disk prepared by FDOT. The descriptive statistics of AADT are provided in Table 5.6.

Table 5.6 Descriptive Statistics for AADT

Statistics	AADT
Mean	20110
Standard Deviation	11534
Minimum	1800
Maximum	78722

The maximum value of AADT was much higher than the median value. These higher values of AADT might be a result of particular characteristics of TWLTL sections. For this reason, sections with extremely high volume were taken out from the data set during the modeling process. AADT values are higher than values for other variables, therefore, the AADT was calculated as 1/10,000 vpd when estimating its parameter during the modeling process to avoid having a much smaller parameter value than the parameters for the other variables.

5.2.1.2 Access Density

Access density is a significant factor to be considered in the safety analysis of TWLTL. The information regarding this specific variable was determined from video logs of all state roads in Florida due to the fact that FDOT crash database did not have information for access density. FDOT provided the video logs for this project. Each side of the roadway was recorded separately in the video logs. Therefore, video logs were reviewed two times for each TWLTL section in order to count the number of driveways in both sides of the road. The number of driveways for each TWLTL section was the result of adding the number of driveways for each side of the road considered. Table 5.7 presents the descriptive statistics for this variable.

Table 5.7 Descriptive Statistics for Access Density

Statistics	Access Density
Mean	32.86
Standard Deviation	25.26
Minimum	0.5
Maximum	149

The values for access density per mile presented a significant difference between the maximum value and the median value. This difference might be a result of particular characteristics of some sections, such as short length of the section. In the modeling process, the sections with extremely high values of access density per mile were taken out from the data set.

5.2.1.3 Number of Lanes

The number of lanes is one of the most important geometric factors in explaining crash occurrence. For the specific case of TWLTL sections, the number of lanes might affect the safety performance of the location. It is different to make a left turn crossing three lanes of through traffic of a road in order to access an area, than crossing a road with two lanes of through traffic. Table 5.8 presents the descriptive statistics for the number of lanes in both directions for the TWLTL sections.

Table 5.8 Descriptive Statistics for Number of Lanes

Number of Lanes	Frequency	Percentage
2	722	42.8
4	833	49.3
6	133	7.9

5.2.1.4 Posted Speed

Posted speed is an important traffic speed control factor for traffic safety analysis. Usually, it is believed that crashes are more likely to occur at higher speed, which actually is not well documented. However, common engineering knowledge is that high speed more likely results in severe crashes. From another point of view, drivers tend to travel at speeds in which they feel comfortable given the prevailing conditions. Therefore, lower posted speed more likely promotes speed differential that is generally more closely associated with crashes. The variable posted speed was transformed from continuous to discrete value because the results of the model would be tabulated to facilitate the use of the model. Based on the distribution, posted speed was divided into two levels, lower speed with sections with posted speed < 45 mph and higher speed with sections with posted speed ≥ 45 mph. Table 5.9 presents the descriptive statistics for the variable posted speed.

Table 5.9 Descriptive Statistics for Posted Speed

Posted Speed	Frequency	Percentage
25	18	1.1
30	124	7.3
35	410	24.3
40	296	17.5
45	583	34.5
50	77	4.6
55	141	8.4
60	39	2.4

5.2.2 The Model for Average Number of Crashes

The effect of TWLTL on roadway crash frequency was studied using statistical models. Even though, distribution fitting indicated that Negative Binomial better fits the data, the Poisson regression was performed as the initial step during the modeling process. Initial Poisson regression results provided the basis to test whether the crash data were over-dispersed or not. The regression-based test for over-dispersion can determine the selection between Poisson regression model and Negative Binomial regression model. Two statistics were used to assess the over-dispersion: mean deviance and Pearson's Chi-square ratio. Generally, the mean deviance and the Pearson's Chi-square ratio should be close to one or within the range between 0.8 and 1.2 in order to consider the Poisson model appropriate to fit the data. If the mean deviance ant the Pearson’s Chi-square ratio values exceed one, the data are considered to display extra variation or over-dispersion relative to the Poisson model. If the values are less than one, the data are said to display under-dispersion relative to the Poisson model. The mean deviance for the initial Poisson model is 6.721, and the Pearson's Chi-square ratio is 6.140, which indicate that extra variation exists in the data. It also shows that the Negative Binomial regression model would be more appropriate than the Poisson regression model to estimate model coefficients.

The Negative Binomial regression was performed as an alternative to Poisson model, and the mean deviance and Pearson's Chi-square ratio were calculated again. The mean deviance and Pearson's Chi-square ratio for the Negative Binomial model are close to one, which indicate that this model was an appropriate choice. Therefore, the Negative Binomial regression model was adopted to develop the predictive model for the crash occurrence at TWLTL sections. A five percent significance level was assumed for the parameters estimated in the model. The models obtained are presented below.

$$\begin{aligned} Ln(\mu_i) = {} & 0.0193 + 0.0082 \text{ Access Density} + 0.5253 \text{ AADT} \\ & - 0.3039 \text{ Posted Speed} + 0.1124 \text{ Number of Lanes} \end{aligned} \tag{5-11}$$

Wheee:

μ_i = average number of crashes/mile/year at the TWLTL section;

Access Density = number of driveways considering both sides of the roadway, driveways/mile;

AADT = average annual daily traffic at the TWLTL section, vpd;

Posted Speed = value equal to 1 is posted speed at the road is greater or equal to 45 mph, and 0 if posted speed at the road is lower than 45 mph;

Number of Lanes = number of lanes considering both sides of the road.

The statistical results for the parameters estimated during the Negative Binomial regression modeling are presented in Table 5.10. Explanations for the contents of Table 5.10 are listed on Table 5.11.

Table 5.10 Estimated Parameters of the Negative Binomial Regression Model

Variable	Coefficient Estimate	Standard Error	Relative Effect	Chi-Square	Pr>Chi-sq
Intercept	0.0193	0.1045		0.0343	0.8532
Access Density	0.0082	0.0013	1.0082	39.7870	<0.0001
Average AADT	0.5253	0.0389	1.6910	182.3541	<0.0001
Posted Speed	-0.3039	0.0633	0.7379	23.0491	<0.0001
Number of Lanes	0.1124	0.0348	1.1190	10.4622	0.0012

Table 5.11 Explanations of Contents of the Results

Column	Explanation
Coefficient Estimate	Estimated parameters.
Standard Error	Estimated standard deviation associated with each parameter.
Relative Effect	Exponent of the estimated parameter of the variable.
Chi-square	Chi-square test statistic for testing that the parameter is 0. This was computed as the square of the ratio of the parameter estimate divided by its standard error.
Pr>Chi-Sq	The probability of obtaining a Chi-square statistic greater than that observed given that the true parameter is zero

The significance of the variables was under 0.05 for each one based on results presented on Table 5.10. This demonstrates that the variables adopted in the model AADT, access density, posted speed and number of lanes, influence crash occurrences at TWLTL sections. Among these variables, AADT, access density and number of lanes have a positive sign. These findings suggest that any increase of these three variables increase the likelihood of crash occurrence. On the other hand, the sign for the parameter of posted speed is negative. The common engineering knowledge is that high speed more likely results in severe crashes. However, drivers tend to travel at speeds in which they feel comfortable given the prevailing conditions. Therefore, lower posted speed more likely promotes speed differential that is generally more closely associated with crashes. Also, lower speed limits most likely indicate that there is a larger speed variance at the TWLTL location.

Once the parameters of the crash predictive model were determined, the average number of crashes per mile per year can be estimated by replacing the regression parameters, β_0, β_1, β_2, … , β_q, with the estimated values, and the variables X_{i1}, X_{i2}, …, X_{iq}, with the corresponding values of the characteristics of the TWLTL sections. The estimated number of crashes obtained from the models would be plotted against AADT values in different curves. From these curves, critical AADT values could be determined for the desired percentile value of average number of crashes for specific characteristics of the site, including access density, number of lanes and posted speed. TWLTL sections that have their AADT value above the critical AADT values for their specific characteristics based on the percentile value selected would be identified as locations that need further study in order to determine if improvements are needed.

Figure 5.5 shows the model curve for a four lane section for the higher speed category. This figure shows that as traffic volume increases, the number of crashes also increases. The figure also presents three curves for different access densities, from these curves it could be observed that as access density increases, crash frequency also increases. The model curves for other TWLTL sections with different characteristics are presented in Appendix A.

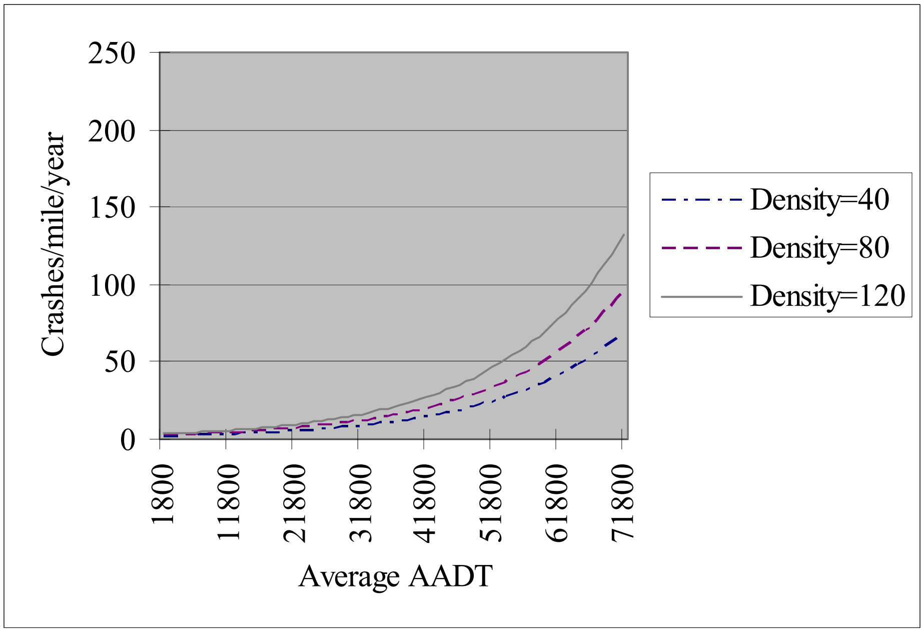


Figure 5.5 Model Curve for Higher Speed Category for Four-lane Sections

5.2.3 Goodness-of-fit of the Model

In order to assess the goodness-of-fit of the model, four statistics including deviance, Pearson's Chi-square, Pearson's R-square and likelihood ratio index, were adopted. Table 5.12 presents the four statistics for the Negative Binomial model considering average number of crashes per mile per year as dependent variable. As show on Table 5.12, the mean deviance and Pearson's Chi-square ratio are close to one indicating that the developed model has a satisfactory capability in fitting the crash data. Pearson's R- square value is around 31% and the likelihood ratio index value is around 45%, indicating that the developed model has a satisfactory capability in explaining the variation of the data.

Table 5.12 Criteria for Assessing the Goodness-of-Fit

Item	Value
Number of Observations	1688
Number of Variables in Model	4
Number of Parameters in Model	4
Degree of Freedom	1684
Log-likelihood Function	-4513.050
Restricted Log likelihood	-8273.891
Deviance	2323.92
Deviance/Degree of Freedom	1.38
Pearson Chi-square	2237.23
Pearson Chi-square/Degree of Freedom	1.33
Pearson R-square	31.81%
Likelihood Ratio Index	45.45%

In addition to statistical justification, the model should also satisfy engineering judgment. This can be assessed by examining the relative effect of each variable. For example, the relative effect of AADT is 1.69, which means that the average number of crashes would increase by 69% if the AADT increase in 10000 vpd given all other variables constant. The relative effect of access density is 1.01, which means that the average number of crashes would increase 1% for every access point on the road. For the number of lanes, the relative effect is 1.12, which means that the average number of crashes would increase 12% for every one lane increase. However, TWLTL sections with higher posted speed (≥ 45 mph) would have 27% fewer crashes than similar sections with lower posted speed (< 45 mph). Even though this result may not be as expected, the analysis of the data indicated this trend, where locations with higher posted speed experienced lower number of crashes than locations with lower posted speed. As mentioned previously, the main reason for this result might be the speed differential for the lower posted speed sections. Another reason may be that the FDOT's policy for using caution in allowing TWLTL on higher speed facilities resulted in the sampled higher speed locations to be skewed from the expected high number of crashes.

5.3 Linear Regression for Crash Rates

Crash rates for segments are defined as crashes per million vehicle miles traveled. Crash rates could be used as a criterion for identifying high crash locations. For this project, crash rates considering all crashes were calculated using Equation 3-2:

$$\text{CRS} = \frac{1{,}000{,}000 \times A}{365 \times T \times V \times L}$$

Where,

CRS = Crash Rate for the Segment,

A = number of police reported crashes,

T = time frame of the analysis (years),

V = average annual daily traffic (AADT),

L = length of the segment (miles).

Once crash rates were calculated, conventional linear regression was used to determine a relationship between crash rates as dependent variable and the independent variables: access density, number of lanes and posted speed. The characteristics and details of the independent variables considered were presented in a previous section of this chapter. AADT was not considered as an independent variable since it was already included in the estimation of crash rates. Poisson regression model and Negative Binomial regression model did not give good results when considering crash rate as dependent variable. The model format used for linear regression was:

$$\begin{aligned}\text{Crash Rate} = \beta_0 + \beta_1 \text{ Access Density} + \beta_2 \text{ Number of Lanes} \\ + \beta_3 \text{ Posted Speed} \qquad (5\text{-}11)\end{aligned}$$

This relationship obtained from linear regression would give an overview of how the independent variables influence crash rates. The effect of the variable considered could be explored without the influence of AADT, since it was not included as an independent variable. Several models were tested, including different forms of the variables and finally three models were selected. The first model considers all crashes together in one group and access density, posted speed, and number of lanes as independent variables. The final equation for this model was:

$$\text{Crash Rate} = 0.836 + 0.006\ \text{Access Density} + 0.108\ \text{Number of Lanes} - 0.015\ \text{Posted Speed} \quad (5\text{-}12)$$

The R-square for this regression model is small, which might indicate that linear regression may not be the best choice to explain crash data. Table 5.13 presents the statistics for the model. In reference to the independent variables, the regression results indicated that the variables considered are significant in the analysis. Access density and number of lanes have a positive sign, which means that as these variables increase, crash rates will increase. Table 5.14 shows the statistics for the parameters of the variables.

Table 5.13 Statistics for the Model with All Crashes

Category	Dependent Variable	Independent Variables	R	R Square	Standard Error
All	Crash Rate	Access Density Posted Speed Number of Lanes	0.332	0.110	0.86

Table 5.14 Estimated Parameters for the Model with All Crashes

Category	Independent Variables	Coefficient	Standard Error	t-stat.	Sig.
All crashes	(Constant)	0.836	0.151	5.544	<0.001
	Access Density	0.006	0.001	6.751	<0.001
	Number of Lanes	0.108	0.019	5.830	<0.001
	Posted Speed	-0.015	0.003	-5.040	<0.001

For the other two models considered, crash data was divided into two groups based on posted speed. The two groups are: lower speed, which considers TWLTL sections with posted speed < 45 mph, and higher speed, with TWLTL sections with posted speed ≥ 45 mph. Since posted speed was used to divide the data, only access density and number of lanes were used as independent variables. The models equations were:

For the lower speed group:

Crash Rate = 0.346 + 0.004 Access Density + 0.115 Number of Lanes (5-13)

For the higher speed group:

Crash Rate = 0.065 + 0.012 Access Density + 0.081 Number of Lanes (5-14)

Similar to the previous model, the R-square for these two regression models were small, which indicates that linear regression might not be the best choice to explain crash data. The statistics for these models are presented in Table 5.15. In reference to the independent variables, access density and number of lanes were significant in the analysis and have a positive sign, which means that crash rates will increase as these variables increase. Table 5.16 shows the statistics for the parameters of the variables of these models.

Table 5.15 Statistics for the Models considering Speed Levels

Category	Dependent Variable	Independent Variables	R	R Square	Standard Error
Posted Speed < 45 mph	Crash Rate	Access Density Number of Lanes	0.198	0.039	0.94
Posted Speed ≥ 45 mph	Crash Rate	Access Density Number of Lanes	0.390	0.152	0.77

Table 5.16 Estimated Parameters for the Models considering Speed Levels

Category	Independent Variables	Coefficient	Standard Error	t-stat.	Sig.

Posted Speed < 45 mph	(Constant)	0.346	0.109	3.167	0.002
	Access Density	0.004	0.001	3.274	0.001
	Number of Lanes	0.115	0.031	3.758	<0.001
Posted Speed ≥ 45 mph	(Constant)	0.065	0.067	0.970	0.332
	Access Density	0.012	0.001	8.272	<0.001
	Number of Lanes	0.081	0.023	3.503	<0.001

For linear regression, it is desirable to maximize the value of R-square in order to better explain the data that is being modeled. Therefore, the small R-squares for the three regression models indicate that linear regression might not be the best choice to explain crash data, which are non-negative, random, discrete and sporadic in nature. In reference to the independent variables, regression results indicated that the variables considered are significant in the analysis. However, these parameters might be over estimated when determined with linear regression due to the nature of the data. Nonetheless, linear regression results show that the variables considered have a significant effect on crash rates. Moreover, access density and number of lanes have a positive sign, which means that as these variables increase, crash rates will increase.

CHAPTER 6. MODELS FOR TWLTL SECTION RANKING AND CASE STUDY

This chapter is intended to present the methodology developed in the research to identify critical TWLTL sections for further improvements and a case study. The following sections explain in detail the different steps of the methodology and how to apply it to identify groups of locations that need further study to determine if improvements are necessaries at those locations. When applying the methodology, only the final steps are required to determine the list of critical locations. The case study would illustrate the application of the methodology.

