2009

Comprehensive Analytical Investigation Of The Safety Of Unsignalized Intersections

Kirolos Haleem
University of Central Florida

COMPREHENSIVE ANALYTICAL INVESTIGATION OF THE SAFETY OF UNSIGNALIZED INTERSECTIONS

by
KIROLOS MAGED HALEEM

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental & Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2009

Major Professor: Mohamed A. Abdel-Aty

ABSTRACT

According to documented statistics, intersections are among the most hazardous locations on roadway systems. Many studies have extensively analyzed safety of signalized intersections, but did not put their major focus on the most frequent type of intersections, unsignalized intersections. Unsignalized intersections are those intersections with stop control, yield control and no traffic control. Unsignalized intersections can be differentiated from their signalized counterparts in that their operational functions take place without the presence of a traffic signal. In this dissertation, multiple approaches of analyzing safety at unsignalized intersections were conducted. This was investigated in this study by analyzing total crashes, the most frequent crash types at unsignalized intersections (rear-end as well as angle crashes) and crash injury severity. Additionally, an access management analysis was investigated with respect to the different median types identified in this study. Some of the developed methodological techniques in this study are considered recent, and have not been extensively applied.

In this dissertation, the most extensive data collection effort for unsignalized intersections was conducted. There were 2500 unsignalized intersections collected from six counties in the state of Florida. These six counties were Orange, Seminole, Hillsborough, Brevard, Leon and Miami-Dade. These selected counties are major counties representing the central, western, eastern, northern and southern parts in Florida, respectively. Hence, a geographic representation of the state of Florida was achieved. Important intersections' geometric and roadway features, minor approach traffic control, major approach traffic flow and crashes were obtained.

The traditional negative binomial (NB) regression model was used for modeling total crash frequency for two years at unsignalized intersections. This was considered since the NB technique is well accepted for modeling crash count data suffering from over-dispersion. The NB models showed several important variables affecting safety at unsignalized intersections. These include the traffic volume on the major road and the existence of stop signs, and among the geometric characteristics, the configuration of the intersection, number of right and/or left turn lanes, median type on the major road, and left and right shoulder widths. Afterwards, a new approach of applying the Bayesian updating concept for better crash prediction was introduced. Different non-informative and informative prior structures using the NB and log-gamma distributions were attempted. The log-gamma distribution showed the best prediction capability.

Crash injury severity at unsignalized intersections was analyzed using the ordered probit, binary probit and nested logit frameworks. The binary probit method was considered the best approach based on its goodness-of-fit statistics. The common factors found in the fitted probit models were the logarithm of AADT on the major road, and the speed limit on the major road. It was found that higher severity (and fatality) probability is always associated with a reduction in AADT, as well as an increase in speed limit.

A recently developed data mining technique, the multivariate adaptive regression splines (MARS) technique, which is capable of yielding high prediction accuracy, was used to analyze rear-end as well as angle crashes. MARS yielded the best prediction performance while dealing with continuous responses. Additionally, screening the covariates using random forest before fitting MARS model was very encouraging.

Finally, an access management analysis was performed with respect to six main median types associated with unsignalized intersections/access points. These six median types were open, closed, directional (allowing access from both sides), two-way left turn lane, undivided and mixed medians (e.g., directional median, but allowing access from one side only). Also, crash conflict patterns at each of these six medians were identified and applied to a dataset including median-related crashes. In this case, separating median-related and intersection-related crashes was deemed significant in the analysis. From the preliminary analysis, open medians were considered the most hazardous median type, and closed and undivided medians were the safest. The binomial logit and bivariate probit models showed significant median-related variables affecting median-related crashes, such as median width, speed limit on the major road, logarithm of AADT, logarithm of the upstream and downstream distances to the nearest signalized intersection and crash pattern.

The results from the different methodological approaches introduced in this study could be applicable to diagnose safety deficiencies identified. For example, to reduce crash severity, prohibiting left turn maneuvers from minor intersection approaches is recommended. To reduce right-angle crashes, avoiding installing two-way left turn lanes at 4-legged intersections is essential. To reduce conflict points, closing median openings across from intersections is recommended. Since left-turn and angle crash patterns were the most dominant at undivided medians, it is recommended to avoid left turn maneuvers at unsignalized intersections having undivided medians at their approach. This could be enforced by installing a left-turn prohibition sign on both major and minor approaches.

ACKNOWLEDGEMENTS

First of all, I would like to thank God, Virgin Mary and all the Saints, whom without their blessings, I could not achieve half of what I achieved till now.

I would like to thank my advisor, Dr. Mohamed Abdel-Aty for all his guidance and assistance during my research work at UCF. He helped me to change my way of thinking from that of a student to that of a highly-academic researcher. I will be grateful to him all my entire life.

I would like to thank all my committee members, in no particular order, Dr. Essam Radwan, Dr. Haitham Al-Deek, Dr. Kevin Mackie and Dr. Xiaogang Su. They always help me with very outstanding advices.

I would like to thank all my friends (at UCF and outside) for their continuous support to me throughout my post-graduate study at UCF. I am proud to have friends like them.

A very special thanks goes to my family (my parents and my brother) who - continuously without seizing - encouraged and supported me a lot throughout my entire life. Without their advices and support, I could achieve nothing. Many thanks to your prayers. I owe you my life.

TABLE OF CONTENTS

LIST OF FIGURES xi
LIST OF TABLES xv
LIST OF ACRONYMS/ABBREVIATIONS xvii
LIST OF ACRONYMS/ABBREVIATIONS xvii
CHAPTER 1. INTRODUCTION 1
1.1 Overview 1
1.2 Research Objectives 4
1.3 Dissertation Organization 6
CHAPTER 2. LITERATURE REVIEW 7
2.1 Significant Factors Contributing to Safety of Unsignalized Intersections 7
2.1.1 Left and Right-Turn Lanes 7
2.1.2 Number of Intersecting Legs 8
2.1.3 Land Use 9
2.1.4 Intersection Skewness 9
2.1.5 Median Width 9
2.1.6 Lighting 10
2.1.7 Channelization 10
2.1.8 Intersection Sight Distance 11
2.1.9 Number of Approach Lanes 11
2.1.10 Shoulder Width 12
2.1.11 Vertical and Horizontal Alignment on Approaches 12
2.1.12 Traffic Flow 12
2.1.13 Traffic Control Type 12
2.1.14 Size of Intersection 13
2.1.15 Minor Road Approach Geometry 13
2.1.16 Grades 13
2.1.17 Signing and Delineation 14
2.1.18 Spacing between Intersections 14
2.1.19 Pedestrian Crossing Facilities 14
2.1.20 Speed Parameters 15
2.1.21 Beacons Use 15
2.1.22 Turn Lanes Configuration 15
2.1.23 Pavement Condition 15
2.2 Some Facts about Unsignalized Intersections 16
2.3 Types of Crashes Occurring at Unsignalized Intersections and their Modeling Scheme 18
2.4 Analysis of Unsignalized Intersections 19
2.5 Safety Effectiveness of Converting Unsignalized Intersections to Signalized Ones 20
2.6 Studies Using the NB Formulation 21
2.7 Studies Analyzing Injury Severity 23
2.8 Studies Related to Crash Prediction 25
2.9 Studies Using Advanced Prediction Techniques 26
2.10 Access Management and Traffic Safety 27

2.10.1 Safety of Median Openings ... 28
2.10.2 Safety of Left-Turn Lanes ... 30
2.10.3 Safety of U-turns ... 31
2.10.4 Studies on Safety of Some Median Types ... 31
2.10.5 Access Management Design Policies ... 32
CHAPTER 3. DATA COLLECTION ... 33
3.1 Introduction ... 33
3.2 Variables Description ... 36
3.3 Median Classification ... 44
3.3.1 Closed Median ... 44
3.3.2 Directional Median ... 45
3.3.3 Open Median ... 46
3.3.4 Two-Way Left Turn Lane Median ... 47
3.3.5 Undivided Median ... 48
3.3.6 Median with Markings ... 49
3.4 Initial Data Collection Procedure ... 51
3.5 Final Data Collection Procedure ... 53
3.5.1 Orange County ... 53
3.5.2 Brevard County ... 55
3.5.3 Hillsborough County ... 56
3.5.4 Miami-Dade County ... 58
3.5.5 Leon County ... 61
3.5.6 Seminole County ... 65
3.6 Summary Table for the Data Collection Procedure throughout the Six Selected Counties ... 68
3.7 Preliminary Categorization of Unsignalized Intersections ... 72
CHAPTER 4. PRELIMINARY ANALYSIS ... 82
4.1 Safety Effect of the Presence of Both Stop Sign and Line, and Stop Sign Only at Intersections in Orange County ... 84
4.2 General Conclusions and Recommendations from the Analysis ... 88
CHAPTER 5. USING A RELIABILITY PROCESS TO REDUCE UNCERTAINTY IN PREDICTING CRASHES ... 90
5.1 Background ... 90
5.2 Methodological Approach ... 92
5.3 Description of Variables ... 99
5.4 Data Preparation ... 105
5.5 Fitted NB Regression Models (Safety Performance Functions) ... 106
5.5.1 Modeling Crash Frequency at 3-Legged Intersections ... 106
5.5.2 Modeling Crash Frequency at 4-Legged Intersections ... 109
5.6 Bayesian Updating Models for 3 and 4-Legged Unsignalized Intersections .. 111
5.6.1 Evaluating NB and Bayesian Models Using Mean Estimates of Parameters 117
5.6.2 Evaluating NB and Bayesian Models Using Uncertainty Estimates ... 120
5.7 General Conclusions from the Reliability Process in Terms of the Bayesian Updating Framework ... 122
CHAPTER 6. CRASH INJURY SEVERITY ANALYSIS ... 124

6.1 Background 124
6.2 Methodological Approach: Probit Model Specification 125
6.3 Data Preparation 128
6.4 Variables' Description 129
6.5 Analysis of the Ordered Probit Framework 133
6.5.1 Three-Legged Model Interpretation 138
6.5.2 Four-Legged Model Interpretation 141
6.6 Analysis of the Binary Probit Framework 142
6.6.1 Three-Legged Model Interpretation 147
6.6.2 Four-Legged Model Interpretation 149
6.7 Comparing the Two Probit Frameworks 150
6.8 Nested Logit Model Estimates 150
6.9 Summary of Results 152
6.10 General Conclusions from the Crash Severity Analysis 154
CHAPTER 7. APPLICATION OF THE MULTIVARIATE ADAPTIVE REGRESSION SPLINES FOR PREDICTING CRASH OCCURRENCE 155
7.1 Introduction 155
7.2 Methodological Approach 156
7.2.1 Multivariate Adaptive Regression Splines Model Characteristics 156
7.2.2 Random Forest Technique 160
7.2.3 Assessing Prediction Performance 161
7.3 MARS Applications 162
7.4 Data Preparation and Variables' Description 163
7.5 Modeling Rear-end Crash Frequency at 3 and 4-Legged Unsignalized Intersections Using the NB Formulation 174
7.5.1 Three-Legged Model Interpretation 179
7.5.2 Four-Legged Model Interpretation 180
7.6 Comparing MARS and NB Models 184
7.7 Examining Fitting MARS Model with Continuous Response 185
7.8 Using MARS in Conjunction with Random Forest 186
7.9 Predicting Angle Crashes Using the MARS Technique 189
7.9.1 Modeling Angle Crash Frequency at 3 and 4-Legged Unsignalized Intersections Using the NB Technique 189
7.9.2 Comparing MARS and NB Models 197
7.9.3 Examining Fitting MARS Model with Continuous Response 198
7.9.4 Using MARS in Conjunction with Random Forest 199
7.10 General Conclusions from the MARS Analysis 202
CHAPTER 8. ACCESS MANAGEMENT ANALYSIS 204
8.1 Introduction 204
8.2 Preliminary Analysis: Comparing Crashes at Different Median Types 205
8.3 Possible Median-related Crashes at Different Median Types 208
8.4 Screening for Median-related Crashes in the Dataset 216
8.5 Preliminary Methodological Approach: Multinomial Logit Framework 225
8.6 Multinomial Logit Model Estimation 226
8.7 Second Methodological Approach: Bivariate Probit Framework 229
8.8 Bivariate Probit Model Estimation 230

8.9 General Conclusions from the Access Management Analysis 232
CHAPTER 9. CONCLUSIONS.. 234
9.1 Summary and Contributions 234
9.2 Research Applications 239
9.3 Further Research 241
LIST OF REFERENCES... 243

LIST OF FIGURES

Figure 3-1: Pie Chart for County Distribution in Each of the 7 Districts in Florida 34
Figure 3-2: Conceptual Road Layout for "Major 1, Major 2, Minor 1 and Minor 2" Approaches 37
Figure 3-3: A 2x2 Unsignalized Intersection with a Closed Median 44
Figure 3-4: Two 2x6 Unsignalized Intersections with a Dual Directional Median 45
Figure 3-5: An Aerial Photo of a One-Way Directional Median 46
Figure 3-6: A 2x6 Four-Legged Unsignalized Intersection with an Open Median 47
Figure 3-7: A 2x4 Three-Legged Unsignalized Intersection with a Two-Way Left Turn Lane Median 48
Figure 3-8: A 2x2 Four-Legged Unsignalized Intersection with a Two-Way Left Turn Lane Median 48
Figure 3-9: A 2x4 Three-Legged Unsignalized Intersection with an Undivided Median 49
Figure 3-10: A 2x2 Four-Legged Unsignalized Intersection with an Undivided Median 49
Figure 3-11: A 2x2 Three-Legged Unsignalized Intersection with Pavement Markings as a Median 50
Figure 3-12: A 2x2 Four-Legged Unsignalized Intersection with Pavement Markings as a Median 50
Figure 3-13: Aerial Image from "Google Earth" for 7 Unsignalized Intersections along SR 50 in Orange County during the Initial Data Collection Procedure 51
Figure 3-14: Screen Shot of the "Video Log Viewer Application" from FDOT's RCI Database 52
Figure 3-15: Aerial Image from "Google Earth" for 3 Unsignalized Intersections in Orange County where Difficulty was Encountered in Identifying their Geometry due to Tree Blockage 55
Figure 3-16: An Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Brevard County 56
Figure 3-17: First Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Hillsborough County 57
Figure 3-18: Second Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Hillsborough County 58
Figure 3-19: First Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Miami-Dade County 59
Figure 3-20: Second Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Miami-Dade County 60
Figure 3-21: Third Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Miami-Dade County 61
Figure 3-22: Fourth Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Miami-Dade County 61
Figure 3-23: First Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Leon County 63
Figure 3-24: Second Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Leon County 63
Figure 3-25: Third Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Leon County 64

Figure 3-26: Fourth Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Leon County 64
Figure 3-27: Fifth Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Leon County 65
Figure 3-28: First Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Seminole County 67
Figure 3-29: A Further View of the Unsignalized Intersection in Figure 3-28 for Better Clarification of the Extension of the Two Left-Turn Lanes to the Signalized Intersection 67
Figure 3-30: Second Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Seminole County 68
Figure 3-31:Conceptual Flow Chart for the Maximum Preliminary Number of Categories at Unsignalized Intersections 77
Figure 3-32: Conceptual Flow Chart for the Aggregated Preliminary Number of Categories at Unsignalized Intersection 79
Figure 4-1: Plot of the Average Total Crash per Intersection Associated with Each Median Type 82
Figure 4-2: Plot of the Average Total Crash per Intersection Associated with Each Median Type 83
Figure 4-3: Plot of the Average Total Crash per Intersection Associated with Each Median Type 84
Figure 4-4: Rumble Strips Installation 89
Figure 5-1: Mixed Median Type on the Major Road (Directional from One Side, and Closed from the Other Side) (Retrieved January 20, 2008, from Google Earth) ... 101
Figure 5-2: Distribution of Intersections by the Natural Logarithm of AADT at 3 and 4-Legged Intersections 104
Figure 5-3: Distribution of Intersections by the Right Shoulder Width (in "feet") at 3 and 4-Legged Intersections 104
Figure 5-4: Plot of the Residuals vs. Log_AADT on the Major Road at 3-Legged Unsignalized Intersections 119
Figure 5-5: Plot of the Residuals vs. Log_AADT on the Major Road at 4-Legged Unsignalized Intersections 120
Figure 6-1: First Attempted Two-level Nesting Structure for the Nested Logit Framework 151
Figure 6-2: Second Attempted Two-level Nesting Structure for the Nested Logit Framework 151
Figure 7-1: Plot of the Distance between Successive Unsignalized Intersections and Rear-end Crashes at 3-Legged Intersections 166
Figure 7-2: Plot of the Upstream Distance to the Nearest Signalized Intersection and Rear-end Crashes at 3-Legged Intersections 167
Figure 7-3: Plot of the Downstream Distance to the Nearest Signalized Intersection and Rear-end Crashes at 3-Legged Intersections 168
Figure 7-4: Plot of the Distance between Successive Unsignalized Intersections and Rear-end Crashes at 4-Legged Intersections 168
Figure 7-5: Plot of the Upstream Distance to the Nearest Signalized Intersection and Rear-end Crashes at 4-Legged Intersections 169

Figure 7-6: Plot of the Downstream Distance to the Nearest Signalized Intersection and Rear-end Crashes at 4-Legged Intersections.. 170
Figure 7-7: Plot of the Distance between Successive Unsignalized Intersections and Angle Crashes at 3-Legged Intersections 170
Figure 7-8: Plot of the Upstream Distance to the Nearest Signalized Intersection and Angle Crashes at 3-Legged Intersections 171
Figure 7-9: Plot of the Downstream Distance to the Nearest Signalized Intersection and Angle Crashes at 3-Legged Intersections 172
Figure 7-10: Plot of the Distance between Successive Unsignalized Intersections and Angle Crashes at 4-Legged Intersections 172
Figure 7-11: Plot of the Upstream Distance to the Nearest Signalized Intersection and Angle Crashes at 4-Legged Intersections 173
Figure 7-12: Plot of the Downstream Distance to the Nearest Signalized Intersection and Angle Crashes at 4-Legged Intersections 174
Figure 7-13: Plot of the Basis Function for "Log_AADT"... 183
Figure 7-14: Plot of the OOB Error Rate against Different Number of Trees 186
Figure 7-15: Variable Importance Ranking Using Node Purity Measure 187
Figure 7-16: Plot of the Basis Function for "Log_AADT"... 197
Figure 7-17: Plot of the OOB Error Rate against Different Number of Trees 199
Figure 7-18: Variable Importance Ranking Using Node Purity Measure 200
Figure 8-1: Plot of the Number of Intersections Associated with Each Median Type (Based on the Collected Data) 205
Figure 8-2: Plot of the Average Total Crash per Intersection in Four Years Associated with Each Median Type (Based on the Collected Data)....................................... 206
Figure 8-3: Plot of the Average Total Crash per Intersection in Four Years for the Five Most Frequent Crash Types Associated with Each Median Type (Based on the Collected Data) 207
Figure 8-4: Possible Median-Related Crash Patterns at Open Medians at 4-legged Intersections 212
Figure 8-5: Possible Median-Related Crash Patterns at Undivided Medians at 4-legged Intersections 213
Figure 8-6: Possible Median-Related Crash Patterns at Two-Way Left Turn Medians at 4-legged Intersections 214
Figure 8-7: Possible Median-Related Crash Patterns at Directional Medians at 3-legged Intersections 215
Figure 8-8: Possible Median-Related Crash Patterns at Mixed Medians at 3-legged Intersections 215
Figure 8-9: A Sketched Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 8 at Two-way Left Turn Lanes (Retrieved from "Hummingbird" Intranet Website) 216
Figure 8-10: Reported Left-turn Crash by the Officer for the Crash in Figure 8-9 217
Figure 8-11: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 4 at Open Medians (Retrieved from "Hummingbird" Intranet Website) ... 217
Figure 8-12: Reported Left-turn Crash by the Officer for the Crash in Figure 8-11 218

Figure 8-13: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 9 at Two-way Left Turn Lanes (Retrieved from "Hummingbird" Intranet Website) 218
Figure 8-14: Reported Left-turn Crash by the Officer for the crash in Figure 8-13 219
Figure 8-15: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 10 for Two-vehicle Crashes other than the Nine Identified Crash Patterns (Retrieved from "Hummingbird" Intranet Website) 220
Figure 8-16: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 10 for Multi-vehicle Crashes other than the Nine Identified Crash Patterns (Retrieved from "Hummingbird" Intranet Website) 220
Figure 8-17: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 11 for Single-vehicle Crashes (Retrieved from "Hummingbird" Intranet Website) 221
Figure 8-18: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 11 for Single-vehicle Crashes (Retrieved from "Hummingbird" Intranet Website) 222

LIST OF TABLES

Table 3-1: County Distribution in Each District in Florida State according to FDOT..... 33
Table 3-2: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Orange County.. 53
Table 3-3: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Brevard County 55
Table 3-4: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Hillsborough County 57
Table 3-5: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Miami-Dade County 59
Table 3-6: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Leon County 62
Table 3-7: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Seminole County 66
Table 3-8: Summary Table for the Data Collection Procedure in Orange County* 69
Table 3-9: Summary Table for the Data Collection Procedure in Brevard County*........ 69
Table 3-10: Summary Table for the Data Collection Procedure in Hillsborough County* 70
Table 3-11: Summary Table for the Data Collection Procedure in Miami-Dade County* 70
Table 3-12: Summary Table for the Data Collection Procedure in Leon County* 71
Table 3-13: Summary Table for the Data Collection Procedure in Seminole County* ... 71
Table 4-1: Summary Descriptive Statistics for Group 1 and Group 2............................. 85
Table 5-1: Variables Description for 3 and 4-Legged Unsignalized Intersections......... 102
Table 5-2: NB Crash Frequency Model at 3-Legged Unsignalized Intersections.......... 107
Table 5-3: NB Crash Frequency Model at 4-Legged Unsignalized Intersections.......... 110
Table 5-4: Updated Parameter Estimates for the 3-Legged Model After Using the Bayesian Updating Framework (for 4 Different Structures) 113
Table 5-5: Updated Parameter Estimates for the 4-Legged Model After Using the Bayesian Updating Framework (for 4 Different Structures) 115
Table 5-6: MOE Values for the Five 3-Legged Models (Before and After Bayesian Updating) 118
Table 5-7: MOE Values for the Five 4-Legged Models (Before and After Bayesian Updating) 118
Table 5-8: Assessing Uncertainty Reduction After Applying the Bayesian Updating Framework for Structure 4 by Comparing Standard Errors Before and After Updating for the 3-Legged Model 121
Table 5-9: Assessing Uncertainty Reduction After Applying the Bayesian Updating Framework for Structure 4 by Comparing Standard Errors Before and After Updating for the 4-Legged Model 121
Table 6-1: Variables Description for 3 and 4-Legged Unsignalized Intersections * 131
Table 6-2: Ordered Probit Estimates for 3 and 4-Legged Unsignalized Intersections ... 134
Table 6-3: Marginal Effects for Fatal Injury Probability for the Fitted Covariates in the 3 and 4-Legged Models 136

Table 6-4: Binary Probit Estimates for 3 and 4-Legged Unsignalized Intersections 143
Table 6-5: Marginal Effects for Severe Injury Probability for the Fitted Covariates in the 3 and 4-Legged Models 145
Table 6-6: Nested Logit Estimates for 3-Legged Unsignalized Intersections (Nesting Structure Shown in Figure 6-2) 152
Table 6-7: Important Factors Affecting Fatal (Severe) Injury at Unsignalized Intersections 153
Table 7-1: Summary Statistics for Rear-end Cashes in the Training and Test Databases in “2003-2006” 165
Table 7-2: Summary Statistics for Angle Crashes in the Training and Test Databases in "2003-2006" 165
Table 7-3: Variables Description for 3 and 4-Legged Unsignalized Intersections 175
Table 7-4: Rear-end Crash Frequency Model at 3 and 4-Legged Unsignalized Intersections 177
Table 7-5: Rear-end Crash Frequency Model at 4-Legged Unsignalized Intersections Using MARS 181
Table 7-6: Comparison between the Fitted MARS and NB Models in terms of Prediction and Fitting 184
Table 7-7: Prediction and Fitting Performance of the Two MARS Models Using a Continuous Response Formulation 185
Table 7-8: MARS Model at 3-Legged Unsignalized Intersections after Screening the Variables Using Random Forest 188
Table 7-9: Prediction and Fitting Assessment Criteria for the Two MARS Models after Screening the Variables Using Random Forest 189
Table 7-10: Angle Crash Frequency Model at 3 and 4-Legged Unsignalized Intersections 190
Table 7-11: Angle Crash Frequency Model at 4-Legged Unsignalized Intersections Using MARS 195
Table 7-12: Comparison between the Fitted MARS and NB Models in terms of Prediction and Fitting 197
Table 7-13: Prediction and Fitting Performance of the Two MARS Models Using a Continuous Response Formulation 198
Table 7-14: MARS Model at 4-Legged Unsignalized Intersections after Screening the Variables Using Random Forest 201
Table 7-15: Prediction and Fitting Assessment Criteria for the Two MARS Models after Screening the Variables Using Random Forest 202
Table 8-1: Possible Median-related Crash Conflicts at 4-legged Unsignalized Intersections 210
Table 8-2: Possible Median-related Crash Conflicts at 3-legged Unsignalized Intersections 211
Table 8-3: A 2x2 Contingency Table for Median Type by Crash Pattern 224
Table 8-4: Binomial Logit Model for Restricted Medians (Baseline is Unrestricted Medians) 228
Table 8-5: Bivariate Probit Model Estimates 231

LIST OF ACRONYMS/ABBREVIATIONS

AADT	Annual Average Daily Traffic
ADT	Average Daily Traffic
AWSC	All-Way Stop Control
CAR	Crash Analysis Reporting System
DHSMV	Department of Highway Safety and Motor Vehicles
FARS	Fatality Analysis Reporting System
FDOT	Florida Department of Transportation
GCV	Generalized Cross-Validation
MARS	Multivariate Adaptive Regression Splines
MOE	Measure of Effectiveness
MS	Microsoft
MUTCD	Manual on Uniform Traffic Control Devices
NB	Negative Binomial
OOB	Out-of-Bag
PDO	Property Damage Only
RCI	Roadway Characteristics Inventory
SAS	Statistical Analysis Software
SR	State Road
TWSC	Two-Way Stop Control
VMT	Vehicle Miles Traveled

CHAPTER 1. INTRODUCTION

1.1 Overview

Transportation is one of the most important aspects in our life. No one can move to another place without using a mode of transportation. Transportation is an important issue in any country's development and progress. The development and progress of any country can be measured by the characteristics of its transportation facilities. Transportation not only includes moving people, but goods as well. Traffic safety analysis is one of the most important applications in transportation. The issue of traffic safety has been of great importance for many researchers in U.S.A. This is because transportation is a mixed-blessing aspect. With the annual increase in the vehicle miles traveled (VMT), many people lost and are still losing their lives on these roadways. So, crashes are the drawback of transportation development. Some of these crashes lead to injuries, and some are fatal. Thus, traffic safety analysts aim to reduce the harm in terms of deaths, injuries, and property damage resulting from vehicle crashes along roadways. Researchers dealing with traffic safety analysis are investigating crashes along arterials (corridors) and at intersections. Of both, intersections are among the most dangerous locations of a roadway network. Crash analysis along arterials is macroscopic, while crash analysis at intersections is microscopic in nature. Macroscopic studies include analytic models that deal with the average traffic stream characteristics, such as flow, speed and density, while microscopic studies consider the characteristics of individual vehicles, and their interactions with other vehicles in the traffic stream (Kang, 2000).

According to the U.S. Census Bureau, Florida was ranked the 17th for traffic fatalities per 100 million VMT in 2003. In 2005, the total number of traffic crashes in Florida was 268,605 (DHSMV, 2005), of which 3185 were fatal crashes, accounting for 1.186% of the total reported crashes. The number of injury crashes was 147,879, accounting for 55.05% of the total reported crashes. In 2006, the total number of reported traffic crashes was 256, 200 (DHSMV, 2006), of which 3084 (1.204%) were fatal crashes. The number of injury crashes was 137,282 (53.58%). Compared to 2005, it is noted that there is a decrease of 4.6% in total reported crashes in 2006, a decrease of 3.17% in fatal crashes and a decrease of 7.16% in injury crashes.

In 2007, the total number of reported traffic crashes in Florida was 256,206 (DHSMV, 2007), of which 2947 (1.15%) were fatal crashes. The number of injury crashes was 135,601 (52.93%). Compared to 2006, it is noted that there is almost the same number of reported and investigated crashes in 2007. Moreover, there is a decrease of 4.44% in fatal crashes, and a decrease of 1.22% in injury crashes.

As indicated by Kuciemba and Cirillo (1992), although intersections constitute a small part of the overall highway system, intersection-related crashes represent more than 50% and 30% of crashes in urban and rural areas, respectively. As indicated in FARS (1999), for the fatal crashes' distribution by location, non-intersection locations constitute the highest percentage (79%), followed by signalized intersections (12%), and finally unsignalized intersections (9%). Of those 9% fatal crashes occurring at unsignalized intersections, 6% occurs in rural areas, and the remaining 3% occurs in urban areas. Moreover, for the fatal crashes' collision manner at unsignalized intersections, 85.6% were angle crashes, 2% were head-on crashes, and 1% were rear-end crashes.

According to DHSMV (2003), there were 96,710 crashes (39.75%) that occurred or were influenced by intersections. According to Wang (2006), "the percentage of injury crashes at intersections was 68.9%, which was much higher than that for all other entities (e.g. road sections), in which the injury crash percentage was 52.4%". This indeed indicates a need for analyzing crashes at intersections more thoroughly for further improvement and reduction in crashes.

Intersections are considered those locations with complex nature for any roadway system. Thus, a thorough understanding of them needs to be achieved in order to design them in the most effective manner. Intersections are classified into two main types, signalized intersections and unsignalized intersections. Crashes at unsignalized intersections are increasing at a high level, and thus, traffic safety at unsignalized intersections needs further study. One important reason for that is the unfamiliarity of drivers to traffic operations at unsignalized intersections, when compared to those of signalized intersections. Very few studies have addressed the safety of unsignalized intersections, which make this issue of an urgent need to be addressed. The study presented deals with traffic safety analysis at unsignalized intersections.

Unsignalized intersections include intersections with stop control, yield control and no traffic control. Unsignalized intersections, which are seen frequently in both rural and urban areas, can be differentiated than their signalized counterparts in that their operational functions take place without the presence of a traffic signal. Though research on them is not highly documented, the contributions from researchers across the nation and the world have proven to be significantly useful.

Crashes at unsignalized intersections are considered complicated incidents involving the interaction between the driver, vehicle, roadway geometry, and traffic-related factors. According to Retting et al. (2003), in U.S.A, around 700,000 reported motor-vehicle crashes by police officers occur annually at stop-controlled intersections, with one third of these crashes involve injuries and more than 3,000 are fatal. This fact was also mentioned in the U.S. Department of Transportation (2002).

Despite the increasing number of crashes at unsignalized intersections (especially, at stop-controlled intersections) and their severe nature, crash patterns at stop-controlled intersections have not been the core of detailed research (Retting et al., 2003). The work done in this study will be focusing on modeling crash frequency and crash severity at unsignalized intersections using advanced statistical and data mining techniques, so as to identify significant factors leading to crashes. Then afterwards, some safety countermeasures are recommended for further safety alleviation.

1.2 Research Objectives

The objectives of this research are six-fold, as follows:

1. Identifying the geometric and traffic factors leading to crashes at unsignalized intersections in the state of Florida using an appropriate statistical approach. For this, the traditional negative binomial (NB) model is used since it accounts for the observed over-dispersion in crash count data, i.e., the variance is greater than the mean.
2. Reducing uncertainty in predicting crash frequency at unsignalized intersections caused by statistical models. Hence, a reliability method based on the full

Bayesian updating concept is used for updating parameter coefficients from the NB model for better prediction performance.

3. Investigating various factors affecting the frequency of the two most dominant types of crashes at unsignalized intersections (rear-end and angle crashes). Then, it is claimed to increase crash prediction performance since researchers rarely develop models for the sole prediction objective. This was done using a very recent data mining technique, which is the multivariate adaptive regression splines (MARS) technique. MARS has superior prediction power, however, it has not been used in safety analysis before. Thus, MARS was used in an attempt to introduce this technique to traffic safety and show its high prediction capability.
4. Identifying the geometric, roadway and traffic factors contributing to crash severity at unsignalized intersections using the ordered probit, binary probit and nested logit frameworks.
5. Analyzing the safety effect of various median types on crash occurrence at unsignalized intersections, in order to get the safest and most hazardous types of medians in terms of safety analysis, as well as the frequent crash patterns at each median type. Hence, a safety remedy is to be applied to alleviate those high crash patterns. This analysis is related to improved access management.
6. Applying the findings from all the statistical modeling approaches to real-life traffic engineering in terms of designing the appropriate countermeasures that can be beneficial to solving any safety deficiencies identified.

By this, the conducted research has covered both the theoretical and implementational aspects in traffic safety analysis at unsignalized intersections.

1.3 Dissertation Organization

Following this chapter, a detailed literature review on previous studies of unsignalized intersections is presented in Chapter 2 of this dissertation. Chapter 3 deals with data collection procedures, variables description, median classification (types of medians) at unsignalized intersections, and an initial perspective for classifying unsignalized intersections. Chapter 4 discusses a preliminary analysis procedure regarding the safety effect of the presence of both stop sign and line, and stop sign only in Orange County. Chapter 5 presents using the reliability method (in terms of the Bayesian updating concept) to reduce the uncertainty from the fitted NB model. Chapter 6 deals with analyzing crash injury severity at unsignalized intersections. Chapter 7 illustrates using the multivariate adaptive regression splines “MARS” technique for analyzing rear-end as well as angle crashes at unsignalized intersections. Chapter 8 presents an access management analysis for the identified median types at unsignalized intersections. The last chapter, Chapter 9 is an application-wise chapter that summarizes the key findings from this research, and accordingly some countermeasures are introduced. Also, some further research avenues are recommended. Finally, the list of references used in this study is presented.