6.1 Overview

The process to develop a methodology to identify locations that need further study to determine if improvements are necessary consists of several steps. The first step of the process includes plotting the crash data to determine the distribution of the actual data for six different groups of locations. The crash data was divided into six different groups based on posted speed and number of lanes in order to facilitate the application of the methodology. The second step consists of the distribution fitting of the crash data. Distribution fitting is used to determine the best fitted distribution for the crash data in order to decide which statistical distribution better represents the actual data. The third step refers to the determination of percentile values for the data according to distribution fitting. A desired percentile value for the number of crashes is set as a threshold. The average numbers of crashes for the percentile values are determined from distribution fitting of the original crash database. The fourth step is the estimation of the critical AADT based on percentile values and modeling results. The average number of crashes per mile per year could be determined for sections with the same characteristics once the parameters of the crash predictive model are estimated. These estimated number of crashes obtained from the model is plotted in different curves. With the percentile value of the average number of crashes and the curves plotted from the models, the critical values for the variables considered in the research could be obtained. The fifth step is the calculation of tables of the critical average AADT values for each one of the six groups considered. These AADT are based on the values of the other variables considered: access density number of lanes and posted speed. The sixth step is the generation of a list of locations identified as critical according to the selected percentile value and the critical average AADT based on access density, posted speed, and number of lanes.

When applying the methodology to identify a group of critical locations, only the final steps are necessary. These steps include the used of the tables of critical average AADT values for each one of the six groups considered and the generation of the list of locations according to the selected percentile value. Details of the methodology application are presented during the case study section.

6.2 Plotting Actual Crash Data

This step involves the plotting of the frequency distribution of the actual crash data. The crash data was divided in six different categories or groups based on posted speed and number of lanes in order to facilitate the application of the methodology. These six categories are presented in Table 6.1. For these categories, posted speed and number of lanes are constant within each category. The category of lower speed includes all TWLTL sections with posted speed < 45 mph, and the category of higher speed includes locations with posted speed ≥ 45 mph. For number of lanes, three levels are used: two-lane sections, four-lane sections and six-lane sections. As mentioned previously, the number of lanes includes the lanes in each direction of the road. . As an example, two-lane sections indicate one lane in each direction of traffic on the road. Figure 6.1 presents the frequency distribution for the group of higher speed and two lanes sections. Appendix B presents all the frequency distributions for the different categories considered.

Table 6.1 Description of Categories for Analysis

Category	Description
1	Higher speed & Two-lane Sections
2	Higher speed & Four-lane Sections
3	Higher speed & Six-lane Sections
4	Lower speed & Two-lane Sections
5	Lower speed & Four-lane Sections
6	Lower speed & Six-lane Sections

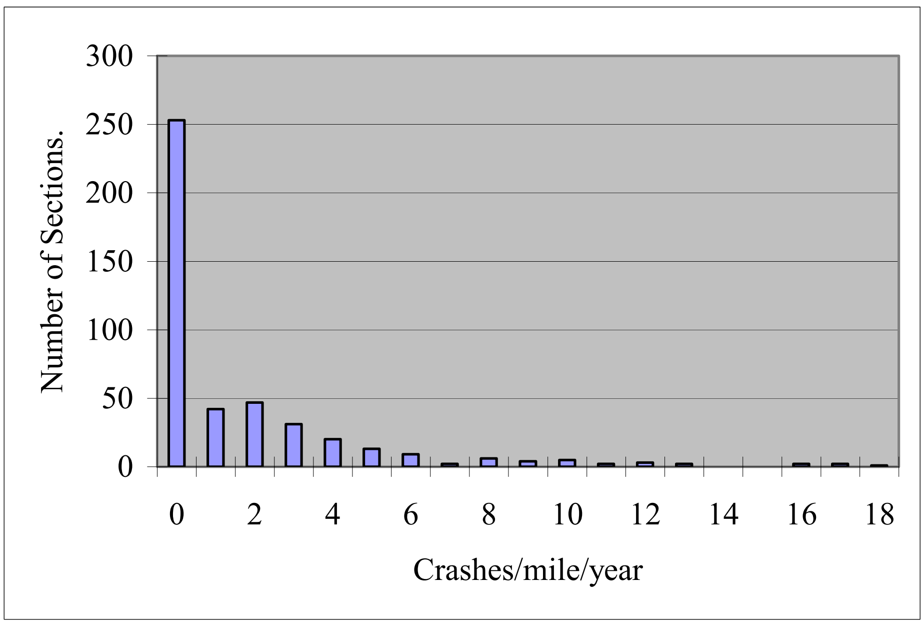


Figure 6.1 Frequency Distribution for a Two-lane Road
in the Higher Speed Category

6.3 Distribution Fitting

For this step, distribution fitting was performed for the average number of crashes per mile per year to determine the best fitted distribution for the crash data in order to decide which statistical distribution better represents the actual data for the six different categories considered. Using the observed mean and variance for each one of the categories, the crash data was fitted to the Poisson, Negative Binomial and Log-normal distributions.

Then, the Chi-square test for average number of crashes per mile per year for each distribution was determined. The detailed process for the Chi-Square test for each one of the categories considered is similar to the one presented in the previous chapter considering all data. As a consequence, the details of all the Chi-Square tests for the different categories are not included and only the Chi-square results are presented in Table 6.2. When selecting the best fitted distribution, the distribution with smaller Chi-square calculation value was selected as the fitted distribution if the three distributions were not rejected by the Chi-square test. If the Chi-square value of at least two distributions were close, the distribution with the biggest difference between the calculated Chi-square value and the critical Chi-square value from statistical tables was selected.

Table 6.2 Chi-square Test for Poisson, Negative Binomial and Lognormal Distribution Fitting

Category	Poisson		Negative Binomial		Lognormal		Distribution Selected
	Chi-square Calculation χ_0^2	Chi-square Table Value $X_{a,k-p-1}^2$	Chi-square Calculation χ_0^2	Chi-square Table Value $X_{a,k-p-1}^2$	Chi-square Calculation X_0^2	Chi-square Table Value $X_{a,k-p-1}^2$	
1	21623860.35	26.2962	0.98	24.9958	0.26	24.9958	NB
2	10010.42	33.9244	0.88	32.6705	0.27	33.9244	NB
3	2421.98	41.3372	3.35	40.1133	0.58	40.1133	Lognormal
4	7385.36	26.2962	0.20	24.9958	0.36	24.9958	NB
5	593.15	36.4151	7.71	35.1752	0.23	35.1752	Lognormal
6	279.78	35.1752	1.89	33.9244	0.80	33.9244	Lognormal

The distribution fitting equations for the different categories are presented below:

For the higher speed 2-lane sections, the Negative Binomial regression equation is:

$$f(x) = \begin{bmatrix} x + 0.39 - 1 \\ x \end{bmatrix} 0.19^{0.39} (1 - 0.19)^x, \qquad x = 0,1,2,...... \qquad (6\text{-}1)$$

For the higher speed 4-lane sections, the Negative Binomial regression equation is:

$$f(x)=\begin{bmatrix} x+1.23-1 \\ x \end{bmatrix} 0.18^{1.23}(1-0.18)^{x}, \qquad x=0,1,2,...... \qquad (6\text{-}2)$$

For the higher speed 6-lane sections, the Lognormal regression equation is:

$$f(x)=\frac{1}{x0.64\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\ln x-2.27}{0.64}\right]^{2}} \qquad x=0,1,2,...... \qquad (6\text{-}3)$$

For the lower speed 2-lane sections, the Negative Binomial regression equation is:

$$f(x)=\begin{bmatrix} x+0.83-1 \\ x \end{bmatrix} 0.22^{0.83}(1-0.22)^{x}, \qquad x=0,1,2,...... \qquad (6\text{-}4)$$

For the lower speed 4-lane sections, the Lognormal regression equation is:

$$f(x)=\frac{1}{x0.77\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\ln x-1.66}{0.77}\right]^{2}} \qquad x=0,1,2,...... \qquad (6\text{-}5)$$

For the lower speed 6-lane sections, the Lognormal regression equation is:

$$f(x)=\frac{1}{x0.64\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\ln x-2.06}{0.64}\right]^{2}} \qquad x=0,1,2,...... \qquad (6\text{-}6)$$

6.4 Percentile Values of Crashes

In this step, percentile values for average number of crashes per mile per year were determined from the fitted distribution in order to set up thresholds. The common percentile values used in transportation are the 85th and 50th percentile values. The 85th percentile value for this project is the point that indicates the average number of crashes per mile per year where 85 percent of the locations considered have this average number of crashes or a lower average value. This value would allow the identification of the top 15 percent locations, considered the top portion of the population, for further analysis. The 50th percentile value is just the middle value of the distribution. Several percentile values, including the 50th and 85th, for each crash distribution for the six categories were obtained. Figure 6.2 shows an example of the process to estimate the percentile values of average number of crashes from the fitted distribution. In this example, 50th and 85th percentile values are estimated for the category higher speed four-lane section. The figures corresponding to the estimation of the 50th and 85th percentile values for the average number of crashes from the fitted distributions for the remaining categories are presented in Appendix C.

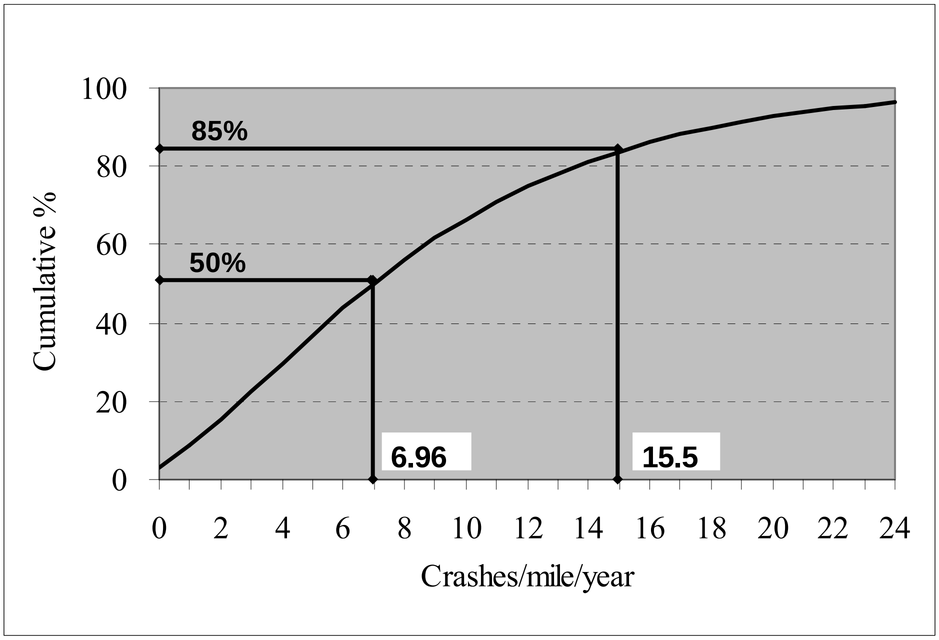


Figure 6.2 The 50% and 85% Percentile Values of the Average Number of Crashes for Higher Speed and Four-lane Sections

Table 6.3 presents the percentile values for the distribution fitting of average number of crashes for each category. With these values, linear regression was performed for the different percentile values for the categories of lower speed and higher speed in order to determine a linear relationship between number of lanes and percentile values. The final percentile values of average number of crashes used to determine the critical values for the different variables were obtained from the linear regressions. Table 6.4 shows the percentile values obtained after the linear regression procedure. Table 6.4 can be used for real applications of this methodology.

Table 6.3 Percentile Values for Distributions of Average Number of Crashes based on Observed Crash Data

Posted Speed	Number of Lanes	Average Number of Crashes per Mile per Year					
		50%	75%	80%	85%	90%	95%
Higher Speed	2	2.35	5.68	6.74	8.11	10.05	13.40
	4	6.96	12.06	13.59	15.5	18.08	22.38
	6	9.22	14.51	16.22	18.46	20.73	27.68
Lower Speed	2	1.83	4.65	5.56	6.72	8.37	11.17
	4	4.74	8.32	9.54	11.16	13.59	18.12
	6	7.37	11.66	13.03	14.86	17.53	22.36

Table 6.4 Final Percentile Values for Average Number of Crashes

Posted Speed	Number of Lanes	Average Number of Crashes per Mile per Year					
		50%	75%	80%	85%	90%	95%
Higher Speed	2	2.14	6.33	7.44	8.85	10.95	14.01
	4	6.18	10.75	12.18	14.00	16.29	21.15
	6	9.62	15.16	16.92	19.20	21.63	28.29
Lower Speed	2	1.88	4.70	5.64	6.84	8.58	11.62
	4	4.65	8.21	9.38	10.91	13.16	17.22
	6	7.42	11.72	13.12	14.98	17.74	22.82

6.5 Critical AADT Values

During this step, the percentile values obtained from crash data are used to determine the critical values of traffic volume for different characteristics of the road. The critical values are determined by plotting percentile values of average number of crashes into the model curves obtained from the predictive model. The model curves are plotted using the average number of crashes per mile per year determined with the crash prediction model for sections with the same characteristics. Figures 6.3 and 6.4 demonstrate the procedure to obtain the critical values of traffic volume for different densities when considering the 50th and 85th percentile values as thresholds. This example is for the category of higher speed and four- lane section. The curves to determine the critical volume based on 50th and 85th percentile values for all the categories are presented in Appendix D.

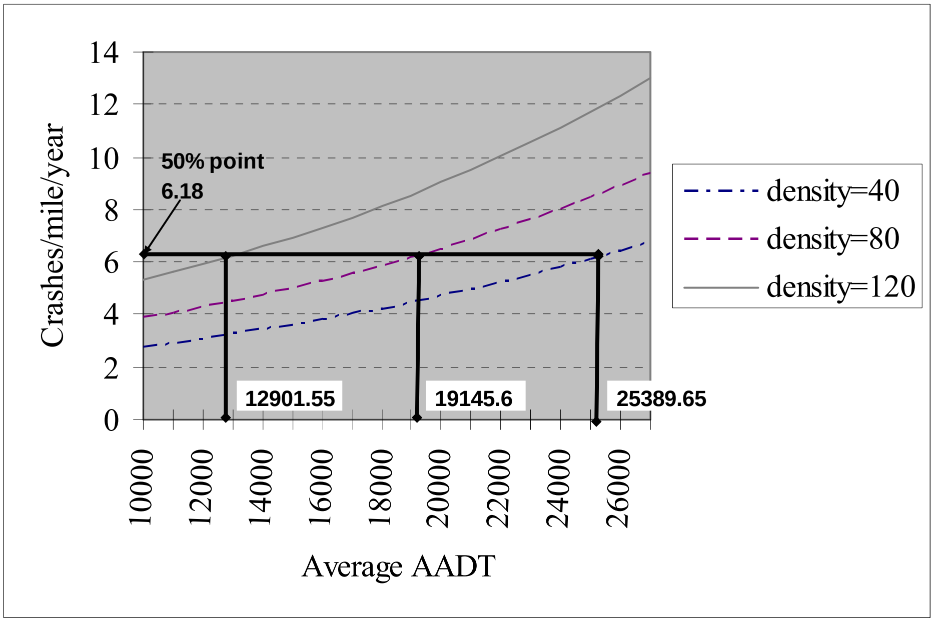


Figure 6.3 Critical Value of AADT according to 50th Percentile Value for a Four-lane Section in the Higher Speed Category

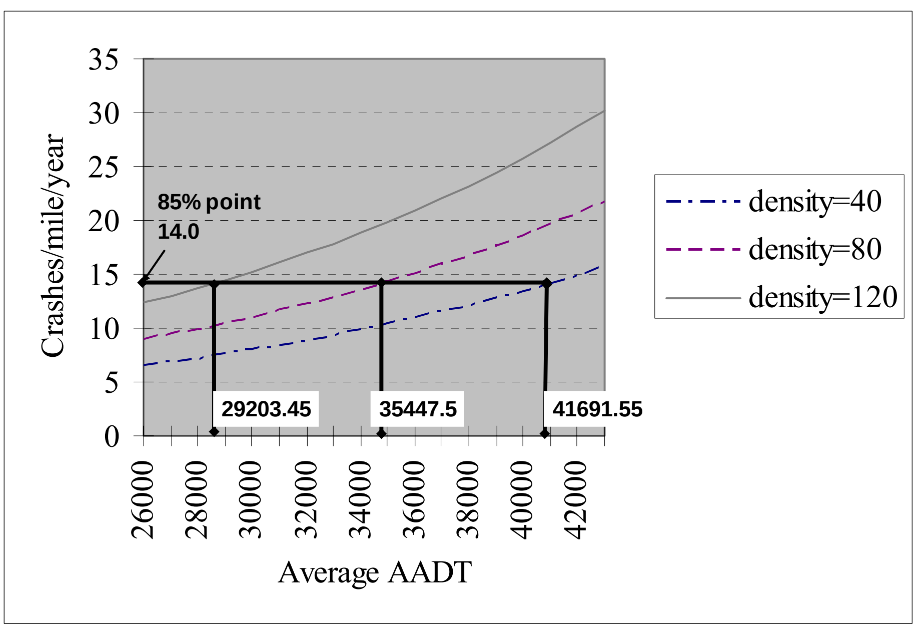


Figure 6.4 Critical Value of AADT according to 85th Percentile Value for a Four-lane Section in the Higher Speed Category

6.6 Tables for Critical AADT Values

In this step, the critical AADT values obtained for different percentile values of average number of crashes for each one of the categories considered are tabulated. Table 6.5 presents critical AADT values for a four-lane road for the higher speed level, different densities and percentile values. Similar tables with the critical AADT values for several percentile values of average number of crashes when considering different characteristics such as access density, number of lanes, and posted speed for TWLTL sections are presented in Appendix E.