CHAPTER 2. LITERATURE REVIEW

Though research done on the safety of unsignalized intersections is not highly documented, the contributions from researchers across the nation and the world have proven to be significantly essential. This chapter indicates a very comprehensive review of literature for studies analyzing safety of unsignalized intersections.

2.1 Significant Factors Contributing to Safety of Unsignalized Intersections

Previous research on the safety of unsignalized intersections focused on topics related to geometric design characteristics such as left and right turn lanes, channelization, number of intersecting legs, intersection skewness, intersection sight distance, approach lanes, approach width, shoulder width, median width and type, vertical and horizontal alignment on approaches, lighting, etc. (Intersection Safety, Nebraska Department of Roads, 2006). The following sections discuss previous studies that addressed the contributing factors to safety at unsignalized intersections. Most of these studies are found in the research "Intersection Safety, Nebraska Department of Roads (2006)".

2.1.1 Left and Right-Turn Lanes

Foody and Richardson (1973) concluded that crash rates decreased by 76 percent at unsignalized intersections when adding a left-turn lane. Moreover, Kulmala (1997) found that the inclusion of a left-turn lane on the major approach reduced the number of rear-end crashes on this approach. Similarly, Vogt (1999) found that the presence of one or more left-turn lanes for four-leg unsignalized intersections resulted in a reduction in total crashes.

Harwood et al. (2002) found a 5 percent reduction in the number of crashes when providing a right-turn lane on one major approach to a rural stop-controlled intersection, and a 10 percent reduction when the provision is done along both major approaches.

Hauer (1988) found that providing left-turn lanes at unsignalized intersections, and at the same time combined with installation of curbs or raised medians, reduced crashes by 70, 65, and 60 percent at urban, suburban, and rural areas, respectively.

California study (1967) indicated larger reductions in crashes at unsignalized intersections given the use of left-turn lanes in a raised medians than with painted left-turn lanes.

2.1.2 Number of Intersecting Legs

David and Norman (1976) found that four-legged stop-controlled intersections in urban areas experienced twice as many crashes as the corresponding three-legged intersections.

Hanna et al. (1976) found that four-legged intersections experienced more crashes than three-legged intersections in rural locations.

Harwood et al. (1995) showed that divided four-legged intersections experienced almost twice as many crashes as three-leg intersections for narrow medians.

Bauer and Harwood (1996) showed that rural and urban stop-controlled four-legged intersections had twice crashes as the three-legged ones.

Leong (1973), Hanna et al. (1976), O'Brien (1976) and David and Norman (1975) have found that 3-legged unsignalized intersections are safer than 4-legged unsignalized intersections, while accounting for the traffic volume variable.

Kulmala (1997) has found that a four-legged intersection is safer than two three-legged intersections for low minor approach traffic volume, but less safe for high minor approach volume. The opposite was concluded by Del Mistro (1979).

2.1.3 Land Use

A recent analysis in California found that an annual average of 1.5 crashes occurs at unsignalized intersections in rural locations, compared with an average of 2.5 crashes per year in urban locations (Bauer and Harwood, 1996).

2.1.4 Intersection Skewness

McCoy et al. (1994) found that as the skew angle increased, crashes at rural two-way stop-controlled (TWSC) three and four-legged intersections increased as well.

2.1.5 Median Width

David and Norman (1975) found that multi-vehicle crashes decreased with the existence of lane dividers (such as raised reflectors, painted lines, barriers and medians).

Harwood et al. (1995) concluded that crashes increased while increasing median width at unsignalized intersections in urban and suburban areas. Likewise, Leong (1973) found that narrow medians on major roads reduced crashes' mean rate at three-leg intersections, but had small effect at four-leg intersections. Moreover, Van Maren (1980) found that median barriers had an increase on crash rates.

Summersgill and Kennedy (1996) found that the presence of an island on the minor approach increased crashes. By contrast, Layfield (1996) found that the presence of an island on the major road had a mixed effect, where some crash types were lower, and others were higher.

Pickering and Hall (1986) found that, at high traffic flow conditions, the presence of painted separation islands resulted in a reduction in crash rates for crashes occurring within 20 m of the intersection's center.

2.1.6 Lighting

The presence of lighting at unsignalized intersections appears to be associated with lower crash rates. For example, Bauer and Harwood (1996) found that lighted rural four-legged stop-controlled intersections experienced fewer crashes than no lighted intersections. In the same trend, Brude (1991) found that in dark hours, there were 30 percent fewer crashes at lighted intersections than unlighted.

The study done by Walker and Roberts (1976) showed night crash reduction after lighting was installed.

2.1.7 Channelization

In general, for intersection safety research, David and Norman (1976) showed that there was an intersection safety improvement when channelization is found.

As shown in "Intersection Safety, Nebraska Department of Roads (2006)", Templer (1980) found that a raised median reduced number of conflicts between both pedestrians and vehicles, however the difference was not significant.

Washington et al. (1991) found that the presence of raised medians on intersection approaches reduced crash rates when compared to other approaches having other median types.

2.1.8 Intersection Sight Distance

Mitchell (1972) concluded that intersection crashes were reduced by removing intersection sight obstructions. Moreover, Poch and Mannering (1996) found that the presence of an intersection sight distance obstruction significantly increased crash frequency.

David and Norman (1975) indicated that unsignalized intersections with an average daily traffic (ADT) greater than 15,000, and with obstructions within the first 20 ft from the stop bar showed more annual crashes than unobstructed intersections within the same recorded distance. Hanna et al. (1976) found that rural unsignalized intersections with poor sight distance tend to have higher crash rates than normal values.

On the other side, Pickering and Hall (1986) found that better visibility resulted in a higher crash frequency. Moreover, Stockton and Bracckett (1981) concluded that at low-volume intersections, sight distance had no observable effect on crash rates.

Thus, it is well noticed that there is inconsistency between the results obtained for the effect of visibility on crash rates.

2.1.9 Number of Approach Lanes

Using an NB regression model, Bauer and Harwood (1996) concluded that crashes at unsignalized intersections were higher on facilities with one approach lane than intersections with two or more approach lanes.

Moreover, studies done by Summersgill and Kennedy (1996) and Layfield (1996) concluded that the increase in the number of approach lanes increased the number of rear-end and lane-change crashes at the analyzed unsignalized intersections.

Weerasuriya and Pietrzyk (1998) developed conflict descriptive tables for Florida's three-legged unsignalized intersections. The introduced tables provided mean, variance, and 90^{th} and 95^{th} percentile conflict rates. The number of lanes was used for classification purposes.

2.1.10 Shoulder Width

The influence of shoulder width on intersection safety was analyzed by Van Maren (1980) as well as Harwood et al. (1995). Both studies concluded that shoulder width has no influence on intersection safety.

2.1.11 Vertical and Horizontal Alignment on Approaches

Fambro (1989) found high crash rates at intersections with crest vertical curves. Moreover, the existence of horizontal curves adds some problems to intersections.

Kuciemba and Cirillo (1992) concluded that the existence of horizontal curves near intersections could affect safety.

2.1.12 Traffic Flow

Studies done by Bauer and Harwood (1996), Huang and May (1991), Del Mistro (1981), Kulmala (1997) and Vogt and Bared (1998) for relating unsignalized intersections' geometry to safety have found that traffic flow is the most important exogeneous variable.

2.1.13 Traffic Control Type

David and Norman (1975) found that signalized intersections showed higher crash rates than stop-controlled intersections. Hanna et al. (1976) concludedthat, for a certain

ADT, rural signalized intersections experienced higher crash rate than those with stop or yield signs.

Van Maren (1980) found that multi-lane unsignalized intersections have lower crashes per million conflicts than the signalized ones. The number of crashes per million conflicts was used as the dependent (or target) variable. Moreover, Leong (1973) found that the presence of traffic signals reduced the average crash rate at four-legged unsignalized intersections, but had negligible effect at the three-legged ones.

2.1.14 Size of Intersection

Van Maren (1980) concluded that large unsignalized intersections (intersections with a large distance across the intersection) had higher crashes per million conflicts than small unsignalized intersections.

2.1.15 Minor Road Approach Geometry

Kulmala (1997) concluded that crash rates are lower than the average at four-legged unsignalized intersections with a curve on the minor road approach just before the intersection.

2.1.16 Grades

Pickering and Hall (1986) found that downhill unsignalized intersections showed higher crash rates than other intersections. On the other hand, Hanna et al. (1976) concluded that intersections with severe grades operate safely than others.

2.1.17 Signing and Delineation

David and Norman (1975) found that unsignalized intersections operating with signs having white lettering on a dark background had more annual crashes than those having dark lettering on a white background. Moreover, they found that unsignalized intersections with raised pavement markers showed fewer crashes than those without raised markers.

Van Maren (1980) found that large-sized stop signs on the minor approaches tended to decrease the number of crashes per million conflicts.

Huang and May (1991) found that intersections with stop signs on major streets had higher crash rates than those with stop signs on minor streets, since drivers did not expect the existence of stop signs on main streets.

A study done by Kitto (1980) showed that unsignalized intersections with yield (or give-way) signs showed almost the same crash rates to those with stop signs.

2.1.18 Spacing between Intersections

A study done by Layfield (1996) concluded that a relatively large spacing between the minor approaches of urban unsignalized intersections resulted in fewer crashes.

2.1.19 Pedestrian Crossing Facilities

Summersgill and Kennedy (1996) as well as Layfield (1996) concluded that the existence of crossing facilities for pedestrians at 3 and 4-legged intersections resulted in higher pedestrian crashes.

2.1.20 Speed Parameters

Summersgill and Kennedy (1996) and Pickering and Hall (1986) found that there was no sufficient evidence that vehicles' speed on both major and minor roads had an influence on crashes. It is to be noted that this result was based on a narrow band of speed data, since Pickering and Hall analyzed only rural unsignalized intersections with speed limits over 50 mph, and Summersgill and Kennedy analyzed only 3-legged unsignalized intersections on 30 and 40 mph roadways. Hence, a significant trend between speed and crash occurrence was difficult to result with such limited speed data.

By contrary, the study done by Brude (1991) showed that lower speeds were found to improve intersection safety.

2.1.21 Beacons Use

King and Goldblatt (1975) found that the installation of flashing beacons to stop-controlled intersections led to favorable effect on safety. However, this result is different from that obtained by Pant and Park (1999).

2.1.22 Turn Lanes Configuration

Poch and Mannering (1996) found that intersection approaches with combined through and left lanes were found to have higher crash frequencies than approaches without this combined configuration.

2.1.23 Pavement Condition

A study done by Chovan et al. (1994) found that around 74% of unsignalized-intersection crashes occurred on dry pavement, around 25% on wet or snowy pavement, and the remaining 1% was misclassified.

2.2 Some Facts about Unsignalized Intersections

According to Marek et al. (1997), under certain traffic volume and geometric characteristics, all-way stop control (AWSC) intersections operate much safer than signalized intersections as well as two-way stop-controlled intersections. Supporting this finding, Briglia (1982) and Hauer and Lovell (1986) showed that AWSC intersections have much lower crash rates than TWSC intersections. Moreover, Byrd and Stafford (1984) showed that traffic flow characteristics for AWSC intersections are different than those controlled by two-way stop signs.

Sayed and Rodriguez (1999) developed an accident prediction model for estimating safety at unsignalized urban junctions using the generalized linear model (GLM) formulation. They estimated the model's parameters based on a methodology presented in the work of Bonneson and McCoy (1997). This methodology was done using the Poisson error structure. For assessing the model goodness-of-fit, Pearson's chi-square was used. The model was useful in some applications such as identifying accident-prone-locations (APLs), ranking identified APLs, and evaluating before-and-after studies.

Sayed and Zein (1999) applied the traffic conflict technique while analyzing safety at unsignalized intersections. The used data were collected from 30 different surveys to establish conflict frequency and severity standard values. These standard values were later applied to compare the relative conflict risk rates between intersections using an intersection conflict index. They developed predictive models to relate traffic conflicts to traffic volumes and crashes.

A study done by Salman and Al-Maita (1995) focused on traffic volume on 18 three-legged unsignalized intersections located in Amman, Jordan. In this study, the authors found that the sum of major and minor volumes were correlated with the number of traffic conflicts.

Vogt (1999) developed a model for four-legged rural stop-controlled intersections. This model showed a crash reduction of 38 percent for total crashes due to the installation of a left-turn lane on the major road.

Lau and May (1988, 1989) used CART (Classification and Regression Trees) analysis, and concluded that left-turn prohibition was a significant factor in predicting injury crashes at unsignalized intersections.

Van Maren (1980) used the number of crashes per million conflicts as the dependent variable, and he found that multi-lane unsignalized intersections have a lower number of crashes per million conflicts than the signalized ones.

As shown by Wang and Abdel-Aty (2006), Poch and Mannering (1996) fitted a rear-end crash frequency model at the approach level. They analyzed 63 four-legged signalized and unsignalized intersections over 7 years (from 1987 till 1993) using the NB model. They used the number of through, right and left-turn lanes on the minor approach as surrogate variables for the magnitude of through, right and left-turning volumes, respectively. They concluded that NB formulation was an appropriate model for isolating traffic and geometric factors influencing crash frequency.

A study done by Retting et al. (2003) who investigated crashes at 4 U.S. cities, Germantown, Tennessee; Oxnard, California; Springfield, Missouri; and Westfield, New Jersey, recommended some countermeasures for an improvement of stop-controlled intersections. They recommended that stop signs should be frequently inspected to ensure they are not obscured by trees or other blockings.

2.3 Types of Crashes Occurring at Unsignalized Intersections and their Modeling Scheme

Summersgill and Kennedy (1996), Layfield (1996), Pickering and Hall (1986), Agent (1988) and Hanna et al. (1976) found that the most common crashes at unsignalized intersections appear to be angle crashes (right-turn or through movement from the minor approach colliding with a through-moving vehicle on the major road) and rear-end crashes. Moreover, they found that single vehicle, head-on, side-swipe and left-turn crashes were common, but were fewer in number.

At unsignalized intersections, McCoy and Malone (1989) found that there was a significant increase in right-angle crashes. However, McCoy et al. (1985) concluded that there was no significant difference in rear-end and left-turn crash rates between unsignalized intersections with and without left-turn lanes.

Chovan et al. (1994) analyzed the crash statistics of stop-controlled intersections having straight-crossing-path crashes. They defined those crashes as crashes in which two vehicles, one with right-of-way and one without, cross each other's path perpendicularly.

Najm et al. (2001) concluded that there were 1.72 million crossing-path crashes. Of these crashes, LTAP (Left Turn Across Path) crashes accounted for the largest percentage (47.2%), followed by SCP (Straight Cross Path) crashes (29.9%). The great majority of these crossing-path crashes occurred at intersections (75.1%), followed by driveways (21.0%). In general, they found that 41.6% of crashes occurred at signalized intersections, 36.3% at stop-signed intersections, and 22.1% at intersections with no controls or other control types.

The research "Strategies to Address Nighttime Crashes at Rural, Unsignalized Intersections, 2008" evaluated crashes for rural unsignalized intersections in the state of Iowa for 2001 to 2005. Results show that 26% of crashes at rural unsignalized intersections occur during nighttime conditions, and another 4% occur during dawn or dusk. Moreover, it was found that 29% of fatal and injury crashes occur during at night.

2.4 Analysis of Unsignalized Intersections

An example of some studies that used stepwise multiple linear regression analysis techniques that assume normal distribution of data is Kitto (1980). Recent studies assumed nonlinear distributions such as the Poisson distribution. An example of this is Agent (1988). Moreover, Vogt (1999) used NB models for analysis, and Bauer and Harwood (1996) used log-normal models in their analysis..

Bauer and Harwood (1996) showed that the use of the Poisson distribution is only relevant when the variance in the crash data is equal to the mean. But, this is not the common case for crash data, as crash data always suffer over-dispersion, where the variance is much greater than the mean. Thus, the use of the Poisson distribution is not valid any longer, as it can result in biased estimated model coefficients and erroneous standard errors. The remedy for this is using the NB model, as it can overcome the over-dispersion issue.

Studies performed by Tijerina et al. (1994), Chovan et al. (1994) and Wang and Knipling (1994) were summarized in a report by Najm et al. (1995). This report provided further insight into the general characteristics of intersection crashes. This report accounted for the following variables:

- Time of day.

- Lighting condition.
- Atmospheric condition.
- Roadway surface condition.
- Roadway alignment.
- Roadway profile.
- Speed limit – the higher-profile road of the intersection is coded.
- Relation to junction.
- Alcohol involvement.
- Maximum severity – police reported severity of worst-injured person.

2.5 Safety Effectiveness of Converting Unsignalized Intersections to Signalized Ones

Studies done by Datta and Dutta (1990), Datta (1991) and King and Goldblatt (1975) as well as the research "Effects of Signalization on Intersection Safety, 1982" found that the number of right-angle crashes decreased at an intersection when the traffic control device was changed from a stop-controlled to a traffic signal. Moreover, Agent (1988) concluded that there was a decrease in right-angle crash rates when a rural stop-controlled intersection with a beacon was changed to a traffic signal.

As for rear-end crashes, research done by Datta (1991), King and Goldblatt (1975) showed rear-end crash increase after signalizing their analyzed intersections. Datta and Dutta (1990) concluded that there was a 53% increase in rear-end crashes after signalization. Other research "Effects of Signalization on Intersection Safety, 1982" found a reduction in rear-end crash frequency after signalization.

2.6 Studies Using the NB Formulation

Traditional NB models are widely used in the prediction of crash frequencies at intersections and have been applied extensively in various types of highway safety studies. These studies varied from the identification of black spots to the development of accident modification factors using the coefficients of the model (Miaou, 1996; Harwood et al., 2000; Vogt, 1999; Lord and Bonneson, 2006). The traditional NB model is developed using a fixed dispersion parameter (Miaou, 1996). However, as shown by Hauer (2001), it is not understood why a constant dispersion parameter could exist. Some other researchers hypothesized that the dispersion parameter has a fixed value (Miaou and Lord, 2003; Heydecker and Wu, 2001; Lord et al., 2005; Miranda-Moreno et al., 2005; El-Basyouny and Sayed, 2006). Heydecker and Wu (2001) estimated varying dispersion parameters as a function of the locations' covariates, such as minor and major traffic volumes at intersections. They concluded that the NB model with a varying dispersion parameter fits data better than the traditional NB model with a fixed dispersion parameter. Later on, it has been found that the estimated dispersion parameter of NB models can be affected when the data are have a small sample size and low sample mean (Piegorsch, 1990; Dean, 1994; Lord, 2006), and crash data are usually characterized by these two criteria (Lord and Bonneson, 2006). Other improvement in the NB formulation was done by Anastasopoulos and Mannering (2009), who examined the random-parameters NB model, and found that it has the potential of providing a fuller understanding of the factors affecting crash frequency.

The NB model is usually characterized by two parameters, the mean μ and the dispersion parameter α. Park and Lord (2008) used simulation to adjust the maximum likelihood estimate of the NB dispersion parameter. Simulation runs were used to develop a relationship between the estimated and the true dispersion parameters. Also, Geedipally and Lord (2008) tested the effects of varying the dispersion parameter on the estimation of confidence intervals of safety performance functions. They concluded that models having a varying dispersion parameter usually produce smaller confidence intervals than those with a fixed dispersion parameter. Hence, varying the dispersion parameter α provides more precise estimates. Zhang et al. (2007) used the bootstrapped maximum likelihood method to estimate the dispersion parameter of the NB distribution while analyzing crash count data.

In traffic safety analysis, the dispersion parameter of NB models introduced the role of empirical Bayes "EB" estimates. Those estimates are used to account for random fluctuation of crash counts. The Bayesian concept was extensively used in crash analysis. The primary application of it is using the EB estimates. The EB approach was originally developed to account for the regression-to-the-mean effect in before-and-after studies (e.g. Powers and Carson, 2004). Moreover, the EB estimates were used for locating black spot locations (Saccomanno et al., 2001). Black spot locations are those locations having high frequency of crashes (and especially severe crashes). Moreover, Persaud et al. (2009) compared the results from the EB and full Bayesian approaches while converting a 4-lane roadway to a 3-lane one (with a two-way left turn lane in the middle). They found that both results are very comparable.

As shown in the abovementioned studies in this section, in spite of the fact that the Bayesian concept was extensively used in traffic safety analysis, using a reliability method based on full Bayesian updating to reduce the uncertainties from the predictive models is not extensively applied. This was the key behind the analysis conducted in Chapter 5, where the NB and log-gamma likelihood functions were examined in the Bayesian updating procedure using informative and non-informative priors.

2.7 Studies Analyzing Injury Severity

Researchers have employed many statistical techniques to analyze injury severity, and those techniques have been used extensively in traffic safety analysis. Examples of those techniques are the multinomial logit, nested logit, and ordered probit models.

Abdel-Aty (2003) used the multinomial logit, nested logit and ordered probit frameworks to identify those factors that affect injury severity at toll plazas. He concluded that the multinomial logit model produced poor results when compared to the ordered probit model. Moreover, it was found that the ordered probit model is better than the nested logit model due to its simplicity. In addition to toll plazas, the author used the ordered probit model to compare those factors that affect injury severity at other roadway locations, including roadway sections and signalized intersections.

For the nested logit model formulation, Savolainen and Mannering (2007) analyzed motorcyclists' injury severities in single and multi-vehicle crashes using nested logit frameworks. The used data were drawn from all police-reported motorcycle crashes in the state of Indiana between 2003 and 2005. They concluded that crashes were less severe under wet pavement conditions, near intersections, and when passengers were on the motorcycle.

Shankar et al. (1996) analyzed single-vehicle injury severity on rural freeways. They found that the nested logit formulation fits the data well. The results showed the significant effect of some rs such as environmental conditions, highway design, accident type, driver characteristics and vehicle characteristics.

Nassar et al. (1994) used three nested logit models to model crash severity. These models were calibrated for three crash situations: single-vehicle, two-vehicle, and multi-vehicle crashes. It was concluded that road surface condition was not significant in the models. They reported that bad weather conditions may alert drivers to slow down and keep enough spacing from other vehicles.

For the ordered probit framework, Quddus et al. (2002) analyzed motorcycle's injury severity resulting from crashes using a 9-year crash data in Singapore. An interesting result found is that a higher road design standard increases the probability of severe injuries and fatalities. Also, the authors did not find that age increase could increase severity.

Hutchinson (1986) used the ordered probit modeling for studying occupants' injury severity involved in traffic crashes. British crash data for 1962–1972 were used in the analysis, and it was concluded that passengers tend to be more seriously injured than drivers in non-overturning crashes, but that there is no significant difference in overturning crashes.

Kockelman and Kweon (2002) used the ordered probit formulation to investigate the risk of different injury levels for single and two-vehicle crashes. They concluded that pickups and SUVs are less safe than passenger cars for single-vehicle crashes. However, in two-vehicle crashes, they were found them to be safer for drivers and more hazardous for passengers.

Duncan et al. (1998) used the ordered probit framework to examine occupant characteristics as well as roadway and environmental conditions influencing injury severity in rear-end crashes involving truck-passenger car crashes. Two models were developed, one with the main-effect exogeneous variables, and the other with interactions among those exogeneous variables. They found that there is an increased severity risk for high speed crashes, those occurring at night, for women, when alcohol is involved, and for crashes when a passenger car rear-ends a truck at a large differential speed between both of them.

From the aforementioned studies in this section, almost no study addressed injury severity at unsignalized intersections. Hence, this was the introductory part for investigating injury severity at unsignalized intersections for exploring the effect of traffic and roadway covariates on crash injury severity, as will be seen in Chapter 6.

2.8 Studies Related to Crash Prediction

Using crash prediction models in safety studies can be found in previous literature (e.g., Hauer et al., 1988; Persaud and Dzbik, 1993; Sawalha and Sayed, 2006 and Abdel-Aty and Radwan, 2000). Miaou (1994) used the NB, Poisson and zero-inflated Poisson models to relate roadway factors to crashes. He recommended the use of NB models when over-dispersion exists in the data. Ivan and O'Mara (1997) applied the Poisson model for predicting traffic crashes. The most significant predictors identified were the speed limit and annual average daily traffic "AADT". Poch and Mannering (1996) used the NB formulation to predict crash frequency on certain sections of principal arterials in Washington State. They concluded that the NB model is a powerful predictive tool and it is strongly recommended to be applied in other crash frequency studies.

2.9 Studies Using Advanced Prediction Techniques

Recently, researchers have proposed new pioneering statistical methods for modeling and predicting crashes that are very comparable to NB and Poisson models. Examples of those methods are neural networks (Mussone et al., 1999 and Abdelwahab and Abdel-Aty, 2002), Bayesian neural networks (Xie et al., 2007 and Riviere et al., 2006) and support vector machine "SVM" (Li et al., 2008). However, neural networks models always suffer from their interpretation complexity, and sometimes they over-fit the data (Vogt and Bared, 1998). For this, Bayesian neural networks were introduced that can accommodate data over-fitting. For example, Xie et al. (2007) applied the Bayesian neural networks in predicting crashes, and found that they are more efficient than NB models. Also, Li et al. (2008) applied a simpler technique than the Bayesian neural networks, which is SVM, to data collected on rural frontage roads in Texas. They fitted several models using different sample sizes, and compared the prediction performance of those models with the NB and Bayesian neural networks models. They found that SVM models are more efficient predictors than both NB and Bayesian neural networks models.

MARS is a multivariate non-parametric regression technique that was introduced by Friedman (1991). MARS is considered a nonparametric technique as it does not require any priori assumption about the form of the relationship between dependent and independent variables, and can reveal the required relationship in a piecewise regression function. This technique is effective when analyzing complex structures in the data such as nonlinearities and interactions. Crash data are those types of data that are characterized by a nonlinear relationship between the predictors and the dependent variable. Also, MARS is a regression-based technique, not suffering from the "black-box" limitation, where the output is easily understood, and can explain the model.

The application of MARS from the methodological point of view can be found in previous studies (e.g., De Veaux et al., 1996; Nguyen-Cong et al., 1996; Lahsen et al., 2001; Put et al., 2004; Sephton, 2001; Leathwick et al., 2005; Francis, 2003 and Attoh-Okine et al., 2003). For example, Put et al. (2004) concluded that MARS has some advantages compared to the more traditionally complicated techniques such as neural networks. Attoh-Okine et al. (2003) used the MARS technique to develop a flexible pavement roughness prediction model. They concluded that MARS allows easy interpretation of the pavement, environmental and traffic predictors found in the model.

From the abovementioned studies in this section, it can be noted that MARS has promising advantages that can be implemented for improving prediction and for accommodating nonlinearities in crash count data, however, there was no research conducted to implement MARS in traffic safety to show its potential characteristics. This was the motivation behind the analysis conducted in Chapter 7.

2.10 Access Management and Traffic Safety

A study done in Ohio (1964) at 316 at-grade intersections on divided highways with partial or no access control analyzed annual crash occurrence as a fraction of divided highway and minor road AADT. It was concluded that crash frequency was more sensitive to minor road traffic (i.e., unsignalized access points) than to divided highway traffic (i.e., arterial corridors). This demonstrates the significant need to deeply analyze access management related to unsignalized intersections.

According to the FDOT Median Handbook (2006), access management is “the location, spacing and design of driveways, medians, median openings, signals and interchanges”. Thus, medians are an application of an access management design. According to the aforementioned handbook, restricted medians (such as directional and closed medians), as well as well designed median openings are known to be very important features in efficient highway system design. The design and placement of those medians and those median openings is an essential part of the access management design.

According to the FDOT Median Handbook (2006), the benefits of installing medians are the following:

1. Safety, i.e. fewer severe crashes, and less motor vehicle/pedestrian conflict.
2. Efficiency, i.e. higher level of service, and less “stop and go” traffic.
3. Aesthetics, i.e. more space for landscaping and pedestrian facilities, and more attractive arterials.

Many studies have shown that restricted medians are of larger safety benefits than those unrestricted medians. One of those studies for evaluating urban multilane highways in Florida in 1993 (FDOT Median Handbook, 2006), revealed that the crash rate for restricted medians is 25% lower than those having a two-way left turn lane. This indeed shows the negative safety effect of installing two-way left turn lane medians.

2.10.1 Safety of Median Openings

A study done by Dissanayake and Lu (2003) showed that the conversion of a full median opening to a directional one reduced the average number of hourly conflicts by around 50%. Moreover, the conflict rate per thousand involved vehicles was also significantly reduced. Additionally, the severity of conflicts was also found to have a reduction after some time period. They also found that the total average travel delay was significantly reduced after the median opening was converted to a directional median.

McDonald (1953) analyzed median openings' safety of 150 at-grade intersections on 180 miles of divided highways in California. He concluded that low crossroad volume intersections experienced higher crash rates per vehicle than did high crossroad volume intersections.

Priest (1964) analyzed at-grade intersections on divided highways with partial or no control access. He found that crash frequency was more sensitive to crossroad traffic than to divided highway traffic, i.e., more sensitive to unsignalized intersection access points. Hence, deep investigation is needed to analyze access management at unsignalized intersections.

Based on their crash data analysis at unsignalized median openings, Levinson et al. (2005) found that crashes related to U-turn and left-turn maneuvers occur infrequently. Hence, they are not of major safety concern. Also, they concluded that the average median opening crash rates for three-legged intersections at urban corridors are lower than the corresponding four-legged intersections.

A study done by FDOT (1995) found that reductions in the number of median openings (i.e., reduction in the number of conflict points) along roadways resulted in crash rate reductions, despite the increased through traffic volume per median opening.

According to Koepke and Levinson (1992), for median openings installation, they recommended that they should be set back far enough from nearby signalized intersections to avoid conflict with intersection queues (backward shock waves).

Cribbins et al. (1967) concluded that median openings do not experience high crash rates under some specific conditions of low vehicle volumes and wide medians. However, as traffic volume increases, the frequency of median openings significantly affects crash risk.

The Florida DOT Median Handbook (1997) identified some important factors that should be considered in determining the spacing of median openings. These are deceleration length, queue storage, turning radius and perception/reaction distance. For urban arterials, Florida identified a "1070 feet" as a minimum median opening spacing.

Harwood et al. (1995) concluded that at rural four-legged unsignalized intersections, crash frequency decreases with the increase in median width. At rural three-legged unsignalized intersections, they found that there is no statistical significance relationship between crash frequency and median width. At urban/suburban three and four-legged unsignalized intersections, they showed that crash frequency increases as median width increases.

From their research, Lu et al. (2005) recommended specific values for the offset distance for median opening (i.e., the distance between the driveway exit and the downstream U-turn location). For four lanes, they recommended an offset distance of 400 feet, whereas for 6 or more lanes, they recommended 500 feet.

2.10.2 Safety of Left-Turn Lanes

As shown in Levinson et al. (2005), an ITE study (Traffic Safety Toolbox, 1987) concluded that there was a crash reduction of around 30% to 65% at unsignalized intersections due to the installation of left-turn lanes. Also, Gluck et al. (1999) found crash reduction of 50% to 77% at unsignalized intersections.

2.10.3 Safety of U-turns

As shown in Levinson et al. (2005), indirect left-turns (or U-turns) are mainly used in many states (e.g., Florida and Michigan) as an alternative to direct left-turn maneuvers. For a study done in Florida, Gluck et al. (1999) concluded that there is around 18% to 22% reduction in crash rate by substituting direct left-turns from driveways with right turns followed by U-turns. In Michigan, they found a 15% to 61% crash rate reduction while replacing direct left-turns from driveways with right turns followed by U-turns.

Potts et al. (2004) concluded that is no statistical regression relationships relating median opening crash frequency to the U-turn and left-turn volumes.

2.10.4 Studies on Safety of Some Median Types

On their analysis on intersections, Bowman and Vecellio (1994) showed that undivided medians are safer than two-way left turn lane medians.

Margiotta and Chatterjee (1995) collected data for 25 highway segments in Tennessee including 12 median-divided segments and 13 segments with two-way left turn lanes. They concluded that medians had fewer crashes than do two-way left turn lanes. Crashes on median divided segments were more frequent at signalized intersections, while those on two-way left turn lane segments were more frequent at unsignalized intersections. Also, they found that rear-end crashes were more likely to occur on a median divided segment, whereas head-on crashes were more probably to occur on a segment with a two-way left turn lane.

2.10.5 Access Management Design Policies

As indicated by the AASHTO Green Book, it reported some essential factors for the design policies of U-turn maneuvers at unsignalized median openings, which were:

- Median width.
- Traffic characteristics that include AADT and truck percentage in the fleet.
- Crash frequency, especially angle and rear-end crashes.
- Spatial covariate in terms of the location of median openings with respect to the signalized intersections. (It is worth mentioning that the spatial covariate in this study was explored in this study in terms of the upstream and downstream distances to the nearest signalized intersection from the unsignalized intersection of interest, as well as the distance between successive unsignalized intersections).
- Presence of exclusive left-turn lanes.

An access management analysis with respect to the identified median types in this study is shown in Chapter 8.

CHAPTER 3. DATA COLLECTION

3.1 Introduction

The data collection process is critical for obtaining good results at the analysis stage and for reaching valuable conclusions, which as a whole fulfill the study's objectives initially specified in the introductory chapter, Chapter 1. The more extensive the data collection process is, the more robust the results will be. Thus, the procedures involved should be done in the most accurate way in order to get a very high confidence level for the results.