Table 6.5 Critical AADT Values for Four-lane Road

for the Higher Speed Category

Access Density (points per mile)	Critical AADT (vpd) Percentile Values					
	50%	75%	80%	85%	90%	95%
10	30,073	40,611	42,989	45,640	48,524	53,494
20	28,512	39,050	41,428	44,079	46,963	51,933
30	26,951	37,489	39,867	42,518	45,402	50,372
40	25,390	35,928	38,306	40,957	43,841	48,811
50	23,829	34,367	36,745	39,396	42,280	47,250
60	22,268	32,806	35,184	37,835	40,719	45,689
70	20,707	31,245	33,623	36,274	39,158	44,128
80	19,146	29,684	32,062	34,713	37,597	42,567
90	17,585	28,123	30,501	33,152	36,036	41,006
100	16,024	26,562	28,940	31,591	34,475	39,445
110	14,463	25,001	27,379	30,030	32,914	37,884
120	12,902	23,440	25,818	28,469	31,353	36,323
130	11,341	21,879	24,257	26,908	29,792	34,762
140	9,780	20,318	22,696	25,347	28,231	33,201

The example table for the four-lane-road for higher speed presented above shows the critical values of traffic volume for different values of access density, and different percentile values. These critical AADT were estimated from the developed model for percentile values from 50^{th} to 95^{th} considering a constant 5% increase. The values selected for access density to determine critical values of traffic volume increase by 10 access points per mile and start from 10 access points per mile up to 140 access points per mile. These number of access points per mile include the access points in both sides of the road. The number of lanes is divided into three groups, 2-lane, 4-lane, and 6-lane sections. As mentioned above, the number of lanes includes the through lanes for each direction of traffic. Posted speed is presented in two categories, higher speed (TWLTL sections with posted speed ≥ 45 mph) and lower speed (TWLTL sections with posted speed < 45 mph). For several cases, no critical values for traffic volume were determined, which indicates that TWLTL sections for those specific conditions for the respective percentile value need further analysis for possible improvements.

6.7 List of Critical TWLTL Sections

In this step, TWLTL sections for which its AADT is greater than the critical AADT value according to its corresponding characteristics (access density, number of lanes and speed level) based on the percentile value selected would be identified as locations that need further study in order to determine if improvements are needed.

6.8 Case Study

This section presents a case study, where the methodology developed in the research is applied to identify the locations that are critical for one of the seven districts of FDOT based on a threshold of 85^{th} percentile value. Critical sections would refer to locations that need further study to determine if improvements are needed at these sections. Appendix F presents the top locations for all the districts considering the same 85^{th} percentile value as threshold.

When applying the methodology, the first two steps will not be used since their function was to determine and to fit the actual crash data to the best statistical distribution for the crash data. Based on that distribution fitting, critical AADT where estimated for different percentile values and will be used in the application of the methodology. From step three, of the methodology only table providing the percentile values for the average number of crashes is necessary. The required steps of the methodology are explained below.

6.8.1 Setting up the Threshold

For this step in the methodology, a threshold to determine the critical AADT values and access density according to a specific number of lanes and posted speed is selected. The threshold is set based on a desired percentile value for the average number of crashes. For this example, the threshold selected is 85%. Based on this percentile value, the average number of crashes for this threshold is determined for the different categories from Table 6.4. Table 6.6 presents the values for the example.

Table 6.6 Average Number of Crashes for Case Study

Posted Speed	Number of Lanes	Average Number of Crashes
Higher Speed	2	8.85
	4	14.00
	6	19.20
Lower Speed	2	6.84
	4	10.91
	6	14.98

6.8.2 Critical AADT Values

The next step of the methodology is to determine the critical AADT values for the different categories based on the selected percentile value. For the example, the critical AADT values determined for the 85^{th} percentile value are summarized on Table 6.7. These

AADT values where obtained from the corresponding tables presented in Appendix E according to the access density, posted speed and number of lanes.

Table 6.7 Critical AADT for the 85th Percentile Value

Access Density (points per mile)	Critical AADT (vpd)					
	Higher Speed			Lower Speed		
	2-lane	4-lane	6-lane	2-lane	4-lane	6-lane
10	NA	45,640	47,373	NA	35,004	36,760
20	NA	44,079	45,812	28,835	33,443	35,199
30	NA	42,518	44,251	27,274	31,882	33,638
40	NA	40,957	42,690	25,713	30,321	32,077
50	NA	39,396	41,129	24,152	28,760	30,516
60	NA	37,835	39,568	22,591	27,199	28,955
70	NA	36,274	38,007	21,030	25,638	27,394
80	NA	34,713	36,446	19,469	24,077	25,833
90	28,700	33,152	34,885	17,908	22,516	24,272
100	27,139	31,591	33,324	16,347	20,955	22,711
110	25,578	30,030	31,763	14,786	19,394	21,150
120	24,017	28,469	30,202	13,225	17,833	19,589
130	22,456	26,908	28,641	11,664	16,272	18,028
140	20,895	25,347	27,080	10,103	14,711	16,467

NA: Not Applicable. The critical AADT is above 30,000 vpd.

6.8.3 Critical TWLTL Sections

This final step of the methodology refers to the identification of TWLTL sections that need further study. Locations that have their AADT above the critical AADT values for their corresponding access density according to the category for which the location corresponds would be identified as a critical location. Table 6.8 presents the locations of District 7 identified as critical for their corresponding category when considering 85th percentile value as the threshold. Appendix F presents the TWLTL sections identified as critical for all the seven districts of FDOT. The threshold for the selection of the locations for all the districts was also the 85th percentile value.

Table 6.8 Identified TWLTL Sections for District 7

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Access Density	AADT	Posted Speed	Avg Crashes
Higher Speed	2	*						
	4	710130000	3.81	3.87	132	29,325	45	98.04
		710130000	4.50	4.80	44	41,036	45	15.77
		715120000	5.65	5.72	27	53,500	50	9.13
	6	710030000	1.55	1.70	112	43,125	45	9.32
		710030000	1.80	2.17	76	43,040	45	22.64
		710030000	2.36	3.01	73	42,771	45	17.84
		710130000	8.83	8.93	61	49,375	45	13.47
		710130000	9.60	9.79	78	64,984	45	55.27
		710160000	0.27	0.65	63	74,224	45	25.37
		710160000	0.90	1.24	62	78,722	45	17.6
		710160000	1.24	1.94	44	77,500	45	0
		715040000	4.69	4.84	13	50,833	45	6.67
		715040000	4.95	5.54	71	52,850	45	16.84
		715040000	5.84	5.91	43	52,167	45	14.29
		715020000	0.00	1.07	79	21,692	30	4.04
		715020000	3.67	4.95	83	25,167	40	1.56
		715020000	7.57	9.23	33	28,174	40	4.62
	4	715007000	3.54	3.80	83	28,100	40	6.56
		715040000	0.80	1.03	82	25,700	35	14.43
		715090000	1.25	2.23	99	24,709	35	3.77
		715090000	2.23	2.34	117	29,667	40	9.01
	6	702000000	14.20	14.55	99	26,450	40	9.75
		710030000	0.05	0.45	86	41,568	40	36.94
		710030000	0.55	0.90	79	41,688	40	30.05
		710030000	1.06	1.44	61	42,921	40	55.7

6.9 Other Considerations

This section presents other considerations in regard to the assessment of TWLTL sections that are not specifically considered in the methodology developed in this research. One of the considerations presented in this section refers to the identification of a section as critical based on the comparison between the average number of crashes based on crash history and crashes estimated from the model. Other considerations included in this section are: assessing a road as a whole and not by sections, and considering future developments to the TWLTL sections.

6.9.1 Number of Crashes

The number of crashes based on the crash history of the locations and the estimated average number of crashes per mile per year from the models could be compared. If the number of crashes from crash history for a specific location is greater than the number of crashes from the model, this would indicate that the location is critical because it is already over the average of number of crashes per mile per year for the group of intersections with the same characteristics. As an example, the section with ID 710030000 between mileposts 1.064 and 1.441 has 55.70 average number of crashes per mile per year based on its crash history. This number of crashes is higher than the average number of crashes per mile per year estimated from the model which is 14.98 for the corresponding lower speed category and six-lane section.

On the other hand, a location might be identified as critical based on the critical AADT value according to its characteristics if the number of crashes based on crash history of the location is below the estimated average number of crashes from the model for the group of intersections with the same characteristics. As an example, the section with ID 710160000 between mileposts 1.240 and 1.939 has no crashes based on its crash history, which is below the average number of crashes per mile per year estimated from the model which is 19.20 for the category of higher speed level and six-lane section. Never the less, this location was identified as critical based on the critical AADT value according to the corresponding access density for its respective category.

6.9.2 Assessing the Road

The TWLTL sections identified as critical in the case study might be part of different roads or might be part of the same road. If one section of a road is identified, this does not necessarily indicate that the remaining sections of the road are also critical or that a road as a whole is critical. For this reason, the following subsection presents a comparison between identified TWLTL sections and the road as a whole.

In order to analyze the road as a whole, the different TWLTL sections of the road, which might have different number of lanes, posted speed limit, and traffic volume, are combined. As an example, the road with ID number 710130000 was selected for this comparison. The characteristics for the different TWLTL sections of the road are presented in Table 6.9, including the identified critical sections. The road as a whole with combined values from the characteristics of the different sections is also presented in the table. Even though, two sections of the road are critical and other sections are close to be critical, the road as a whole would not be identified as critical if evaluated with the combined characteristics of its TWLTL sections. Therefore, an evaluation of the road is also necessary when analyzing critical TWLTL sections. The developed methodology does not have this capability of evaluating the road as a whole at this point. However, further study could be performed to develop models that can rank roadways too.

Table 6.9 Evaluation of TWLTL Sections and the Road as a Whole

Posted Speed Level	Number of Lanes	Road ID (critical)	Begin Milepost	End Milepost	Access Density	AADT	Posted Speed	Actual Average Number of Crashes
Higher Speed	4	710130000	3.54	3.69	52	29,500	45	8.71
		710130000	3.69	3.81	26	33,500	45	0.00
		710130000 (critical)	3.81	3.87	132	29,325	45	98.04
		710130000	3.87	4.22	91	29,386	45	21.44
		710130000	4.22	4.50	32	42,000	45	0.00
		710130000 (critical)	4.50	4.80	44	41,036	45	15.77
Higher Speed	4	710130000 (combined)	3.54	4.80	63	34,125	45	23.99

6.9.3 Assessing Future Development

Finally, if one or more of the TWLTL sections of a road would have future developments, the developments' impacts on safety can be projected for each section. In this sense, development of any of the sections could put the section in a critical level based on projections. This section should be considered for further study to determine if improvements are desirable before or concurrent with the future development. This type of proactive safety approach can facilitate the orderly placement of median openings. As an example, several sections of the road showed in Table 6.9 are close to being identified as critical. This means that if some development occurs on these sections, the road sections would likely become critical. The software discussed in the next section can be used to assess changes caused by future development and their impact on safety.

6.10 Software for the Developed Methodology

Software to apply the methodology to identify critical TWLTL sections was developed. This software has two specific functions that allow the used of the methodology considering two different aspects. The first function allows the estimation of the critical AADT value for a TWLTL section based on a selected percentile value and its characteristics: access density, posted speed and number of lanes. If the AADT of the TWLTL section is greater than the critical AADT value estimated by the software, the section is included in the list of critical locations for the specify percentile group. The second function of the software allows the determination of the percentile group to which a TWLTL section belongs based on its characteristics: access density, posted speed, number of lanes and AADT. Depending on the desired critical percentile value, the percentile group to which the section belongs could be identified as critical or not.

CHAPTER 7 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

7.1 Summary

The purpose of this research was to develop a procedure to identify the factors that are influential in the safety experience of TWLTL sections. This procedure also allows the identification of groups of TWLTL locations that present existing and future safety concerns and need further analysis to determine possible improvements.

A three-year crash history database including traffic crashes from totally 1688 TWLTL sections all over Florida was used in this research. The FDOT crash database from 1996 to 1998 was used because of the relative ease in the gathering of crash and other related data. Statistical models were used to determine the relationship between number of crashes per mile per year and several factors such as traffic volume, access density, posted speed and number of lanes. Percentile values were utilized as thresholds to determine TWLTL sections that might need further analysis for possible improvements.

In the analysis, distribution fitting for the Poisson, Negative Binomial and Lognormal distributions was performed for the crash data. The results of comparing the Chi-square test at a 95% confidence level for each distribution indicated that crash data best fitted the negative binomial regression. Based on this, a Negative Binomial regression model was developed to estimate the number of crashes per mile per year for the TWLTL sections. During the modeling process, Poisson regression was performed as the initial step and Negative Binomial regression was finally applied because over-dispersion was tested existing in the crash data. The regression parameters were estimated by using the maximum likelihood method with SAS and SPSS. The goodness-of-fit for the developed model was evaluated based on Pearson's R-square and likelihood ratio index. After that, the average number of crashes per year per mile was determined for sections with the same characteristics.

The process to develop the methodology to identify locations that need further study to determine if improvements are necessary consists of several steps. One step was plotting crash data to determine the distribution of the actual data for six different groups of locations. These groups were based on posted speed and number of lanes. Then, distribution fitting of the crash data was performed to determine the statistical distribution that better represents the actual data. In the next step, the percentile values for the average numbers of crashes from distribution fitting of the original crash database are determined. Then, the average number of crashes per mile per year was determined for sections with the same characteristics using the parameters of the crash predictive model. These estimated number of crashes obtained from the model is plotted in different curves. With the percentile value of the average number of crashes and the curves plotted from the models, the critical values for the variables considered in the research were obtained. The next step consists of the calculation of tables of critical average AADT values for each one of the six groups considered. These AADT are based on values of access density number of lanes and posted speed. The final step of the methodology is the generation of a list of locations identified as critical according to the selected percentile value and critical average AADT based on access density, posted speed, and number of lanes. When applying the methodology to identify a group of critical locations, only the final steps mentioned above are necessary. These steps include the used of tables of critical average AADT values for each one of the six groups considered and the generation of the list of locations according to the selected percentile value.

7.2 Conclusions

After reviewing over 1600 TWLTL sections located all over Florida, the results indicated that increases in AADT, access density, number of lanes and posted speed all increase influence the crash occurrence at these locations. Also based on modeling results and percentile values obtained from crash distribution fitting, a group of TWLTL locations could be identified for further analysis to determine if improvements are necessary.

According to the analysis and results of the previous chapters the following conclusions can be obtained:

- For the statistical modeling to determine the relationship between the average number of crashes per mile per year at TWLTL sections and several factors such as AADT, access density, number of lanes and posted speed, Negative Binomial regression was proved to be appropriate to model the crash data. The crash data showed extra-variations relative to Poisson distribution

- All the factors considered during modeling, AADT, access density, number of lanes and posted speed had estimated parameters significant at the 5% significance level for the developed model. This indicates that these factors influence the crash occurrence at TWLTL sections. Among these variables, the parameters for the variables AADT, access density and number of lanes had a positive sign. These findings suggested that an increase on any of these three variables increases the likelihood of crash occurrence on TWLTL sections. On the other hand, the sign for the parameter for posted speed was negative. This could be attributed to the fact that drivers tend to travel at speeds in which they feel comfortable given the prevailing conditions. Thus, lower posted speed more likely promotes speed differential that is generally more closely associated with crashes.

- Regarding the goodness-of-fit, the statistical model developed has satisfactory capability in fitting crash data and explaining the variation of the crash data based the four statistics considered: deviance, Pearson’s Chi-square, Pearson's R- square and likelihood ratio index. According to likelihood ratio index, the model explained more than 45% variation in the crash data

- The Chi-square tests indicated that average number of crashes for six different categories selected according to posted speed levels and number of lanes followed either a Negative Binomial or Log-normal distribution at a 95% confidence level. Then, the 50th, 75th, 80th, 85th, 90th, and 95th percentile values for each category were estimated based on these crash distribution fitting.

- Critical values for AADT for different characteristics of TWLTL sections were estimated by plotting the percentile values of the distribution fitting for the six categories selected into the curves obtained from the average number of crashes estimated from the developed model. TWLTL sections that have an AADT higher that the critical value for specific characteristics of the section would be identified as a section that needs further study to determine if improvements are necessary.

- Based on the methodology developed in the research, locations for each one of the seven districts of FDOT where identified based on the critical values of AADT for different values of access density, number of lanes and posted speed. The locations identified need further study in order to determine if improvements are needed.

7.3 Recommendations

The methodology developed in this research identified TWLTL sections that require further study in order to determine if improvements are needed. Future research would allow the extension of the methodology in order to include the evaluation of the road as a whole, not only the critical TWLTL section that belongs to that road, and the evaluation of future developments on a TWLTL section and its effect on the section and the road.

Also, other factors should be incorporated into the analysis of the safety of TWLTL sections in future studies. These factors include time of the crash and gap acceptance, which would allow a more complete assessment of the TWLTL sections. Also, a comparison of this median treatment to other median treatments would allow the estimation of the conditions under which TWLTL work better.

LIST OF REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*. John Wiley & Sons, New York.

American Association of State Highway and Transportation Officials (2001). *A Policy on the Geometric Design of Highways and Streets*, Washington D.C.

Azzeh, J.A., B.A. Thorson, J. J.Valenta, J. C. Glennon. and C.J. Wilton (1975) *Evaluation of Techniques for the Control of Direct Access to Arterial Highways.* Report FHWA-RD-76-87, FHWA, U.S. Department of Transportation.

Bauer, K.M. and D.W. Harwood (1996). *Statistical Models for At-grade Intersection Accidents.* Report FHWA-RD-96-125. Federal Highway Administration, McLean, VI.

Ben-Akiva, M. and S.R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Mass.

Bonneson, J.A. and McCoy, P.T. (1997). *Capacity and Operation Effects of Midblock Left Turn Lanes*, National Cooperative Highway Research Program Report 395, Transportation Research Board, National Research Council, Washington, DC.

Bonneson, J.A. and McCoy, P.T. (March 1998). *Median Treatment Selection for Existing Arterial Streets*, Institute of Transportation Engineers Journal, pp. 26-34.

Council, F.M. and J.R. Stewart (2000) *Safety Effects of The Conversion of Rural Two-lane to Four-Lane Roadways (Based on Cross-sectional Models*), U.S. Department of Transportation Federal Highway Administration, Institute of Transportation Engineers Journal, pp. 37-41.

Greene, W. (1997). *Econometric Analysis*. Prentice-Hall, New Jersey.

Harwood, D.W. and A.D. St. John (1985) *Passing Lanes and Other Operational Improvements on Two-lane Highways*. Report FHWA-RD-85.028. FHWA, U.S. Department of Transportation.

Harwood, D.W. (1986) *Multilane Design Alternatives for Improving Suberban*

Highways. National Cooperative Highway Research Program Report No. 282, TRB Transportation Research Record, Washington D.C.

Hauer, E., J.C.N.Ng, and J. Lovell (1988) *Estimation of Safety at Signalized Intersections.* Transportation Research Record 1185, TRB, National Research Council, Washington, D.C.

Hummer, J. E. (1994). *Traffic Accident Studies. Manual of Transportation Engineering Studies*. Institute of Transportation Engineers. Prentice Hall.

Joshua, S. C. and N. J. Garber (1990) *Estimating Truck Accident Rate and Involvement Using Linear and Poisson Regression Models.* Transportation Planning and Technology, Vol. 15.

Jovanis, P. and H. Chang (1985). *Modeling the Relationship of Accidents to Miles Traveled.* Transportation Research Record 1068, TRB, National Research Council, Washington, D.C.

King, G. F. and R.B. Goldblatt (1975), KLD Associates, Inc. *Relationship of Accident Patterns to Type of Intersection Control*. Transportation Research Record, 540. Transportation Research Board. Washing D. C.

Lawless, J. F. (1987). *Negative Binomial and Mixed Poisson Regression*. The Canadian Journal of Statistics, Vol15, No. 3.

McCullagh, P., and J. A. Nelder (1983). *Generalized Linear Models*. Chapman and Hall, London.

Miaou, S., and H. Lum (1993). *Modeling Vehicle Accident and Highway Geometric Design Relationships.* Accidents Analysis and Prevention, Vol. 25, No. 6.

Miaou, S., A. Lu, and H. S. Lum (1996). *Pitfalls of Using* R^2 *to Evaluate Goodness of Fit of Accident Prediction Models*..Transportation Research Record 1542, TRB, Washington, D.C.

Miaou, S. (1994). *The Relationship between Truck Accidents and Geometric Design of Road Section: Poisson versus Negative Binomial Regression.* Accident Analysis and Prevention. Vol. 26(4).

Nemeth, Z. A. (1978) *Two-way Left-Turn Lanes: A State-of-the-Art Overview and an Implementation Guide*, Ohio State University, Research Record 681, Transportation Research Record.

Nicholson, A.J., (1985) *The Variability of Accident Counts*, Accident Analysis & Prediction, Volume. 17. No. 1 pp. 47-56

Parker, M.R. (1983). *Design Guidelines for Raised and Traversable Medians in Urban Areas*, Virginia Highway and Transportation Research Council, Charlottesville, Virginia.

Parsonson, P.S.. (1990) *Development of Policies and Guidelines Governing Median Selection*. School of Civil Engineering, Georgia Institute of Technology, Georgia Tech Sponsored Research Project No. E-20-841, 1990

Poch, M. and F. Mannering (1996), *Negative Binomial Analysis of Intersection-Accident Frequencies*, Journal of Transportation Engineering, Vol. 122, No. 2. pp.105-113.

Resende, P. and R.F. Benekohal (1997) *Effects of Roadway Section Length on Accident Modeling*, Challenges, Innovations and Opportunities: Proceedings on Traffic Congestion and Traffic Safety in the 21st Century, New York, pp. 403-409.

Robertson, H.D., J.E. Hummer, and D.C. Nelson (1998) *Manual of Transportation Engineering Studies*, Institute of Transportation Engineers, Prentice Hall, Englewood Cliffs, NJ.

Saccomanno, F., and C. Buyco (1988) *Generalized Loglinear Models of Truck Accident Rates*. Paper presented at Transportation Research Board 67th Annual Meeting, Washington, D.C.

Shankar, V., F. Mannering, and W. Barfield (1995). *Effect of Roadway Geometric and Environmental Factors on Rural Accident Frequencies*. Accident Analysis and Prevention, Vol. 27, No. 3.

Squires, C.A., and P.S. Parsonson (1989). *Accident Comparison of Raised Median and Two-way Left-turn Lane Median Treatments*, Transportation Research Record 1239, Transportation Research Board, Washington, D.C.

APPENDICES

APPENDIX A

MODEL CURVES FOR DIFFERENT

ROADWAY SECTION CATEGORIES

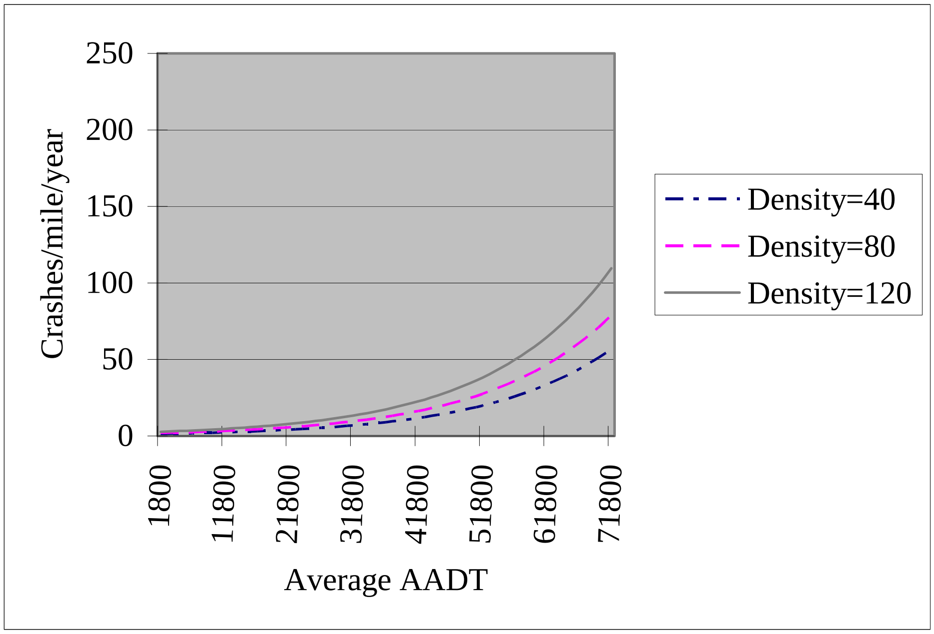


Figure A.1 Model Curve for a Two-lane Road in the Higher Speed Category

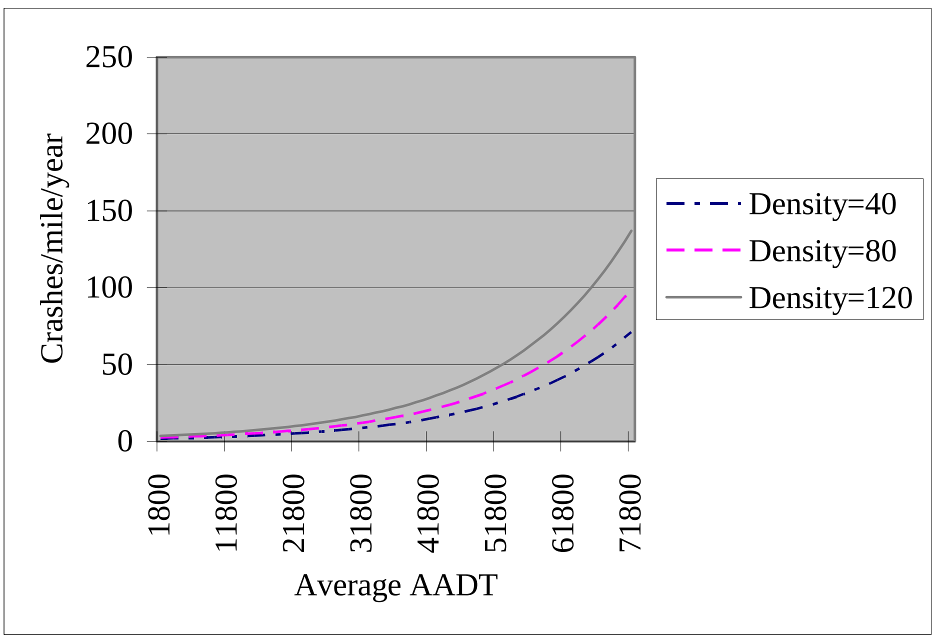


Figure A.2 Model Curve for a Four-lane Road in the Higher Speed Category

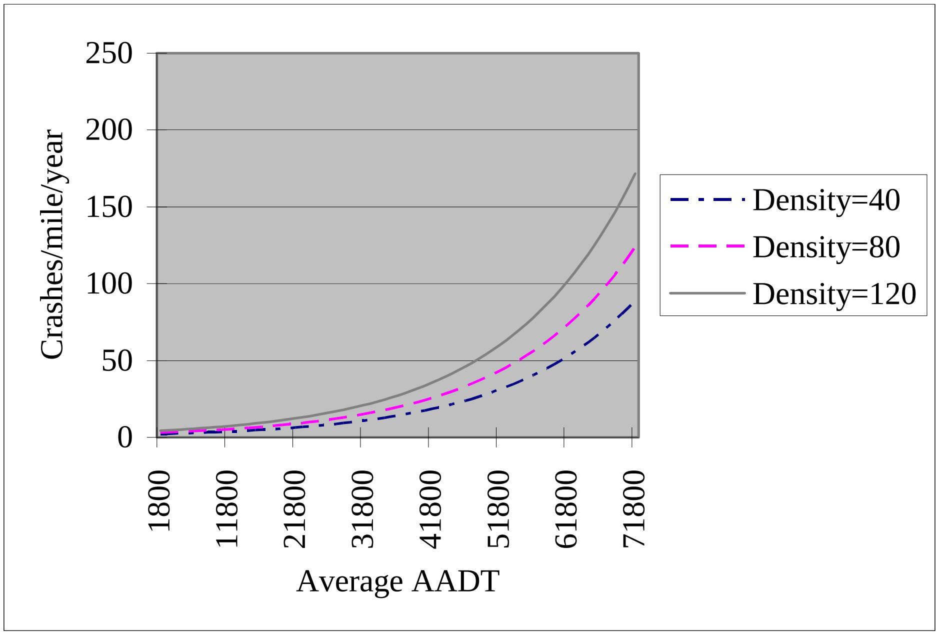


Figure A.3 Model Curve for a Six-lane Road in the Higher Speed Category

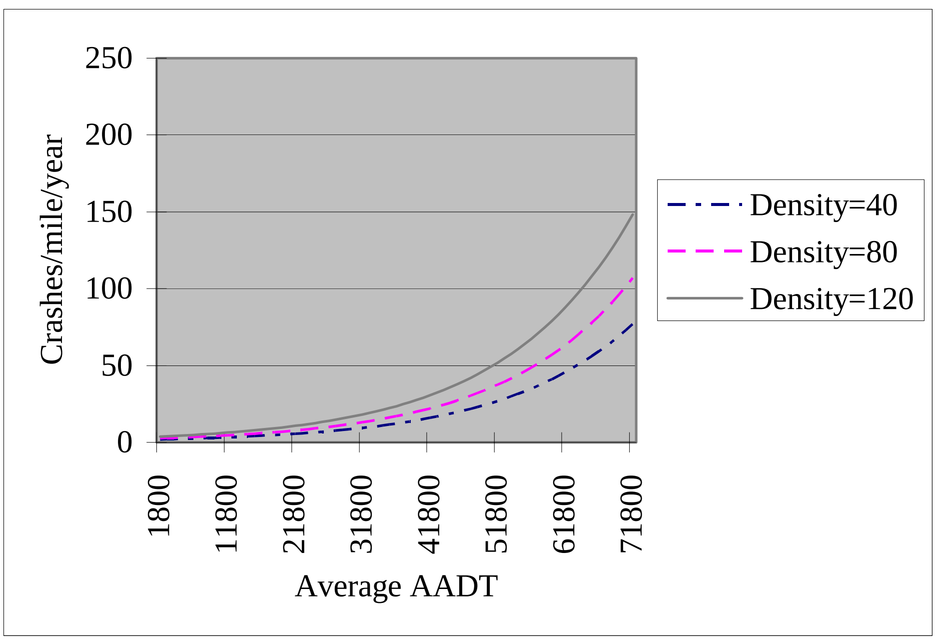


Figure A.4 Model Curve for a Two-lane Road in the Lower Speed Category

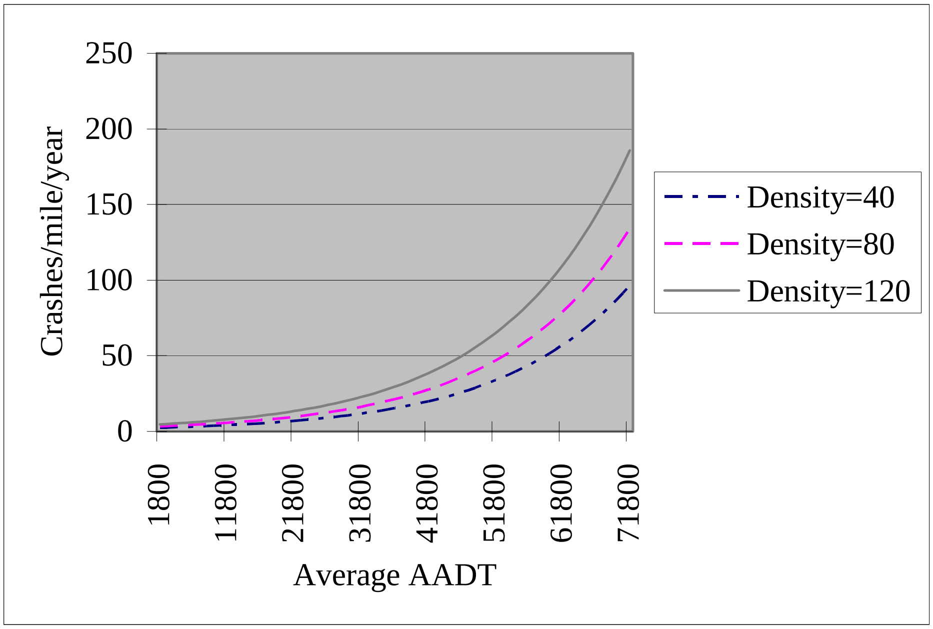


Figure A.5 Model Curve for a Four-lane Road in the Lower Speed Category

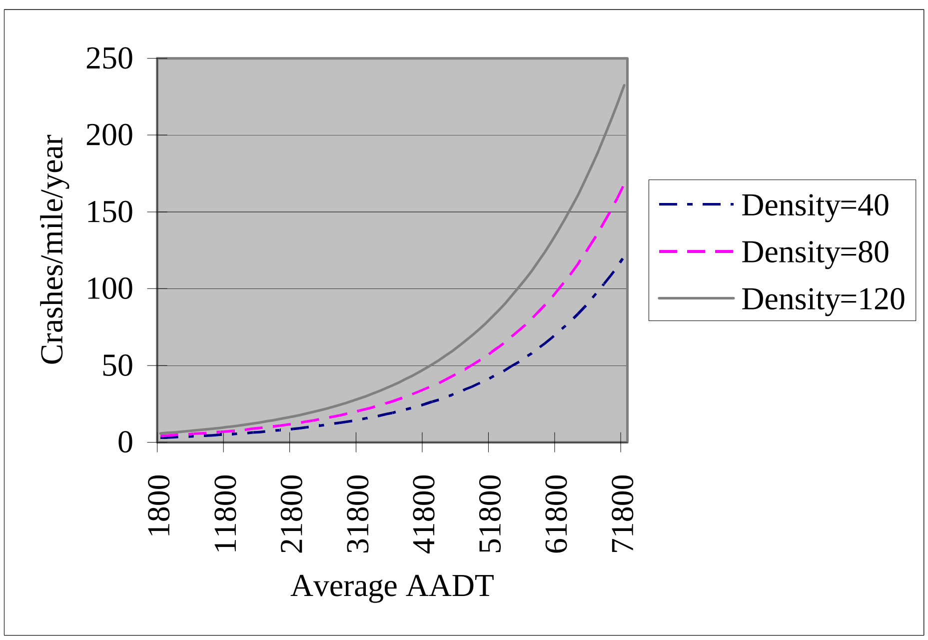


Figure A.6 Model Curve for a Six-lane Road in the Lower Speed Category

APPENDIX B

CRASH FREQUENCY DISTRIBUTIONS

FOR DIFFERENT ROADWAY SECTION CATEGORIES

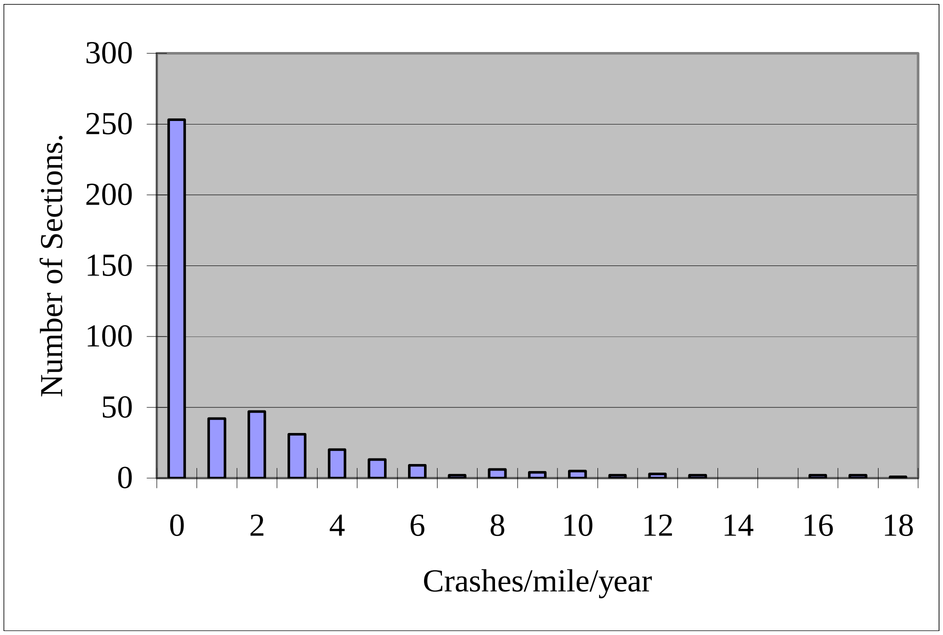


Figure B.1 Frequency Distribution for a Two-lane Road in the Higher Speed Category