In order to start the data collection procedure, it is first better to understand the FDOT'S (FDOT Map, 2007) procedure for classifying the districts in the state of Florida. In Florida, there are 7 districts, and 67 counties. A district is the major entity classified by the FDOT after the state. The second big entity is the county, which is mainly a region with borders, consisting of cities, towns, villages, and so on. The distribution of counties in each district in the state of Florida is shown in Table 3-1. Moreover, Figure 3-1 shows a pie chart for this distribution.

Table 3-1: County Distribution in Each District in Florida State according to FDOT

District	Number of counties	County name
1	12	Charlotte, Collier, DeSoto, Glades, Hardee, Hendry, Highlands, Lee, Manatee, Okeechobee, Polk and Sarasota
2	18	Alachua, Baker, Bradford, Clay, Columbia, Dixie, Duval, Gilchrist, Hamilton, Lafayette, Levy, Madison, Nassau, Putnam, St. Johns, Suwannee, Taylor and Union

District	Number of counties	County name
3	16	Bay, Calhoun, Escambia, Franklin, Gadsden, Gulf, Holmes, Jackson, Jefferson, Leon, Liberty, Okaloosa, Santa Rosa, Wakulla, Walton and Washington
4	5	Broward, Indian River, Martin, Palm Beach and St. Lucie
5	9	Brevard, Flagler, Lake, Marion, Orange, Osceola, Seminole, Sumter and Volusia
6	2	Miami-Dade and Monroe
7	5	Citrus, Hernando, Hillsborough, Pasco and Pinellas

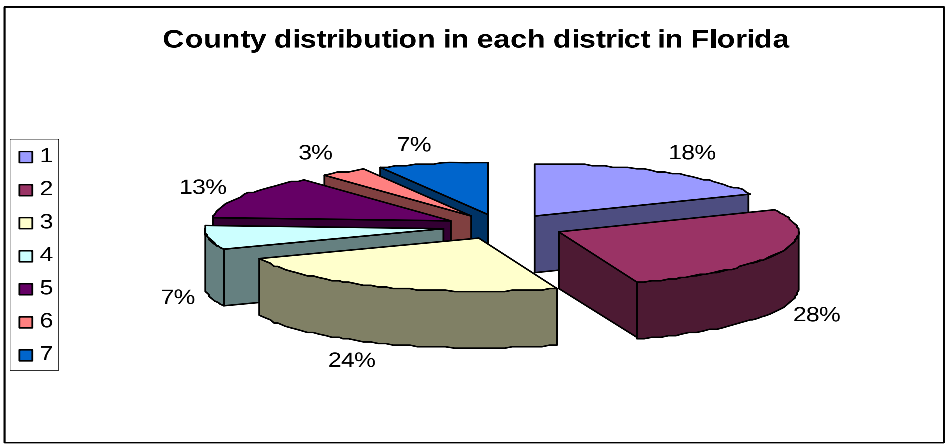


Figure 3-1: Pie Chart for County Distribution in Each of the 7 Districts in Florida

From this chart, it is clear that district 2 has the highest percentage of counties (28%), followed by district 3 (24%), and district 1 (18%). District 6 has the smallest percentage (3%).

Despite the fact that unsignalized intersections have less number of crashes compared to signalized intersections, unsignalized intersections are more frequent than the signalized ones. This makes the process of data collection much more difficult in the essence that the required sample size should be much more than that of the signalized intersections to accurately depict the population size.

In order to represent the population of 67 counties in Florida, a sample of 6 counties was selected to represent this population. This selection was not based on the random selection, but was based on the geographic location in Florida, so as to represent the Northern, Southern, Central, Eastern and Western parts in Florida. Leon County was selected to represent the Northern part, Miami-Dade was selected to represent the Southern part, Orange and Seminole Counties were selected to represent the Central part, Brevard County was selected to represent the Eastern part, and finally Hillsborough County was selected to represent the Western part. Moreover, the selection was based on having a combination of both urban and rural areas, so as to make the conclusion from the analysis procedure valid to all types of land use, and not only leaned to a specific type. This indeed will lead to more robust and accurate results. It is known that Leon County has a high percentage of rural roads, and in addition, it has the capital of Florida, Tallahassee. It is to be noted that those selected counties concur with the selected counties in the study done by Wang (2006), who analyzed the spatial and temporal effect of signalized intersections in the state of Florida.

It was decided to collect 2500 unsignalized intersections from those 6 selected counties. This sample was deemed sufficient for the analysis procedure. Moreover, it was decided to collect 500 unsignalized intersections from Orange County, and 400 unsignalized intersections from the other 5 counties. The following sections explain the initial and final data collection procedures, the list of variables (representing the geometric, traffic and control fields) used in data collection, some difficulties encountered during the data collection procedure in each selected county and some unfamiliar intersections captured.

3.2 Variables Description

A "MS Excel" spreadsheet that lists all the required geometric, traffic and control fields required for getting a full understanding of the identified unsignalized intersections was created. There was a total of 46 variables listed in this table. It is to be noted that these 46 variables were not defined all at once, but the table was expanding until these 46 variables were captured. Below is a detailed description of these 46 variables:

I. Geometric fields:

1. District: This variable shows the district number as indicated in the FDOT database.
2. Roadway ID: This variable shows the state road (SR) ID as indicated in the FDOT database.
3. Intersection Node: This variable shows the intersection node number as indicated in the FDOT database.
4. Mile Point: This variable shows the mile post for each intersection (i.e. node) as indicated in the FDOT database.
5. County: This variable shows the county name to which each analyzed state road belongs.
6. County ID: This variable shows the ID of the county to which each analyzed state road belongs as indicated in the FDOT database.
7. Major Road Name: This variable shows the name of the major road in the intersection.
8. Minor Road Name: This variable shows the name of the minor road in the intersection.

9. Stop Sign Minor 1: This variable shows whether there is a stop sign on Minor 1 ("1" if it exists, "0" if it does not exist and "N/A" if not applicable). The main difference between "0" and "N/A" is that "0" is used when Minor 1 leg exists, but there is no stop sign existing, while "N/A" means that Minor 1 leg does not exist.

Figure 3-2 shows the concept for identifying the 4 approaches; Major 1, Major 2, Minor 1 and Minor 2 while collecting data on unsignalized intersections.

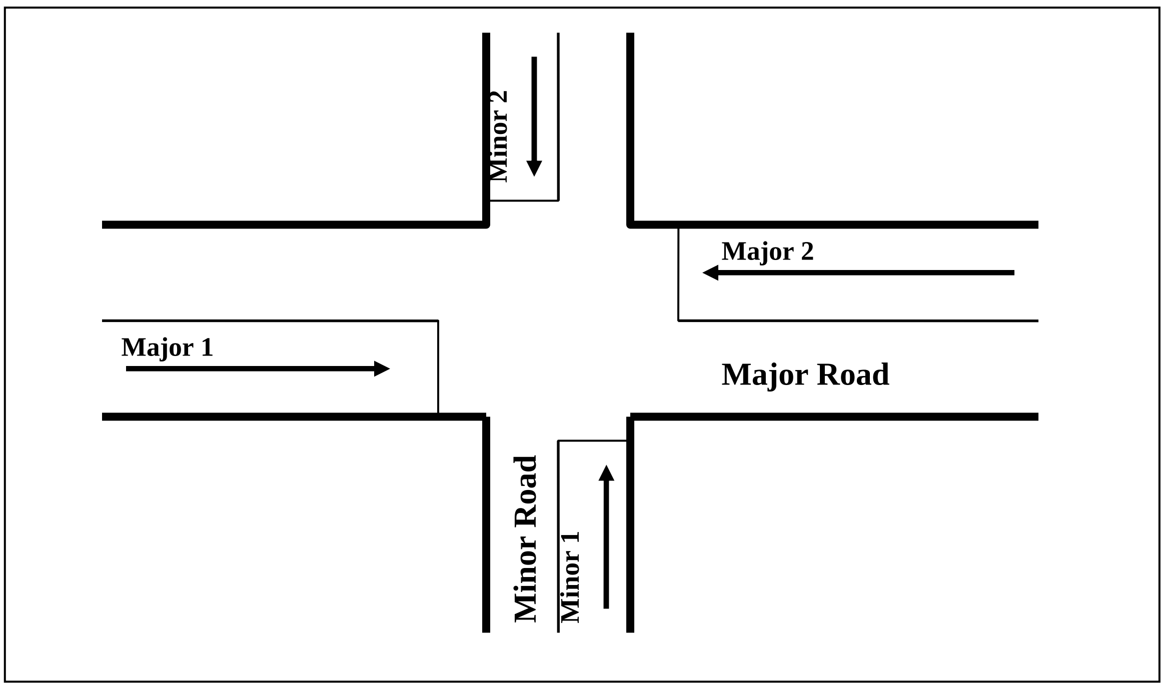


Figure 3-2: Conceptual Road Layout for "Major 1, Major 2, Minor 1 and Minor 2" Approaches

10. Stop Sign Minor 2: This variable shows whether there is a stop sign on Minor 2 ("1" if it exists, "0" if it does not exist and "N/A" if not applicable). The difference between "0" and "N/A" is the same as that mentioned in variable "9".

11. Stop Sign Major 1: This variable shows whether there is a stop sign on Major 1 (“1” if it exists, “0” if it does not exist and “N/A” if not applicable).
12. Stop Sign Major 2: This variable shows whether there is a stop sign on Major 2 (“1” if it exists, “0” if it does not exist and “N/A” if not applicable).
13. Stop Line Minor 1: This variable shows whether there is a stop line (i.e. stop bar) on Minor 1 (“1” if it exists, “0” if it does not exist and “N/A” if not applicable).
14. Stop Line Minor 2: This variable shows whether there is a stop line (i.e. stop bar) on Minor 2 (“1” if it exists, “0” if it does not exist and “N/A” if not applicable).
15. Stop Line Major 1: This variable shows whether there is a stop line (i.e. stop bar) on Major 1 (“1” if it exists, “0” if it does not exist and “N/A” if not applicable).
16. Stop Line Major 2: This variable shows whether there is a stop line (i.e. stop bar) on Major 2 (“1” if it exists, “0” if it does not exist and “N/A” if not applicable).
17. Crosswalk Minor 1: This variable shows whether there is a crosswalk for pedestrians on Minor 1 (“1” if it exists, “0” if it does not exist and “N/A” if not applicable).

18. Crosswalk Minor 2: This variable shows whether there is a crosswalk for pedestrians on Minor 2 ("1" if it exists, "0" if it does not exist and "N/A" if not applicable).
19. Crosswalk Major 1: This variable shows whether there is a crosswalk for pedestrians on Major 1 ("1" if it exists, "0" if it does not exist and "N/A" if not applicable).
20. Crosswalk Major 2: This variable shows whether there is a crosswalk for pedestrians on Major 2 ("1" if it exists, "0" if it does not exist and "N/A" if not applicable).
21. Size of Intersection: This variable shows the number of through lanes for both the major and minor roads, based on the normal cross-section of each (e.g., 2x2, 2x3 and 2x4). The first number represents the number of through lanes for the minor approach for both directions, and the second number represents the number of through lanes for the major approach for both directions.
22. Type: This variable was listed as "the total number of approach lanes for the minor approach x the total number of through lanes for the major approach". An example for this, if the minor approach configuration for a three-legged unsignalized intersection has 1 right-turn approach lane, 1 left-turn approach lane and 1 receiving lane, and the major approach has 6 through lanes for both directions, then the type of this intersection is "3x6". This variable was captured so as to relate the geometric configuration of the intersection to the crash pattern occurring at that specific intersection.

23. Number of Intersecting Legs: This variable shows the number of legs of the intersection (e.g. 3 legs and 4 legs).

24. Number of Through Lanes for Major 1: This variable shows the number of through lanes for Major 1 approach.

25. Number of Through Lanes for Major 2: This variable shows the number of through lanes for Major 2 approach.

26. Number of Through Lanes for Minor 1: This variable shows the number of through lanes for Minor 1 approach.

27. Number of Through Lanes for Minor 2: This variable shows the number of through lanes for Minor 2 approach.

28. Number of Right Turn Lanes for Major 1: This variable shows the number of right turn lanes for Major 1 approach.

29. Number of Right Turn Lanes for Major 2: This variable shows the number of right turn lanes for Major 2 approach.

30. Number of Right Turn Lanes for Minor 1: This variable shows the number of right turn lanes for Minor 1 approach.

31. Number of Right Turn Lanes for Minor 2: This variable shows the number of right turn lanes for Minor 2 approach.

32. Number of Left Turn Lanes for Major 1: This variable shows the number of left turn lanes for Major 1 approach.

33. Number of Left Turn Lanes for Major 2: This variable shows the number of left turn lanes for Major 2 approach.

34. Number of Left Turn Lanes for Minor 1: This variable shows the number of left turn lanes for Minor 1 approach.

35. Number of Left Turn Lanes for Minor 2: This variable shows the number of left turn lanes for Minor 2 approach.

36. Median Type for Major 1: This variable shows the type of median for Major 1 (e.g. open, directional, closed, two-way left turn lane and undivided). A detailed explanation of those median types is shown in the following sections, accompanied with some snap shots for better understanding.

37. Median Type for Major 2: This variable shows the type of median for Major 2 (e.g. open, directional, closed, two-way left turn lane and undivided).

38. Adjacent Upstream Signalized Intersection Distance for Major 1: This variable determines the closest upstream signalized intersection distance (in miles) to the specified unsignalized one with respect to Major 1. This distance can be written as "not applicable" (N/A) if the distance exceeds 1 mile. Also, this variable was listed in the table with the attempt to test the spatial correlation of the unsignalized intersections with the nearest signalized intersections.

39. Adjacent Downstream Signalized Intersection Distance for Major 1: This variable determines the closest downstream signalized intersection distance (in miles) to the specified unsignalized one with respect to Major 1.

40. Adjacent Upstream Signalized Intersection Distance for Major 2: This variable determines the closest upstream signalized intersection distance (in miles) to the specified unsignalized one with respect to Major 2. It is to be noted that this distance is exactly the same distance as variable “38”.

41. Adjacent Downstream Signalized Intersection Distance for Major 2: This variable determines the closest downstream signalized intersection distance (in miles) to the specified unsignalized one with respect to Major 2. It is to be noted that this distance is exactly the same distance as variable “37”.

42. Distance between Successive Unsignalized Intersections: This distance was specific for each roadway ID. So, the first intersection within each assigned roadway ID always takes a distance value of zero, and the second intersection takes a value of the smaller distance from the upstream or downstream intersection (to account for both stream sides), and so on until the last intersection within the same roadway ID. Then the first intersection in another roadway ID takes a distance value of zero, and the procedure continues for all the collected roadway IDs.

43. Skewness: This variable shows the angle between the centerlines of both major and minor roads (e.g. 45, 60 and 90 degrees). Also, if both minor approaches have different angles with the major approach, it was decided to take the smallest angle as the skewness, so as to get the worst possible case.

II. **Control fields:**

44. Major Control Type: This variable shows the traffic control type on the major road being considered. For unsignalized intersections on state roads, there will be always no traffic control on the major approach, i.e. no stop or yield traffic control, as this major approach always represents a traffic stream with no stops. However, for unsignalized intersections on non-state roads (i.e., three and four-way stopped-controlled intersections), there exists a stop sign on one or both major approaches.
45. Minor Control Type: This variable shows the traffic control type on the minor road being considered (e.g. 1-way stop, 2-way stop, yield traffic control and no traffic control).

Finally, the last variable listed in the table is:

46. Important (Useful) Note: This indicates an important note to be included in the table for some unsignalized intersections that have uncommon characteristics. Also, it indicates special notes for some unsignalized intersections that have been noticed through the data collection procedure.

III. **Traffic fields:**

Traffic fields like AADT on the major approach as well as speed limit on the major approach were collected after merging the previously collected fields with the Roadway Characteristic Inventory (RCI) and Crash Analysis Reporting System (CAR) databases, as it was impossible to collect these data from "Google Earth". Further explanation of how the merging procedure was done is shown as well.

3.3 Median Classification

Median classification was the hardest issue before starting the data collection procedure, and before coming up with the list of variables that represent all the required fields that will be used for collecting data. For the scope of this study, the median type on the major approach is to be considered for classification and analysis purposes. After going back and forth, it was decided to include 6 main types of medians, these are: open, directional, closed, two-way left turn lane, undivided and markings. It is to be noted that open, two-way left turn lane, undivided and markings medians are unrestricted medians; i.e. the vehicle from both major and minor approaches can pass through those median types. On the other hand, directional and closed are restricted medians, i.e. the vehicle on the minor approach can never pass through those two medians.

3.3.1 Closed Median

For the scope of this study, the unsignalized intersection that has a closed median on the major road is always treated as 3-legged intersection with a one-way direction on the major road. An example of a closed median is shown in Figure 3-3.

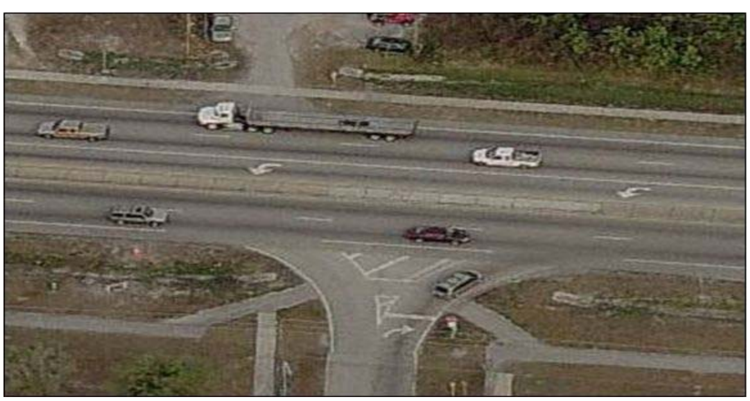


Figure 3-3: A 2x2 Unsignalized Intersection with a Closed Median

From this figure, we can note that the size of the intersection is 2x2 because the major road just near the minor road has 2 through lanes (one-way), and the minor road has 2 lanes on both directions.

3.3.2 Directional Median

The unsignalized intersection that has a directional median on the major road is always treated as two 3-legged intersections. But, it is to be noted that the directional median can be dual (from both major directions) or one-way (from one major direction only). So, for the dual directional median, both major approaches in addition to one of the minor approaches are to be considered. An example of a dual directional median is shown in Figure 3-4.

Figure 3-4: Two 2x6 Unsignalized Intersections with a Dual Directional Median

From this figure, we can note that there are two 3-legged intersections; i.e. the two sides; "a" and "b. For side "a", the minor road on that side in addition to the 2 major road approaches are considered. So, for side "a" (the first 3-legged intersection), the size of the intersection is 2x6. For side "b", the minor road on that side in addition to the 2 major road approaches are considered. So, for side “b” (the second 3-legged intersection), the size of the intersection is 2x6 as well.

An aerial photo of a one-way directional median is shown in Figure 3-5.

Figure 3-5: An Aerial Photo of a One-Way Directional Median

From this figure, we can note that there are two 3-legged intersections; i.e. sides “a” and “b. Side “a” can be treated as if there is a closed median, while side “b” can be treated as if there is a directional median. For side “a”, the minor road on that side in addition to the major road just near that minor road are considered. So, for side “a” (the first 3-legged intersection), the size of the intersection is 2x3. While for side “b”, the minor road on that side in addition to the 2 major road approaches are considered. So, for side “b” (the second 3-legged intersection), the size of the intersection is 2x6.

3.3.3 Open Median

The open median was the hardest type of median for classification. There was a confusion on how to classify an unsignalized intersection that has an open median on the major road and the two minor roads are existing. That is whether to classify this intersection as 2 “3-legged” intersections or one 4-legged intersection. Finally, it was agreed to consider this type of intersection as one 4-legged intersection from the geometry point of view, even if the number of lanes on both major approaches exceeds 6 lanes because there is no geometric restriction for vehicles to pass from the first minor road to the second minor road, crossing the whole major road width. This scope was considered although it was found that drivers do not intend to do this maneuver so often. Drivers usually risk to do this maneuver at late night when the roads are nearly empty. Thus, this scope was considered although this type of maneuver is very rare at daylight.

So, unsignalized intersections with two minor roads and an open median on the major road are treated as a four-legged intersection from the geometric point of view. An aerial photo of a four-legged unsignalized intersection with an open median on the major road is shown in Figure 3-6.

Figure 3-6: A 2x6 Four-Legged Unsignalized Intersection with an Open Median

3.3.4 Two-Way Left Turn Lane Median

An unsignalized intersection having a two-way left turn lane median on the major road is either treated as a 4-legged intersection if both minor roads exist, or a 3-legged intersection if only one minor road exists. Two aerial photos for a 3-legged unsignalized intersection and a 4-legged unsignalized intersection are shown in Figures 3-7 and 3-8, respectively.

Figure 3-7: A 2x4 Three-Legged Unsignalized Intersection with a Two-Way Left Turn Lane Median

Figure 3-8: A 2x2 Four-Legged Unsignalized Intersection with a Two-Way Left Turn Lane Median

3.3.5 Undivided Median

The fifth type of medians are undivided medians. Those undivided medians are mainly two solid yellow lines separating directional traffic, and are most common on two-lane roadways. So, an unsignalized intersection having an undivided median on the major road is treated as 3-legged intersection with both major road approaches in addition to one of the minor road approaches, and as 4-legged intersection if both minor road approaches exist. Two aerial photos for 3 and 4-legged unsignalized intersections with an undivided median on the major road are shown in Figures 3-9 and 3-10, respectively.

Figure 3-9: A 2x4 Three-Legged Unsignalized Intersection with an Undivided Median

Figure 3-10: A 2x2 Four-Legged Unsignalized Intersection with an Undivided Median

3.3.6 Median with Markings

The last type of medians are medians having yellow pavement markings. The main difference between those markings and undivided medians is that for markings, there is a yellow restricted region just in front of the intersection, which acts as a storage area for left turning vehicles to stop by in case there is heavy traffic on the opposing direction. Those markings can act as a storage area for broken down vehicles as well. So, an unsignalized intersection having markings as a median on the major approach is treated as 3-legged intersection with both major road approaches in addition to one of the minor road approaches, and as 4-legged intersection if both minor road approaches exist. Two aerial photos for 3 and 4-legged unsignalized intersections having markings as a median on the major road are shown in Figures 3-11 and 3-12, respectively.

Figure 3-11: A 2x2 Three-Legged Unsignalized Intersection with Pavement Markings as a Median

Figure 3-12: A 2x2 Four-Legged Unsignalized Intersection with Pavement Markings as a Median

3.4 Initial Data Collection Procedure

The initial data collection procedure was done by randomly selecting some unsignalized intersections along randomly selected SRs in Orange County using the “Google Earth” software. These randomly selected unsignalized intersections were chosen on the basis of having as many types of unsignalized intersections as possible. The first chosen road for the data collection process was “SR 50”, and the starting intersection (i.e. node) was the “SR 434/SR 50” signalized intersection. Then afterwards, there was an agreement to move in the westbound direction, heading towards downtown. While moving in the westbound direction of “SR 50”, 25 unsignalized intersections (including access points and driveway intersections) were randomly identified. A sample of these intersections is shown in Figure 3-13.

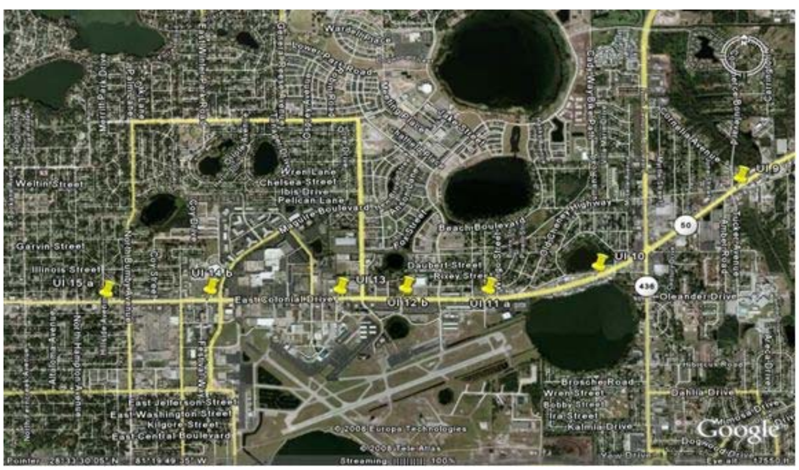


Figure 3-13: Aerial Image from “Google Earth” for 7 Unsignalized Intersections along SR 50 in Orange County during the Initial Data Collection Procedure

After identifying the 25 randomly selected unsignalized intersections along SR 50, it was concluded that it would be extremely hard to identify the respective roadway ID, mile point and node number for each. As a solution, it was decided to think in the reverse manner (i.e. to first identify the unsignalized intersections with their corresponding roadway ID, mile point and node number as indicated in the RCI database using the "Video Log Viewer Application", and then to assign those intersections on "Google Earth". A screen shot of the "Video Log Viewer Application" from the RCI database is shown in Figure 3-14. This application is an advanced tool developed by the FDOT, and has the advantage of capturing the driving environment through any roadway. Moreover, this advanced application has two important features, which are the "right view" and the "front view". The "right view" option provides the opportunity of identifying whether a stop sign and a stop line exist or not. The "front view" feature provides the opportunity of identifying the median type as well as the number of lanes per direction more clearly.

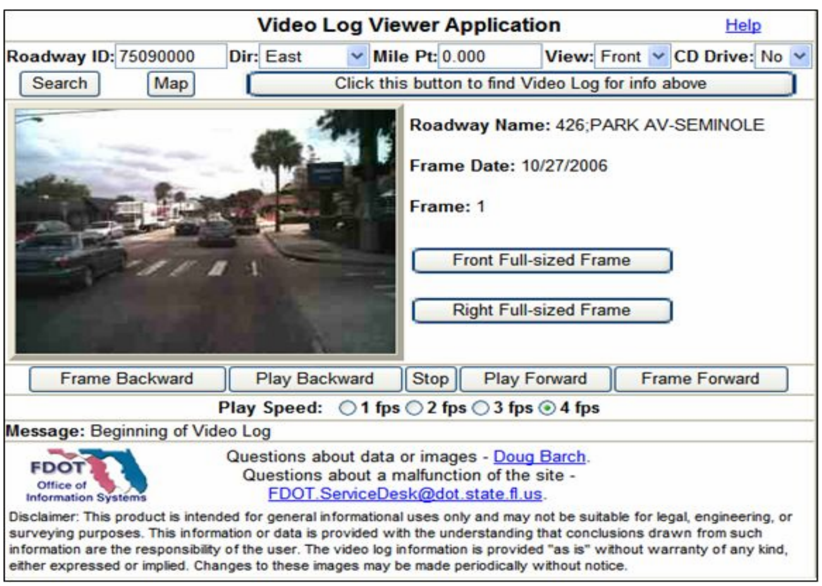


Figure 3-14: Screen Shot of the "Video Log Viewer Application" from FDOT's RCI Database

Thus, the reverse thinking just described led to the last procedure of data collection, which will be detailed in the next section.

3.5 Final Data Collection Procedure

3.5.1 Orange County

As previously mentioned, this procedure came up after deciding to use the RCI database first for identifying unsignalized intersections along state roads. The procedure started with Orange County; and it was noted that there are 31 state roads in Orange County. So, the random selection method was used for choosing some state roads until ending up with 500 unsignalized intersections in this county. The randomly selected state roads were 10, which are: SR 50, SR 434, SR 436, SR 414, SR 423, SR 426, SR 438, SR 424, SR 482 and SR 551. The number of selected intersections on each state road is shown in Table 3-2.

Table 3-2: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Orange County

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
50	201	426	42
434	39	438	35
436	65	424	29
414	8	482	9
423	42	551	30

Then, using the “Video Log Viewer Application” accompanied with an “MS Excel” spreadsheet that has all the unsignalized nodes in the whole state of Florida with their respective roadway ID, mile point and node number, the final data collection procedure was introduced. This procedure was used afterwards for collecting data throughout the remaining 5 counties as well, as this procedure proved to be the most efficient and fastest way. It is to be noted that the previously listed variables in the "MS Excel" spreadsheet was used for collecting data from all the selected 6 counties.

3.5.1.1 Difficulties Faced during the Data Collection Process

The first difficulty encountered was the difficulty of collecting some traffic fields in all the 6 selected counties like AADT on the major approach as well as speed limit on the major approach, as previously mentioned. Thus, it has been decided that these fields are to be filled later on after importing the used "MS Excel" spreadsheet into the "SAS" software, and also importing another "MS Excel" spreadsheet from the RCI database that has all the required characteristics for every roadway ID and mile point. Then, a "SAS" code was used to merge these 2 databases by roadway ID and mile point; thus, all the blank fields will be filled in automatically after the merging procedure in "SAS".

Another difficulty encountered in Orange County was while observing the aerial images from "Google Earth" (e.g. visibility was not too clear to determine the required number of lanes, presence or lack of stop signs, stop bars, extensive presence of trees that blocked the vision, etc.). As a solution, it was decided to use the website "http://www.live.com"; this was used when there was a difficulty in defining some fields that could not be identified through "Google Earth". A screen shot of 3 unsignalized intersections that present some difficulties in defining their geometric characteristics is shown in Figure 3-15.

Figure 3-15: Aerial Image from "Google Earth" for 3 Unsignalized Intersections in Orange County where Difficulty was Encountered in Identifying their Geometry due to Tree Blockage

3.5.2 Brevard County

After collecting the 500 unsignalized intersections in Orange County, the data collection procedure proceeded in the same manner. The second selected county is Brevard County. There were 10 arterials used for collecting the 401 intersections in Brevard County. The used arterials were SRs. The used state roads, and the number of selected unsignalized intersections on each road are shown in Table 3-3.

Table 3-3: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Brevard County

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
3	44	46	24
507	34	50	33
514	30	5	80
518	18	405	24
519	55	A1A	59

3.5.2.1 *Some Unfamiliar Intersections Collected*

After illustrating the SRs used in Brevard County, as well as the number of intersections collected on each, this section discusses some unfamiliar unsignalized intersections collected. Figure 3-16 shows a roundabout just on a 4-legged unsignalized intersection, which is an unfamiliar type. The size of the intersection in this case is “2x4”, and the type of median on the major approach is a two-way left turn lane.

Figure 3-16: An Aerial Image from “Google Earth” for an Unfamiliar Unsignalized Intersection in Brevard County

3.5.3 Hillsborough County

The third selected county is Hillsborough County. There were 10 arterials (SRs) used for collecting 485 intersections in Hillsborough County. The used state roads, and the number of selected unsignalized intersections on each road are shown in Table 3-4.

Table 3-4: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Hillsborough County

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
60	65	574	70
39	28	580	31
45	95	597	33
43	68	676	10
39	31	45	54

3.5.3.1 Some Unfamiliar Intersections Collected

Figure 3-17 shows a 4-legged unsignalized intersection, where both minor approaches are not on the same line, which is an unfamiliar type. The size of the intersection in this case is "2x4", and the type of median on the major approach is an open median.

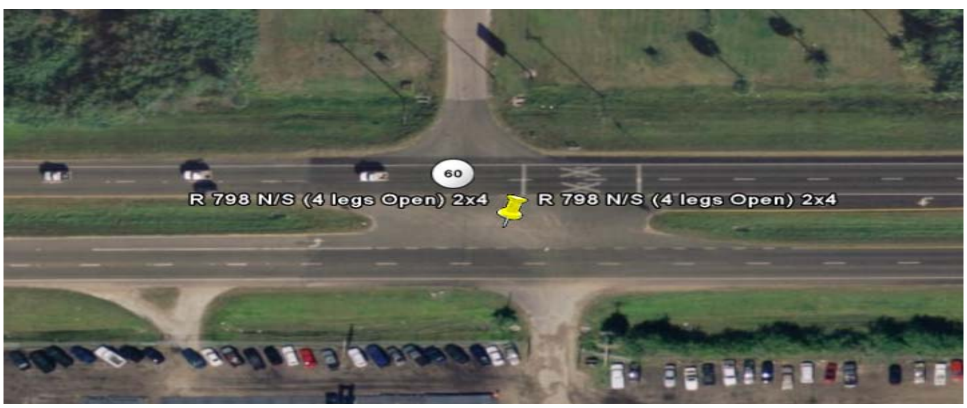


Figure 3-17: First Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Hillsborough County

Figure 3-18 shows a 4-legged unsignalized intersection, where the major approach has a total of 8 lanes (4 lanes per direction), which is an unfamiliar, as it is rarely to see a total of 8 lanes on both directions on arterials. Usually 4 lanes per direction exist on interstate roads. The size of the intersection in this case is “2x8”, and the type of median on the major approach is an open median.

Figure 3-18: Second Aerial Image from “Google Earth” for an Unfamiliar Unsignalized Intersection in Hillsborough County

3.5.4 Miami-Dade County

The fourth selected county is Miami-Dade County. There were 10 arterials (SRs) used for collecting 488 intersections in Miami-Dade County. The state roads used and the number of selected unsignalized intersections on each road are shown in Table 3-5.

Table 3-5: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Miami-Dade County

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
5	36	826	22
9	128	25	49
817	29	90	75
823	64	916	23
94	35	953	27

3.5.4.1 Some Unfamiliar Intersections Collected

Figure 3-19 shows a 3-legged unsignalized intersection on a signalized one. It is clear that there is channelized lane, having a stop sign, for making right on the major approach. Thus, this intersection is not that familiar, where it is very rare to find a stop sign on a signalized intersection. The size of the intersection is “1x3”, and the type of median on the major approach is a closed median.

Figure 3-19: First Aerial Image from “Google Earth” for an Unfamiliar Unsignalized Intersection in Miami-Dade County

Figure 3-20 shows a 4-legged unsignalized intersection. It can be noticed that there is a subway bridge just above the median, so, this intersection is not familiar. The size of the intersection is “2x4”, and the type of median on the major approach is an open median. It is noted also that there is a crosswalk crossing the major approach, without stopping the major road traffic, thus, it is expected to have high percentage of pedestrian crashes at those types of intersections.