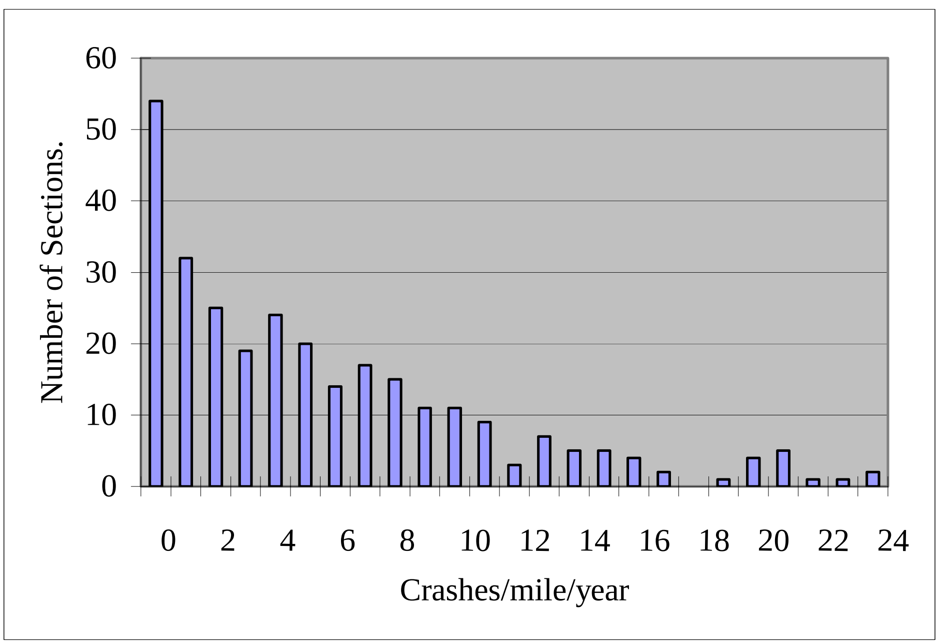


Figure B.2 Frequency Distribution for a Four-lane Road in the Higher Speed Category

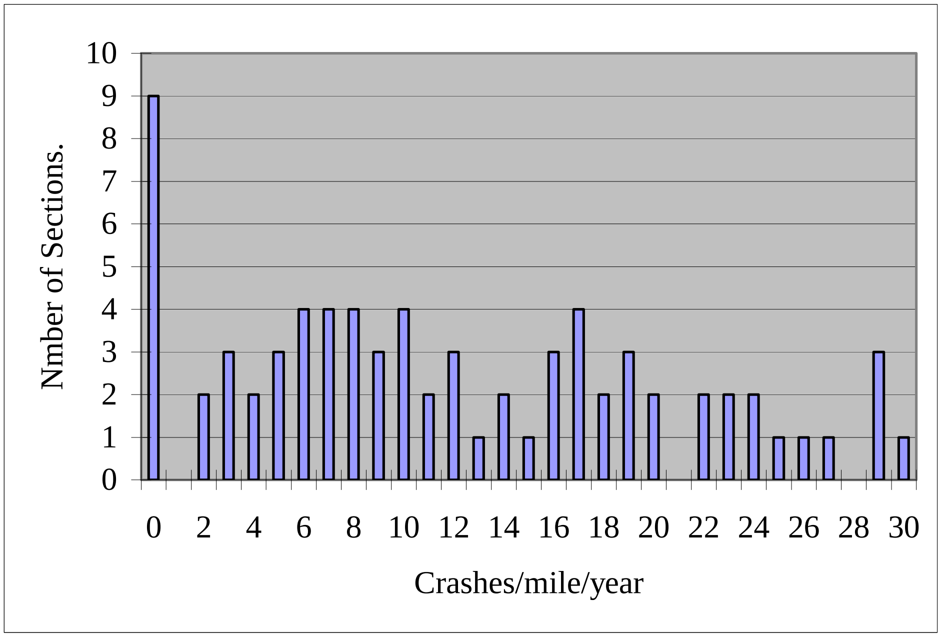


Figure B.3 Frequency Distribution for a Six-lane Road in the Higher Speed Category

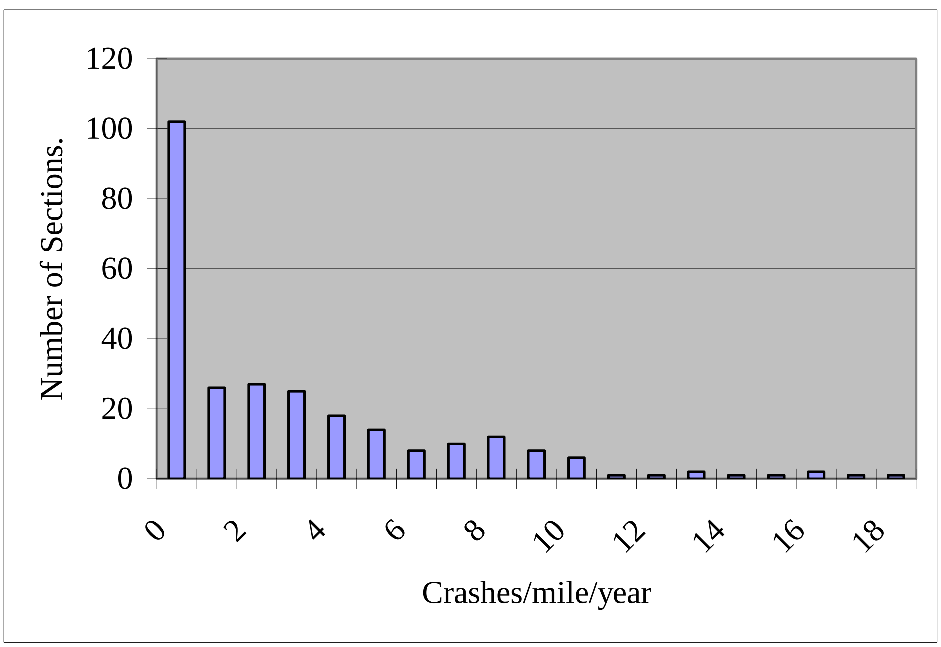


Figure B.4 Frequency Distribution for a Two-lane Road in the Lower Speed Category

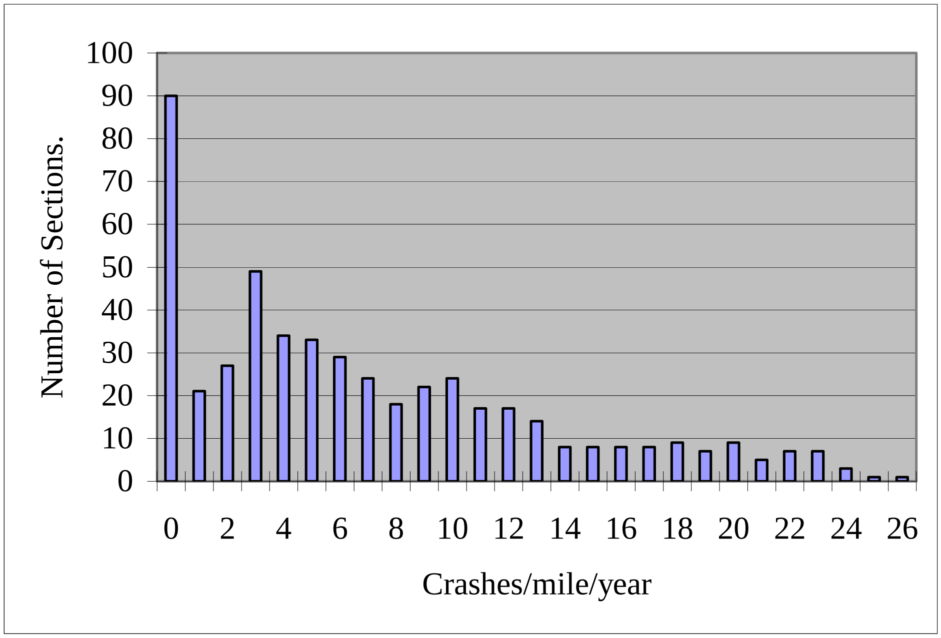


Figure B.5 Frequency Distribution for a Four-lane Road in the Lower Speed Category

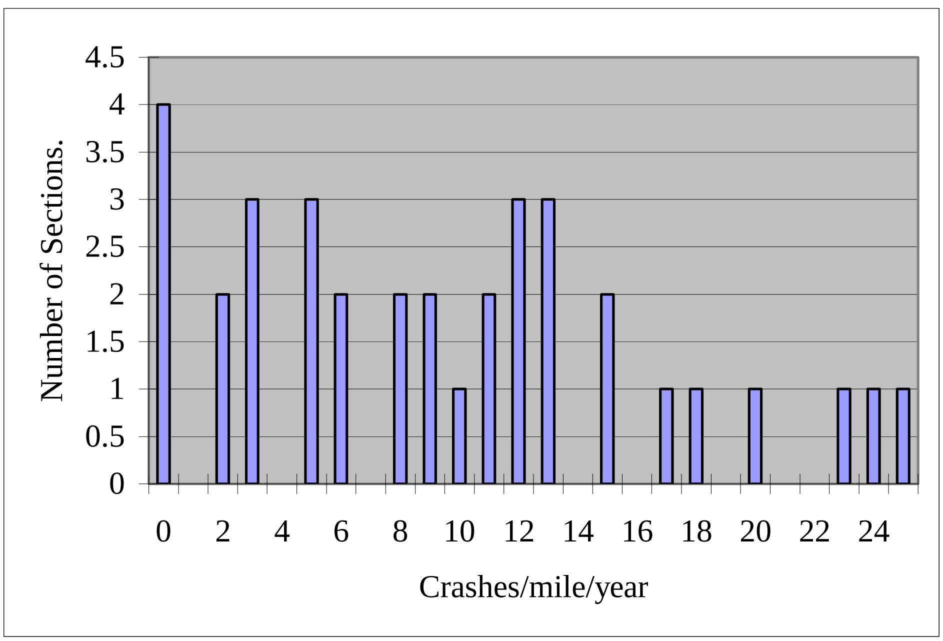


Figure B.6 Frequency Distribution for a Six-lane Road in the Lower Speed Category

APPENDIX C

ESTIMATION OF PERCENTILE VALUES OF CRASHES

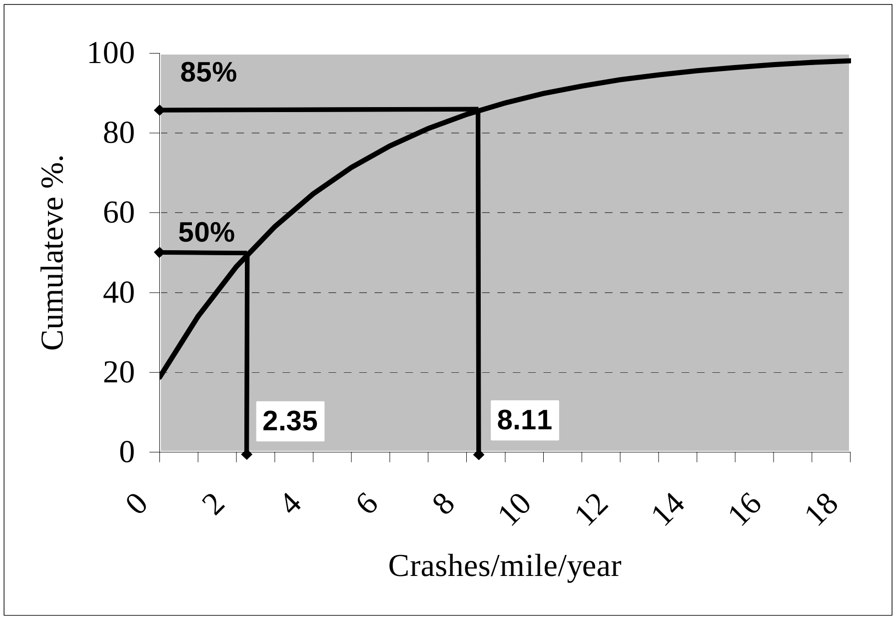


Figure C.1 The 50% and 85% Percentile Value of the Average Number of Crashes for Two-lane Road for the Higher Speed Category

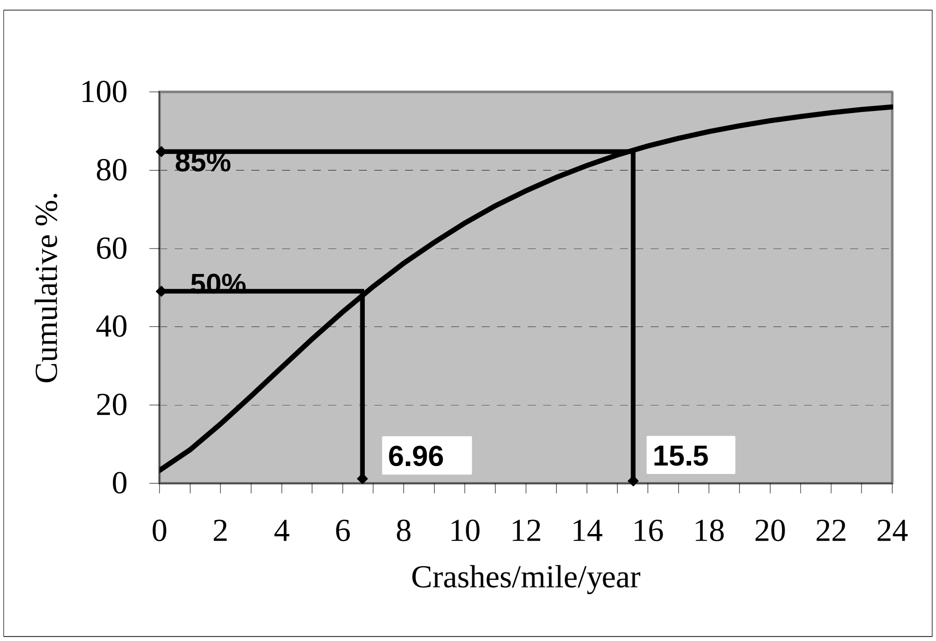


Figure C.2 The 50% and 85% Percentile Value of the Average Number of Crashes for Four-lane Road for the Higher Speed Category

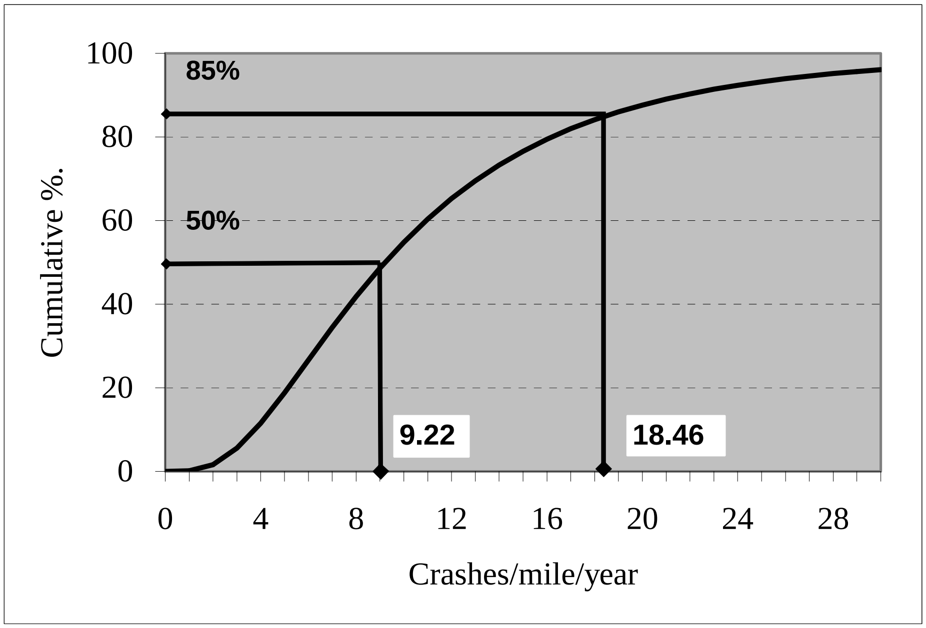


Figure C.3 The 50% and 85% Percentile Value of the Average Number of Crashes for Six-lane Road for the Higher Speed Category

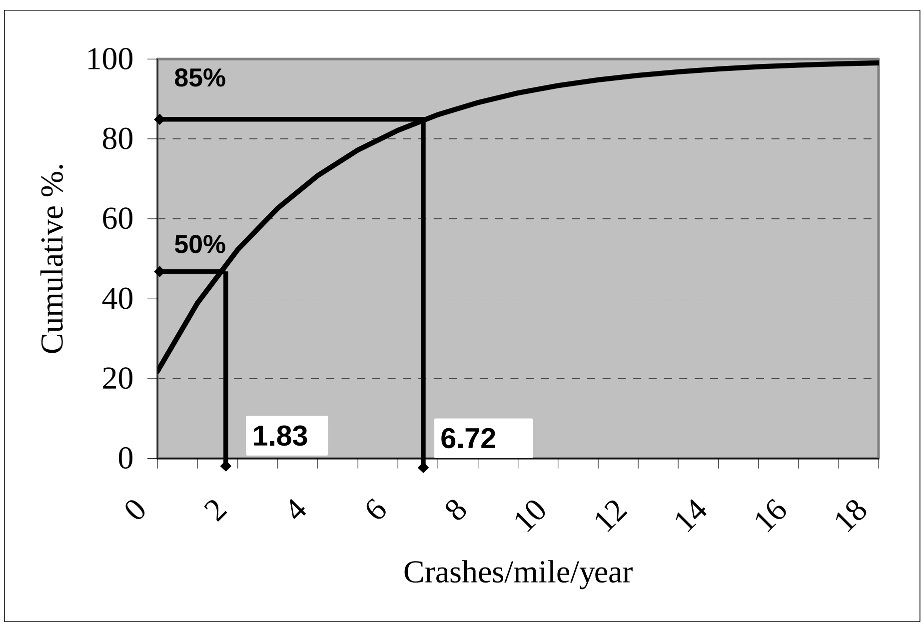


Figure C.4 The 50% and 85% Percentile Value of the Average Number of Crashes for Two-lane Road for the Lower Speed Category