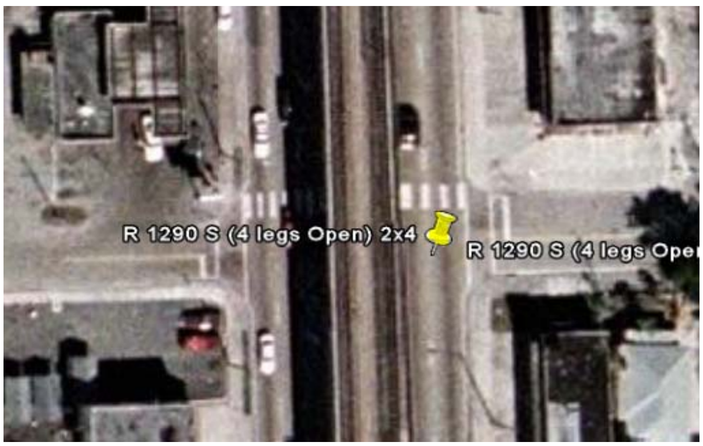


Figure 3-20: Second Aerial Image from “Google Earth” for an Unfamiliar Unsignalized Intersection in Miami-Dade County

Figure 3-21 shows a 3-legged unsignalized intersection. The strange thing in this intersection is the very wide grassed median, as well as having two stop signs on this median for both maneuvers The size of the intersection is “2x4”, and the type of median on the major approach is an open median.

Figure 3-21: Third Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Miami-Dade County

Figure 3-22 shows a 3-legged unsignalized intersection, where the major approach has a total of 8 lanes (4 lanes per direction), and the minor approach has 4 lanes (two approaching lanes and two receiving lanes), so, the size of the intersection is "4x8". It is to be noted that this large size of intersection is an unfamiliar type. The type of median on the major approach is an open median.

Figure 3-22: Fourth Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Miami-Dade County

3.5.5 Leon County

As previously illustrated, Leon County was selected to be representative of a rural county. There were 7 arterials selected for collecting 364 unsignalized intersections in Leon County. Those used arterials were SRs. The used state roads, and the number of selected unsignalized intersections on each road are shown in Table 3-6. It is to be noted that the total number of collected intersections is 364, and not 400. This is attributed to the fact that Leon is a small county, which is much smaller than the previous 4 counties (Orange, Brevard, Hillsborough and Miami-Dade), so it was extremely hard to capture more than those 364 intersections.

Table 3-6: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Leon County

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
10	67	261	31
20	79	263	34
61	83	363	43
63	27		

3.5.5.1 Some Unfamiliar Intersections Collected

Figure 3-23 shows a signalized intersection, where a stop sign exists for the right channelized lane. This type of intersection is uncommon. The size of the intersection in this case is “1x2”, and the type of median on the major approach is closed.

Figure 3-23: First Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Leon County

Figure 3-24 shows a 3-legged unsignalized intersection with a stop sign on the minor leg, where the major approach has a three-direction traffic. Each direction is separated from the other by a median. The size of the intersection in this case is "2x4", and the type of median on the major approach is closed.

Figure 3-24: Second Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Leon County

Figure 3-25 shows a 3-legged unsignalized intersection with a stop sign on the minor leg, and the type of median on the major approach is directional. This directional median is uncommon, where the maneuver is only allowed for a left turn from the minor leg. The common shape of the directional median is to allow only the left-turn maneuver from the major approach, and not the minor one. The size of the intersection in this case is “2x4”.

Figure 3-25: Third Aerial Image from “Google Earth” for an Unfamiliar Unsignalized Intersection in Leon County

Figure 3-26 shows a 3-legged unsignalized intersection with a stop sign on the minor leg, and the type of median on the major approach is open. This open median is uncommon, as there is a small-sized middle median at the centre of the median opening. The traditional way of designing any open median is to have a full median opening. Still, this median type is open, as the left-turn maneuver from both major and minor approaches is permitted. The size of the intersection is “2x4.

Figure 3-26: Fourth Aerial Image from “Google Earth” for an Unfamiliar Unsignalized Intersection in Leon County

Figure 3-27 shows a 3-legged unsignalized intersection, where the major approach has a total of 6 lanes (3 lanes per direction), and the minor approach has 4 lanes (two approaching lanes and two receiving lanes), so, the size of the intersection is “4x6”. It is to be noted that this large size of intersection is an uncommon type. The type of median on the major approach is an open median. Moreover, the exclusive left-turn lane on the northbound major approach is mainly used for U-turns, while the opposing left-turn lane on the southbound major approach can be used for either making a U-turn, or entering the minor leg.

Figure 3-27: Fifth Aerial Image from “Google Earth” for an Unfamiliar Unsignalized Intersection in Leon County

3.5.6 Seminole County

The last selected county is Seminole County. There were 5 arterials selected for collecting 267 unsignalized intersections in Seminole County. Those used arterials were SRs. The used state roads, and the number of selected unsignalized intersections on each road are shown in Table 3-7. Once more, it is to be noted that the total number of collected intersections is 267, and not 400. This is attributed to the fact that Seminole is a small county, which is much smaller than the previous 4 counties (Orange, Brevard, Hillsborough and Miami-Dade), so it was extremely hard to capture more than those 267 intersections.

Table 3-7: The Used SRs, and the Corresponding Number of Unsignalized Intersections on Each of them in Seminole County

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
15	68	426	27
434	73	419	21
46	78		

3.5.6.1 Some Unfamiliar Intersections Collected

Figure 3-28 shows a 3-legged unsignalized intersection (at the yellow pin), where the major road has a total of 6 lanes (3 lanes per direction), and the minor road has 4 lanes (two approaching lanes and two receiving lanes), so, the size of the intersection is "4x6, which is an uncommon type. The type of median on the major approach is an open median. Moreover, with the aid of both Figures 3-28 and 3-29, it can be seen that there are two left-turn lanes on the southbound major approach. Actually, those two left-turn lanes are the extension of the exclusive left-turn lanes of the upstream signalized intersection. So, having 2 left-turn lanes in front of the unsignalized intersection is very dangerous, and can encourage many drivers to use the outer left-turn lane, which is a risky maneuver. It is expected to have large number of angle (left-turn) and side-swipe crashes at this intersection. Angle crashes can result from the conflict between the left-turn maneuver from the southbound major approach with that through maneuver from the northbound major approach. Side-swipe crashes can result from the conflict between the two left-turn maneuvers from the southbound major approach.

Figure 3-28: First Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Seminole County

Figure 3-29: A Further View of the Unsignalized Intersection in Figure 3-28 for Better Clarification of the Extension of the Two Left-Turn Lanes to the Signalized Intersection

Figure 3-30 shows a 4-legged unsignalized intersection, where both minor approaches are not on the same line, which is an unfamiliar type as well. The size of the intersection in this case is "2x4", and the type of median on the major approach is a two-way left turn lane.

Figure 3-30: Second Aerial Image from "Google Earth" for an Unfamiliar Unsignalized Intersection in Seminole County

3.6 Summary Table for the Data Collection Procedure throughout the Six Selected Counties

Tables 3-8 to 3-13 present summary tables for the number of unsignalized intersections collected on each state road in each of the six counties, as well as the total number of intersections collected in each county.

Table 3-8: Summary Table for the Data Collection Procedure in Orange County*

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
50	201	426	42
434	39	438	35
436	65	424	29
414	8	482	9
423	42	551	30

*** Total number of intersections is 500**

Table 3-9: Summary Table for the Data Collection Procedure in Brevard County*

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
3	44	46	24
507	34	50	33
514	30	5	80
518	18	405	24
519	55	A1A	59

*** Total number of intersections is 401**

Table 3-10: Summary Table for the Data Collection Procedure in Hillsborough County*

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
60	65	574	70
39	28	580	31
45	95	597	33
43	68	676	10
39	31	45	54

*** Total number of intersections is 485**

Table 3-11: Summary Table for the Data Collection Procedure in Miami-Dade County*

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
60	65	574	70
39	28	580	31
45	95	597	33
43	68	676	10
39	31	45	54

*** Total number of intersections is 488**

Table 3-12: Summary Table for the Data Collection Procedure in Leon County*

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
10	67	261	31
20	79	263	34
61	83	363	43
63	27		

*** Total number of intersections is 364**

Table 3-13: Summary Table for the Data Collection Procedure in Seminole County*

SR	Number of unsignalized intersections on each SR	SR	Number of unsignalized intersections on each SR
15	68	426	27
434	73	419	21
46	78		

*** Total number of intersections is 267**

3.7 Preliminary Categorization of Unsignalized Intersections

In order to classify unsignalized intersections, various steps were followed in order to fulfill this categorization. Following are the details of these procedures as well as the final categories obtained. It is to be noted that this categorization was not based on any data, but rather from a perspective approach.

1) First of all, unsignalized intersections were classified based on **five main factors.** These five categories are:
   i. Classification based on the number of legs (**3 and 4-legged intersections).**
   ii. Classification based on the size of the intersection (the number of total approach through lanes on the major approach and the number of through lanes on the minor approach) (**2x2, 2x4, 2x6 and 4x4**).
   iii. Classification based on land use (**urban and rural**).
   iv. Classification based on median type on the major approach (**divided and undivided**).
   v. Classification based on type of control on the minor approach (**no control, yield control and stop control**).

It is to be noted that the stop control can be a "1-way stop control" and a "3-way stop control" on a 3-legged unsignalized intersection, and a "2-way stop control" and a "4-way stop control" on a 4-legged unsignalized intersection. Moreover, 2x2, 2x4, 2x6 and 4x4 intersections were used in the categorization procedure, as they were thought to be the dominant intersection sizes. So, the number of possible combinations is: 2 x 4 x 2 x 2 x 2 x 2 = 128 categories.

2) Secondly, after searching for previous literature on modeling crashes at unsignalized intersections, it was found that AADT was a significant factor while modeling crash frequencies occurring at unsignalized intersections. Examples of those studies are those done by Bauer and Harwood (1996), Huang and May (1991), Del Mistro (1981), Kulmala (1997) and Vogt and Bared (1998). Moreover, the posted speed limit on the major approach was an important factor, as indicated by Summersgill and Kennedy (1996), Pickering and Hall (1986) and Brude (1991). Hence, unsignalized intersections were further classified based on **seven main factors.** These seven categories are:

   i. Classification based on the number of legs (**3 and 4-legged intersections**).
   ii. Classification based on the size of the intersection (the number of total approach through lanes on the major approach and the number of through lanes on the minor approach) (**2x2, 2x4, 2x6 and 4x4**).
   iii. Classification based on land use (**urban and rural**).
   iv. Classification based on median type on the major approach (**divided and undivided**).
   v. Classification based on type of control on the minor approach (**no control, yield control and stop control**).
   vi. Classification based on AADT per lane on the major approach.

vii. Classification based on the speed limit on the major approach.

If the AADT to be classified into two categories (high AADT and low AADT) by cutting the AADT at its median, as well as classifying the posted speed into two categories (high speed and low speed limits), the number of possible combinations is: 2 x 4 x 2 x 2 x 2 x 2 x 2 x 2 = 512 categories.

3) In order to be more specific, it was found that classifying median types on the major approach into two categories only was not sufficient. Thus, divided medians were further classified into more specific categories. These categories are open, directional, two-way left turn lane and closed medians. Thus, the seven main categories of unsignalized intersections now become:

    i. Classification based on the number of legs (**3 and 4-legged intersections**).
    ii. Classification based on the size of the intersection (the number of total approach through lanes on the major approach and the number of through lanes on the minor approach) (**2x2, 2x4, 2x6 and 4x4**).
    iii. Classification based on land use (**urban and rural**).
    iv. Classification based on median type on the major approach (**open, directional, two-way left-turn lane, closed and undivided**).
    v. Classification based on type of control on the minor approach (**no control, yield control and stop control**).
    vi. Classification based on AADT per lane on the major approach.
    vii. Classification based on the speed limit on the major approach.

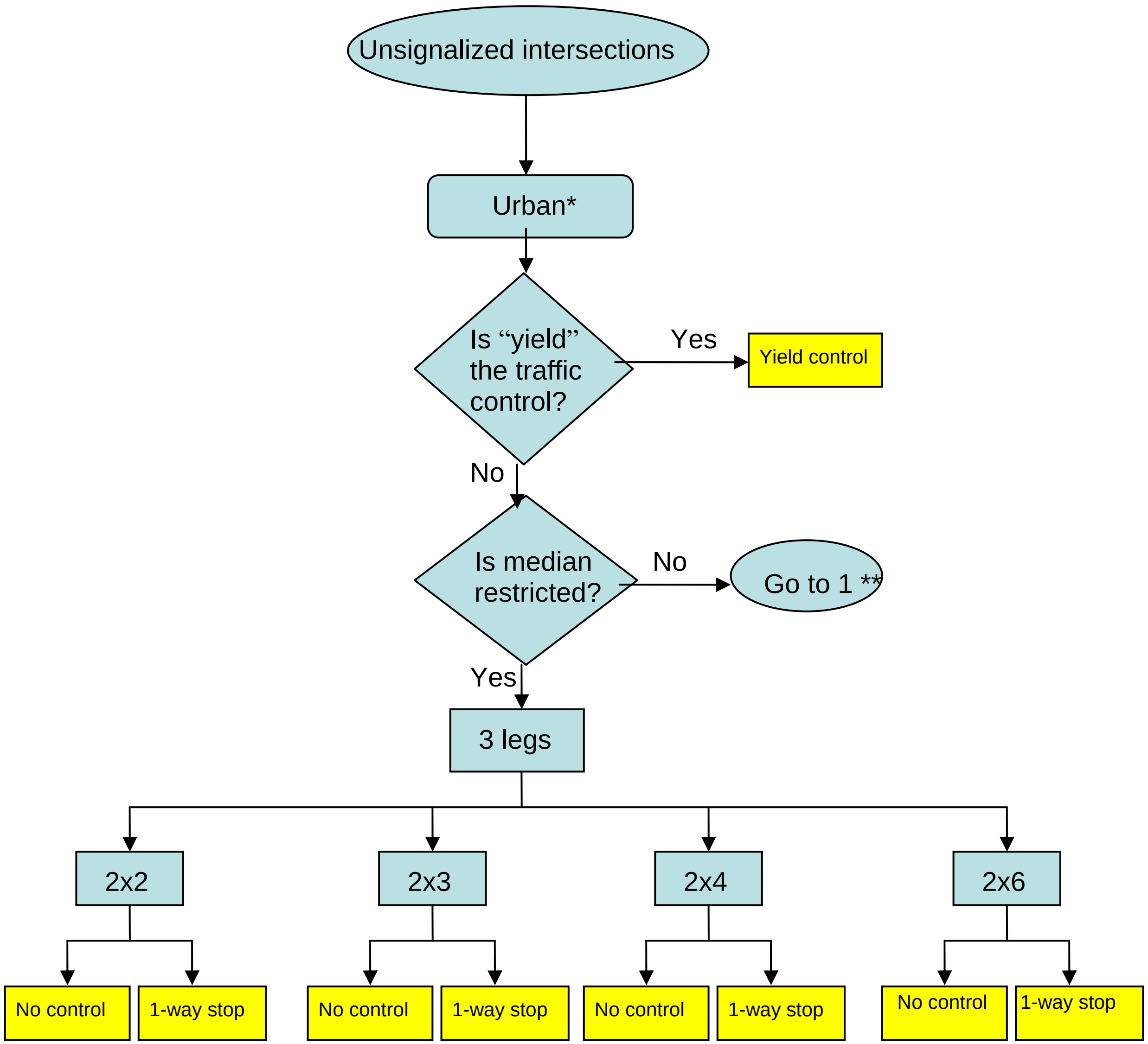
So, the number of possible combinations is: 2 x 4 x 2 x 5 x 2 x 2 x 2 x 2 = 1280 categories.

4) Then, afterwards, it was realized that this introduced number of categories (1280) was a relatively large number. So, this leads to the final step of categorization, which reveals using as few **general** categories as possible to describe nearly all dominant types of unsignalized intersections.

Summarizing the categorization process, the final classification possibilities have been defined as follows:

- Aggregated number of possible combinations = **34 categories.**
- Maximum number of possible combinations = **52 categories.**

Figures 3-31 and 3-32 show a conceptual flow diagram for the final classification of unsignalized intersections based on the maximum and the aggregated categories, respectively.



* The same categorization is done for rural unsignalized intersections

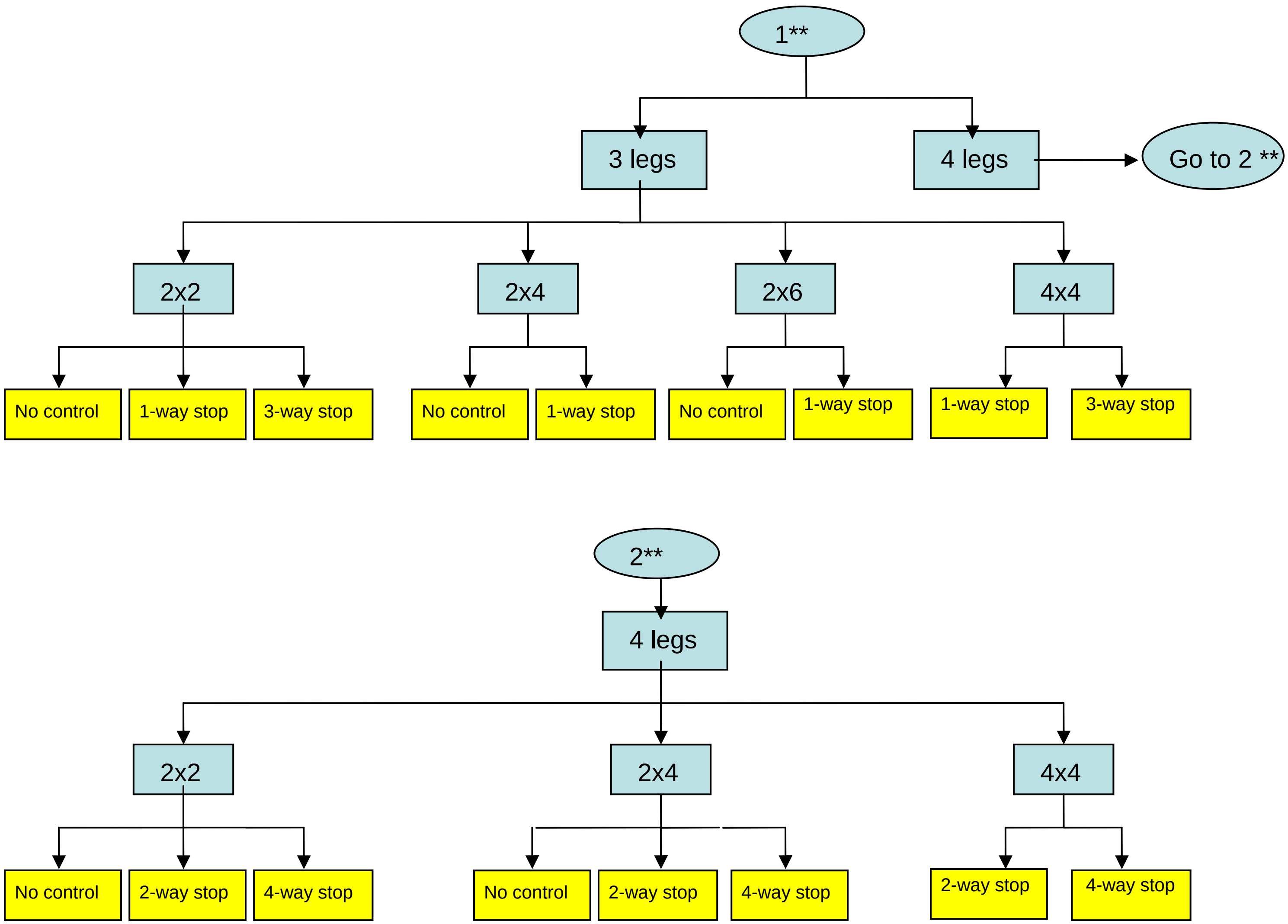
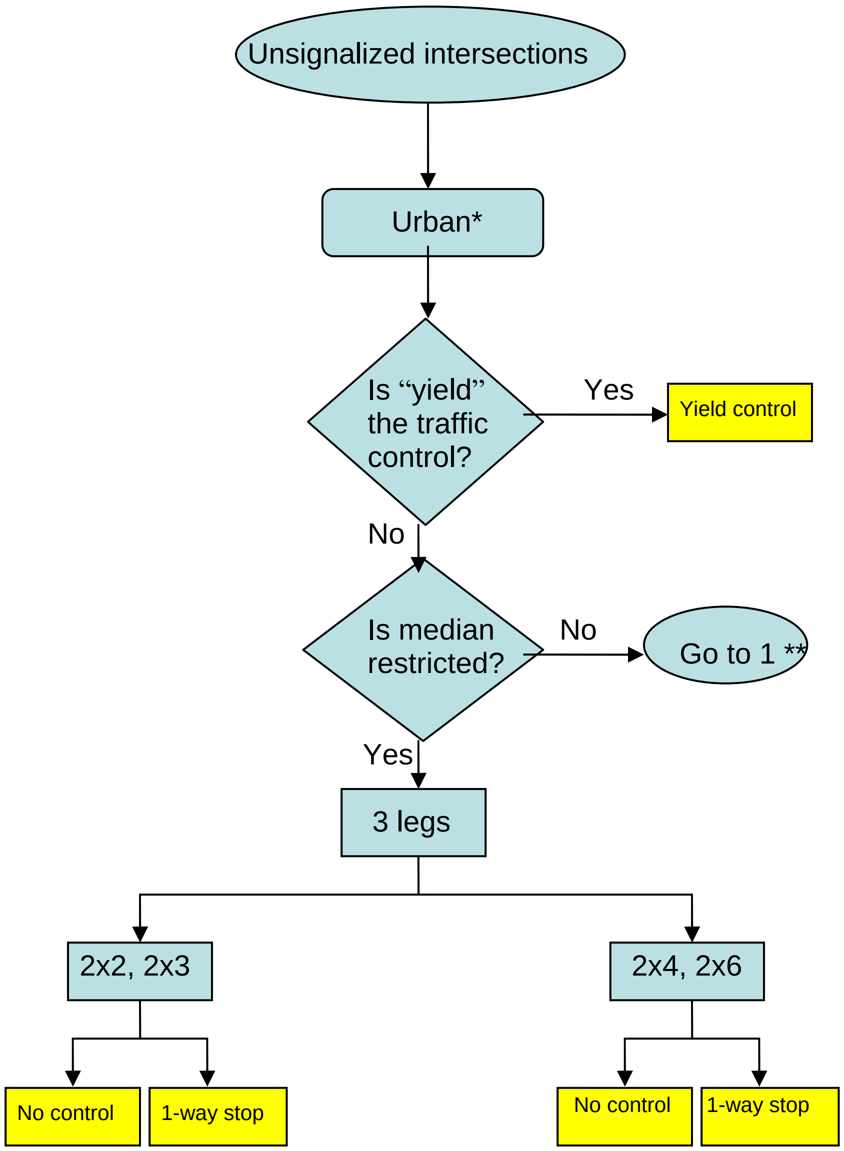


Figure 3-31:Conceptual Flow Chart for the <u>Maximum</u> Preliminary Number of Categories at Unsignalized Intersections



* The same categorization is done for rural unsignalized intersections

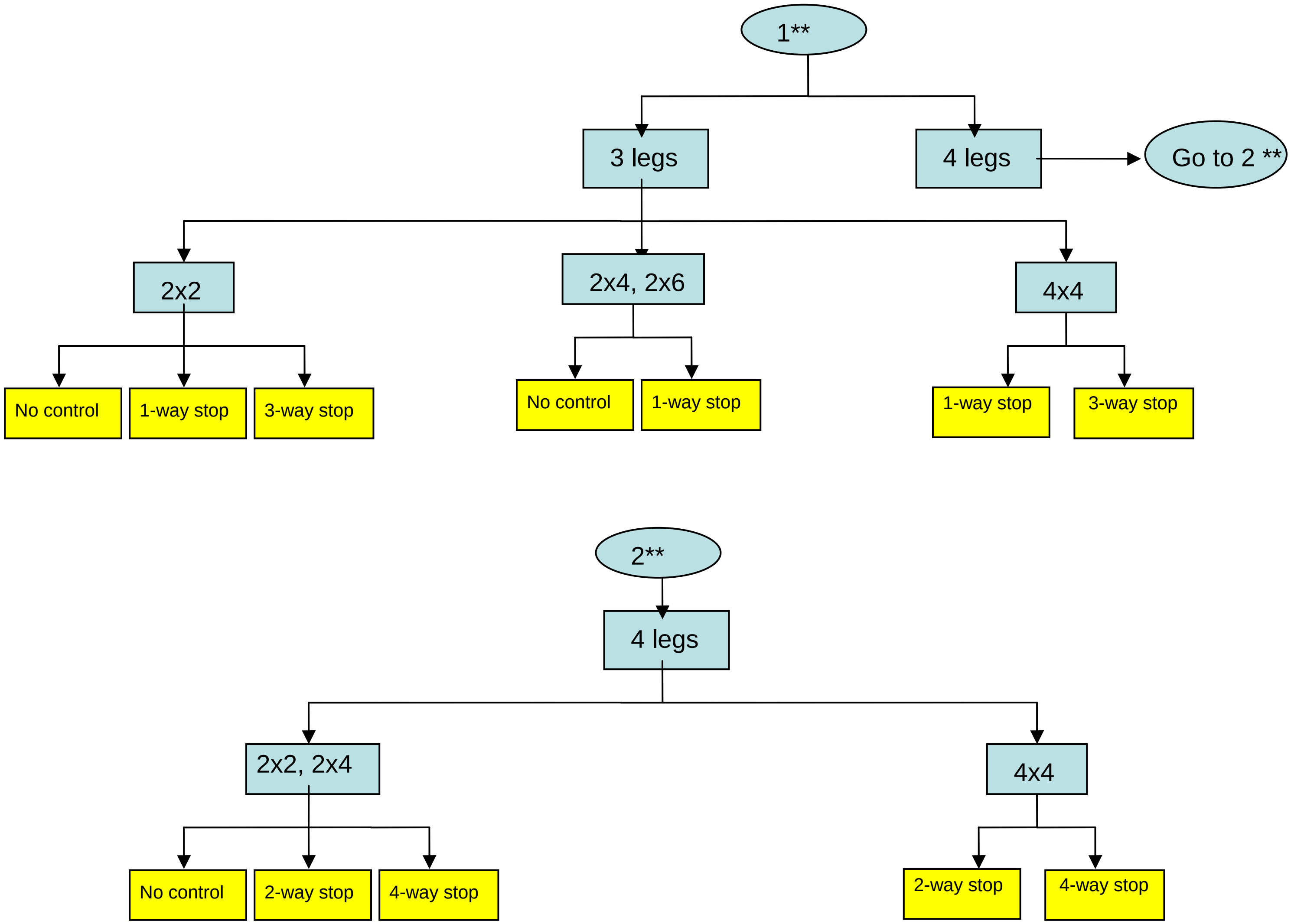


Figure 3-32: Conceptual Flow Chart for the Aggregated Preliminary Number of Categories at Unsignalized Intersection

For the maximum preliminary number of categories (as shown in Figure 3-31), the number of terminal nodes (leaves) in "yellow" color for urban unsignalized intersections is 26, but the same categorization is done for rural unsignalized intersections, so the total number of categorization is 26 * 2 = 52 categories.

For the aggregated preliminary number of categories (as shown in Figure 3-32), the number of terminal nodes (leaves) for urban unsignalized intersections is 17, but the same categorization is done for rural unsignalized intersections, so the total number of categorization is 17 * 2 = 34 categories. It is to be noted that for the 3-legged median-restricted unsignalized intersections, the "2x2" and the "2x3" classification is aggregated together since both of them are usually found on closed and undivided medians, and there is not that much difference between both intersection sizes. Also, the "2x4" and the "2x6" classification is aggregated together since both of them are usually found on two-way left turn lane, open and directional medians.

From Figures 3-31 and 3-32, it is well noticed that the "YIELD" control is used only as one category to summarize all crashes occurring at unsignalized intersections having a "YIELD" control. This "YIELD" control can be either the first traffic control or the second traffic control. This classification of all "YIELD" control crashes as one unsignalized intersection category was concluded through a detailed inspection of crash data for 3 years (2002 – 2004) at all the "YIELD" control types. These crash data include all the crashes occurring in the whole state of Florida with the exception of crashes occurring at freeways. The most important notes from these crash data are summarized as follows:

- Total number of crashes analyzed = 745,342 crash records.

- Number of "YIELD" crashes ("YIELD" is the 1^{st} traffic control) = 3312 crashes (0.49%).
- Number of "YIELD" crashes ("YIELD" is the 2^{nd} traffic control) = 507 crashes (0.08%).
- Number of crashes for a "YIELD" sign as the 2^{nd} traffic control, and a traffic signal as the 1^{st} traffic control = 238 crashes (0.04%).
- Number of crashes for a "YIELD" sign as the 1^{st} traffic control, and a traffic signal as the 2^{nd} traffic control = 84 crashes (0.01%).
- The highest number of "YIELD" crashes ("YIELD" is the 2^{nd} traffic control) occurs at intersections. The number of crashes is 267 crash records (percentage = 267 / 507 = 52.66%). The second highest number occurs at driveway accesses (15.38%).
- The highest number of "YIELD" crashes ("YIELD" is the 1^{st} traffic control) occurs at intersections. The number of crashes is 1299 crash records (percentage = 1299 / 3312 = 39.22%). The second highest number occurs at driveway accesses (16.79%).
- The number of "YIELD" crashes occurring at ramps (entrance or exit ramps) is very small, and can be neglected.

Thus - from the above mentioned points - it is very obvious that crashes occurring at "YIELD" control types are very rare. That is why crashes occurring at "YIELD" traffic control are only categorized as one unsignalized intersection category. This category accounts for crashes occurring at a "YIELD" traffic control on on and off-ramps, on signalized intersections, and on unsignalized intersections (which is very rare).

CHAPTER 4. PRELIMINARY ANALYSIS

The first part of the preliminary analysis conducted in this chapter deals with descriptive statistics plots for the 2500 collected unsignalized intersections. The used crash data were the 4-year crash “2003-2006” aggregated over the 4 years for each intersection. Figure 4-1 shows the plot of the average total crash per intersection in 4 years “from 2003 until 2006” associated with each median type for 3 and 4-legged intersections. The new identified median type is the mixed median (same as the directional one, but allows access from one side only). It is noticed that directional, closed and mixed medians do not exist across from 4-legged intersections. A fast glance at this plot shows that the 4-legged average crashes are much higher than those for 3-legged intersections. Also, the highest average total crashes exist at 3 and 4-legged intersections having open medians across from their approach. This is mainly due to the relatively large number of conflict patterns at open medians, when compared to other types.

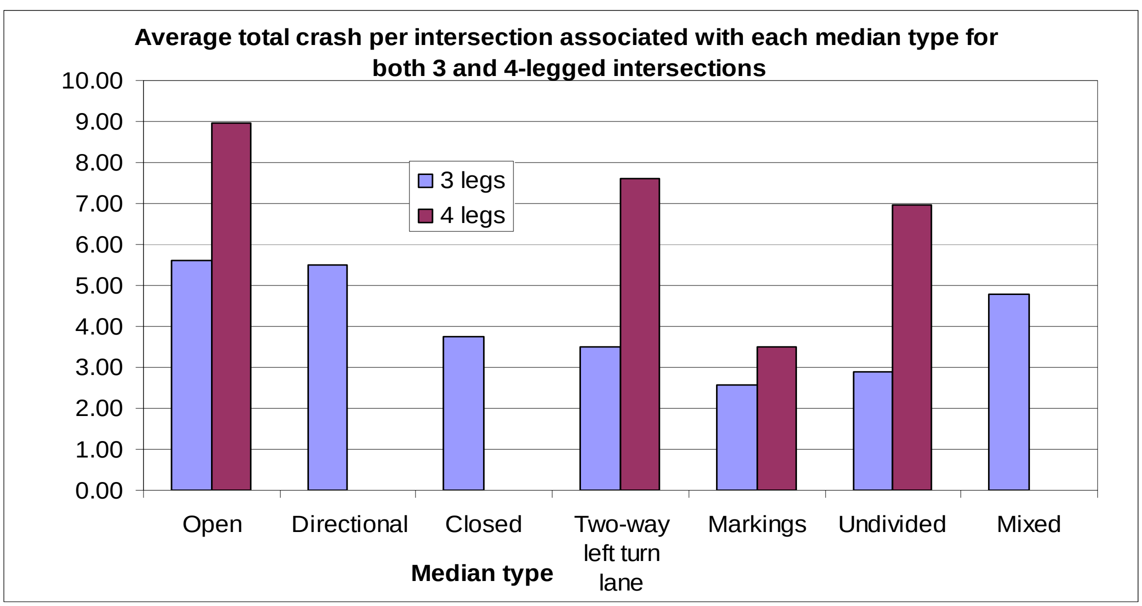


Figure 4-1: Plot of the Average Total Crash per Intersection Associated with Each Median Type

A second plot for the average total crash per intersection in 4 years “from 2003 until 2006” associated with each major road configuration for 3 and 4-legged intersections is shown in Figure 4-2. In fact, the number of 4-legged intersections existing on 8-lane arterials in the dataset was very limited, thus they were excluded. From this plot, it is noticed that as the lane configuration on the major road increases, the average crashes at both 3 and 4-legged intersections increase as well. This shows the hazardous effect of large unsignalized intersection sizes on safety, and this result conforms to the study done by Van Maren (1980).

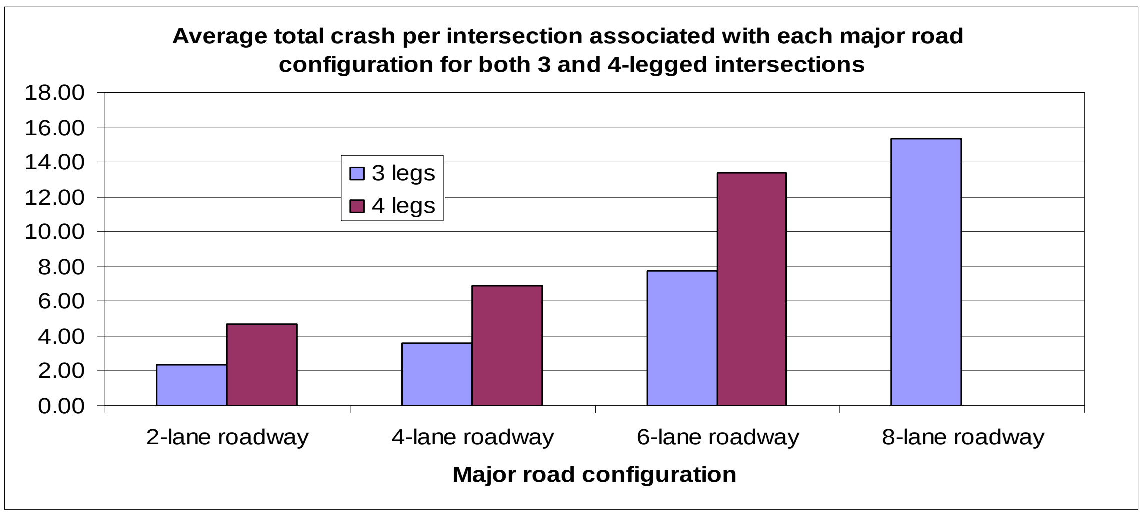


Figure 4-2: Plot of the Average Total Crash per Intersection Associated with Each Median Type

A third plot for the average total crash per intersection type in 4 years “from 2003 until 2006” for 3 and 4-legged intersections is shown in Figure 4-3. Intersections were categorized into four main types, access points or driveways, ramp junctions, regular intersections and intersections close to railroad crossings. Regular intersections are those intersections with distant minor road stretches. Intersections in the vicinity of railroad crossings can exist either upstream or downstream the crossing. It is noted that ramp junctions are always 3-legged. From this plot, it is noticed that intersections close to railroad crossings have the highest average, followed by regular intersections, ramp junctions and finally access points. Also, 4-legged intersections experience higher averages than the 3-legged ones.

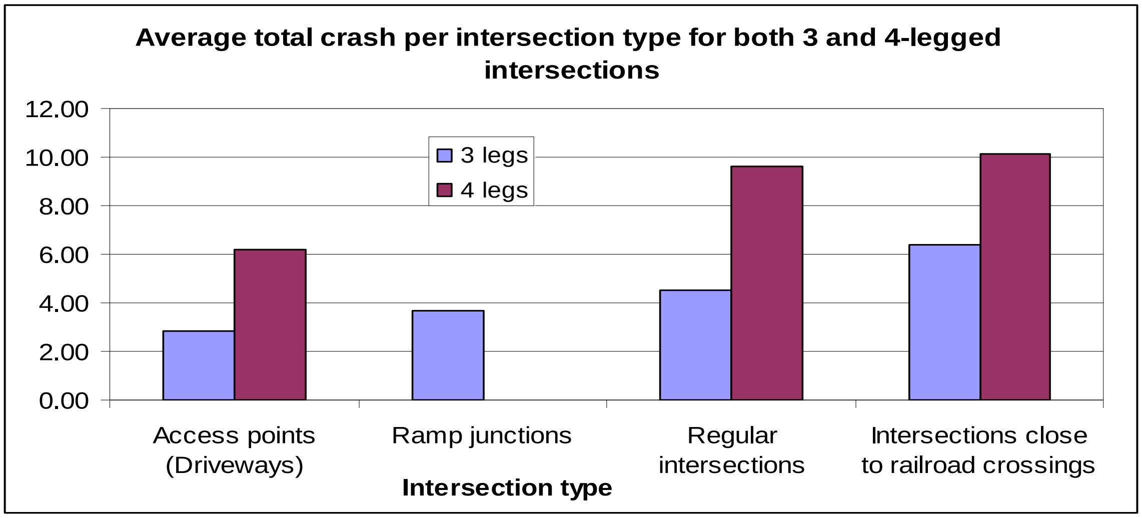


Figure 4-3: Plot of the Average Total Crash per Intersection Associated with Each Median Type

4.1 Safety Effect of the Presence of Both Stop Sign and Line, and Stop Sign Only at Intersections in Orange County

This section is testing the safety effect of the presence of both stop sign and line vs. a stop sign only. The main objective of this analysis is to determine whether the presence of both stop sign and line would help increase or decrease crash frequency at unsignalized intersections. The used county is Orange, since it was the first county collected.

In order to perform this analysis, 4 years of data from (2003 till 2006) were used. Each of the 4-year data includes geometric, traffic and control fields, as previously indicated in Chapter 3. The total number of unsignalized intersections used is 433 intersections, which is deemed a sufficient sample size to perform this type of analysis.

As previously mentioned in Chapter 3, some of those geometric, traffic and control fields were collected using “Google Earth” and “Video Log” applications, and the remaining fields were collected by merging those fields with the RCI database and the CAR database for each year separately. Then afterwards, all these 4 databases representing the 4 years were appended with each other in one database.

Since the collected unsignalized intersections contain both 3 and 4-legged intersections, thus this analysis was done for each type separately. It is to be noted that only stop-controlled intersections were used in the analysis. Thus, any intersection having yield control sign or no control was excluded from this analysis. The number of 3-legged unsignalized intersections for the 4-year database after excluding yield and non-controlled intersections was 237. Of those 237 intersections, 160 intersections have both stop signs and lines (group 1), and 77 intersections have stop signs only, with no stop lines (group 2). For 4-legged unsignalized intersections, the number of unsignalized intersections for the 4-year database after excluding yield and non-controlled intersections was 58. Of those 58 intersections, 25 intersections have both stop signs and lines (group 1), and 33 intersections have stop signs only, with no stop lines (group 2).

Table 4-1 shows a summary descriptive statistics for both groups for the 3 and 4-legged stop-controlled intersections.

Table 4-1: Summary Descriptive Statistics for Group 1 and Group 2

	3-legged stop-controlled intersections		4-legged stop-controlled intersections	
	Group 1	Group 2	Group 1	Group 2
Sample size	160	77	25	33
Total number of crashes through all the selected intersections in 4 years	1348	485	336	319
Average number of crashes per intersection in 4 years	8.425	6.299	13.44	9.667

From this table, it is noticed that the total number of crashes for all the intersections in the 4 years as well as the average number of crashes per intersection in the 4 years for group 1 is more than that for group 2, for both 3 and 4-legged stop-controlled intersections. Moreover, it is well noticed that the average number of crashes per intersection for 4-legged stop-controlled intersections is much higher than the corresponding 3-legged stop-controlled ones. This indicates that 4-legged stop-controlled intersections are much more hazardous than 3-legged stop-controlled intersections, as more conflicts are found for the 4-legged intersections, especially for through maneuvers crossing the whole major road width.

This finding concurs with many studies dealing with safety of unsignalized intersections. For example, David and Norman (1976) as well as Bauer and Harwood (1996) found that four-leg intersections experienced twice as many crashes as three-leg intersections. On the same pattern, Harwood et al. (1995) showed that divided highway intersections with four legs experienced about twice as many crashes as three-leg intersections for narrow medians and more than five times as many crashes as for wide medians. Also, Hanna et al. (1976) found that in rural areas, four-leg intersections experienced 69 percent more crashes than three-leg intersections. Moreover, Leong (1973), O'Brien (1976) and David and Norman (1975) have shown that 3-legged unsignalized intersections are much safer than 4-legged unsignalized intersections, while taking into account the traffic volume parameter.

In order to statistically compare the 2 groups, a student's t-test (for two independent samples) is used to achieve this comparison. However, there are two types of tests existing, which are the student's t-test assuming equal variances for 2 independent samples, and the student's t-test assuming unequal variances. In order to choose one of them, an F-test is initially used to test whether the 2 samples have equal variances or not.

Following this aspect, the aforementioned procedure was done for 3-legged and 4-legged stop-controlled intersections. For 3-legged stop-controlled intersections, the F-test indicated unequal variances for the two tested groups (1 and 2), as the resulted p-value was 0.000215. Then afterwards, the student's t-test assuming unequal variances was used, and the resulted p-value (for a two-tailed distribution) was 0.042976. Thus, there is a sufficient evidence to indicate that there is a significant difference between the 2 groups at the 95% confidence level (5% significance level). This in turn indicates that group 1 (both stop signs and stop lines exist) has a significant higher crash frequency than group 2 (stop signs only exists). So, having both stop signs and lines for 3-legged stop-controlled intersections is much riskier than having stop signs only.

Although this finding is unexpected, this is mainly attributed to the fact that taking care of the existence of both stop signs and stop lines is always done at hazardous intersections. Another reason is that there were some trees blocking stop signs' visibility in group's 1 sample while collecting geometric fields in the data collection procedure. Those dense trees can act as visibility blockage for motorists approaching the intersection. Thus, in spite of having a stop line on the pavement, motorists could not make a full stop due to the inexistence of stop sign (from their perspective), and thus a crash happens, as most motorists do not consider an intersection as a stop-controlled intersection unless a stop sign is provided.

For 4-legged stop-controlled intersections, the F-test indicated equal variances for the two tested groups (1 and 2), as the resulted p-value was 0.22262. Then, the student's t-test assuming equal variances was used, and the resulted p-value (for a two-tailed distribution) was 0.21175. Thus, there is not a sufficient evidence to indicate that there is a significant difference between the 2 groups at the 95% confidence level. So, the existence of both stop signs and lines for 4-legged stop-controlled intersections has significantly the same safety effect (in terms of crash frequency at those selected intersections) as having stop signs only.

4.2 General Conclusions and Recommendations from the Analysis

This analysis has examined the safety effect of the existence of both stop signs and stop lines, and stop signs only for 3-legged and 4-legged stop-controlled intersections in Orange County. Although it was concluded that having both stop signs and lines for 3-legged stop-controlled intersections is significantly much riskier than having stop signs only, this should not be a misleading finding. The reason is that this analysis is based on simple statistics, and also the minor road AADT was not reflected in this analysis (since it was not available). And for 4-legged stop-controlled intersections, it was concluded that there is no significant difference between those two categories in terms of the safety pattern.

Thus, as a recommendation, installing another stop sign on the left side of the minor road (or minor driveway, or access point) at those 3-legged stop-controlled intersections with both stop signs and lines is one of the safety countermeasures for alleviating that significant high crash occurrence. This countermeasure was examined by Polaris (1992), who found it to be effective in some cases.

Also, in order to increase drivers' awareness of the existence of stop signs, rumble strips can be installed at intersection approaches in order to call their attention. Figure 4-4 shows how rumble strips are installed on the pavement. Rumble strips are usually recommended for application when measures such as pavement markings or flashers were tried and showed failure to alleviate high crash occurrence. Moreover, rumble strips can be coordinated with a "STOP AHEAD" device, i.e. when the driver crosses the rumble strip, this control device starts flashing. More literature review about rumble strip usage can be found in Harwood (1993). He suggests that installing rumble strips on stop-controlled approaches can provide a reduction of at least 50 percent in rear-end crashes as well as crashes involving running through a stop sign. Moreover, installing advance stop sign rumble strips was one of the countermeasures recommended by the research “Strategies to Address Nighttime Crashes at Rural, Unsignalized Intersections, 2008”.

Figure 4-4: Rumble Strips Installation

Finally, maintenance of stop signs should be performed at a high standard to ensure that their effectiveness is obtained. According to the Manual on Uniform Traffic Control Devices “MUTCD” criteria, stop signs should be kept clean, and visible at all times (at day and night). Improper signs should be replaced without delay. Special care should be taken to make sure that trees, shrubs, and other vegetations do not block stop signs.

CHAPTER 5. USING A RELIABILITY PROCESS TO REDUCE UNCERTAINTY IN PREDICTING CRASHES

5.1 Background

In spite of the fact that intersections constitute only a small part of the overall highway system, intersection-related crashes are considered high. According to the Florida Department of Transportation (2006), there is an average of 5 crashes at intersections every minute and one person dies every hour at an intersection somewhere in the nation. Additionally, almost one in every four fatal crashes occurs at or near an intersection. In 2004, Florida led the nation in intersection fatalities, where 30% of fatalities occurred at intersections, and in 2006, around 43% of fatalities occurred at or were influenced by intersections.

This chapter deals with investigating and predicting crash frequency at 3 and 4-legged unsignalized intersections using the NB statistical model, which helps to identify those geometric and traffic factors leading to crashes at those intersections. In addition, reducing uncertainty developed from the probabilistic NB model was explored using the full Bayesian updating approach by updating the estimated coefficients from the fitted NB models for better prediction. For the scope of this analysis, the 2-year "2003-2004" crash data were used for modeling purposes, and the 2-year "2005-2006" crash data were used for predictions and assessments.

Statistical models are common tools for estimating safety performance functions of many transportation systems (Abbess et al., 1981; Kulmala, 1995; Lord, 2000; Miaou and Lord, 2003; Oh et al., 2003; Miaou and Song, 2005; Caliendo et al., 2007). The most common probabilistic models used by transportation safety analysts for modeling motor vehicle crashes are the traditional Poisson and Poisson-gamma (or NB) distributions. NB regression models are usually favored over Poisson regression models because crash data are usually characterized by over-dispersion (Lord et al., 2005), which means that the variance is greater than the mean. The NB distribution takes care of the over-dispersion criterion (Hauer, 1997). Other advantages of using the NB model can be found in Park and Lord (2007) and Miaou and Lord (2003).

A Bayesian formulation combines prior and current information to derive an estimate for the expected safety performance of that site being evaluated (Persaud et al., 2009). Empirical Bayes and full Bayes are the two types of Bayesian approaches. According to Persaud et al. (2009), “the full Bayesian approach has been suggested lately as a useful, though complex alternative to the empirical Bayes approach in that it is believed to better account for uncertainty in analyzed data, and it provides more detailed causal inferences and more flexibility in selecting crash count distributions”.

The analysis in the chapter aims at achieving the following objectives:

1) Providing a crash frequency model (safety performance function) for 3 and 4-legged unsignalized intersections using the NB statistical model. Detailed data are collected to identify significant factors contributing to crashes at unsignalized intersections.
2) Applying the Bayesian updating approach to update not only the best estimates of the parameter coefficients, but also to generate full probability distributions for the coefficients.

3) Evaluating and comparing the fitted NB models before updating and the Bayesian-structure models after updating using several criteria, like the capability of reducing uncertainty ("standard errors" of the fitted models were used as surrogate measure for "uncertainty"), the Akaike information criterion (AIC), the mean absolute deviance (MAD), the mean square prediction error (MSPE), and the overall prediction accuracy.

5.2 Methodological Approach

NB regression is widely used to model otherwise over-dispersed Poisson models. Over-dispersion can lead to biased standard errors, resulting in predictors appearing to significantly contribute to the model, when in fact they do not. The NB methodological approach can be found in previous studies (e.g., Miaou, 1994; Poch and Mannering, 1996; Park and Lord, 2008; Saha and Paul, 2005).

The following discussion addresses the Bayesian updating concept. For applying the Bayesian updating framework with the log-gamma likelihood function, the following equation describes the log-gamma distribution of crash frequency.

$$C_i = \exp(X_i^T \Theta)\exp(\varepsilon_i) = \hat{\lambda}\ h_i \tag{5.1}$$

where C_i is the number of crashes, X_i is the vector of variables or uncertain parameters considered in the analysis, Θ is the vector of coefficients to these parameters, $\hat{\lambda}$ is the best estimate of the crash prediction model, and $h_i = \exp(\varepsilon_i)$ is the error term that has the one parameter gamma distribution with mean = 1, and variance σ^2 equals the over-dispersion parameter (= $1/\theta_g$).

$$f(h_i) = \frac{\theta_g^{\theta_g} \, h_i^{\theta_g - 1} \exp(-\theta_g \, h_i)}{\Gamma(\theta_g)} \tag{5.2}$$

with $h_i > 0$, $\theta_g > 0$ and $p(\theta_g) = 1/\theta_g$.

$\varepsilon_i = \ln(h_i)$ so that ε_i has a log-gamma distribution, as shown in Equation (5.3).

$$f(\varepsilon_i) = \frac{\theta_g^{\theta_g} \; \exp[-\theta_g \exp(\varepsilon_i)] \, [\exp(\varepsilon_i)]^{\theta_g}}{\Gamma(\theta_g)} \tag{5.3}$$

with $-\infty < h_i < \infty$, and $\theta_g > 0$, and $\Gamma(\,)$ as the gamma function.

Bayesian updating provides a framework for including subjective data, rather than objective data, into a probabilistic or reliability analysis. An existing state of knowledge regarding uncertain parameters in a model can be updated by observations that may take the form of actual data points, upper or lower bounds, and ranges of values. The vector of uncertain parameters considered in the model is denoted as Θ. The input to and the result after Bayesian updating are both joint probability distributions $f(\Theta)$.

The prior distribution of Θ is updated using the following formula:

$$f(\Theta) = c \, L(\Theta) \, p(\Theta) \tag{5.4}$$

where $L(\Theta)$ is the likelihood function that contains observations regarding the model that are used to update the prior joint distribution of parameters $p(\Theta)$. The resulting posterior distribution of the parameters is obtained after determination of the normalization constant c that guarantees the joint posterior distribution normalizes to a unit value.

For the log-gamma model considered in this chapter, the likelihood function is:

$$L(\Theta, \theta_g) \propto \prod p[C_i = \hat{\lambda}(\Theta) \; \exp(\varepsilon_i)] \tag{5.5}$$

Thus, $L(\Theta,\theta_g) \propto \prod f_{h_i}\left(\frac{C_i}{\hat{\lambda}(\Theta)}\right)$ (5.6)

Applying Equation (5.6) in Equation (5.2) yields the following:

$$f(h_i)=\frac{\theta_g{}^{\theta_g}[\frac{C_i}{\hat{\lambda}(\Theta)}]^{\theta_g-1}\exp[-\theta_g\frac{C_i}{\hat{\lambda}(\Theta)}]}{\Gamma(\theta_g)} \quad (5.7)$$

The NB Bayesian model uses the same functional form as shown in Equation (5.1); however, the error term is described by the NB distribution.

There are many choices of possible prior distributions $p(\Theta)$, making the use of Bayesian updating less intuitive for some. In this study, it was assumed that the error term was statistically independent of other parameter estimates of the NB model for estimating the prior distribution. In the absence of prior information, one can make use of a non-informative prior. Depending on the domain of parameters, Box and Tiao (1992) have suggested non-informative priors. As the parameters Θ considered in this study are diffuse, the non-informative prior is a constant and absorbed by the normalization constant *c*, except for the parameters describing the one-parameter log-gamma distribution (θ_g) and negative binomial distribution (α). These two parameters are limited to the positive domain; therefore the non-informative prior takes the form of $1/\theta_g$ and $1/\alpha$, respectively.

In this study, both non-informative and informative priors were explored. For both priors, two likelihood distributions were examined, the NB and log-gamma distributions. The non-informative prior reflects a lack of information at the beginning of the analysis and can be used to estimate the joint distribution of the parameters $f(\Theta)$.

Informative priors use known information and often result in lower uncertainty in the posterior distributions for each of the parameters being updated. As the informative prior distribution need not be exact to obtain accurate posterior results, it is assumed in this paper that the parameters follow a multinormal distribution. This was applied for the NB and log-gamma likelihood functions with informative priors.

The multinormal prior distribution is specified according to:

$$p(\Theta)=\frac{1}{(2\pi)^{n/2}\left|\Sigma_{\Theta\Theta}\right|^{1/2}}\exp[-\frac{1}{2}(\Theta-M_{\Theta})^{T}\,\Sigma_{\Theta\Theta}{}^{-1}(\Theta-M_{\Theta})] \tag{5.8}$$

where M_{θ} is the mean parameter vector, $\Sigma_{\theta\theta}$ is the covariance matrix so as to have a desirable confidence interval for the updated parameters, and n is the total number of parameters being estimated.

For the case of the NB likelihood function updated using an informative prior in Equation (5.8), the mean parameter values were selected based on expert traffic engineering judgment and opinion. To illustrate this point, for example, it is expected that the logarithm of AADT increases crash frequency at intersections, as shown in Wang and Abdel-Aty (2006), hence, it was assigned a high positive sign (e.g., +1). Other new variables that were not examined before such as the presence of right and left turn lanes on the major approach were based on the engineering assessment. The presence of a right turn lane on each major approach is expected to reduce crash more than the existence on one approach only. Hence, the presence of one right turn lane on each approach was assigned a value of -1, and on one approach only was assigned a value of -0.5. The covariance matrix was assigned values that could yield a 70% confidence interval, by assuming the standard deviation is equal to or greater than the parameter estimate (i.e., a coefficient of variation of one).

For the log-gamma likelihood with informative prior, M_{θ} is the mean parameter vector determined from the posterior estimates of the log-gamma likelihood with non-informative prior, and $\Sigma_{\theta\theta}$ is the covariance matrix determined from the same posterior estimates. As the unsignalized intersection data were used to estimate these posterior statistics, a set of additional 66 three and four-legged intersection data was collected from a neighboring county (Seminole County) and used to populate the log-gamma likelihood function for the second updating (to avoid using same data twice). As an illustration, for the 3-legged model, the values 0, 1, 1 and 10.31 correspond to an intersection having no stop sign on the minor approach, one right turn lane on each major approach, one left turn lane on each major approach and a natural logarithm of AADT of 10.31.

The difficulty in applying the Bayesian updating formula "Equation (5.4)" for a value of parameters Θ with order higher than 3 is the determination of the normalization constant and posterior statistics. The normalization constant is computed according to,

$$c = \left[\int L(\Theta)p(\Theta)d\Theta\right]^{-1} \tag{5.9}$$

where the integral is over as many dimensions as the order of Θ. Standard numerical integration techniques are cumbersome and not well behaved in terms of convergence, especially when the domain of the integrals is from $-\infty$ to ∞. There are many numerical approaches for computing the posterior statistics, including crude Monte Carlo simulation, importance sampling, directional simulation, and Markov chain Monte Carlo (MCMC) simulation.

An importance sampling method of computing the Bayesian integrals is adopted in this study based on the approach taken by Gardoni et. al. (2002). The Bayesian integral is rewritten in a more general form as,

$$I = \int W(\Theta)\, L(\Theta)\, p(\Theta)\, d\Theta \qquad (5.10)$$

The normalization constant (c) can be calculated by setting $W(\Theta) = 1$ in Equation (5.10). Similarly, the posterior statistics are easily found by setting $W(\Theta)$ equal to $c * \Theta$ for mean of the posterior (M_Θ), and setting $W(\Theta)$ equal to $c * \Theta * \Theta^T$ for the mean square. Thus, the covariance for the posterior ($\Sigma_{\Theta\Theta}$) = $E[\Theta\Theta^T]$ - M_Θ M_Θ.

Equation (5.10) is solved used importance sampling by letting:

$$B(\Theta) = W(\Theta)\; L(\Theta)\; p(\Theta) \qquad (5.11),\ \text{and}$$

$$I = \int \frac{B(\Theta)}{S(\Theta)} S(\Theta)\, d\Theta \qquad (5.12)$$

where $S(\Theta)$ is the importance sampling density function, and $S(\Theta) \neq 0$ whenever $B(\Theta) \neq 0$. The solution of the integration in Equation (5.12) is equal to the expectation of $\frac{B(\Theta)}{S(\Theta)}$ relative to the sampling density.

$\bar{I} = \frac{1}{N}\sum_{i=1}^{N}\frac{B(\Theta)}{S(\Theta)}$ with random iterations Θ_i , i = 1, 2, 3, - - - , N

Also following the recommendations of Gardoni et al. (2002), the joint sampling distribution is calculated based on a Nataf distribution (Nataf, 1962) with marginals and correlations specified for each one of the model parameters. Only the first two parameters of the marginals are estimated both using the method of maximum likelihoods. The mean is obtained from the maximum likelihood estimate of Θ and the covariance is obtained from the negative inverse of the Hessian of the log-likelihood function, evaluated at the maximum likelihood estimate.

To examine the effect before and after updating, there were two evaluations performed. One was based on comparing the mean estimates of the parameters, and four MOE criteria were used for assessment, AIC, MAD, MSPE and the overall prediction accuracy. AIC offers a relative measure of the information lost when a given model is used to describe reality. Also, AIC is used to describe the tradeoff between bias and variance in model construction, as well as between precision and complexity of the model. So, the lower the AIC, the better is the model. The MAD and MSPE criteria were also used in the study done by Lord and Mahlawat (2009) for assessing the goodness-of-fit of the fitted models. Moreover, the same MOE criteria were used by Jonsson et al. (2009) to assess the fitted models for both three and four-legged unsignalized intersections. The overall prediction accuracy is estimated by dividing the total predicted crashes by the total observed crashes at the collected intersections. Equations (5.13) and (5.14) show how to evaluate MAD and MSPE, respectively.

$$MAD = \frac{1}{n}\sum\left|y_i - \mu_i\right| \tag{5.13}$$

$$MSPE = \frac{1}{n}\sum(y_i - \mu_i)^2 \tag{5.14}$$

where n is the sample size in the prediction dataset (2005-2006); y_i is the observed crash frequency for intersection i; and μ_i is the predicted crash frequency for intersection i.

The second evaluation was done for comparing uncertainty reduction before and after updating. For measuring uncertainty, standard errors of the estimated and updated parameters were used as surrogated measure for uncertainty.

5.3 Description of Variables

The data collection process was performed after identifying unsignalized intersections in Orange County. The CAR database retrieved from the FDOT was used to identify all the SRs in Orange County, and it was noted that there are 31 SRs in Orange County. The random selection method was used for choosing some state roads. Then, unsignalized intersections were identified along these randomly selected SRs using "Google Earth" and "Video Log Viewer Application", hence leading to identifying 328 unsignalized intersections in Orange County (257 three-legged and 71 four-legged). The "Video Log Viewer Application" requires the roadway ID for the SR, the mile point and the direction of travel. This application is an advanced tool developed by the FDOT, and has the advantage of capturing the feeling of driving along this roadway. Moreover, this advanced application has two important features, which are the "right view" and the "front view". The "right view" option provides the opportunity of identifying whether a stop sign and a stop line exist or not, as these are important variables, as will be discussed later. The "front view" feature provides the opportunity of identifying the median type as well as the number of lanes per direction more clearly.

Afterwards, all the geometric, traffic and control fields of these 328 intersections were filled out in a spreadsheet. These collected fields were merged with the RCI database for the 4 years (2003, 2004, 2005 and 2006) separately. The RCI database – which is developed by the FDOT - includes physical and administrative data, such as functional classification, pavement, shoulder and median data related to the roadway (the New Web-based RCI Application). Each of these facilities is indexed by a roadway ID number with beginning and ending mile points. The used criteria for merging the data are the roadway ID and the mile point. The crash frequency for those identified unsignalized intersections was determined from the CAR database. Then once more, this crash frequency database for the 4 years was merged with the already merged database (geometric, traffic and control fields with RCI database) for the 4 years separately. In this case, the used criterion for merging is the intersection ID. All these merging procedures were done using SAS (2002). In Florida, a distance of 250 feet – measured from the centre of the intersection - is set as the default value for “influenced by intersection” crashes. A full description of the important variables used in the modeling procedure for 3 and 4-legged unsignalized intersections is shown in Table 5-1.

From Table 5-1, it is important to note the existence of mixed medians on the major road (level “7”). This is shown in Figure 5-1, and it depicts one of the cases encountered in the data collection procedure. This case is presented to illustrate the complexity of the data collection phase. The intersection on the right side of the figure is an unsignalized intersection, where the median type on the major road is a directional median. The intersection on the left side of the figure is an unsignalized intersection, where the median type on the major road is a closed median. Since no vehicles can cross from one side to the other, we consider these two intersections as two 3-legged unsignalized intersections.

Figure 5-1: Mixed Median Type on the Major Road (Directional from One Side, and Closed from the Other Side) (Retrieved January 20, 2008, from Google Earth)

From Table 5-1, the cutoff value for classifying the skewness angle into 2 categories is 75 degrees. This value is based on previous studies (Gattis and Low, 1998; Wang, 2006). Wang (2006) has found that a minimum of 37 to 75 degrees will offer an improved line of sight. For the size of the intersection (e.g. “2x4” intersections), the first number indicates the number of lanes for both directions in the minor road, and the second number indicates the number of through lanes for both directions in the major road.

Table 5-1: Variables Description for 3 and 4-Legged Unsignalized Intersections

Variable Description	Variable Levels for 3 Legs	Variable Levels for 4 Legs
Existence of stop sign on the minor approach	= 0; if no stop sign exists; and = 1; if stop sign exists	= 1; if only one stop sign exists on one of the minor approaches; and = 2; if one stop sign exists on each minor approach
Existence of stop line on the minor approach	= 0; if no stop line exists; and = 1; if stop line exists	= 1; if only one stop line exists on one of the minor approaches; and = 2; if one stop line exists on each minor approach
Existence of crosswalk on the minor approach	= 0; if no crosswalk exists; and = 1; if crosswalk exists	= 0; if no crosswalk exists; and = 1; if only one crosswalk exists on one of the minor approaches; and
Size of the intersection (the first number represents total number of approach lanes for the minor approach, and the second number represents total number of through lanes for the major approach)	= 1; for “2x2” and “2x3” intersections; = 2; for “2x4” intersections; and = 3; for “2x6” intersections	= 1; for “2x2” intersections; = 2; for “2x4” intersections; and = 3; for “2x6” intersections
Number of right turn lanes on the major approach	= 0; if no right turn lane exists; = 1; if one right turn lane exists on only one direction; and = 2; if one right turn lane exists on each direction*	= 0; if no right turn lane exists; = 1; if one right turn lane exists on only one direction; and = 2; if one right turn lane exists on each direction
Number of left turn lanes on the major approach	= 0; if no left turn lane exists; = 1; if one left turn lane exists on only one direction; and = 2; if one left turn lane exists on each direction**	= 0; if no left turn lane exists; = 1; if one left turn lane exists on only one direction; and = 2; if one left turn lane exists on each direction
Number of through movements on the minor approach	N/A***	= 1; if one through movement exists on one minor approach only; and = 2; if one through movement exists on each minor approach
Median type on the major approach	= 1; for open median; = 2; for directional median; = 3; for closed median; = 4; for two-way left turn lane; = 5; for markings in front of the intersection; = 6; for undivided median; and = 7; for mixed median (directional from one side, and closed from the other side)	= 1; for open median; and = 4; for two-way left turn lane
Median type on the minor approach	= 1; for undivided median, two-way left turn lane and markings; and = 2; for any type of divided median	= 1; for undivided median, two-way left turn lane and markings; and = 2; for any type of divided median

Variable Description	Variable Levels for 3 Legs	Variable Levels for 4 Legs
Skewness level	= 1; if skewness angle <= 75 degrees; and = 2; if skewness angle > 75 degrees	= 1; if skewness angle <= 75 degrees; and = 2; if skewness angle > 75 degrees
Natural logarithm of the section annual average daily traffic "AADT" on the major road	---****	---
Natural logarithm of the upstream distance (in feet) to the nearest signalized intersection from the unsignalized intersection of interest	---	---
Natural logarithm of the downstream distance (in feet) to the nearest signalized intersection from the unsignalized intersection of interest	---	---
Left shoulder width near the median on the major road (in feet)	---	---
Right shoulder width on the major road (in feet)	---	---
Percentage of trucks on the major road	---	---

* One right turn lane on each major road direction for 3-legged unsignalized intersections: Two close unsignalized intersections, one on each side of the roadway, and each has one right turn lane. The extended right turn lane of the first is in the influence area of the second.

** One left turn lane on each major road direction for 3-legged unsignalized intersections: One of these left turn lanes is only used as U-turn.

*** N/A means not applicable

**** A continuous variable

Figures 5-2 and 5-3 show a distribution for the frequency of intersections by some variables (e.g., natural logarithm of AADT and right shoulder width, respectively) at both 3 and 4-legged intersections.