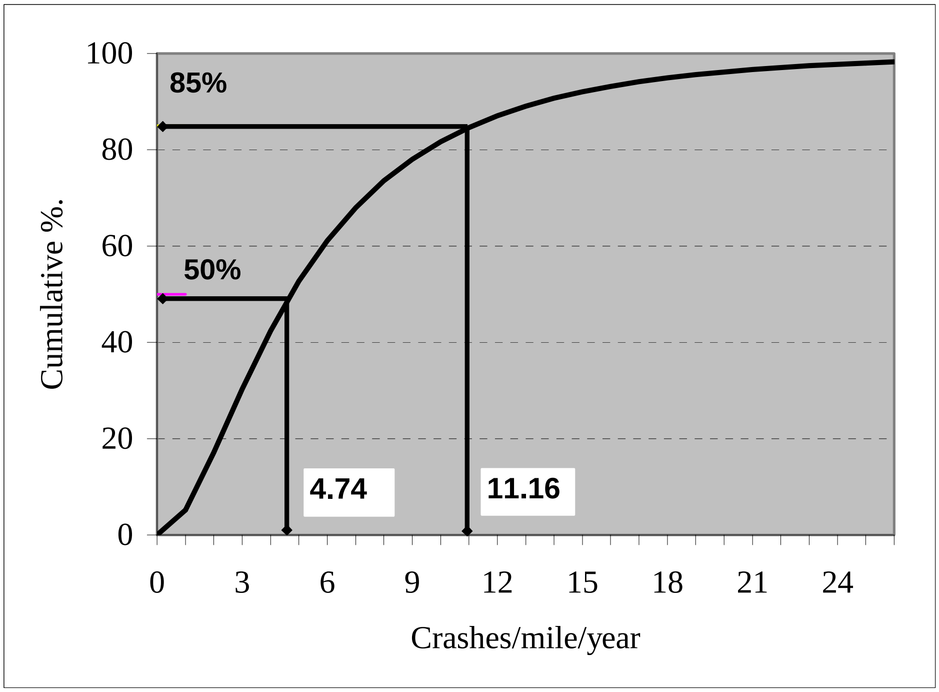


Figure C.5 The 50% and 85% Percentile Value of the Average Number of Crashes for Four-lane Road for the Lower Speed Category

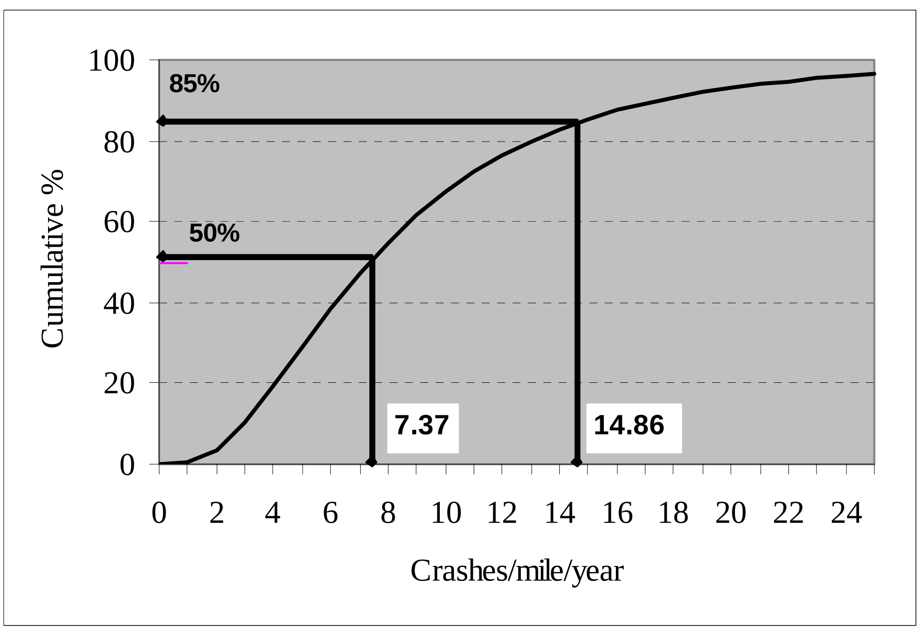


Figure C.6 The 50% and 85% Percentile Value of the Average Number of Crashes for Six-lane Road for the Lower Speed Category

APPENDIX D

ESTIMATION OF CRITICAL VALUES OF AADT

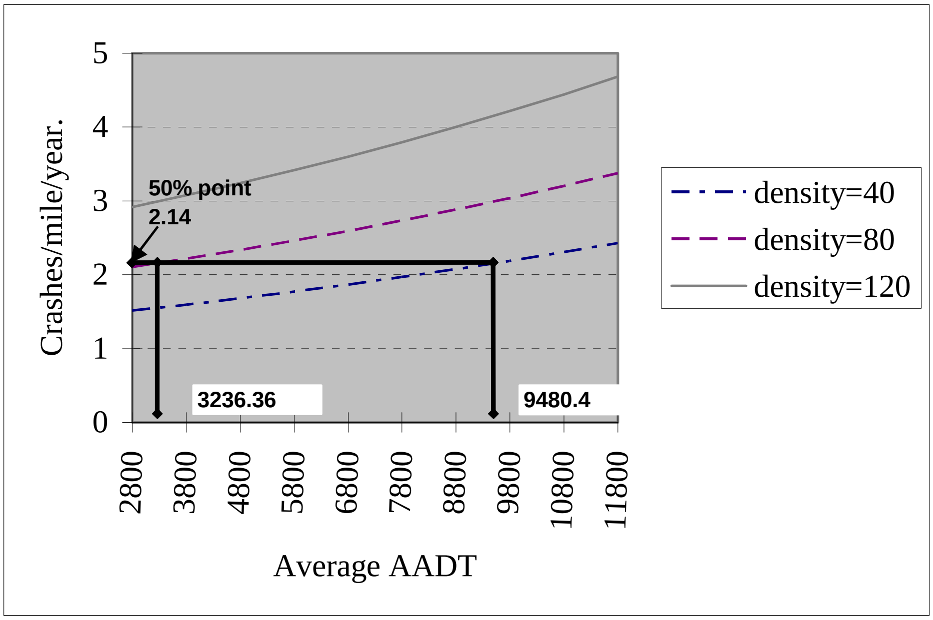


Figure D.1 Critical Values of AADT according to 50th Percentile Value for a Two-lane Road in the Higher Speed Category

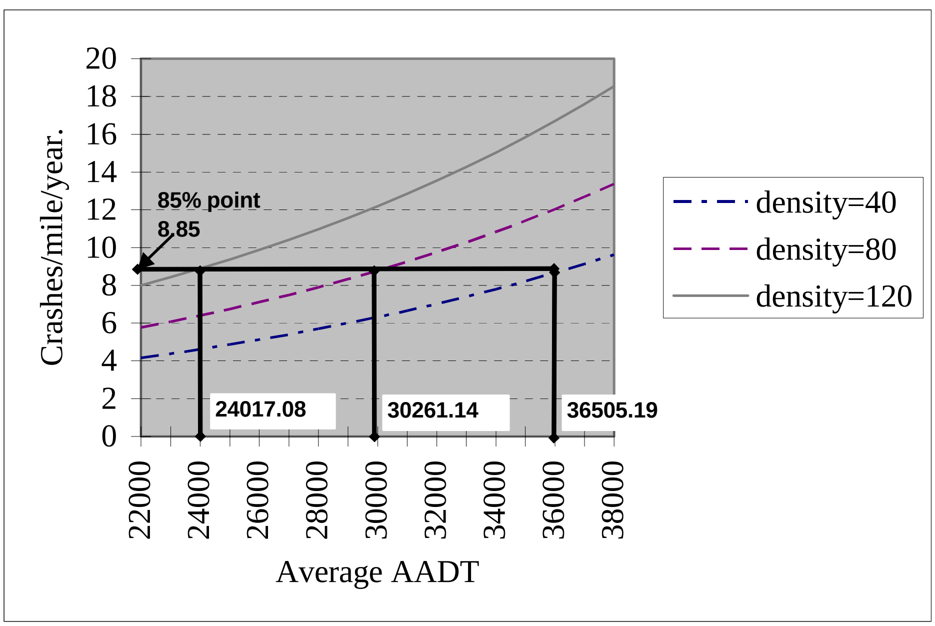


Figure D.2 Critical Values of AADT according to 85th Percentile Value for a Two-lane Road in the Higher Speed Category

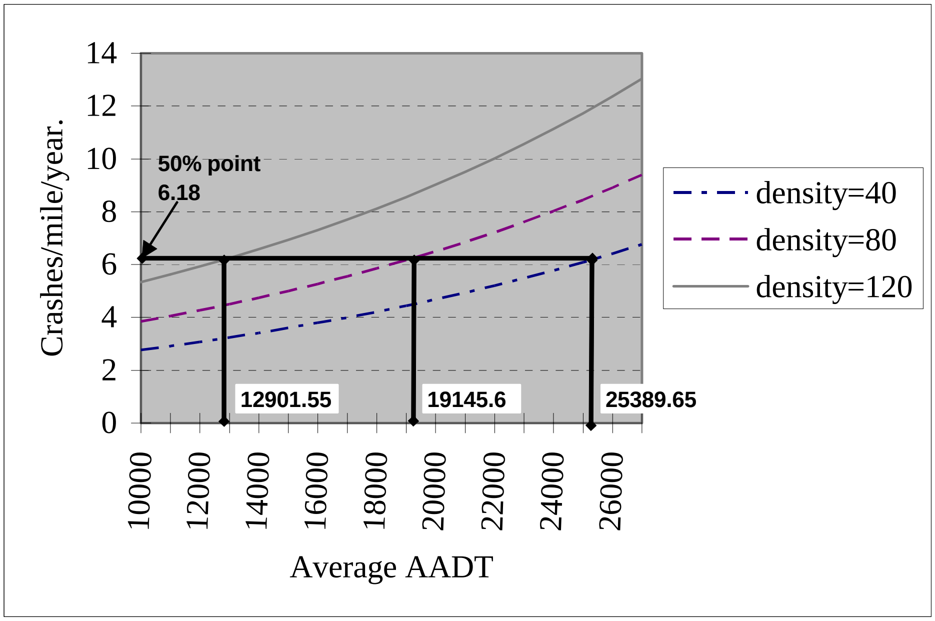


Figure D.3 Critical Values of AADT according to 50th Percentile Value for a Four-lane Road in the Higher Speed Category

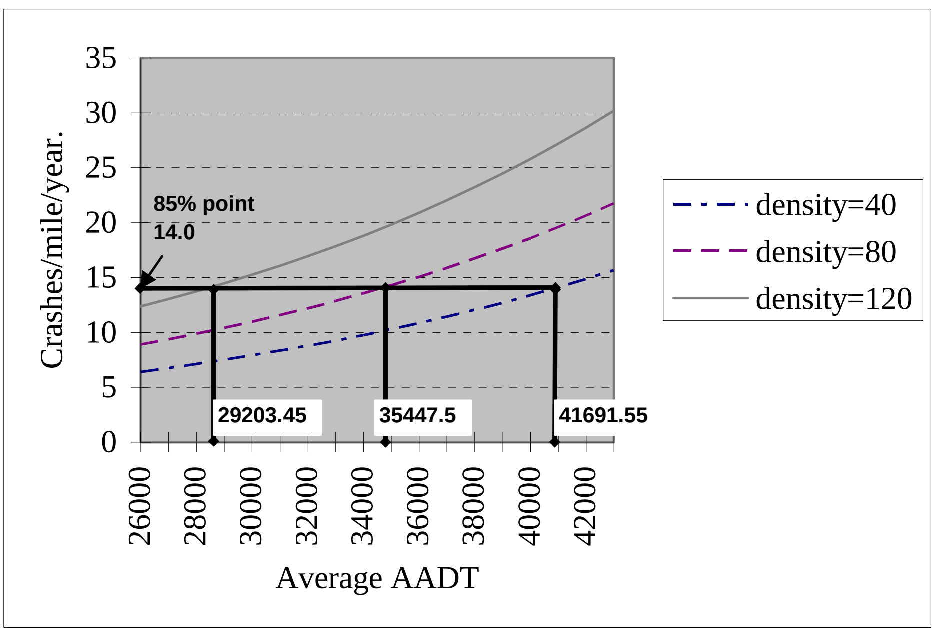


Figure D.4 Critical Values of AADT according to 85th Percentile Value for a Four-lane Road in the Higher Speed Category

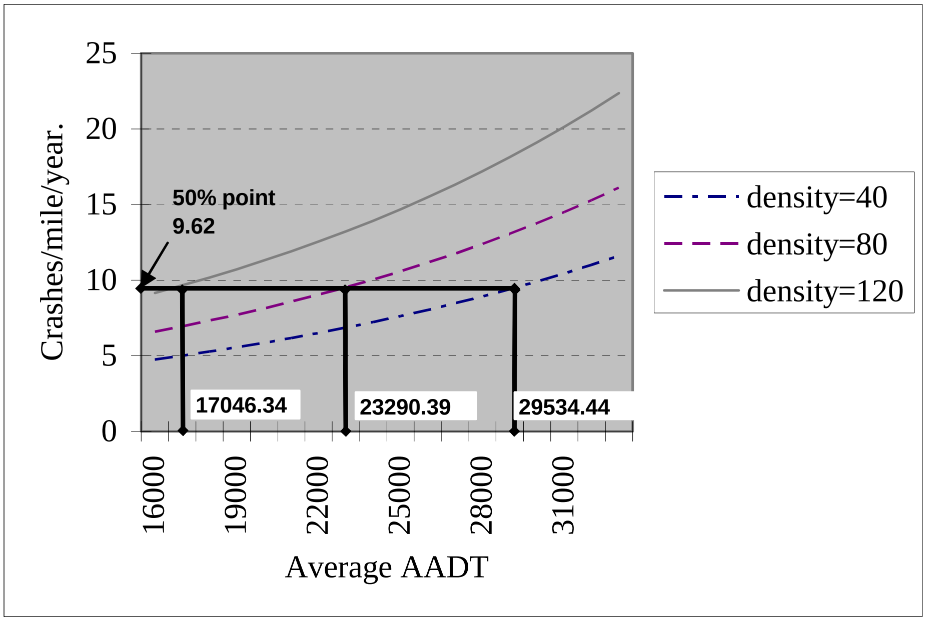


Figure D.5 Critical Values of AADT according to 50th Percentile Value for a Six-lane Road in the Higher Speed Category

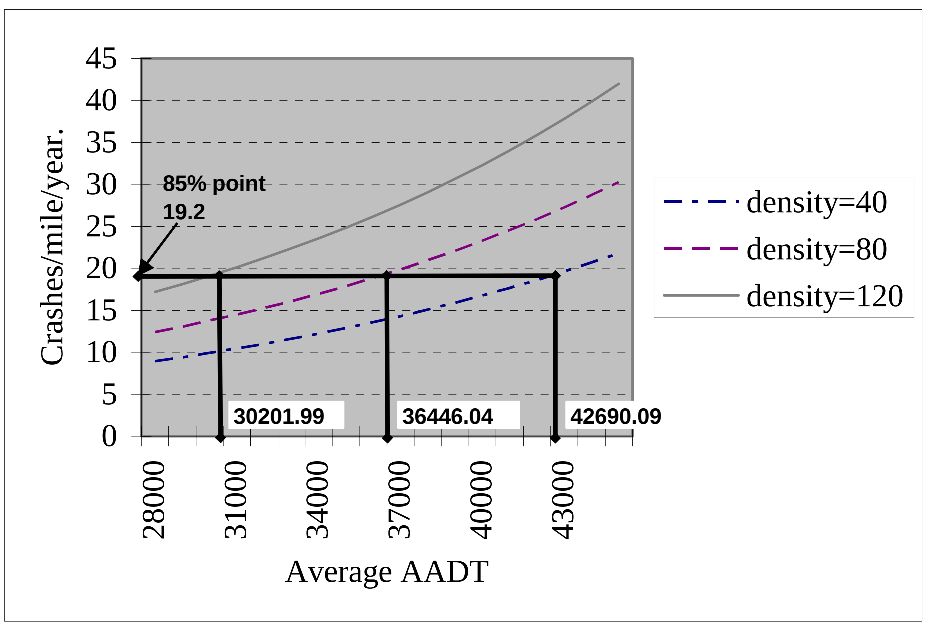


Figure D.6 Critical Values of AADT according to 85th Percentile Value for a Six-lane Road in the Higher Speed Category

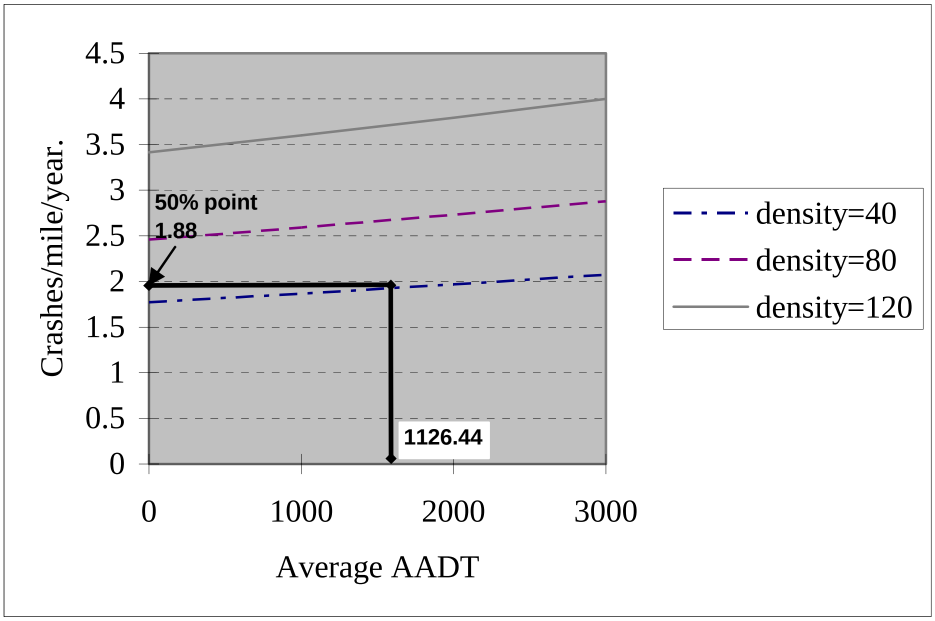


Figure D.7 Critical Values of AADT according to 50th Percentile Value for a Two-lane Road in the Lower Speed Category

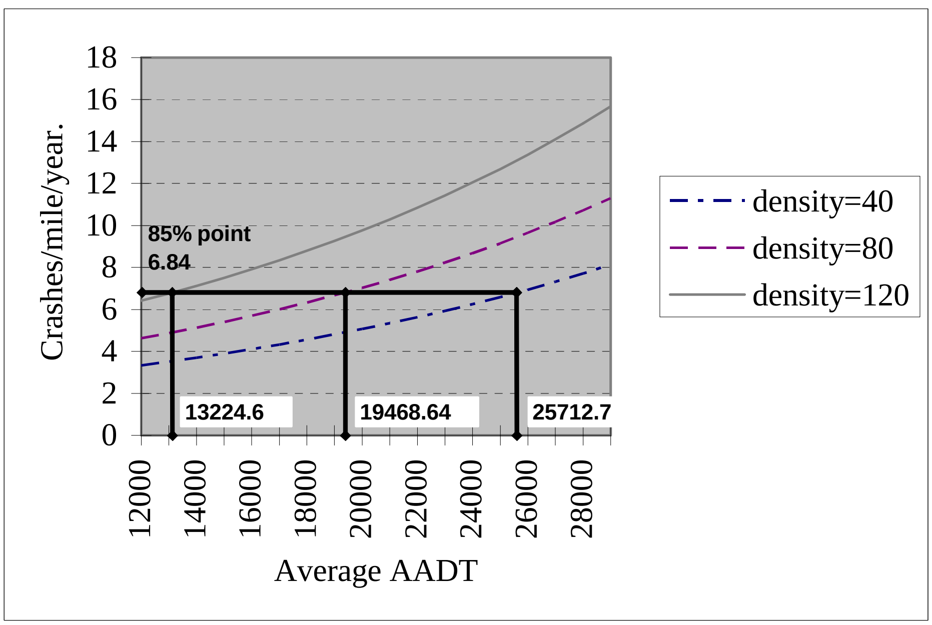


Figure D.8 Critical Values of AADT according to 85th Percentile Value for a Two-lane Road in the Lower Speed Category

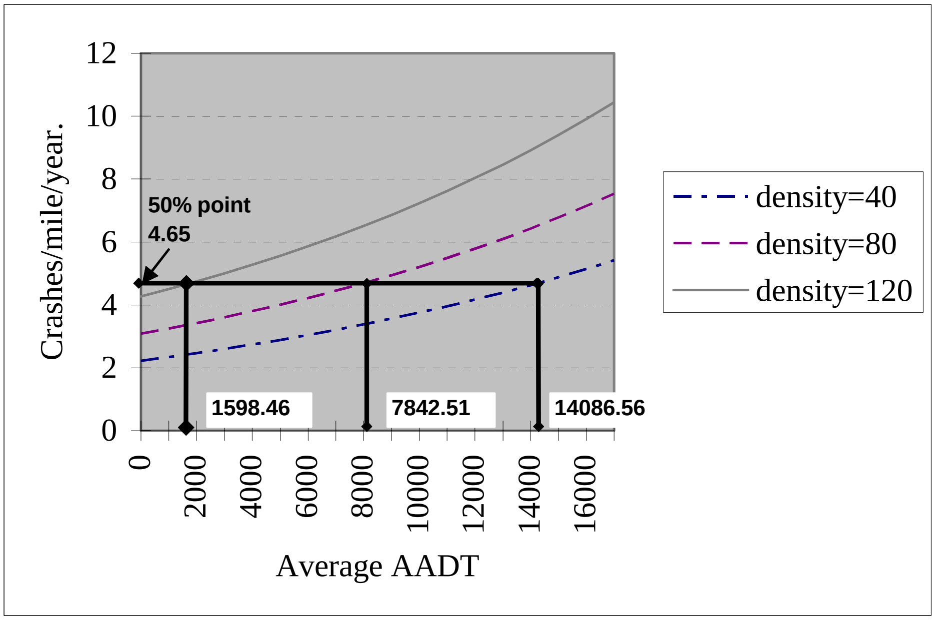


Figure D.9 Critical Values of AADT according to 50th Percentile Value for a Four-lane Road in the Lower Speed Category

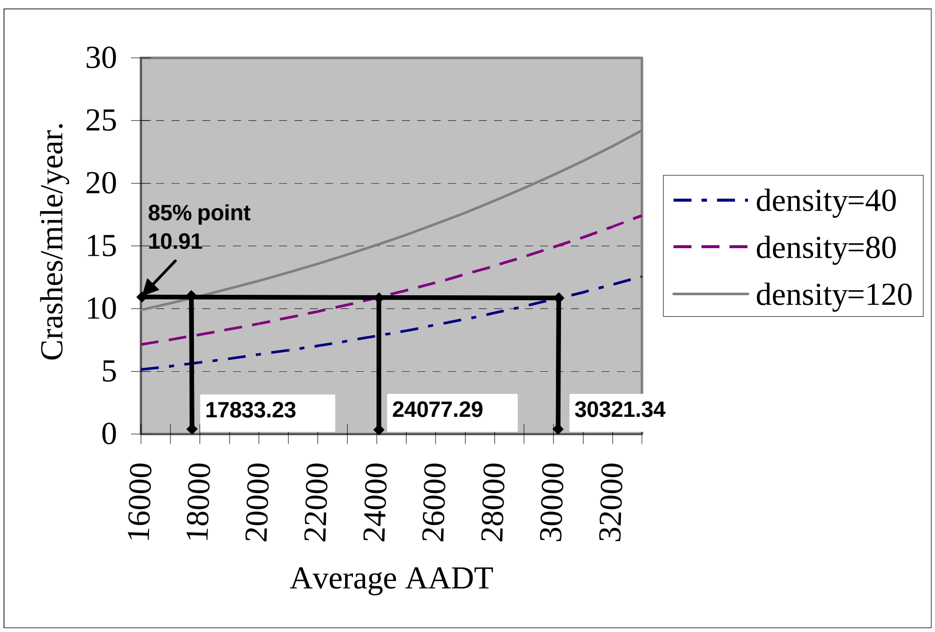


Figure D.10 Critical Values of AADT according to 85th Percentile Value for a Four-lane Road in the Lower Speed Category

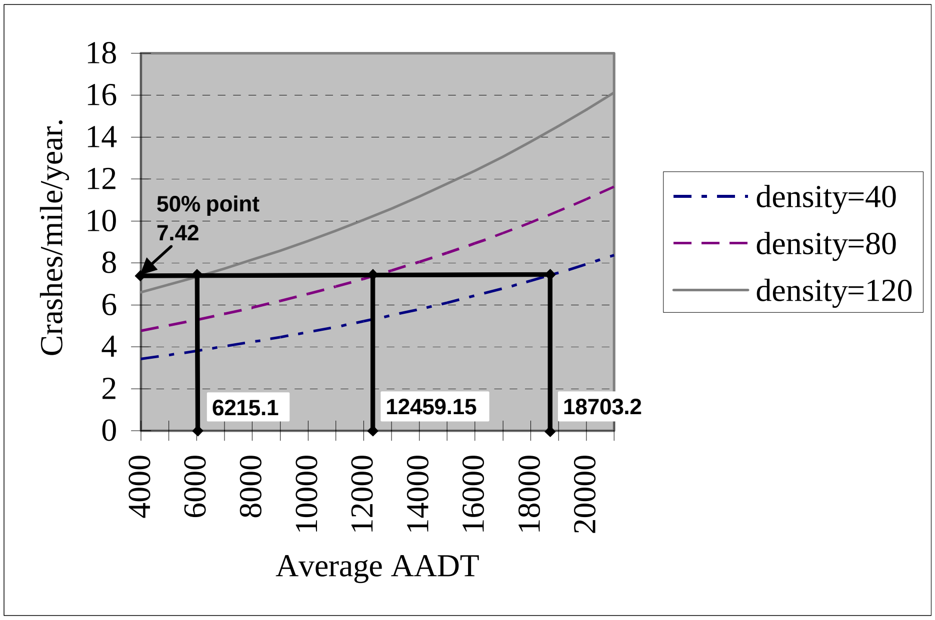


Figure D.11 Critical Values of AADT according to 50th Percentile Value for a Six-lane Road in the Lower Speed Category

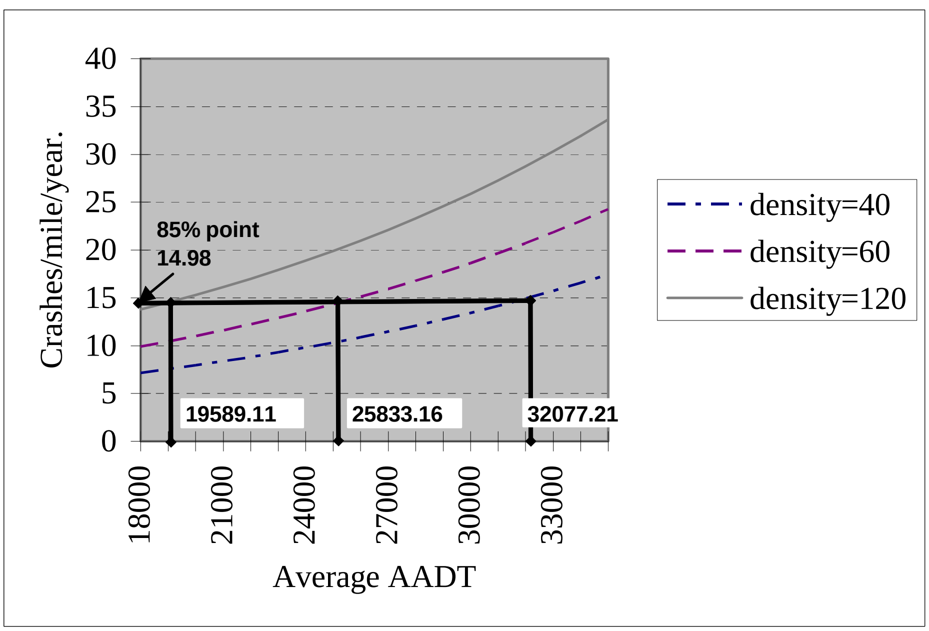


Figure D.12 Critical Values of AADT according to 85th Percentile Value for a Six-lane Road in the Lower Speed Category

APPENDIX E

CRITICAL VALUES FOR AADT

Table E.1 Critical AADT Values for Two-lane Road

for the Higher Speed Category

Access Density (points per mile)	Critical AADT (vpd) Percentile Values					
	50%	75%	80%	85%	90%	95%
10	14,163	NA	NA	NA	NA	NA
20	12,602	NA	NA	NA	NA	NA
30	11,041	NA	NA	NA	NA	NA
40	9,480	NA	NA	NA	NA	NA
50	7,919	28,565	NA	NA	NA	NA
60	6,358	27,004	NA	NA	NA	NA
70	4,797	25,443	28,518	NA	NA	NA
80	3,236	23,882	26,957	NA	NA	NA
90	1,675	22,321	25,396	28,700	NA	NA
100	114	20,760	23,835	27,139	NA	NA
110	-	19,199	22,274	25,578	29,631	NA
120	-	17,638	20,713	24,017	28,070	NA
130	-	16,077	19,152	22,456	26,509	NA
140	-	14,516	17,591	20,895	24,948	29,640

NA: Not Applicable. The critical AADT is above 30,000 vpd.

Table E.2 Critical AADT Values for Four-lane Road for the Higher Speed Category

Access Density (points per mile)	Critical AADT (vpd) Percentile Values					
	50%	75%	80%	85%	90%	95%
10	30,073	40,611	42,989	45,640	48,524	53,494
20	28,512	39,050	41,428	44,079	46,963	51,933
30	26,951	37,489	39,867	42,518	45,402	50,372
40	25,390	35,928	38,306	40,957	43,841	48,811
50	23,829	34,367	36,745	39,396	42,280	47,250
60	22,268	32,806	35,184	37,835	40,719	45,689
70	20,707	31,245	33,623	36,274	39,158	44,128
80	19,146	29,684	32,062	34,713	37,597	42,567
90	17,585	28,123	30,501	33,152	36,036	41,006
100	16,024	26,562	28,940	31,591	34,475	39,445
110	14,463	25,001	27,379	30,030	32,914	37,884
120	12,902	23,440	25,818	28,469	31,353	36,323
130	11,341	21,879	24,257	26,908	29,792	34,762
140	9,780	20,318	22,696	25,347	28,231	33,201

Table E.3 Critical AADT Values for Six-lane Road

for the Higher Speed Category

Access Density (points per mile)	Critical AADT (vpd) Percentile Values					
	50%	75%	80%	85%	90%	95%
10	34,217	42,876	44,967	47,373	49,642	54,752
20	32,656	41,315	43,406	45,812	48,081	53,191
30	31,095	39,754	41,845	44,251	46,520	51,630
40	29,534	38,193	40,284	42,690	44,959	50,069
50	27,973	36,632	38,723	41,129	43,398	48,508
60	26,412	35,071	37,162	39,568	41,837	46,947
70	24,851	33,510	35,601	38,007	40,276	45,386
80	23,290	31,949	34,040	36,446	38,715	43,825
90	21,729	30,388	32,479	34,885	37,154	42,264
100	20,168	28,827	30,918	33,324	35,593	40,703
110	18,607	27,266	29,356	31,763	34,032	39,142
120	17,046	25,705	27,795	30,202	32,471	37,581
130	15,485	24,144	26,234	28,641	30,910	36,020
140	13,924	22,583	24,673	27,080	29,349	34,459

Table E.4 Critical AADT Values for Two-lane Road for the Lower Speed Category

Access Density (points per mile)	Critical AADT (vpd) Percentile Values					
	50%	75%	80%	85%	90%	95%
10	5,809	23,253	26,723	NA	NA	NA
20	4,248	21,692	25,162	28,835	NA	NA
30	2,687	20,131	23,601	27,274	NA	NA
40	1,126	18,570	22,040	25,713	NA	NA
50	-	17,009	20,479	24,152	28,466	NA
60	-	15,448	18,918	22,591	26,905	NA
70	-	13,887	17,357	21,030	25,344	NA
80	-	12,326	15,796	19,469	23,783	29,557
90	-	10,765	14,235	17,908	22,222	27,996
100	-	9,204	12,674	16,347	20,661	26,435
110	-	7,643	11,113	14,786	19,100	24,874
120	-	6,082	9,552	13,225	17,539	23,313
130	-	4,521	7,991	11,664	15,978	21,752
140	-	2,959	6,430	10,103	14,417	20,191

NA: Not Applicable. The critical AADT is above 30,000 vpd.