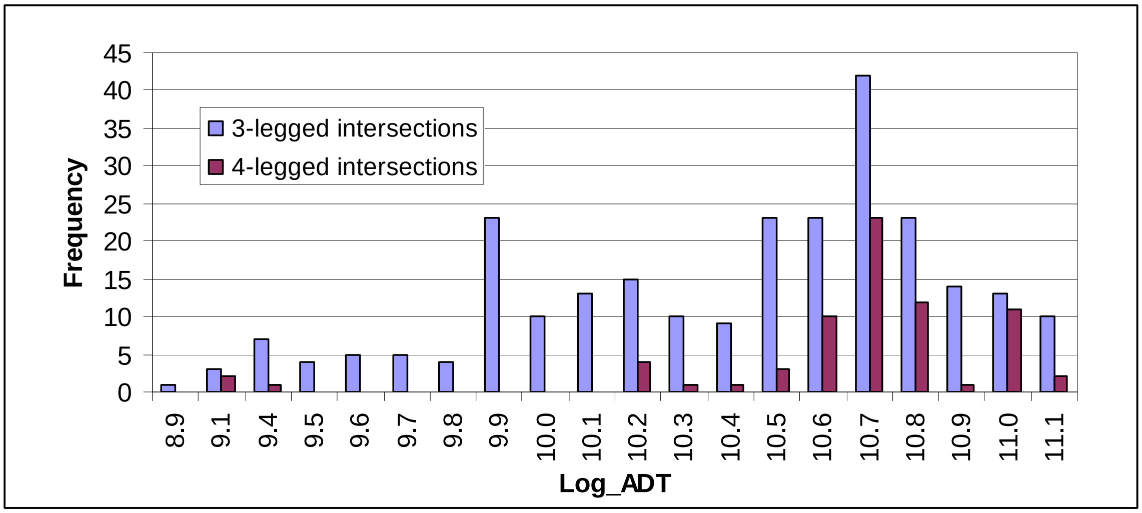


Figure 5-2: Distribution of Intersections by the Natural Logarithm of AADT at 3 and 4-Legged Intersections

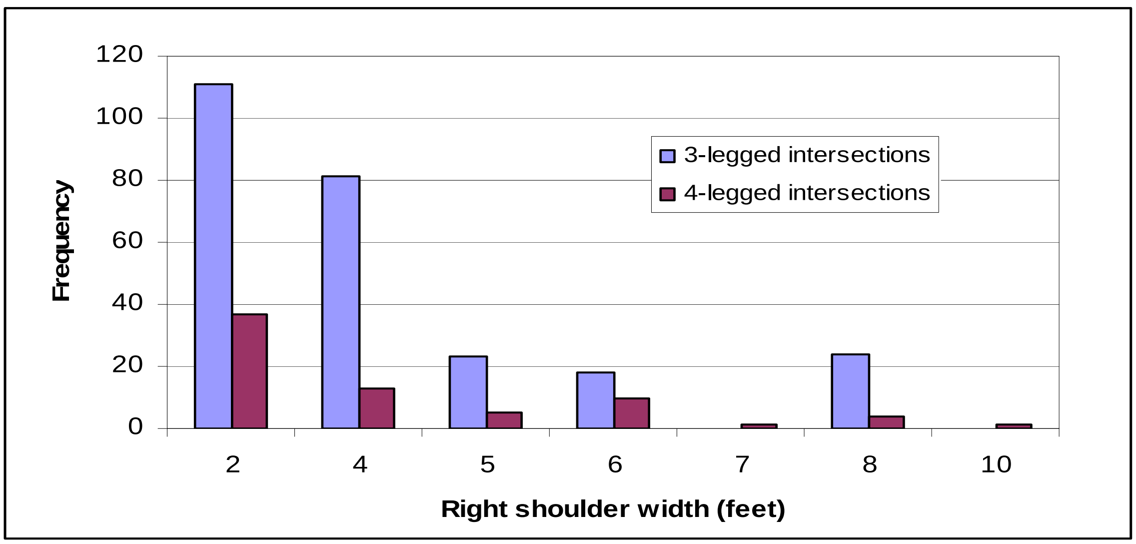


Figure 5-3: Distribution of Intersections by the Right Shoulder Width (in “feet”) at 3 and 4-Legged Intersections

5.4 Data Preparation

The four-year databases (from 2003 until 2006) were merged into one dataset. The 2 initial years (2003 and 2004) of data were used for modeling the frequency of crashes during this period, and for predicting the frequency of crashes in 2005 and 2006 (combined together). For the calibration dataset (2003-2004), there were 257 three-legged intersections with 497 total crashes, and 71 four-legged intersections with 176 total crashes. It was decided to use two separate models for 3-legged and 4-legged intersections as the attempt of having one model that includes both of them as a dummy variable did not show good results. Moreover, other studies (e.g. Jonsson et al., 2009) modeled total crash frequency and specific crash types at three and four-legged unsignalized intersections separately. For the scope of this chapter's analysis, the modeled unsignalized intersections include only intersections having a stop sign or no control. So, intersections having a yield sign as the control type were not used in the model, as they were very rare (mostly at ramps).

The use of the NB framework was very appropriate in this study, as it was found that crash frequency variance was greater than the mean (i.e., over-dispersion exists) for both 3 and 4-legged datasets. For the 3-legged calibration dataset, the crash frequency mean per intersection was 1.93, and the standard deviation was 2.35 (i.e., variance equals 5.52). As for the 4-legged dataset, the crash frequency mean was 2.48, and the standard deviation was 2.88 (i.e., variance equals 8.29).

5.5 <u>Fitted NB Regression Models (Safety Performance Functions)</u>

5.5.1 Modeling Crash Frequency at 3-Legged Intersections

After using SAS (2002), the NB crash frequency model for 3-legged unsignalized intersections is shown in Table 5-2. It is to be noted that the model parameters (Θ) were obtained using the maximum likelihood estimation.

From Table 5-2, it can be noticed that having a stop sign on the minor road increases the frequency of crashes significantly when compared to the case of having no stop sign (no control). While this result seems questionable, it can be explained that stop signs are possibly installed at hazardous intersections with relatively higher AADT (especially on the minor approach), but traffic volume on the minor approach was not used in this study due to data limitations. Therefore, intersections having a stop sign on the minor approach might be considered more hazardous than those with no stop sign, and thus crash frequency could be higher.

Table 5-2: NB Crash Frequency Model at 3-Legged Unsignalized Intersections

Parameter		Variable Description	Estimate	Standard Error	P-value
Intercept			-14.4469	2.3492	<.0001
stop_sign_mnr	1	A stop sign exists on the minor approach	0.5107	0.2297	0.0262
stop_sign_mnr	0	No stop sign exists on the minor approach	--- [a]		
major_RT	2	One right turn lane exists on each major road direction	-0.6699	0.2659	0.0118
major_RT	1	One right turn lane exists on only one major road direction	-0.3909	0.1555	0.0119
major_RT	0	No right turn lane exists	--- [a]		
major_LT	2	One left turn lane exists on each major road direction	0.0648	0.1872	0.7293
major_LT	1	One left turn lane exists on only one major road direction	-0.5351	0.1814	0.0032
major_LT	0	No left turn lane exists	--- [a]		
dir	2	One-way major road is related to the intersection	0.3182	0.2095	0.1288
dir	1	Two-way major road is related to the intersection	--- [a]		
log_AADT		Natural logarithm of the section annual average daily traffic on the major road	1.4084	0.2245	<.0001
Dispersion			0.3494	0.0809	
AIC[b]			898.29		
Rho-squared[c]			0.09		

[a] Used base case in SAS
[b] Akaike Information Criterion
[c] Pseudo R-squared or McFadden's Log-likelihood Ratio Index

Having 2 right turn lanes on both major road directions is much safer than having only 1 right turn lane, as shown in the higher negative coefficient for 2 right turn lanes than that for 1 right turn lane, and both decrease the frequency of crashes when compared to having no right turn lanes. Moreover, both coefficients are statistically significant.

In contrast to the previous finding, having 2 left turn lanes on both major road directions is more dangerous than having only 1 left turn lane, as shown in the positive coefficient for the 2 left turn lanes and the negative coefficient for the 1 left turn lane, when compared to having no left turn lanes. Mainly, this is because one of those 2 left turn lanes is used as a U-turn, which creates more traffic conflict. Moreover, the coefficient for the 1 left turn lane is statistically significant, but it is not significant for the 2 left turn lanes.

An odd finding is that when two-way major road is related to the unsignalized intersection, it is safer than when one-way major road is related to the intersection. It is worth mentioning that intersections with one-way major road always occur when closed medians exist on the major approach. The only explanation for this is that drivers coming from the minor approach are more attentive when both major approaches exist, while the opposite can happen when one of the major approaches do not exist. However, the coefficient is not statistically significant at the 90% confidence interval.

As expected, increasing the logarithm of AADT on the major road increases the frequency of crashes. Other studies (Wang and Abdel-Aty, 2006; Lord and Persaud, 2000; Anastasopoulos and Mannering, 2009; Chin and Quddus, 2003; Maher and Summersgill, 1996; Mountain et al., 1998) reached the same outcome while analyzing crashes at other locations. For example, Wang and Abdel-Aty (2006) has found that the logarithm of the AADT per lane increases the rear-end crash frequency at signalized intersections, and it was one of the most significant variables. The remaining aforementioned studies found that AADT (or AADT in thousands) increases crash frequency. Lord and Persaud (2000) analyzed crashes at 4-legged signalized intersections in Toronto, Canada from 1990 till 1995, and found AADT to be significant. Moreover, Anastasopoulos and Mannering (2009) used the random-parameters NB model, and found that AADT (in thousand vehicles) significantly increases crash frequency at rural interstate highways in Indiana.

Finally, for the dispersion parameter, it is noticed that the standard error is much less than the coefficient itself, indicating that the data are statistically over-dispersed (i.e. variance is much higher than the mean). Accordingly, choosing the NB model was appropriate for the data.

5.5.2 Modeling Crash Frequency at 4-Legged Intersections

After using SAS (2002), the NB crash frequency model for 4-legged unsignalized intersections is shown in Table 5-3.

From Table 5-3, the results show that having 2 stop signs on both minor approaches increases the frequency of crashes significantly when compared to the case of having only 1 stop sign on one of the minor approaches (similar result obtained and explained in Table 5-2).

Having 2 right turn lanes on both major road directions is more dangerous than having only 1 right turn lane, as shown in the positive coefficient for 2 right turn lanes and the negative coefficient for 1 right turn lane. This is attributed to the fact that in the case of 4-legged intersections, there is a possible conflict between through (from the minor approach) and right turn (on the major approach) maneuvers. Thus, as the number of right turn lanes increases, the probability of having a conflict increases, and this indeed increases the crash risk. Moreover, both coefficients are statistically significant, with having 2 right lanes more significant.

Having 2 through lanes on both minor approaches significantly decreases the frequency of crashes due to the fact that much care is given by drivers while crossing the major road.

Table 5-3: NB Crash Frequency Model at 4-Legged Unsignalized Intersections

Parameter		Variable Description	Estimate	Standard Error	P-value
Intercept			-13.4765	5.1569	0.009
stop_sign_mnr	2	One stop sign exists on each minor approach	0.5636	0.2606	0.0306
stop_sign_mnr	1	One stop sign exists on one of the minor approaches	--- [a]		
major_RT	2	One right turn lane exists on each major road direction	0.6775	0.3673	0.0651
major_RT	1	One right turn lane exists on only one major road direction	-0.6274	0.3725	0.0921
major_RT	0	No right turn lane exists	--- [a]		
minor_through	2	Two through movements exist on both minor approaches (one on each minor approach)	-1.0664	0.3851	0.0056
minor_through	1	One through movement exists on one minor approach only	--- [a]		
major_MDT	4	A two-way left turn lane median on the major road	0.4737	0.2412	0.0495
major_MDT	1	An open median on the major road	--- [a]		
log_AADT		Natural logarithm of the section annual average daily traffic on the major road	1.3501	0.4761	0.0046
SLDWIDTH_num		Right shoulder width on the major road (in feet)	0.0818	0.0503	0.104
ISLDWDTH_num		Left shoulder width near the median on the major road (in feet)	-0.1443	0.1023	0.1585
Dispersion			0.2889	0.1321	
AIC[b]			283.65		
Rho-squared[c]			0.11		

[a] Used base case in SAS
[b] Akaike Information Criterion
[c] Pseudo R-squared or McFadden's Log-likelihood Ratio Index

Having a two-way left turn lane as a median on the major road significantly increases the frequency of crashes, when compared to the case of having an open median. This indeed shows the dangerous effect of having two-way left turn lanes.

Increasing the logarithm of AADT on the major road increases the frequency of crashes. Moreover, the coefficient is statistically significant.

Increasing the shoulder width increases the frequency of crashes, as shown by the positive sign, and the coefficient is statistically significant at the 90% confidence. Increasing the shoulder width by 0.33 m (1 feet) increases crash frequency by e0.0818 (1.085 crashes).

As the left shoulder width near the median on the major approach increases, the frequency of crashes decreases. However, the coefficient is not statistically significant at the 90% confidence interval. Thus, increasing the left shoulder width near the median by 0.33 m (1 feet) decreases crash frequency by e0.1443 (1.155 crashes). This illustrates the importance of having a relatively large width beside the median, so that vehicles do not hit medians immediately.

5.6 Bayesian Updating Models for 3 and 4-Legged Unsignalized Intersections

Mathematica (Wolfram Mathematica 6) was used to perform the Bayesian updating procedure for both the 3 and 4 legs NB model. There is no built-in code for executing the Bayesian updating concept. Hence, this was done by writing a code for estimating the posterior estimates of the parameters using the method described previously. The main objective was to update the distribution of the parameters in the NB model for more accurate prediction of crashes in 2005 and 2006, to reduce the uncertainty of the associated predicted crashes across all the selected intersections and to generate a full probability distribution for the parameters' coefficients. Two types of priors were used while performing the Bayesian updating framework, the non-informative and the informative priors, to give clear insight of various types of priors on the prediction performance of crashes. This will indeed lead to a more concrete conclusion than only attempting a specific type.

In this study, two different likelihood functions for each of the non-informative and informative priors were used, the log-gamma and NB likelihood functions. For the informative prior using the NB likelihood function, engineering judgment was used to provide values for the different parameters to be used as a starting prior. For the informative prior using the log-gamma likelihood function, a second iteration for performing the full Bayesian updating was done using the posterior estimates from the non-informative prior with the log-gamma likelihood function. However, to avoid using data twice, additional intersection data from Seminole County were used. The updated estimates (based on the posterior mean) using those 4 Bayesian updating structures for both the 3 and 4-legged models are shown in Tables 5-4 and 5-5, respectively.

Table 5-4: Updated Parameter Estimates for the 3-Legged Model After Using the Bayesian Updating Framework (for 4 Different Structures)

			Non-informative Prior (Log-gamma*) (Structure 1)		Non-informative Prior (NB*) (Structure 2)		Informative Prior (NB*) (Structure 3)		Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma *) (Structure 4)	
Parameter		**Variable Description**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
Intercept			-7.6744	1.6073	-14.6336	2.3163	-10.449	0.8152	-9.1249	**1.2869**
stop_sign_mnr	1	A stop sign exists on the minor approach	0.1969	0.1687	0.526	0.2311	0.2814	0.1269	0.1626	**0.1471**
stop_sign_mnr	0	No stop sign exists on the minor approach	--- [a]		--- [a]		--- [a]		--- [a]	
major_RT	2	One right turn lane exists on each major road direction	0.0002	0.2462	-0.7055	0.2664	-0.8919	0.2155	0.0041	**0.1999**
major_RT	1	One right turn lane exists on only one major road direction	-0.2317	0.108	-0.4037	0.1499	-0.3641	0.1181	-0.2193	**0.1022**
major_RT	0	No right turn lane exists	--- [a]		--- [a]		--- [a]		--- [a]	
major_LT	2	One left turn lane exists on each major road direction	-0.004	0.1454	0.07	0.1955	-0.0029	0.1647	-0.0200	**0.1317**
major_LT	1	One left turn lane exists on only one major road direction	-0.4962	0.1315	-0.5375	0.1803	-0.6118	0.1497	-0.4863	**0.1177**
major_LT	0	No left turn lane exists	--- [a]		--- [a]		--- [a]		--- [a]	
dir	2	One-way major road is related to the intersection	0.1585	0.1532	0.3253	0.2129	-0.0371	0.0638	0.2373	**0.1311**
dir	1	Two-way major road is related to the intersection	--- [a]		--- [a]		--- [a]		--- [a]	

			Non-informative Prior (Log-gamma*) (Structure 1)		Non-informative Prior (NB*) (Structure 2)		Informative Prior (NB*) (Structure 3)		Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma *) (Structure 4)	
Parameter		**Variable Description**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
log_AADT		Natural logarithm of the section annual average daily traffic on the major road (for both directions)	0.8189	0.1538	1.4284	0.2216	1.0523	0.0727	0.9551	**0.1236**
Dispersion			0.3576	0.0361	0.3738	0.0838	0.3214	0.0523	0.3212	0.0269
AIC[b]			106.85		540.12		549.17		**48.69**	
DIC[c]			106.40		539.75		542.91		**35.69**	

[*] Used likelihood function
[a] Base case
[b] Akaike Information Criterion
[c] Deviance Information Criterion

Table 5-5: Updated Parameter Estimates for the 4-Legged Model After Using the Bayesian Updating Framework (for 4 Different Structures)

			Non-informative Prior (Log-gamma*) (Structure 1)		Non-informative Prior (NB*) (Structure 2)		Informative Prior (NB*) (Structure 3)		Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma *) (Structure 4)	
Parameter		**Variable Description**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
Intercept			-5.3639	2.6324	-14.0285	5.6094	-11.4059	3.3946	-9.8359	**2.0104**
stop_sign_mnr	2	One stop sign exists on each minor approach	0.5567	0.2225	0.5799	0.2748	0.5012	0.2143	0.3355	**0.1984**
stop_sign_mnr	1	One stop sign exists on one of the minor approaches	--- [a]	---	--- [a]		--- [a]		--- [a]	
major_RT	2	One right turn lane exists on each major road direction	0.4828	0.3193	0.6839	0.4379	1.083	0.2358	0.5388	**0.2755**
major_RT	1	One right turn lane exists on only one major road direction	-0.2075	0.3027	-0.6384	0.391	-1.028	0.2774	-0.286	**0.2473**
major_RT	0	No right turn lane exists	--- [a]	---	--- [a]		--- [a]		--- [a]	
minor_through	2	Two through movements exist on both minor approaches (one on each minor approach)	-1.2026	0.3475	-1.0578	0.4379	-0.602	0.2025	-1.1567	**0.2807**
minor_through	1	One through movement exists on one minor approach only	--- [a]	---	--- [a]		--- [a]		--- [a]	
major_MDT	4	A two-way left turn lane median on the major road	0.5101	0.2049	0.4874	0.2486	0.2217	0.1217	0.3732	**0.1464**
major_MDT	1	An open median on the major road	--- [a]	---	--- [a]		--- [a]		--- [a]	
log_AADT		Natural logarithm of the section annual average daily traffic on the major road (for both directions)	0.6270	0.2412	1.3995	0.5127	1.1693	0.3118	1.0465	**0.1838**

Parameter		Variable Description	Non-informative Prior (Log-gamma*) (Structure 1) Estimate	Standard Error	Non-informative Prior (NB*) (Structure 2) Estimate	Standard Error	Informative Prior (NB*) (Structure 3) Estimate	Standard Error	Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma *) (Structure 4) Estimate	Standard Error
SLDWIDTH_num		Right shoulder width on the major road (in feet)	0.0802	0.0402	0.0826	0.0567	-0.0029	0.0092	0.0936	**0.0325**
ISLDWDTH_num		Left shoulder width near the median on the major road (in feet)	-0.0657	0.0763	-0.1728	0.1145	-0.1618	0.072	-0.0912	**0.0631**
Dispersion			0.3521	0.066	0.4218	0.1779	0.3717	0.1582	0.2696	0.0345
AIC[b]			111.26		184.15		190.1		**42.96**	
DIC[c]			112.67		184.02		180.37		**30.52**	

[*] Used likelihood function
[a] Base case
[b] Akaike Information Criterion
[c] Deviance Information Criterion

5.6.1 Evaluating NB and Bayesian Models Using Mean Estimates of Parameters

When comparing Tables 5-2 and 5-4, and Tables 5-3 and 5-5, the coefficients' estimates using the non-informative and informative priors with NB as the likelihood function are close to those before updating. On the other hand, using the non-informative prior with log-gamma as the likelihood function, as well as the second Bayesian iteration structure led to different coefficients, due to using a different likelihood function (log-gamma) other than the NB link function initially utilized to fit the parameters before updating. Moreover, from Tables 5-4 and 5-5, it is noticed that structure 4 led to the least standard errors (bolded values) for all the parameters, compared to the other three structures.

It is noted that the Bayesian model with structure 4 (with log-gamma as the likelihood function) is the best Bayesian-structure model, since it has the lowest AIC value. The AIC value for structure 4 after applying the Bayesian updating framework equals 48.69 for the 3-legged model, and 42.96 for the 4-legged model. The other three MOE values (MAD, MSPE and the overall prediction accuracy) for the 5 models (before and after updating) for both 3 and 4-legged models are shown in Tables 5-6 and 5-7, respectively.

Table 5-6: MOE Values for the Five 3-Legged Models (Before and After Bayesian Updating)

	Before Bayesian Updating	After Bayesian Updating			
MOE	**NB Model Before Updating**	**Non-informative Prior (Log-gamma) (Structure 1)**	**Non-informative Prior (NB) (Structure 2)**	**Informative Prior (NB) (Structure 3)**	**Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma) (Structure 4)**
MAD	**1.39**	**1.41**	**1.38**	**1.35**	**1.38**
MSPE	**3.50**	**3.03**	**3.33**	**3.41**	**2.99**
Overall Prediction Accuracy	**0.78**	**0.95**	**0.85**	**0.74**	**0.97**

Table 5-7: MOE Values for the Five 4-Legged Models (Before and After Bayesian Updating)

	Before Bayesian Updating	After Bayesian Updating			
MOE	**NB Model Before Updating**	**Non-informative Prior (Log-gamma) (Structure 1)**	**Non-informative Prior (NB) (Structure 2)**	**Informative Prior (NB) (Structure 3)**	**Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma) (Structure 4)**
MAD	**1.79**	**1.79**	**1.80**	**1.80**	**1.71**
MSPE	**5.55**	**4.98**	**5.52**	**6.33**	**4.98**
Overall Prediction Accuracy	**0.68**	**0.92**	**0.71**	**0.81**	**0.84**

From Table 5-6, the best overall model for prediction is the Bayesian model with structure 4, as it has the second least MAD, least MSPE and the highest overall prediction accuracy. Also, the second best model in prediction accuracy is the Bayesian model with structure 1. From Table 5-7, the two best models are structures 1 and 4 (for the log-gamma likelihood function). It can be noted that structure 1 has the highest overall prediction accuracy (0.92), followed by structure 4 (0.84); however, structure 4 was deemed the best Bayesian-structure model, as it has a lower MAD value and there is little difference between both prediction accuracies. This indeed demonstrates the importance of using the log-gamma likelihood function as a valid distribution for updating the parameters. Moreover, these results show the significant effect of applying the Bayesian updating approach to increase the prediction accuracy, and reduce the AIC, MAD and MSPE.

The plot of the residuals (the difference between the actual and predicted crash frequencies at each intersection) against one of the key covariates (Log_AADT on the major road) for both 3 and 4-legged intersections is shown in Figures 3 and 4, respectively. This plot was obtained by arranging the residuals in an increasing order for the "Log_AADT" covariate. The indication that the model has a good fit for the key covariate happens when the residuals oscillate around the value of zero, and the residuals are not widely spread. From these two plots, it is noticed that the structures 1 and 4 (structure 4, shown in thicker line weight) have the least spread among all other structures. For example, for 3-legged intersections, the residuals for structures 1 and 4 range from around -3 till 4.8. These results show the significant effect of applying the Bayesian updating approach to reduce the spread of the residuals, with the log-gamma likelihood function being the best Bayesian updating structure.

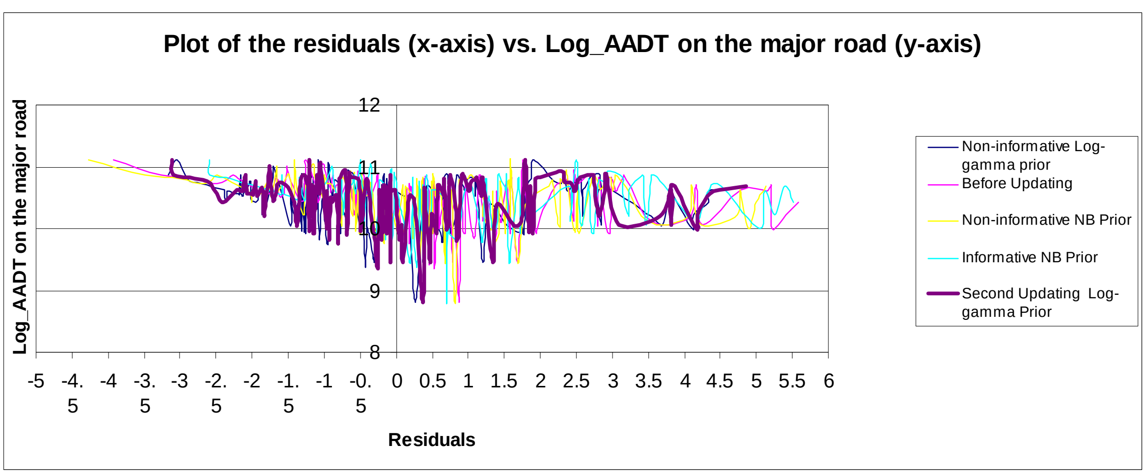


Figure 5-4: Plot of the Residuals vs. Log_AADT on the Major Road at 3-Legged Unsignalized Intersections

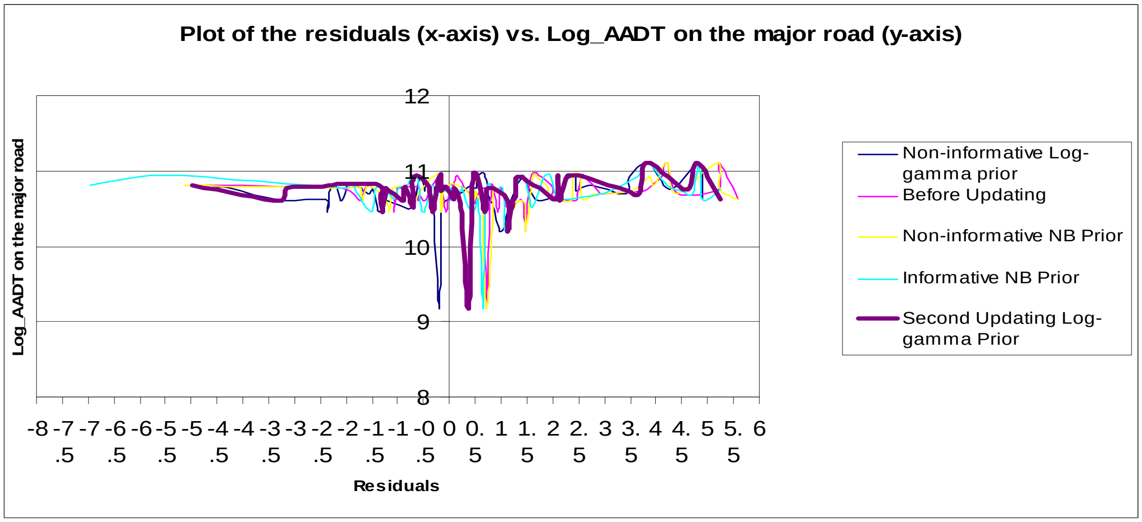


Figure 5-5: Plot of the Residuals vs. Log_AADT on the Major Road at 4-Legged Unsignalized Intersections

5.6.2 Evaluating NB and Bayesian Models Using Uncertainty Estimates

To assess one of the main objectives in this chapter, which is reducing uncertainty from those probabilistic models, a comparison between the standard errors (a surrogate measure for uncertainty) for those fitted parameters from the fitted NB model before updating and those for the best Bayesian-structure model (structure 4) after updating the parameters for both 3 and 4-legged models is shown in Tables 5-8 and 5-9, respectively.

Table 5-8: Assessing Uncertainty Reduction After Applying the Bayesian Updating Framework for Structure 4 by Comparing Standard Errors Before and After Updating for the 3-Legged Model

Parameter		Variable Description	Standard Error (Before Updating)	Standard Error for Structure 4 (After Updating)	% Change*
Intercept			**2.3492**	**1.2869**	**-45.22**
stop_sign_mnr	1	A stop sign exists on the minor approach	**0.2297**	**0.1471**	**-35.96**
major_RT	2	One right turn lane exists on each major road direction	**0.2659**	**0.1999**	**-24.82**
major_RT	1	One right turn lane exists on only one major road direction	**0.1555**	**0.1022**	**-34.28**
major_LT	2	One left turn lane exists on each major road direction	**0.1872**	**0.1317**	**-29.65**
major_LT	1	One left turn lane exists on only one major road direction	**0.1814**	**0.1177**	**-35.12**
dir	2	One-way major road is related to the intersection	**0.2095**	**0.1311**	**-37.42**
log_AADT		Natural logarithm of the section annual average daily traffic on the major road	**0.2245**	**0.1236**	**-44.94**

*Negative sign indicates an uncertainty reduction after applying the Bayesian updating framework

Table 5-9: Assessing Uncertainty Reduction After Applying the Bayesian Updating Framework for Structure 4 by Comparing Standard Errors Before and After Updating for the 4-Legged Model

Parameter		Variable Description	Standard Error (Before Updating)	Standard Error for Structure 4 (After Updating)	% Change*
Intercept			**5.1569**	**2.0104**	**-61.02**
stop_sign_mnr	2	One stop sign exists on each minor approach	**0.2606**	**0.1984**	**-23.87**
major_RT	2	One right turn lane exists on each major road direction	**0.3673**	**0.2755**	**-24.99**
major_RT	1	One right turn lane exists on only one major road direction	**0.3725**	**0.2473**	**-33.61**
minor_through	2	Two through lanes exist on both minor approaches (one on each minor approach)	**0.3851**	**0.2807**	**-27.11**
major_MDT	4	A two-way left turn lane median on the major road	**0.2412**	**0.1464**	**-39.30**
log_AADT		Natural logarithm of the section annual average daily traffic on the major road	**0.4761**	**0.1838**	**-61.39**
SLDWIDTH_num		Right shoulder width on the major road (in feet)	**0.0503**	**0.0325**	**-35.39**
ISLDWDTH_num		Left shoulder width near the median on the major road (in feet)	**0.1023**	**0.0631**	**-38.32**

*Negative sign indicates an uncertainty reduction after applying the Bayesian updating framework

From Tables 5-8 and 5-9, it is noticed that there is an uncertainty reduction after updating those parameters for both the 3 and 4-legged models. There is always a standard error reduction for all the fitted parameters in both models. The highest uncertainty reduction (highlighted) for the parameters (other than the intercept) in the 3-legged model is 44.94%, whereas for the 4-legged model is 61.39%. Thus, in conjunction with previous findings in this study, the importance of using the log-gamma likelihood function as a valid distribution for updating the parameters is assessed.

5.7 General Conclusions from the Reliability Process in Terms of the Bayesian Updating Framework

The analysis performed in this chapter used a coordinated application of the NB model, as well as a reliability method (in terms of the full Bayesian updating framework) for reducing uncertainty in predicting crash frequency at 3-legged and 4-legged unsignalized intersections. A broad exploration of both non-informative and informative priors was conducted using both the NB and the log-gamma likelihood functions. Moreover, a second Bayesian updating iteration was explicitly investigated in terms of the informative prior with the log-gamma likelihood function.

The fitted NB regression models (before updating) showed several important variables that affect the safety of unsignalized intersections. These include the traffic volume on the major road and the existence of stop signs, and among the geometric characteristics, the configuration of the intersection, number of right and/or left turn lanes, median type on the major road, and left and right shoulder widths.

It was concluded that all the four Bayesian-structure models (after updating) perform much better than before updating (NB model). Measuring uncertainty was done using a surrogate measure of the parameters' standard error. The second Bayesian updating using the log-gamma likelihood function (structure 4) was deemed the best structure by having the least standard error values. The highest uncertainty reduction for structure 4 for the 3-legged model is around 45%, and that for the 4-legged model is around 61%.

Other assessment criteria (such as the AIC, MAD, MSPE and overall prediction accuracy) demonstrated the significant effect of structure 4 as being the best Bayesian-structure model. Structure 4 succeeded in producing a 97% prediction accuracy for the 3-legged model, and an 84% prediction accuracy for the 4-legged model. The plot of the "residuals" against "Log_AADT on the major road" showed that structure 4 has the least spread when compared to the other three structures, and before updating as well.

Thus, the findings from this chapter point to that the log-gamma likelihood function is strongly recommended as a robust distribution for updating the parameters of the NB probabilistic models. Also, results from this study show that the full Bayesian updating framework for updating parameter estimates of probabilistic models is promising. However, the use of the estimates from the NB regression models (without updating) led to favorable results, where the prediction accuracy was 78% for the 3-legged model, and 68% for the 4-legged model. Thus, traffic safety researchers and professionals are recommended to use parameter estimates from the NB regression model for prediction purposes, but the prediction accuracy will not be as high as after updating those estimates using the full Bayesian method with the log-gamma likelihood function.

CHAPTER 6. CRASH INJURY SEVERITY ANALYSIS

6.1 Background

According to the Florida Department of Transportation (2006), almost one in every four fatal crashes occurs at or near an intersection. In 2004, Florida led the nation in intersection fatalities, where 30% of fatalities occurred at intersections. Identifying those geometric and traffic factors leading to severe crashes at unsignalized intersections is an essential task of traffic safety analysts. This helps identify appropriate countermeasures for any observed safety deficiency. Crash injury severity is considered the most serious crash outcome, which is the core of this paper.

The analysis conducted in this chapter focuses on analyzing crash injury severity with respect to its inherently ordered nature, not its frequency. The most common statistical frameworks for analyzing crash severity are multinomial logit, ordered probit and nested logit models (Abdel-Aty, 2003; Savolainen and Mannering, 2007 and Chang and Mannering, 1999). The use of the ordered probit model formulation in this study was deemed more beneficial than multinomial logit and probit models for while accounting for the categorical nature of the dependent variable, they do not account for the ordinal nature of the modeled response categories (Duncan et al., 1998), which can be a serious issue.

In this chapter, crash injury severity is analyzed at 1547 three-legged and 496 four-legged unsignalized intersections (i.e., a total of 2043 intersections, including stop-controlled, yield-controlled and non-controlled intersections) in the state of Florida in 4 years (from 2003 till 2006) using the ordered probit, binary probit and nested logit methodologies. Florida's six counties (Orange, Hillsborough, Brevard, Seminole, Leon and Miami-Dade) were used in the analysis.