Table E.5 Critical AADT Values for Four-lane Road for the Lower Speed Category

Access Density (points per mile)	Critical AADT (vpd) Percentile Values					
	50%	75%	80%	85%	90%	95%
10	18,770	29,592	32,128	35,004	38,574	43,693
20	17,209	28,031	30,567	33,443	37,013	42,132
30	15,648	26,470	29,006	31,882	35,452	40,571
40	14,087	24,909	27,445	30,321	33,891	39,010
50	12,526	23,348	25,884	28,760	32,330	37,449
60	10,965	21,787	24,323	27,199	30,769	35,888
70	9,404	20,226	22,762	25,638	29,208	34,327
80	7,843	18,665	21,201	24,077	27,647	32,765
90	6,282	17,104	19,640	22,516	26,086	31,204
100	4,720	15,543	18,079	20,955	24,525	29,643
110	3,159	13,982	16,518	19,394	22,964	28,082
120	1,598	12,421	14,957	17,833	21,403	26,521
130	37	10,860	13,396	16,272	19,842	24,960
140	-	9,299	11,835	14,711	18,281	23,399

Table E.6 Critical AADT Values for Six-lane Road

for the Lower Speed Category

Access Density (points per mile)	Critical AADT (vpd) Percentile Values					
	50%	75%	80%	85%	90%	95%
10	23,386	32,088	34,236	36,760	39,979	44,773
20	21,825	30,527	32,675	35,199	38,418	43,212
30	20,264	28,966	31,114	33,638	36,857	41,651
40	18,703	27,405	29,553	32,077	35,296	40,090
50	17,142	25,844	27,992	30,516	33,735	38,529
60	15,581	24,283	26,431	28,955	32,174	36,968
70	14,020	22,722	24,870	27,394	30,613	35,407
80	12,459	21,161	23,309	25,833	29,052	33,846
90	10,898	19,600	21,748	24,272	27,491	32,285
100	9,337	18,039	20,187	22,711	25,930	30,724
110	7,776	16,478	18,626	21,150	24,369	29,163
120	6,215	14,917	17,065	19,589	22,808	27,602
130	4,654	13,356	15,504	18,028	21,247	26,041
140	3,093	11,795	13,943	16,467	19,686	24,480

APPENDIX F

TWLTL SECTIONS IDENTIFIED AS CRITICAL

Table F.1 TWLTL Sections Identified as Critical for District 1

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Higher Speed	2	No sections identified for this group							
	4	No sections identified for this group							
	6	113010000	0	0.13	0.13	38	44,500	50	0
		113010000	2.14	2.38	0.24	61	43,900	45	33.88
		113010000	2.38	3.01	0.63	84	43,063	45	17.15
		113010000	3.01	4.22	1.21	67	43,364	45	33.33
		113010000	4.22	5.06	0.88	71	42,376	45	35.63
		117020000	15.14	16.37	1.23	56	50,477	45	17.34
		117040000	1.24	1.49	0.25	28	41,667	45	3.97
Lower Speed	2	112040000	5.66	5.90	0.24	58	24,167	40	4.17
	4	113010000	5.37	5.46	0.09	116	29,688	40	31.01
		113150000	6.59	8.17	1.58	76	37,977	40	9.25
		113150000	8.17	8.31	0.14	37	36,500	40	19.9
		116030000	27.77	27.89	0.12	53	32,500	40	0
		116250000	25.75	27.35	1.60	91	36,398	40	12.3
		116250000	27.35	27.50	0.15	113	26,500	30	8.83
		116250000	27.50	28.65	1.15	85	25,293	30	8.13
		116300000	0.60	0.67	0.07	81	28,000	40	13.51
		116300000	0.67	0.76	0.09	81	30,500	35	15.5
		117020000	18.69	19.00	0.31	57	34,625	40	12.74
		117120000	0.13	0.56	0.43	73	32,732	35	21.86
		117120000	0.56	1.15	0.59	64	36,652	35	12.97
		191070000	9.63	9.84	0.21	84	29,063	35	12.46
	6	112010000	21.05	23.38	2.33	84	43,259	40	16.6
		112010000	23.38	23.46	0.08	36	44,429	40	27.78
		117020000	16.37	16.98	0.61	69	51,592	40	41.33
		117020000	16.98	17.31	0.33	34	52,196	35	23.45
		117040000	0.65	0.99	0.34	74	35,000	40	1.97

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.2 TWLTL Sections Identified as Critical for District 2

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Higher Speed	2	No sections identified for this group							
	4	272230000	0.98	1.21	0.23	34	47,000	45	10.06
		272230000	1.21	1.36	0.15	53	48,833	45	13.33
	6	272100000	6.47	7.58	1.11	59	58,210	45	43.82
		272230000	0.44	0.98	0.54	58	44,422	45	19.98
Lower Speed	2	226070000	18.54	18.71	0.17	46	27,600	35	9.63
		229010000	6.34	6.58	0.24	99	26,643	35	9.64
		272110000	0.24	0.58	0.34	69	34,000	40	17.91
		272110000	0.58	1.59	1.01	78	28,625	35	5.25

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.2 TWLTL Sections Identified as Critical for District 2 (Contd.)

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Lower Speed	4	226010000	13.62	14.62	1.00	64	36,175	30	26.67
		226010000	14.62	15.33	0.71	62	34,319	35	22.13
		226010000	15.35	15.81	0.46	72	37,451	35	29.71
		226070000	19.42	20.02	0.60	75	29,892	30	33.28
		228010000	7.55	8.07	0.52	69	27,357	30	17.85
		272014000	1.12	1.74	0.62	49	43,441	35	21.24
		272014000	2.31	4.15	1.84	69	39,176	40	28.29
		272014000	4.33	4.60	0.27	11	35,500	40	0
		272014000	4.72	5.06	0.34	32	33,000	40	0
		272014000	5.06	5.29	0.23	48	33,000	35	0
		272014000	5.29	5.43	0.14	64	33,000	35	0
		272014000	5.43	5.49	0.06	97	27,833	35	16.13
		272014000	5.56	5.96	0.40	38	33,000	35	0
		272015000	1.08	2.12	1.04	27	35,052	40	15.5
		272015000	2.12	2.30	0.18	33	38,077	40	23.55
		272015000	2.30	2.37	0.07	27	36,500	40	9.13
		272028000	1.42	1.91	0.49	4	36,500	35	0
		272100000	3.20	3.28	0.08	84	41,233	40	60.24
		272100000	3.28	5.57	2.29	57	40,118	40	18.54
		272150000	2.14	3.00	0.86	35	31,991	35	21.81
		272170000	4.84	5.85	1.01	55	35,735	35	22.51
		272190000	0.50	3.20	2.70	50	31,266	40	15.06
		272190000	13.70	14.77	1.07	74	28,255	35	9.65
		272291000	2.90	3.24	0.34	79	28,417	35	11.7
	6	229010000	8.11	9.08	0.97	72	40,345	35	14.51
		272014000	1.74	1.95	0.21	105	50,960	40	39.68
		272014000	1.95	2.25	0.30	26	48,000	40	0
		272018000	6.00	6.80	0.80	61	35,412	35	35.15

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.3 TWLTL Sections Identified as Critical for District 3

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Higher Speed	2	No sections identified for this group							
	4	346020000	0.75	1.15	0.40	2	53,328	45	30.76
	6	346010000	16.18	16.68	0.50	4	53,422	45	23.81
Lower Speed	2	355100000	1.42	1.60	0.18	146	17,300	40	0
	4	348004000	8.24	8.79	0.55	38	35,507	40	41.21
		348012000	3.57	3.81	0.24	21	36,429	35	9.8
		348012000	4.08	4.44	0.36	25	36,943	35	32.14
		348012000	4.58	4.68	0.10	10	40,833	35	80.81
		348040000	9.01	9.49	0.48	2	35,600	40	14.01
		348070000	4.20	4.47	0.27	52	31,500	35	0
		348070000	4.47	6.03	1.56	42	40,500	35	0
		348070000	6.60	6.88	0.28	32	53,500	35	0
		355005000	0.75	1.25	0.50	44	31,400	40	6.64
		355040000	11.55	11.84	0.29	21	36,250	25	4.66
		355090000	1.60	2.78	1.18	45	33,601	40	33.62
		357030000	10.91	11.15	0.24	61	33,000	35	8.13
		357030000	12.24	12.48	0.24	17	43,689	35	29.91
		357040000	12.38	12.98	0.60	17	44,500	35	0
	6	348020000	10.49	10.98	0.49	55	35,759	40	18.33
		348020000	10.98	11.19	0.21	61	34,891	35	35.83
		355060000	7.68	8.54	0.86	25	42319	30	43.65
		357040000	0.66	2.22	1.56	46	45,301	30	14.52
		357040000	2.22	3.42	1.20	54	45,467	40	12.49
		357040000	3.42	4.66	1.24	31	43,500	40	2.15
		357040000	12.98	13.93	0.95	37	38,800	35	3.49

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.4 TWLTL Sections Identified as Critical for District 4

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Higher Speed	2	No sections identified for this group							
	4	486100000	8.40	8.50	0.10	10	56,125	45	51.78
		486100000	4.20	5.09	0.89	44	41,265	45	19.1
	6	486016000	5.01	5.21	0.20	59	44,000	45	8.25
Lower Speed	2	No sections identified for this group							
		486010000	0.83	2.46	1.63	63	27,467	35	6.92
		486010000	2.72	5.95	3.23	65	32,560	35	11.15
		486100000	0.00	0.69	0.69	45	43,875	40	11.54
		486100000	0.97	2.57	1.60	41	41,996	40	28.89
		486100000	2.57	4.20	1.63	49	41,948	40	25.75
		486100000	8.50	10.03	1.53	2	46,674	40	20.07
		486210000	3.01	3.15	0.14	21	28,500	40	4.69
		486210000	3.15	3.33	0.18	64	28,500	40	3.88
		488010000	4.79	5.61	0.82	67	28,364	35	8.94
		488010000	5.61	5.81	0.20	77	28,000	35	1.71
		489090000	13.85	14.58	0.73	25	41,292	35	5.5
		489092000	0.00	0.26	0.26	23	34,600	35	6.51
		493040000	0.00	0.39	0.39	82	27,909	35	9.38
		493120000	20.33	20.40	0.07	56	29,500	35	0
		493200000	8.23	9.13	0.90	73	27,545	35	14.83
		494010000	10.25	11.78	1.53	51	39,936	40	10.26
		494010000	11.78	12.23	0.45	42	36,705	40	16.19
		494010000	12.23	12.73	0.50	68	35,267	40	9.98
		494010000	12.73	13.50	0.77	47	31,786	30	6.1
		494010000	13.50	13.96	0.46	61	28,750	35	10.12
		494120000	7.85	9.29	1.44	42	35,967	40	3.49
	6	486040000	14.17	15.38	1.21	50	36,155	35	7.98
		486100000	10.22	10.32	0.10	10	49,944	40	29.41
		486200000	3.51	3.64	0.13	8	53,582	40	5.33

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.5 TWLTL Sections Identified as Critical for District 5

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Higher Speed	2	No sections identified for this group							
	4	575010000	2.03	2.30	2.27	7	51,500	55	0
		575050000	0.23	0.49	2.26	12	46,500	55	0
		575050000	5.44	5.80	0.36	53	39,600	45	4.63
		575060000	5.34	5.85	0.51	43	48,583	45	19.61
		575060000	5.85	6.79	0.94	41	48,409	45	23.5
		577120000	4.23	4.45	0.22	18	54,950	45	15.02
		577120000	4.45	4.83	0.38	16	53,450	45	8.89
		577120000	4.83	4.99	0.16	37	54,891	45	47.33
	6	511040000	4.83	5.29	0.46	37	44,688	45	11.72
		570100000	10.14	10.59	0.45	38	47,209	45	20.36
		575003000	5.00	7.08	2.08	56	56,037	45	30.4
		575003000	7.08	7.20	0.12	67	51,594	45	44.82
		575003000	7.64	7.83	0.19	48	45,091	50	39.22
		575010000	5.97	6.75	0.78	45	47,453	45	40.7
		575010000	7.25	7.58	0.33	43	44,273	45	11.25
		575010000	7.58	8.02	0.44	72	44,000	45	15.7
		575010000	8.27	8.56	0.29	79	44,353	45	19.34
		575010000	8.79	10.05	1.26	71	45,560	45	22.08
		575010000	10.05	11.07	1.02	62	50,398	45	28.76
		577010000	0.96	1.39	0.43	45	62,000	45	15.8
		592030000	0.29	0.51	0.22	64	43,000	45	9.09
		592030000	0.51	0.63	0.12	43	45,000	45	2.9
		592090000	12.76	12.87	0.11	28	60,838	45	46.3
		592090000	12.87	13.37	0.50	64	60,906	45	18.56
		592090000	13.37	13.77	0.40	52	59,952	45	18.98

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.5 TWLTL Sections Identified as Critical for District 5 (Contd.)

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Lower Speed	2	No sections identified for this group							
	4	511040000	2.32	2.72	0.40	84	30,333	35	5.06
		511040000	4.02	4.10	0.08	63	28,500	35	16.67
		536010000	14.71	14.82	0.11	116	28,000	40	2.98
		536010000	14.82	15.23	0.41	94	29,222	40	7.44
		536080000	0.47	0.81	0.34	79	27,400	35	9.8
		570140000	1.06	1.46	0.40	63	29,750	40	20
		575006000	0.14	0.93	0.79	71	33,833	35	6.32
		575006000	0.93	1.03	0.10	114	31,333	35	19.05
		575010000	12.35	12.90	0.55	41	31,385	35	23.42
		575030000	4.88	5.98	1.10	63	34,581	35	11.25
		575040000	11.86	12.28	0.42	35	34,332	35	22.59
		575040000	12.28	12.53	0.25	58	30,000	35	2.75
		575050000	15.44	15.79	0.35	63	33,000	40	0
		575060000	0.00	1.88	1.88	43	49,227	40	26.95
		575060000	1.88	2.18	0.30	63	45,444	40	20
		575060000	2.18	2.65	0.47	59	46,776	40	20.44
		575060000	19.65	20.31	0.66	82	46,000	35	0
		575060000	20.31	20.80	0.49	57	48,000	35	0
		575260000	3.00	3.08	0.08	88	26,000	40	0
		579030000	3.23	4.28	1.05	66	30,290	40	9.85
		579030000	4.30	5.52	1.22	73	27,444	40	2.47
		579040000	7.52	8.20	0.68	66	30,219	40	7.84
		579040000	8.20	8.27	0.07	62	31,000	40	10.26
		592010000	11.73	12.23	0.50	111	25,967	40	9.94

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.5 TWLTL Sections Identified as Critical for District 5 (Contd.)

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Lower Speed	6	511040000	4.62	4.83	0.21	106	45,000	35	12.82
		536001000	24.07	24.96	0.89	79	29,563	35	12
		536003000	0.00	0.67	0.67	67	36,050	35	4.95
		570010000	16.24	17.17	0.93	71	36,938	40	5.75
		570020000	0.00	0.41	0.41	68	41,643	40	5.65
		570020000	0.41	0.83	0.42	62	40,500	40	3.2
		570020000	2.72	3.87	1.15	69	47,250	40	11.07
		570020000	3.96	4.22	0.26	85	46,250	35	5.13
		570020000	4.22	4.51	0.29	103	37,938	40	9.16
		575250000	5.90	6.02	0.12	34	40,000	40	11.3
		592030000	0.00	0.29	0.29	76	42,875	40	9.2
		592090000	13.77	14.07	0.30	71	60,982	40	12.39
		592090000	14.07	14.95	0.88	74	49,337	40	29.55
		592090000	14.95	15.38	0.43	55	48,935	40	23.7

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.6 TWLTL Sections Identified as Critical for District 6

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Higher Speed	2	No sections identified for this group							
	4	No sections identified for this group							
	6	No sections identified for this group							
Lower Speed	2	No sections identified for this group							
	4	687001000	7.65	7.87	0.22	57	37,143	40	10.23
		687008000	7.90	10.12	2.22	66	27,787	40	11.26
		687030000	11.72	11.86	0.14	15	35,214	30	17.28
		687030000	12.89	13.84	0.94	33	36,225	30	17.99
		687030000	15.57	16.14	0.57	34	33,375	35	14.13
		687030000	16.14	16.50	0.36	47	39,688	35	7.41
		687030000	18.06	19.26	1.20	41	31,256	35	24.92
		687044000	7.98	8.47	0.49	59	37,720	35	17.08
		687053000	1.66	6.03	4.37	57	36,269	40	15.04
		687062000	4.57	5.06	0.49	51	35,522	40	15.52
		687072000	3.86	5.16	1.30	57	29,377	40	18.76
		687090000	10.41	10.51	0.10	20	42,125	40	40
		687090000	10.51	12.26	1.75	27	40,305	40	21.51
		687090000	12.26	13.49	1.23	17	39,628	40	20.05
		687120000	10.25	13.16	2.91	59	39,591	35	21.96
		687120000	13.39	14.39	1.00	43	33,525	35	27.05
		687140000	5.65	5.71	0.06	82	29,500	40	60.11
		687281000	0.00	2.62	2.62	21	35,924	40	4.2
		687281000	5.84	6.66	0.82	25	36,281	35	6.46
		687281000	6.66	8.19	1.53	45	35,966	40	9.64
	6	687030000	11.42	11.72	0.30	20	38,571	30	7.7
		687281000	5.65	5.82	0.17	23	40,000	35	13.18

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.

Table F.7 TWLTL Sections Identified as Critical for District 7

Threshold selected: 85 percentile value

Posted Speed Level	Number of Lanes	Road ID	Begin Milepost	End Milepost	Section Length	Access Density	AADT	Posted Speed	Actual Average Crashes
Higher Speed	2	No sections identified for this group							
	4	710130000	3.81	3.87	0.06	132	29,325	45	98.04
		710130000	4.50	4.80	0.30	44	41,036	45	15.77
		715120000	5.65	5.72	0.07	27	53,500	50	9.13
	6	710030000	1.55	1.70	0.15	112	43,125	45	9.32
		710030000	1.90	2.17	0.37	76	43,040	45	22.64
		710030000	2.36	3.01	0.65	73	42,771	45	17.84
		710130000	8.83	8.93	0.10	61	49,375	45	13.47
		710130000	9.60	9.79	0.19	78	64,984	45	55.27
		710160000	0.27	0.65	0.38	63	74,224	45	25.37
		710160000	0.90	1.24	0.34	62	78,722	45	17.6
		710160000	1.24	1.94	0.70	44	77,500	45	0
		715040000	4.69	4.84	0.15	13	50,833	45	6.67
		715040000	4.95	5.54	0.59	71	52,850	45	16.84
		715040000	5.84	5.91	0.07	43	52,167	45	14.29
Lower Speed	2	715020000	0.00	1.07	1.07	79	21,692	30	4.04
		715020000	3.67	4.95	1.28	83	25,167	40	1.56
		715020000	7.57	9.23	1.66	33	28,174	40	4.62
	4	715007000	3.54	3.80	0.26	83	28,100	40	6.56
		715040000	0.80	1.03	0.23	82	25,700	35	14.43
		715090000	1.25	2.23	0.97	99	24,709	35	3.77
		715090000	2.23	2.34	0.11	117	29,667	40	9.01
	6	702000000	14.20	14.55	0.35	99	26,450	40	9.75
		710030000	0.05	0.45	0.40	86	41,568	40	36.94
		710030000	0.55	0.90	0.35	79	41,688	40	30.05
		710030000	1.06	1.44	0.38	61	42,921	40	55.7

Note: The evaluation of any TWLTL section on each side of the sections identified, at a lightly lower percentage level, may provide the identification of total segment that are longer and better suited for median treatment projects.