Thus, the main objective of the analysis in this chapter is to identify the significant factors contributing to injury severity at unsignalized intersections. This helps identify those geometric, traffic and driver-related factors leading to severe crashes at those intersections. A comparison between the three formulations is attempted to select the best modeling scheme for analyzing crash injury severity at unsignalized intersections. Finally, some countermeasures are recommended as a remedy for alleviating some safety problems identified.

6.2 Methodological Approach: Probit Model Specification

Bliss (1935) introduced Probit models. Ordered probit models are types of probit models, which assume standard normal distribution for the parameters. Similar to many models for qualitative dependent variables, the ordered probit model is originated from bio-statistics (Aitchison and Silvey, 1957). It was brought into the social sciences by the two political scientists, McKelvey and Zavoina (1975).

The modeled response variable (crash injury severity) is inherently ordered with five main categories, no injury or property damage only (PDO), possible injury, non-incapacitating injury, incapacitating injury and fatal (within 30 days). Thus, the response variable y takes the following ordered values:

$$y = \begin{cases} 1; \textit{if the accident injury severity level is a PDO} \\ 2; \textit{if the accident injury severity level is a possible injury} \\ 3; \textit{if the accident injury severity level is a non-incapacitating injury} \\ 4; \textit{if the accident injury severity level is an incapacitating injury} \\ 5; \textit{if the accident injury severity level is a fatal injury (within 30 days after the accident)} \end{cases}$$

For the aggregated binary probit model, incapacitating injury and fatal injury were combined to represent severe injuries, whereas non-severe crash level included PDO, possible injury and non-incapacitating injury. The reason for this aggregation is to increase the number of observations to reduce the variability caused by random effects (Chang and Mannering, 1999). This is essential since the data used in this study had too few observations on incapacitating and fatal injuries to set apart their individual effects. Thus, the response variable y takes the following binary values:

$$y = \begin{cases} 0; \textit{if the accident is non}-\textit{severe} \\ 1; \textit{if the accident is severe} \end{cases}$$

The ordered probit models have come into fairly wide use as a framework for analyzing such response variables. The ordered choice model assumes the relationship to be as shown in Equations (6.1) and (6.2).

$$\sum_{j=1}^{j} P_n(j) = F(\alpha_j - \beta X_n, \theta), \quad j = 1, 2, ..., J-1 \tag{6.1}$$

$$P_n(J) = 1 - \sum_{j=1}^{J-1} P_n(j) \tag{6.2}$$

where: $P_n(j)$ is the probability that subject (intersection) n (n = 1, 2, ---, N) belongs to category j (with N = total number of intersections);

J is the total number of categories;

α_j is a specific parameter (to be estimated with β);

X_n is a vector of measurable characteristics specific to subjects (intersections);

β is a vector of estimated coefficients; and

θ is a shape parameter parameter that controls the cumulative probability distribution F. (For an ordered probit model, the assumed cumulative probability distribution F is the cumulative standard normal distribution Φ).

The marginal effects (also called elasticities, as shown by Chang and Mannering, 1999) are equivalent to the partial derivative of the expectation of the targeted response variable with respect to the vector of covariates (*X*). Assuming that the used model is:

$$Y = X^T \beta + \varepsilon \tag{6.3}$$

Thus, the expectation of the target response variable Y is F($X^T \beta$), i.e. E(Y) = F($X^T \beta$). For ordered probit models, E(Y) = Φ($X^T \beta$), and the marginal effects can be estimated as shown in Equation (6.4).

$$\frac{\partial E(Y)}{\partial X} = \left\{ \frac{d\,F(X^T \beta)}{d\,(X^T \beta)} \right\} \beta = \mathrm{f}(X^T \beta) * \beta = \phi(X^T \beta) * \beta \tag{6.4}$$

It is worth mentioning that the nested logit framework was also examined in this chapter, and the nested logit model formulation can be found in previous literature (e.g., Chang and Mannering, 1999; Ben-Akiva and Lerman, 1985; Abdel-Aty and Abdelwahab, 2004; McFadden, 1978 and McFadden, 1981).

6.3 Data Preparation

The analysis done in this chapter was performed on 2043 unsignalized intersections collected from six counties in the state of Florida. The CAR database maintained by the FDOT was used to identify all SRs in those 6 counties. Then, a random selection method was used for choosing some state roads. Unsignalized intersections were then identified along these randomly selected SRs using “Google Earth” and “Video Log Viewer Application”. In order to use the “Video Log Viewer Application”, the roadway ID for the used SR, the mile point and the direction of travel should be specified. This application is an advanced tool developed by FDOT, and has the advantage of capturing the driving environment through the roadway. Moreover, this advanced application has two important features allowing different video perspectives, the “right view” and the “front view”. The “right view” feature provides the opportunity of identifying whether a stop sign and a stop line exist or not. The “front view” feature provides the opportunity of identifying the median type as well as the number of lanes per direction more clearly.

Afterwards, all the geometric and control fields of the collected intersections were identified and added to the database. These collected fields were then merged with the RCI database to capture those important traffic (such as annual average daily traffic and percentage of trucks) and roadway (such as right shoulder width, left shoulder width and median width) features. The RCI database – which is developed by the FDOT - includes physical and administrative data, such as functional classification, pavement, shoulder and median data related to the roadway (the New Web-based RCI Application). Each of these facilities is indexed by a roadway ID number with beginning and ending mile points. The criteria used for merging the two databases (intersections and RCI) were the roadway ID and the mile point. The merging procedure was done using SAS (2002).

Crash data for the 4 years used in the analysis were collected from the CAR database. In order to capture the most important crash variables (e.g., crash injury severity), the 2043 intersections from the 6 counties were merged with crash data from 2003 till 2006. The criteria used for merging purposes were the roadway ID, mile point, and intersection node number. The final merged dataset has 10722 observations, with 6808 observations (63.5%) representing 3-legged unsignalized intersections, and 3914 observations (36.5%) representing 4-legged unsignalized intersections.

For the 3-legged dataset, the percentages of the five injury levels were as follows, 42.14% PDO, 27.7% possible injury, 21.71% non-incapacitating injury, 7.48% incapacitating injury, and 0.97% fatal. This means that there are 91.55% non-severe injuries, and 8.45% severe injuries.

For the 4-legged dataset, the percentages of the five injury levels were as follows, 47.34% PDO, 25.52% possible injury, 19.09% non-incapacitating injury, 7.15% incapacitating injury, and 0.89% fatal. In other words, there are 91.96% non-severe injuries, and 8.04% severe injuries.

6.4 Variables' Description

It was decided to use two separate models for 3-legged and 4-legged intersections as both intersection types have different operating characteristics. A full description of the important variables used in the ordered and binary probit, and nested logit modeling procedure for 3 and 4-legged unsignalized intersections is shown in Table 6-1.

This analysis conducted in this chapter is considered comprehensive since it explores new important roadway and traffic covariates that were not examined before. Examples of those new roadway covariates are the existence of crosswalks on the minor and major approaches, effect of various minor approach control types (e.g., stop sign, no control and yield sign), various sizes of intersections, intersection type (whether it is a regular unsignalized intersection, access point or ramp junction), various median types on the major approach (open, closed, two-way left turn lane, etc.), distance between unsignalized intersections and signalized ones (from both the upstream and downstream aspects), distance between successive unsignalized intersections, and left (or median) shoulder width.

Regular unsignalized intersections are those intersections having longer segments (distant stretches) on the minor approaches, whereas access points include parking lots at plazas and malls, and driveways that are feeding to the major approach. An important traffic covariate explored is the surrogate measure for AADT on the minor approach, which is represented by the number of through lanes on this approach. The AADT on the minor approaches was not available for most of the cases, since they are mostly non-state roads. Another traffic covariate explored is the percentage of trucks in the fleet.

Table 6-1: Variables Description for 3 and 4-Legged Unsignalized Intersections [*]

Variable Description	Variable Levels for 3 Legs	Variable Levels for 4 Legs
Crash location in any of the 6 counties	Orange, Brevard, Hillsborough, Miami-Dade, Leon and Seminole	Orange, Brevard, Hillsborough, Miami-Dade, Leon and Seminole
Existence of stop sign on the minor approach	= 0; if no stop sign exists; = 1; if stop sign exists	= 0; if no stop sign exists; = 1; if only one stop sign exists on one of the minor approaches; = 2; if one stop sign exists on each minor approach
Existence of stop line on the minor approach	= 0; if no stop line exists; = 1; if stop line exists	= 0; if no stop line exists; = 1; if only one stop line exists on one of the minor approaches; = 2; if one stop line exists on each minor approach
Existence of crosswalk on the minor approach	= 0; if no crosswalk exists; = 1; if crosswalk exists	= 0; if no crosswalk exists; = 1; if only one crosswalk exists on one of the minor approaches; = 2; if one crosswalk exists on each minor approach
Existence of crosswalk on the major approach	= 0; if no crosswalk exists; = 1; if one crosswalk exists on one of the major approaches; = 2; if one crosswalk exists on each major approach	= 0; if no crosswalk exists; = 1; if one crosswalk exists on one of the major approaches; = 2; if one crosswalk exists on each major approach
Control type on the minor approach	= 1; if stop sign exists (1-way stop); = 3; if no control exists; = 5; if yield sign exists	= 2; if stop sign exists on each minor approach (2-way stop); = 3; if no control exists on both minor approaches; = 4; if stop sign exists on the first minor approach, and no control on the other
Size of the intersection [a]	= 1; for "1x2", "1x3" and "1x4" intersections; = 2; for "2x2" and "2x3" intersections; = 3; for "2x4", "2x5" and "2x6" intersections; = 4; for "2x7" and "2x8" intersections; = 5; for "3x2", "3x3", "3x4", "3x5", "3x6" and "3x8" intersections; = 6; for "4x2", "4x4", "4x6" and "4x8" intersections	= 2; for "2x2" and "2x3" intersections; = 3; for "2x4", "2x5" and "2x6" intersections; = 4; for "2x7" and "2x8" intersections; = 5; for "3x2", "3x3", "3x4", "3x5", "3x6" and "3x8" intersections; = 6; for "4x2", "4x4", "4x6" and "4x8" intersections
Type of unsignalized intersection [b]	= 1; for access point (driveway) intersections; = 2; for ramp junctions; = 3; for regular intersections; = 4; for intersections close to railroad crossings	= 1; for access point (driveway) intersections; = 3; for regular intersections; = 4; for intersections close to railroad crossings
Number of right turn lanes on the major approach	= 0; if no right turn lane exists; = 1; if one right turn lane exists on only one direction; = 2; if one right turn lane exists on each direction [c]	= 0; if no right turn lane exists; = 1; if one right turn lane exists on only one direction; = 2; if one right turn lane exists on each direction
Number of left turn lanes on the major approach	= 0; if no left turn lane exists; = 1; if one left turn lane exists on only one direction; = 2; if one left turn lane exists on each direction [d]	= 0; if no left turn lane exists; = 1; if one left turn lane exists on only one direction; = 2; if one left turn lane exists on each direction
Number of left turn movements on the minor approach	= 0; if no left turn movement exists; = 1; if one left turn movement exists	= 0; if no left turn movement exists; = 1; if one left turn movement exists on one minor approach only; = 2; if one left turn movement exists on each minor approach

Variable Description	Variable Levels for 3 Legs	Variable Levels for 4 Legs
Land use at the intersection area	= 1; for rural area; = 2; for urban/suburban areas	= 1; for rural area; = 2; for urban/suburban areas
Median type on the major approach	= 1; for open median; = 2; for directional median; = 3; for closed median; = 4; for two-way left turn lane; = 5; for markings; = 6; for undivided median; = 7; for mixed median [e]	= 1; for open median; = 4; for two-way left turn lane; = 5; for markings; = 6; for undivided median
Median type on the minor approach	= 1; for undivided median, two-way left turn lane and markings; = 2; for any type of divided median	= 1; for undivided median, two-way left turn lane and markings; = 2; for any type of divided median
Skewness level	= 1; if skewness angle <= 75 degrees; = 2; if skewness angle > 75 degrees	= 1; if skewness angle <= 75 degrees; = 2; if skewness angle > 75 degrees
Lighting condition	= 1; for daylight; = 2; for dusk; = 3; for dawn; = 4; for dark (street light); = 5; for dark (no street light)	= 1; for daylight; = 2; for dusk; = 3; for dawn; = 4; for dark (street light); = 5; for dark (no street light)
Road surface type	= 1; if gravel or brick/block; = 2; if concrete; = 3; if blacktop	= 1; if gravel or brick/block; = 2; if concrete; = 3; if blacktop
Road surface condition	= 1; if dry; = 2; if wet; = 3; if slippery	= 1; if dry; = 2; if wet; = 3; if slippery
Posted speed limit on the major road	= 1; if posted speed limit < 45 mph; = 2; if posted speed limit >= 45 mph	= 1; if posted speed limit < 45 mph; = 2; if posted speed limit >= 45 mph
Number of through lanes on the minor approach [f]	= 1; if one through lane exists; = 2; if two through lanes exist; = 3; if three through lanes exist; = 4; if four through lanes exist	= 2; if two through lanes exist; = 3; if more than two through lanes exist
At-fault driver's age category	= 1; if 15 <= age <= 19 (very young) = 2; if 20 <= age <= 24 (young) = 3; if 25 <= age <= 64 (middle) = 4; if 65 <= age <= 79 (old) = 5; if age >= 80 (very old)	= 1; if 15 <= age <= 19 (very young) = 2; if 20 <= age <= 24 (young) = 3; if 25 <= age <= 64 (middle) = 4; if 65 <= age <= 79 (old) = 5; if age >= 80 (very old)

[a] The first number represents total number of approach lanes for the minor approach, and the second number represents total number of through lanes for the major approach

[b] Regular unsignalized intersections are those intersections having distant stretches on the minor approaches; whereas access points include parking lots at plazas and malls as well as driveways that are feeding to the major approach; and railroad crossing can exist upstream or downstream the intersection of interest

[c] One right turn lane on each major road direction for 3-legged unsignalized intersections: Two close unsignalized intersections, one on each side of the roadway, and each has one right turn lane. The extended right turn lane of the first is in the influence area of the second.

[d] One left turn lane on each major road direction for 3-legged unsignalized intersections: One of these left turn lanes is only used as U-turn.

[e] Mixed median is directional from one side, and closed from the other side (i.e., allows access from one side only)

[f] Surrogate measure for AADT on the minor approach

[*] The continuous variables are the natural logarithm of AADT on the major road, the natural logarithm of the upstream and downstream distances to the nearest signalized intersection, the left shoulder width near the median on the major road, the right shoulder width on the major road, percentage of trucks on the major road, and the natural logarithm of the distance between 2 successive unsignalized intersections

6.5 Analysis of the Ordered Probit Framework

The fitted ordered probit model for both 3 and 4-legged unsignalized intersections using the five crash injury levels of the response variable is shown in Table 6-2, which includes some goodness-of-fit statistics as well, such as log-likelihood at convergence, log-likelihood at zero and AIC. The marginal effects for the estimated models for both 3 and 4-legged intersections are shown in Table 6-3.

The marginal effects depict the effect of change in a certain explanatory variable on the probability of an injury severity level. Since, the main concern is on fatal injuries (as they are the most serious), the interpretation will be focused on them. Also, the interpretations for both the three and four-legged models are discussed separately.

Table 6-2: Ordered Probit Estimates for 3 and 4-Legged Unsignalized Intersections

	Three-Legged Model		Four-Legged Model	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
Intercept 1	-1.6936 (0.5295)	0.0014	-0.1144 (0.6773)	0.8659
Intercept 2	0.9914 (0.0451)	<0.0001	1.0151 (0.0629)	<0.0001
Intercept 3	1.8849 (0.0476)	<0.0001	1.8539 (0.0659)	<0.0001
Intercept 4	2.6427 (0.0486)	<0.0001	2.5772 (0.0672)	<0.0001
Natural logarithm of AADT on the major road	-0.0807 (0.0332)	0.0151	-0.2447 (0.0518)	<0.0001
Natural logarithm of the upstream distance to the nearest signalized intersection	0.0442 (0.0153)	0.0039	0.0457 (0.0255)	0.0731
Natural logarithm of the downstream distance to the nearest signalized intersection	N/S [b]		0.0383 (0.0250)	0.1262
Posted speed limit on major road < 45 mph	-0.1096 (0.0337)	0.0011	-0.0818 (0.0496)	0.0994
Posted speed limit on major road >= 45 mph	--- [c]		--- [c]	
Skewness angle <= 75 degrees	N/S		0.1563 (0.0826)	0.0586
Skewness angle > 75 degrees	N/S		--- [c]	
No right turn lane exists on the major approach	-0.1725 (0.0935)	0.0654	N/S	
One right turn lane exists on only 1 major road direction	-0.1710 (0.0968)	0.0776	N/S	
One right turn lane exists on each major road direction	--- [c]		N/S	
No left turn movement exists on the minor approach	-0.0536 (0.0350)	0.1258	N/S	
One left turn movement exists on the minor approach	--- [c]		N/S	
One through lane exists on the minor approach	0.7919 (0.3917)	0.0432	N/A [d]	
Two through lanes exist on the minor approach	0.5098 (0.2827)	0.0713	N/S	
Three through lanes exist on the minor approach	0.5658 (0.3264)	0.0831	N/S	
Four through lanes exist on the minor approach	--- [c]		N/A	
15 <= At-fault driver's age <= 19 (very young)	-0.1391 (0.0954)	0.1448	N/S	
20 <= At-fault driver's age <= 24 (young)	-0.1705 (0.0946)	0.0716	N/S	
25 <= At-fault driver's age <= 64 (middle)	-0.1646 (0.0900)	0.0674	N/S	
65 <= At-fault driver's age <= 79 (old)	-0.0473 (0.1016)	0.6414	N/S	

	Three-Legged Model		Four-Legged Model	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
At-fault driver's age >= 80 (very old)	--- [c]		N/S	
Left shoulder width near the median on the major road	0.0323 (0.0126)	0.0105	0.0807 (0.0194)	<0.0001
Right shoulder width on the major road	N/S		-0.0189 (0.0076)	0.0130
Daylight lighting condition	-0.2718 (0.0615)	<0.0001	N/S	
Dusk lighting condition	-0.3030 (0.0999)	0.0024	N/S	
Dawn lighting condition	-0.3372 (0.1477)	0.0225	N/S	
Dark (street light) lighting condition	-0.1428 (0.0678)	0.0353	N/S	
Dark (no street light) lighting condition	--- [c]		N/S	
"1x2", "1x3" and "1x4" intersections	-0.4077 (0.3135)	0.1935	N/A	
"2x2" and "2x3" intersections	-0.2897 (0.1329)	0.0293	N/S	
"2x4", "2x5" and "2x6" intersections	-0.1482 (0.1281)	0.2474	N/S	
"2x7" and "2x8" intersections	-0.0383 (0.1532)	0.8024	N/S	
"3x2", "3x3", "3x4", "3x5", "3x6" and "3x8" intersections	-0.1384 (0.1367)	0.3113	N/S	
"4x2", "4x4", "4x6" and "4x8" intersections	--- [c]		N/S	
Dummy variable for Brevard County	-0.0378 (0.0796)	0.6346	0.2636 (0.0983)	0.0074
Dummy variable for Hillsborough County	-0.4935 (0.0664)	<0.0001	-0.2668 (0.0757)	0.0004
Dummy variable for Leon County	-0.5359 (0.0678)	<0.0001	-0.1392 (0.0884)	0.1153
Dummy variable for Miami-Dade County	-0.6560 (0.0659)	<0.0001	-0.4452 (0.0805)	<0.0001
Dummy variable for Orange County	-0.0060 (0.0663)	0.9277	0.3314 (0.0852)	0.0001
Dummy variable for Seminole County	--- [c]		--- [c]	
Log-likelihood at convergence	-8514		-4696	
Log-likelihood at zero [e]	-8783.5		-4890.6	
AIC	17091		9423	

[a] Standard error in parentheses [b] N/S means not significant [c] Base case [d] N/A means not applicable [e] Likelihood while fitting the intercept only

Table 6-3: Marginal Effects for Fatal Injury Probability for the Fitted Covariates in the 3 and 4-Legged Models

	Three-Legged Model	Four-Legged Model
Variable Description	Probability of fatal injury	Probability of fatal injury
Natural logarithm of AADT on the major road	-0.002	-0.006
Natural logarithm of the upstream distance to the nearest signalized intersection from the unsignalized intersection of interest	0.001	0.001
Natural logarithm of the downstream distance to the nearest signalized intersection from the unsignalized intersection of interest	N/S [a]	0.001
Posted speed limit on major road < 45 mph	-0.003	-0.002
Skewness angle <= 75 degrees	N/S	0.004
No right turn lane exists on the major approach	-0.004	N/S
One right turn lane exists on only 1 major road direction	-0.004	N/S
No left turn movement exists on the minor approach	-0.001	N/S
One through lane exists on the minor approach	0.021	N/A [b]
Two through lanes exist on the minor approach	0.013	N/S
Three through lanes exist on the minor approach	0.015	N/S
15 <= At-fault driver's age <= 19 (very young)	-0.004	N/S
20 <= At-fault driver's age <= 24 (young)	-0.004	N/S
25 <= At-fault driver's age <= 64 (middle)	-0.004	N/S
65 <= At-fault driver's age <= 79 (old)	-0.001	N/S
Left shoulder width near the median on the major road	0.001	0.002
Right shoulder width on the major road	N/S	0.000
Daylight lighting condition	-0.007	N/S
Dusk lighting condition	-0.008	N/S
Dawn lighting condition	-0.009	N/S
Dark (street light) lighting condition	-0.004	N/S

	Three-Legged Model	Four-Legged Model
Variable Description	**Probability of fatal injury**	**Probability of fatal injury**
“1x2”, “1x3” and “1x4” intersections	-0.011	N/A
“2x2” and “2x3” intersections	-0.008	N/S
“2x4”, “2x5” and “2x6” intersections	-0.004	N/S
“2x7” and “2x8” intersections	-0.001	N/S
“3x2”, “3x3”, “3x4”, “3x5”, “3x6” and “3x8” intersections	-0.004	N/S
Dummy variable for Brevard County	-0.001	0.006
Dummy variable for Hillsborough County	-0.013	-0.006
Dummy variable for Leon County	-0.014	-0.003
Dummy variable for Miami-Dade County	-0.017	-0.011
Dummy variable for Orange County	0.000	0.008

[a] N/S means not significant [b] N/A means not applicable

6.5.1 Three-Legged Model Interpretation

From Table 6-3, increasing the natural logarithm of AADT on the major road by unity (which inherently means increasing AADT) significantly reduces fatal injury probability by 0.2%. As the AADT increases, speed decreases, and hence fatal crashes decrease as well, whereas crashes occurring at higher AADT (like rear-end and sideswipe crashes) are not generally fatal. This result is consistent with that done by Klop and Khattak (1999), who found a significant decrease in bicycle injury severity with the increase in AADT.

The spatial effect for the upstream distance to the nearest signalized intersection from the unsignalized intersection of interest showed that there is a 0.1% increase in the fatal injury probability for a unit increase in the natural logarithm of the distance. This could be attributed to the fact that as the distance between intersections increases, drivers tend to drive at (or above) the speed limit on that stretch (which is mostly high), and thus accident severity increases at high speeds, which is an expected outcome. This was also examined by Malyshkina and Mannering (2008), and Klop and Khattak (1999), as previously illustrated. Moreover, its probit coefficient is statistically significant at the 95% confidence.

Lower speed limits (less than 45 mph) significantly reduce fatal injury probability by 0.3%, when compared to speed limits greater than 45 mph. This result is consistent with the previous finding, and is very reasonable, as fatal crashes always occur at higher speeds. This conforms to the study done by Malyshkina and Mannering (2008) and Renski et al. (1998), who examined the safety effect of speed limits on severe accidents, and found that high speed limits are associated with high accident severities. Also, the study by Klop and Khattak (1999) found a significant increase in bicycle and passenger car injury severity with increase in speed limits.

An interesting finding is that having no right turn lanes or 1 right turn lane on the major road decreases fatal injury probability by 0.4% when compared to having 2 right turn lanes. Their probit estimates are statistically significant at the 90% confidence.

Having no left turn movement on the minor approach decreases the probability of fatal injury by 0.1%, when compared to having 1 left turn movement. This is mainly due to the reduction of conflict points while prohibiting the left turn maneuver. This result is consistent with the study done by Liu et al. (2007) and Lu et al. (2001 a; 2001 b; 2004 and 2005), who found that there is a reduction in total crashes and fatality for right turns followed by U-turns, as an alternative to direct left turn maneuvers from driveways. However, the probit estimate is not statistically significant at the 90% confidence.

Having one, two and three through lanes on the minor approach always increase the fatal injury probability when compared to having 4 though lanes. The highest increase is 2.1% where one through lane existed. One through lanes could exist at ramp junctions with yield signs, where merging and diverging maneuvers always occur, thus these traffic conflicts result in traffic problems and serious injuries. Its estimate is statistically significant at the 95% confidence.

The highest significant reduction in the probability of having a fatal injury occurs in middle, young and very young at-fault drivers, which is 0.4% less than that at very old drivers. This result is consistent with the study by Abdel-Aty et al. (1998), who concluded that young and very young drivers are associated with fatal injury reduction as well. Although very old drivers tend to drive slowly and carefully, their weak physical condition, as well as their higher reaction time could explain the higher fatality risk.

Increasing the inside (left or median) shoulder width by 1 feet significantly increases fatal injury by 0.1%. This finding contradicts with the finding of Noland and Oh (2004), who found that there is no statistical association with changes in safety for inside shoulder widths. The use of the inside shoulder width was not explored extensively in traffic safety analysis in terms of severe crashes. For example, Klop and Khattak (1999) did not use the inside shoulder width in their analysis due to the unrealistic values documented in their dataset.

The highest significant reduction in the probability of having a fatal injury occurs at dawn, which is 0.9% less than that at dark with no street lights. This might be attributed to the low traffic volume at dawn time (i.e., lower conflict risk).

The only significant reduction in the probability of having a fatal injury occurs at “2x2” and “2x3” intersections, which is 0.8% less than that at “4x2”, “4x4”, “4x6” and “4x8” intersections. This result is considered reasonable, given the complexity of large intersections for some drivers.

The highest reduction in the probability of having a fatal injury occurs at Miami-Dade County, which is 1.7% (0.017) less than that at Seminole County. Miami-Dade County is the heaviest-populated and most urbanized county used in this study (U.S. Census, 2000), thus, more crash frequency is expected to occur, however, less fatal injuries could happen due to high-dense roadways (relatively high AADT). Moreover, its probit estimate is statistically significant, as shown in Table 6-2.

6.5.2 Four-Legged Model Interpretation

From Table 6-3, as anticipated, increasing the natural logarithm of AADT on the major road by unity significantly reduces fatal injury probability by 0.6%.

As expected, there is a 0.1% increase in the fatal injury probability for a unit increase in the natural logarithm of the upstream and downstream distances to the nearest signalized intersections. This is consistent with that at 3-legged unsignalized intersections.

Lower speed limits (less than 45 mph) reduce fatal injury probability by 0.2%, when compared to speed limits greater than 45 mph. This finding is consistent with that at 3-legged unsignalized intersections.

Intersection's skewness angle less than or equal to 75 degrees significantly increases fatal injury probability by 0.4%, when compared to skewness angle greater than 75 degrees. This is a very reasonable outcome, as the sight distance is a problem. This illustrates the significant importance of designing intersections with skewness angle around 90 degrees, to reduce severe crashes.

As found in the three-legged model, increasing the inside (left or median) shoulder width by 1 feet significantly increases fatal injury by 0.2%.

An increase in the right shoulder width by 1 feet has almost no effect on the probability of fatal injuries. This finding is consistent with that of Klop and Khattak (1999), who examined the effect on the right shoulder width on bicycle crash severity on two-lane, undivided roadways in North Carolina, and found that the right shoulder width has no statistical effect on severity compared to the absence of a shoulder.

The highest significant reduction in the probability of having a fatal injury occurs at Miami-Dade County, which is 1.1% less than that at Seminole County. This finding is consistent with the three-legged model. This might also be related to varying reporting thresholds at different counties.

6.6 Analysis of the Binary Probit Framework

The fitted binary probit model for both 3 and 4-legged unsignalized intersections using the two levels (severe vs. non-severe) of the response variable is shown in Table 6-4. The marginal effects for the estimated models for both 3 and 4-legged intersections are shown in Table 6-5.

Table 6-4: Binary Probit Estimates for 3 and 4-Legged Unsignalized Intersections

	Three-Legged Model		Four-Legged Model	
Variable Description	Estimate [a]	P-value	Estimate [a]	P-value
Intercept	-0.5872 (0.8890)	0.5089	0.6682 (0.6980)	0.3384
Natural logarithm of AADT on the major road	-0.1015 (0.0592)	0.0866	-0.1643 (0.0651)	0.0117
Natural logarithm of the upstream distance to the nearest signalized intersection	0.0528 (0.0255)	0.0383	N/S [b]	
Natural logarithm of the downstream distance to the nearest signalized intersection	0.0639 (0.0265)	0.0161	N/S	
No stop line exists on the minor approach	0.1133 (0.0629)	0.0718	N/S	
A stop line exists on the minor approach	--- [c]		N/S	
Posted speed limit on major road < 45 mph	-0.1252 (0.0633)	0.0481	-0.2547 (0.0722)	0.0004
Posted speed limit on major road >= 45 mph	--- [c]		--- [c]	
Skewness angle <= 75 degrees	N/S		0.3183 (0.1178)	0.0069
Skewness angle > 75 degrees	N/S		--- [c]	
No right turn lane exists on the major approach	-0.2139 (0.1413)	0.1302	-0.1964 (0.1106)	0.0758
One right turn lane exists on only 1 major road direction	-0.2363 (0.1464)	0.1066	0.0133 (0.1236)	0.9142
One right turn lane exists on each major road direction	--- [c]		--- [c]	
No left turn lane exists on the major approach	0.0036 (0.0751)	0.9613	N/S	
One left turn lane exists on only 1 major road direction	0.1124 (0.0607)	0.0641	N/S	
One left turn lane exists on each major road direction	--- [c]		N/S	
15 <= At-fault driver's age <= 19 (very young)	-0.2720 (0.1496)	0.0692	N/S	
20 <= At-fault driver's age <= 24 (young)	-0.2360 (0.1480)	0.1109	N/S	
25 <= At-fault driver's age <= 64 (middle)	-0.1837 (0.1391)	0.1867	N/S	
65 <= At-fault driver's age <= 79 (old)	-0.1401 (0.1591)	0.3785	N/S	
At-fault driver's age >= 80 (very old)	--- [c]		N/S	
Right shoulder width on the major road	0.0209 (0.0113)	0.0651	N/S	
Daylight lighting condition	-0.4425 (0.0864)	<0.0001	N/S	
Dusk lighting condition	-0.6063 (0.1696)	0.0004	N/S	

	Three-Legged Model		Four-Legged Model	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
Dawn lighting condition	-0.3626 (0.2316)	0.1175	N/S	
Dark (street light) lighting condition	-0.2314 (0.0971)	0.0172	N/S	
Dark (no street light) lighting condition	--- [c]		N/S	
Access point unsignalized intersections	0.4426 (0.2853)	0.1209	N/S	
Ramp junctions	-4.1439 (0.1987)	<0.0001	N/A [d]	
Regular unsignalized intersections	0.4640 (0.2798)	0.0972	N/S	
Unsignalized intersections close to railroad crossings	--- [c]		N/S	
"1x2", "1x3" and "1x4" intersections	4.8632 (0.1987)	<0.0001	N/A	
"2x2" and "2x3" intersections	-0.1546 (0.2140)	0.4701	N/S	
"2x4", "2x5" and "2x6" intersections	0.0419 (0.2064)	0.8391	N/S	
"2x7" and "2x8" intersections	0.1258 (0.2489)	0.6132	N/S	
"3x2", "3x3", "3x4", "3x5", "3x6" and "3x8" intersections	0.0174 (0.2199)	0.9367	N/S	
"4x2", "4x4", "4x6" and "4x8" intersections	--- [c]		N/S	
Dummy variable for Brevard County	-0.1314 (0.1216)	0.2798	0.1706 (0.1460)	0.6467
Dummy variable for Hillsborough County	-0.1444 (0.1018)	0.1562	-0.0534 (0.1166)	0.0975
Dummy variable for Leon County	-0.6443 (0.1109)	<0.0001	-0.2390 (0.1442)	0.0109
Dummy variable for Miami-Dade County	-0.4746 (0.1070)	<0.0001	-0.3263 (0.1281)	0.6467
Dummy variable for Orange County	-0.2244 (0.1041)	0.0312	-0.0477 (0.1331)	0.7198
Dummy variable for Seminole County	--- [c]		--- [c]	
Percentage of trucks on the major road	-0.0096 (0.0085)	0.2612	N/S	
Log-likelihood at convergence	-1869		-1039	
Log-likelihood at zero [e]	-1971.1		-1095.7	
AIC	3804		2100	

[a] Standard error in parentheses [b] N/S means not significant [c] Base case [d] N/A means not applicable [e] Likelihood while fitting the intercept only

Table 6-5: Marginal Effects for Severe Injury Probability for the Fitted Covariates in the 3 and 4-Legged Models

	Three-Legged Model	Four-Legged Model
Variable Description	Probability of severe injury	Probability of severe injury
Natural logarithm of AADT on the major road	-0.015	-0.023
Natural logarithm of the upstream distance to the nearest signalized intersection from the unsignalized intersection of interest	0.008	N/S [a]
Natural logarithm of the downstream distance to the nearest signalized intersection from the unsignalized intersection of interest	0.009	N/S
No stop line exists on the minor approach	0.017	N/S
Posted speed limit on major road < 45 mph	-0.018	-0.036
Skewness angle <= 75 degrees	N/S	0.045
No right turn lane exists on the major approach	-0.031	-0.028
One right turn lane exists on only 1 major road direction	-0.035	0.002
No left turn lane exists on the major approach	0.001	N/S
One left turn lane exists on only 1 major road direction	0.017	N/S
15 <= At-fault driver's age <= 19 (very young)	-0.040	N/S
20 <= At-fault driver's age <= 24 (young)	-0.035	N/S
25 <= At-fault driver's age <= 64 (middle)	-0.027	N/S
65 <= At-fault driver's age <= 79 (old)	-0.021	N/S
Right shoulder width on the major road	0.003	N/S
Daylight lighting condition	-0.065	N/S
Dusk lighting condition	-0.089	N/S
Dawn lighting condition	-0.053	N/S
Dark (street light) lighting condition	-0.034	N/S
Access point unsignalized intersections	0.065	N/S
Ramp junctions	-0.650	N/A

	Three-Legged Model	Four-Legged Model
Variable Description	**Probability of severe injury**	**Probability of severe injury**
Regular unsignalized intersections	0.068	N/S
"1x2", "1x3" and "1x4" intersections	0.716	N/A [b]
"2x2" and "2x3" intersections	-0.023	N/S
"2x4", "2x5" and "2x6" intersections	0.006	N/S
"2x7" and "2x8" intersections	0.019	N/S
"3x2", "3x3", "3x4", "3x5", "3x6" and "3x8" intersections	0.003	N/S
Dummy variable for Brevard County	-0.019	0.024
Dummy variable for Hillsborough County	-0.021	-0.008
Dummy variable for Leon County	-0.095	-0.034
Dummy variable for Miami-Dade County	-0.070	-0.046
Dummy variable for Orange County	-0.033	-0.007
Percentage of trucks on the major road	-0.001	N/S

[a] N/S means not significant [b] N/A means not applicable

6.6.1 Three-Legged Model Interpretation

From Table 6-5, as expected, increasing the natural logarithm of AADT on the major road by unity reduces fatal injury probability by 1.5%.

There is a 0.8 and 0.9% significant increase in severity probability for a unit increase in the natural logarithm of the upstream and downstream distances to the nearest signalized intersection, respectively.

Having no stop lines on the minor approach increases severity probability by 1.7%, when compared to having stop lines. This is a reasonable outcome, emphasizing the importance of marking stop lines at unsignalized intersections for reducing severity. Moreover, their probit estimates are statistically significant at the 90% confidence.

Lower speed limits (less than 45 mph) significantly reduce severe injury probability by 1.8%, when compared to speed limits greater than 45 mph.

As concluded from the ordered probit model, having no right turn lanes or 1 right turn lane on the major road decreases severe injury probability when compared to having 2 right turn lanes. However, their probit estimates are not statistically significant at the 90% confidence.

An interesting finding is that having 1 left turn lane on one of the major approaches increases severe injury probability by 1.7%, when compared to having 2 left turn lanes. The estimate is statistically significant at the 90% confidence.

As previously found, the highest reduction in the severity probability occurs in young and very young at-fault drivers.

An increase in the right shoulder width by 1 feet increases the severity probability by 0.3%. This can be attributed to the fact that wide shoulders encourage to inappropriately using this shoulder, hence, there is a high sideswipe and rear-end crash risk, which might be severe at relatively high speeds. This finding indeed conforms to that of Noland and Oh (2004), who found that increasing the right shoulder width increases severity.

The highest significant reduction in the probability of having a severe injury occurs at dusk, which is 8.9% less than that at dark with no street lights. This might be attributed to the relatively lower conflict risk.

Although ramp junctions are usually controlled by a yield sign, and merging maneuvers are more dominant, those intersection types significantly reduce severe injury probability by 65% than intersections nearby railroad crossings.

The highest significant increase in the probability of severe injury occurs at "1x2", "1x3" and "1x4" intersections, which is 71.6% higher than that at "4x2", "4x4", "4x6" and "4x8" intersections. Intersection's configurations ("1x2", "1x3" and "1x4") could exist at ramp junctions with yield signs, where merging and diverging maneuvers occur, hence traffic conflicts and serious injuries are more likely, especially at higher speeds.

The second highest significant reduction in the probability of severe injury occurs at Miami-Dade County, which is 7% less than that at Seminole County. This assesses the previous finding that highly-urbanized areas experience less severity.

Increasing the percentage of trucks on the major road by unity reduces the probability of severe injury. This could be interpreted as drivers are very attentive while overtaking or driving behind trucks. However, the probit estimate is not statistically significant at the 90% confidence.

6.6.2 Four-Legged Model Interpretation

From Table 6-5, as expected, increasing the natural logarithm of AADT on the major road by unity significantly reduces severe injury probability by 2.3%.

Lower speed limits (less than 45 mph) significantly reduce severe injury probability by 3.6%, when compared to speed limits greater than 45 mph. This finding is consistent with that at 3-legged unsignalized intersections.

As previously found, having skewness angle less than or equal to 75 degrees significantly increases severity probability, when compared to skewness angle greater than 75 degrees.

As concluded from the three-legged model, having no right turn lanes on the major road decreases severe injury probability when compared to having 2 right turn lanes. However, the probit estimate is not statistically significant at the 90% confidence.

As previously found, the highest significant reduction in the probability of severe injury occurs at Miami-Dade County, which is 4.6% less than that at Seminole County. This finding is consistent with that from the three-legged model.

6.7 Comparing the Two Probit Frameworks

By comparing the AIC and the log-likelihood values in the four fitted 3 and 4-legged probit models, it is obvious that the aggregated binary probit models fit the data better than the disaggregated ordered probit models (lower AIC and higher log-likelihood at convergence). This demonstrates that the aggregate model works better in analyzing crash severity at unsignalized intersections.

6.8 Nested Logit Model Estimates

The last approach performed in this chapter is fitting a nested logit model for both 3 and 4-legged intersections. Figures 6-1 and 6-2 show the two attempted nesting structures. For example, Figure 6-2 describes the analysis of crash injury level (PDO, possible injury, and non-incapacitating injury) conditioned on non-severe injury, as well as the analysis of crash injury level (incapacitating injury and fatal) conditioned on severe injury. The shown nesting structure has 2 levels. The first level (at the bottom of the nest) contains the five crash injury levels, whereas the second level (at the top of the nest) contains the two crash injury levels, severe and non-severe injuries.

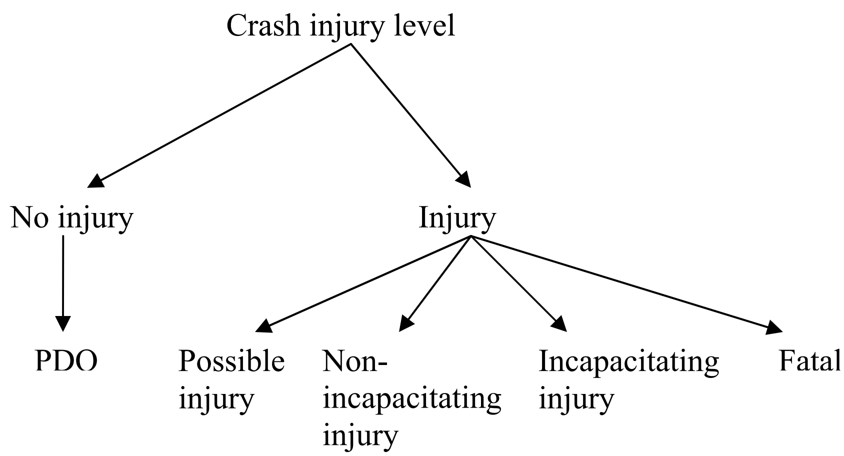


Figure 6-1: First Attempted Two-level Nesting Structure for the Nested Logit Framework

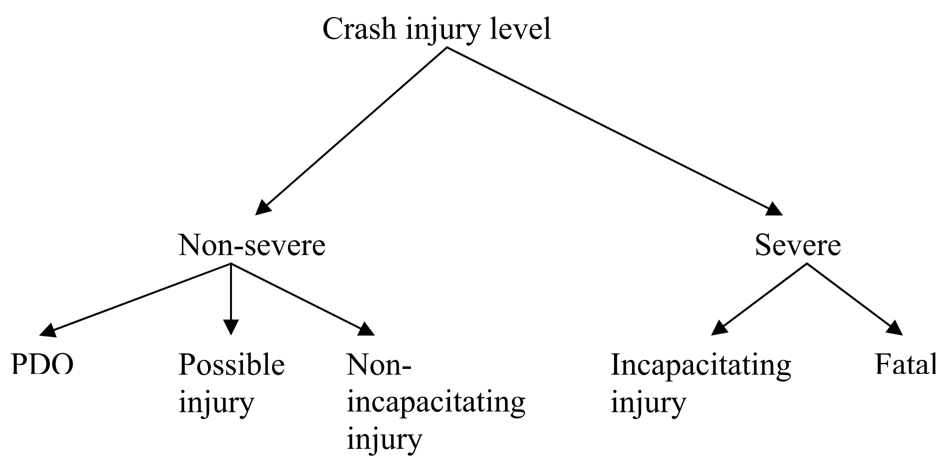


Figure 6-2: Second Attempted Two-level Nesting Structure for the Nested Logit Framework

The nesting structure shown in Figure 6-2 showed better results than Figure 6-1. This was concluded from the resulted AIC and log-likelihood values. The fitted nested logit model for 3-legged intersections using the nesting structure sketched in Figure 6-2 is shown in Table 6-6.

Table 6-6: Nested Logit Estimates for 3-Legged Unsignalized Intersections (Nesting Structure Shown in Figure 6-2)

Variable Description	Estimate	Standard Error	P-value
Posted speed limit on the major road	-0.0100	0.0015	<0.0001
At-fault driver's age	-0.0011	0.0004	0.0173
Left shoulder width near the median on the major road	0.0173	0.0084	0.0396
Natural logarithm of the upstream distance to the nearest signalized intersection from the unsignalized intersection of interest	-0.0110	0.0096	0.2532
Size of the intersection	-0.0136	0.0097	0.1657
Inclusive parameter of the "severity" nest	4.8495	0.3695	<0.0001
Log-likelihood at convergence	-9182		
AIC	18375		
Number of observations	34040		

From this table, the inclusive parameter is significantly greater than one, hence the nested logit model is not accepted for the modeling purpose of these data. It is obvious that fewer variables are significant in the model and the goodness-of-fit criterion (e.g., AIC) is not as favorable as the ordered or binary probit models. Variables like the natural logarithm of the upstream distance and the speed limit have unexpected negative coefficients, as opposed to the corresponding probit estimates, hence, they are difficult to interpret.

6.9 Summary of Results

The important geometric, traffic, driver and demographic factors from this chapter's analysis affecting fatal (severe) injury at unsignalized intersections are summarized in Table 6-7. The effect of the shown continuous variables is estimated based on an increase of unity in each of them, while the effect of those categorical variables is estimated with respect to the base case for each.

Table 6-7: Important Factors Affecting Fatal (Severe) Injury at Unsignalized Intersections

Factors	Effect on fatal (severe) injury (Statistical significance)
Geometric and roadway factors	
Right shoulder width on the major approach (in feet)	Increase*
Left shoulder width near the median on the major approach (in feet)	Increase**
Natural logarithm of the upstream distance to the nearest signalized intersection from the unsignalized intersection of interest	Increase*
Natural logarithm of the downstream distance to the nearest signalized intersection from the unsignalized intersection of interest	Increase*
Posted speed limit on major road < 45 mph (Base is speed limit >= 45 mph)	Decrease*
No stop line exists on the minor approach (Base is 1 stop line)	Increase*
Skewness angle <= 75 degrees (Base is skewness angle > 75 degrees)	Increase**
Ramp junctions (Base are intersections close to railroad crossings)	Decrease**
One left turn lane on the major approach (Base is 2 left turn lanes)	Increase*
Traffic factors	
Natural logarithm of AADT on the major approach	Decrease*
One, two and three through lanes on the minor approach (Surrogate measure for AADT on the minor approach) (Base is 4 through lanes)	Increase***
Driver-related factors	
Young at-fault drivers (Base is very old at-fault drivers)	Decrease*
Demographic factors	
Heavily-populated and highly-urbanized area (Base is less-populated area)	Decrease**

*Statistical significance at the 90% confidence

**Statistical significance at the 95% confidence

***Existence of one through lane is the only statistically significant at the 90% confidence

6.10 General Conclusions from the Crash Severity Analysis

The analysis conducted in this chapter attempted to put deep insight into factors affecting crash injury severity at 3 and 4-legged unsignalized intersections using the most comprehensive data collected by using the ordered probit, binary probit and nested logit frameworks. The common factors found in the fitted probit models are the logarithm of AADT on the major road, and the speed limit on the major road. It was found that higher severity (and fatality) probability is always associated with a reduction in AADT, as well as an increase in speed limit. The fitted probit models also showed several important traffic, geometric and driver-related factors affecting safety at unsignalized intersections. Traffic factors include AADT on the major approach, and the number of through lanes on the minor approach (a surrogate for AADT on the minor approach). Geometric factors include the upstream and downstream distance to the nearest signalized intersection, existence of stop lines, left and right shoulder width, number of left turn movements on the minor approach, and number of right and left turn lanes on the major approach. As for driver factors, young and very young at-fault drivers were always associated with the least fatal/severe probability compared to other age groups. Also, heavily-populated and highly-urbanized areas experience lower fatal/severe injury.

Comparing the aggregated binary probit model and the disaggregated ordered probit model showed that the aggregate probit model produces comparable if not better results, thus for its simplicity the binary probit models could be used to model crash injury severity at unsignalized intersections. The nested logit models did not show any improvement over the probit models.

CHAPTER 7. APPLICATION OF THE MULTIVARIATE ADAPTIVE REGRESSION SPLINES FOR PREDICTING CRASH OCCURRENCE

7.1 Introduction

Statistical models (or safety performance functions) are mainly used for identifying some relationships between the dependent variable and a set of explanatory covariates. Also, predicting crashes is another important application of safety performance functions. Those predicted crashes can help identify hazardous sites, hence significant countermeasures can be applied for further safety remedy. The most common probabilistic models used by transportation safety analysts for modeling vehicle crashes are the traditional Poisson and NB distributions. NB regression models are usually favored over Poisson regression models since crash data are usually characterized by over-dispersion (Lord et al., 2005), which means that the variance is greater than the mean.

Transportation safety analysts usually focus on comparing various statistical models based on some goodness-of-fit criteria (e.g., Miaou and Lord, 2003 and Shankar et al., 1997). Since prediction is an essential objective of crash models, some studies that focused on developing models for mainly predicting vehicle crashes are Lord (2000), Xie et al. (2007) and Li et al. (2008). Researchers are always trying to introduce and develop statistical tools for effectively predicting crash occurrence.

Thus, one of the main objectives of the analysis in this chapter is to explore the potential of applying a recently developed data mining technique, the multivariate adaptive regression splines (MARS), for a precise and efficient crash prediction. This was demonstrated in this chapter through various applications of MARS via data collected at unsignalized intersections. Another objective is to explore the significant factors that contribute to specific crash type occurrence (rear-end as well as angle crashes) at unsignalized intersections by utilizing a recently collected extensive dataset of 2475 unsignalized intersections.

7.2 Methodological Approach

7.2.1 Multivariate Adaptive Regression Splines Model Characteristics

Most of the methodology described here is found in Put et al. (2004). According to Abraham et al. (2001), splines are defined as "an innovative mathematical process for complicated curve drawings and function approximation". To develop any spline, the X-axis representing the space of predictors is broken into number of regions. The boundary between successive regions is known as a knot (Abraham et al., 2001). While it is easy to draw a spline in two dimensions (using linear or quadratic polynomial regression models), manipulating the mathematics in higher dimensions is best-accomplished using the "basis functions", which are the elements of fitting a MARS model.

According to Friedman (1991), the MARS method is a local regression method that uses a series of basis functions to model complex (such as nonlinear) relationships. The global MARS model is defined as shown in Equation (7.1) (Put et al., 2004).

$$\hat{y} = a_0 + \sum_{m=1}^{M} a_m B_m(x) \tag{7.1}$$

where: $\hat{y}$ is the predicted response;

a_0 is the coefficient of the constant basis function;

$B_m(x)$ is the mth basis function, which can be a single spline function or an interaction of two (or more) spline functions;

a_m is the coefficient of the mth basis function; and

M is the number of basis functions included in the MARS model.

According to Put et al. (2004), there are three main steps to fit a MARS model. The first step is a constructive phase, in which basis functions are introduced in several regions of the predictors and are combined in a weighted sum to define the global MARS model (as shown in Equation (7.1)). This global model usually contains many basis functions, which can cause an over-fitting. The second step is the pruning phase, in which some basis functions of the over-fitting MARS model are deleted. In the third step, the optimal MARS model is selected from a sequence of smaller models.

In order to describe in details the three MARS steps, the first step is created by continually adding basis functions to the model. The introduced basis functions consist either of a single spline function or a product (interaction) of two (or more) spline functions (Put et al., 2004). Those basis functions are added in a "two-at-a-time" forward stepwise procedure, which selects the best pairs of spline functions in order to improve the model. Each pair consists of one left-sided and one right-sided truncated function defined by a given knot location, as shown in Equations (7.2) and (7.3), respectively. For this, spline functions in MARS are piecewise polynomials.

$$[-(x-t)_+^q] = \begin{cases} (t-x)^q\,;\; x < t \\ 0;\ \text{otherwise} \end{cases} \tag{7.2}$$

$$[+(x-t)_+^q] = \begin{cases} (x-t)^q\,;\; x > t \\ 0;\ \text{otherwise} \end{cases} \tag{7.3}$$

Also, from (Put et al., 2004), it is to be noted that the search for the best predictor and knot location is performed in an iterative way. The predictor, as well as knot location which contribute most to the model, are selected first. Also, at the end of each iteration, the introduction of an interaction is checked so as to improve the model. As shown by Put et al. (2004), the order of any fitted MARS model indicates the maximum number of basis functions that interact (for example, in a second-order MARS model, the interaction order is not more than two). The iterative building procedure continues until a maximum number of basis functions "M_{max}" is included. The value of "M_{max}" should be considerably larger than the optimal model size "M*" produced by MARS. According to Friedman (1991), the order of magnitude of "M_{max}" is twice that of "M*".

From Put et al. (2004), the second step is the pruning step, where a "one-at-a-time" backward deletion procedure is applied in which the basis functions with the lowest contribution to the model are excluded. This pruning is mainly based on the generalized cross-validation (*GCV*) criterion (Friedman, 1991), and in some cases, the n-fold cross validation can be used for pruning. The *GCV* criterion is used to find the overall best model from a sequence of fitted models. While using the *GCV* criterion, a penalty for the model complexity is incorporated. A larger *GCV* value tends to produce a smaller model, and vice versa. The *GCV* criterion is estimated using Equation (7.4) (Put et al., 2004).

$$GCV(M) = \frac{1}{N} \frac{\sum_{i=1}^{N} (y_i - \hat{y})^2}{(1 - C(M)/N)^2} \tag{7.4}$$

where: N is the number of observations;

y_i is the response for observation i;

$\hat{y}$ is the predicted response for observation i; and

$C(M)$ is a complexity penalty function, which is defined as shown in Equation (7.5).

$$C(M) = M + dM \tag{7.5}$$

where: M is the number of non-constant basis functions (i.e., all terms of Equation (7.1) except for "a_0"); and d is a defined cost for each basis function optimization. As shown by Put et al. (2004), the higher the cost d is, the more basis functions will be excluded. Usually, d is increased during the pruning step in order to obtain smaller models. Along with being used during the pruning step, the increase in the *GCV* value while removing a variable from the model is also used to evaluate the importance of the predictors in the final fitted MARS model.

As shown by Xiong and Meullenet (2004), the term "$\sum_{i=1}^{N}(y_i - \hat{y})^2$" measures the lack of fit on the *M* basis functions in the MARS model, which is the same as the sum of squared residuals, and "$(1 - C(M)/N)^2$" is a penalty term for using *M* basis functions.

Finally, the third step is mainly used for selecting the optimal MARS model. The selection is based on an evaluation of the prediction characteristics of the different fitted MARS models. For more details on MARS formulation, Friedman (1991), Put et al. (2004) as well as Sekulic and Kowalski (1992) are relevant references.

7.2.2 Random Forest Technique

Since the random forest technique was attempted in this study in conjunction with MARS, a brief description of this technique is discussed. Random forest is one of the most recent promising machine learning techniques proposed by Breiman (2001) that is well known for selecting important variables from a set of variables. In this technique, a number of trees are grown by randomly selecting some observations from the original dataset with replacement, then searching over only a randomly selected subset of covariates at each split (Harb et al., 2009 and Kuhn et al., 2008).

As well known, for each grown tree, the important covariates are shown on the root (top) of the tree, and leaves (terminal nodes) are shown on the bottom of the tree. Terminal nodes have no further splitting. For each split on the grown tree, rules are assigned for selecting other important covariates, and so on. For each tree, the prediction performance (based on the misclassification rate) is done on the terminal nodes.

As shown by Grimm et al. (2008), random forest is robust to noise in the covariates. The main advantages of random forest are that it usually yields high classification accuracy, and it handles missing values in the covariates efficiently.

To test whether the attempted number of trees is sufficient enough to reach relatively stable results, the plot of the out-of-bag (OOB) error rate against various tree numbers is generated, as recommended by the R package. The best number of trees is that having the minimum error rate, as well as a constant error rate nearby.

To select the important covariates, the R package provides the mean decrease Gini "IncNodePurity" diagram. This diagram shows the node purity value for every covariate (node) of a tree by means of the Gini index (Kuhn et al., 2008). A higher node purity value represents a higher variable importance, i.e., nodes are much purer.

7.2.3 Assessing Prediction Performance

To examine the significant prediction performance of the MARS technique (for example, while comparing with the NB model), there were two main evaluation criteria used, the MAD and the MSPE. The MAD and MSPE criteria were also used in the study done by Lord and Mahlawat (2009) for assessing the goodness-of-fit of the fitted models. The same criteria were used by Jonsson et al. (2009) to assess the fitted models for both three and four-legged unsignalized intersections. Also, Li et al. (2008) used the MAD and MSPE criteria while comparing NB to SVM models, as well as while comparing SVM to the Bayesian neural networks models. Equations (5.13) and (5.14) - previously mentioned in Chapter 5 - show how to evaluate MAD and MSPE, respectively. However, the estimated MAD and MSPE values in this chapter are normalized by the average of the response variable. This was done because crash frequency has higher range, hence error magnitude is relatively higher. However, normalizing crash frequency by the logarithm of AADT or considering the logarithm of crash frequency results in having smaller range, hence error magnitude is relatively lower. By this, the comparison between the MARS models using discrete and continuous responses is valid.

7.3 MARS Applications

There were three main applications performed in this study using the MARS technique. Each application was performed separately for analyzing each of the rear-end and angle crashes. These crash types were specifically selected, as they are the most frequent crash types occurring at unsignalized intersections (Summersgill and Kennedy, 1996; Layfield, 1996; Pickering and Hall, 1986; Agent, 1988 and Hanna et al., 1976).

The first application dealt with a comparison between the fitted NB and MARS models while treating the response in each of them as a discrete variable (crash frequency). For the scope of this analysis, the traditional NB framework was used, and the training dataset used for calibration was 70% of the total data, while the remaining 30% was used for prediction. Thus, two NB rear-end crash frequency models were developed for 3 and 4-legged unsignalized intersections using a training dataset (1735 intersections) for four-year crash data from 2003 till 2006. Also, two NB angle crash models were developed for 3 and 4-legged unsignalized intersections using a training dataset (1732 intersections) for the same four years. Afterwards, using the same significant predictors in each of the NB models, MARS models were fitted and compared to the corresponding NB models. The prediction assessment criteria were performed on a test dataset (740 intersections for rear-end crashes analysis, and 743 intersections for angle crashes analysis) for the four-year crash data as well.

The second application dealt with treating the response in the fitted MARS models as a continuous variable. For rear-end crashes analysis, this was considered while normalizing the crash frequency by the natural logarithm of AADT. As for angle crashes analysis, the natural logarithm of AADT was considered as the response. The same training and test datasets were used as well. This application was proposed due to the high prediction capability of the MARS technique while dealing with continuous responses, as shown by Friedman (1991).

The third application dealt with combining MARS with the random forest technique for screening the variables before fitting a MARS model. This was investigated, because the attempt to fit a MARS model using all possible covariates did not improve the prediction. Thus, important covariates were identified using random forest, then fitted in a MARS model, and a comparison between MARS models (with the covariates initially screened using random forest) and MARS models (with the covariates initially screened using the NB model) was held.

7.4 Data Preparation and Variables' Description

The analysis conducted in this study was performed on 2475 unsignalized intersections collected from six counties in the state of Florida. The CAR database maintained by the FDOT was used to identify all SRs in those six counties. Then, the random selection method was used for choosing some state roads. Unsignalized intersections were then identified along these randomly selected SRs using "Google Earth" and "Video Log Viewer Application". In order to use the "Video Log Viewer Application", the roadway ID for the used SR, the mile point and the direction of travel should be specified. This application is an advanced tool developed by FDOT, and has the advantage of capturing the driving environment through the roadway. Moreover, this advanced application has two important features allowing different video perspectives, the "right view" and the "front view". The "right view" option provides the opportunity of identifying whether a stop sign and a stop line exist or not. The "front view" feature provides the opportunity of identifying the median type as well as the number of lanes per direction more clearly.

Afterwards, all the geometric, traffic and control fields of the collected intersections were filled out in a spreadsheet. These collected fields were merged with the RCI database for the 4 years (2003, 2004, 2005 and 2006). The RCI database – which is developed by the FDOT - includes physical and administrative data, such as functional classification, pavement, shoulder and median data related to the roadway (the New Web-based RCI Application). Each of these facilities is indexed by a roadway ID number with beginning and ending mile points. The used criteria for merging the data are the roadway ID and the mile point. The rear-end as well as the angle crash frequency for those identified unsignalized intersections were determined from the CAR database. The crash frequency database for the 4 years was merged with the already merged database (geometric, traffic and control fields with RCI database) for the 4 years. In this case, the used criterion for merging is the intersection ID. All these merging procedures were done using SAS (2002).

A summary statistics for rear-end and angle crashes in the modeling (training) and validation (test) databases for both 3 and 4-legged intersections is shown in Tables 7-1 and 7-2, respectively. From both tables, it can be noticed that there is an over-dispersion exists in the training datasets, hence, the use of the NB framework was appropriate for the scope of the analysis.

Table 7-1: Summary Statistics for Rear-end Cashes in the Training and Test Databases in "2003-2006"

	Three-legged training dataset in 4 years "2003-2006"	Four-legged training dataset in 4 years "2003-2006"	Three-legged test dataset in 4 years "2003-2006"	Four-legged test dataset in 4 years "2003-2006"
Number of observations	1338	397	599	141
Total number of rear-end crashes	1588	636	678	230
Mean rear-end crash frequency per intersection	1.186	1.602	1.131	1.631
Rear-end crash standard deviation per intersection	1.934	2.216	1.788	2.352

Table 7-2: Summary Statistics for Angle Crashes in the Training and Test Databases in "2003-2006"

	Three-legged training dataset in 4 years "2003-2006"	Four-legged training dataset in 4 years "2003-2006"	Three-legged test dataset in 4 years "2003-2006"	Four-legged test dataset in 4 years "2003-2006"
Number of observations	1341	391	596	147
Total number of angle crashes	1197	1008	585	312
Mean angle crash frequency per intersection	0.892	2.578	0.981	2.122
Angle crash standard deviation per intersection	1.734	3.856	2.079	2.808

It was decided to use two separate models for 3 and 4-legged intersections as both intersection types have different operating characteristics. For example, for 4-legged unsignalized intersections, there is an additional maneuver, which is vehicles crossing the whole major road width from the first minor approach to the second minor approach, thus leading to a right-angle crash risk. Other studies (e.g., Jonsson et al., 2009) modeled total crash frequency and specific crash types at three and four-legged intersections separately. A full description of the important variables used in the NB and MARS modeling procedures for 3 and 4-legged unsignalized intersections is shown in Table 7-3.

From Table 7-3, regular unsignalized intersections are those intersections having distant stretches on the minor approaches, whereas access points include parking lots at plazas and malls, and driveways that are feeding to the major approach. Due to the unavailability of AADT on most minor roads, an important traffic covariate explored in this study is the surrogate measure for AADT on the minor approach, which is represented by the number of through lanes on this approach.

The three MARS applications are shown for the analysis of rear-end crashes first, then are presented for angle crashes afterwards.

To explore the three spatial covariates (logarithm of upstream and downstream distances to the nearest signalized intersection, and logarithm of the distance between successive unsignalized intersections) on rear-end and angle crashes, Figures 7-1 to 7-12 are presented.

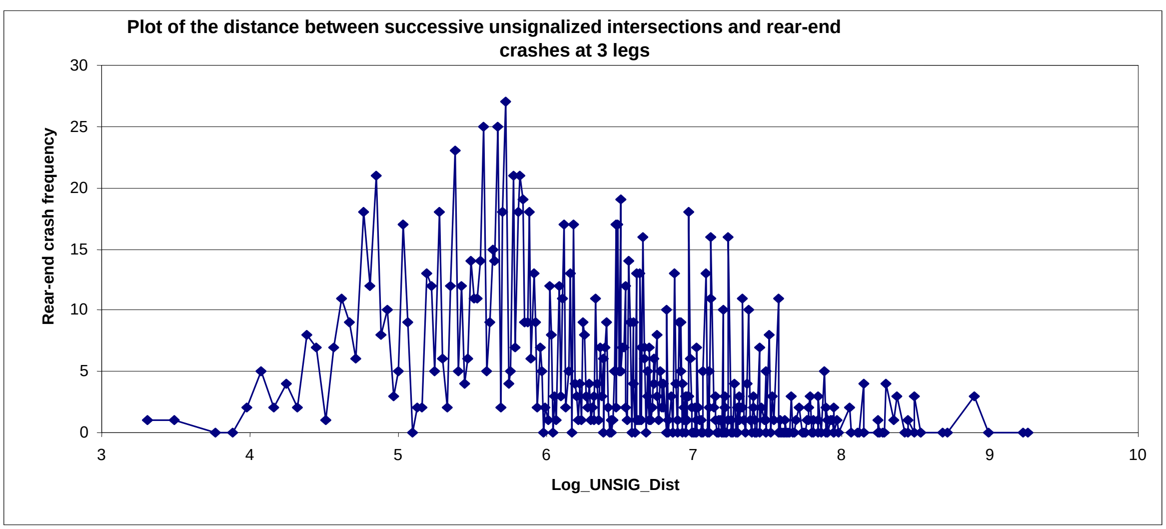


Figure 7-1: Plot of the Distance between Successive Unsignalized Intersections and Rear-end Crashes at 3-Legged Intersections

From Figure 7-1, it is noticed that there is a fluctuation in the trend, and it is difficult to determine the effect of the distance between successive unsignalized intersections on rear-end crashes at 3 legs from this plot.

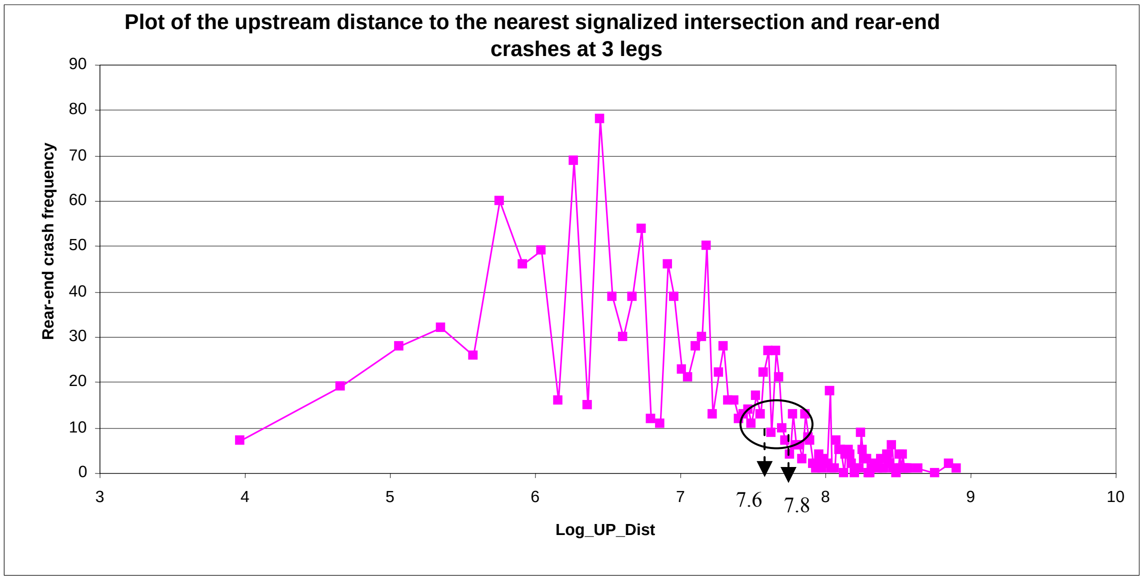


Figure 7-2: Plot of the Upstream Distance to the Nearest Signalized Intersection and Rear-end Crashes at 3-Legged Intersections

From Figure 7-2, it is noticed that rear-end crashes at 3 legs tend to decrease after a range of 7.6 to 7.8 for the log upstream distance (i.e., 0.38 to 0.46 miles), and there is no more trend fluctuation after this cut-off range. The highest rear-end crash frequency nearly occurs at a log upstream distance of around 6.5 (0.13 miles). Also, it can be deduced that rear-end crashes decrease with relatively large upstream distance at 3 legs.

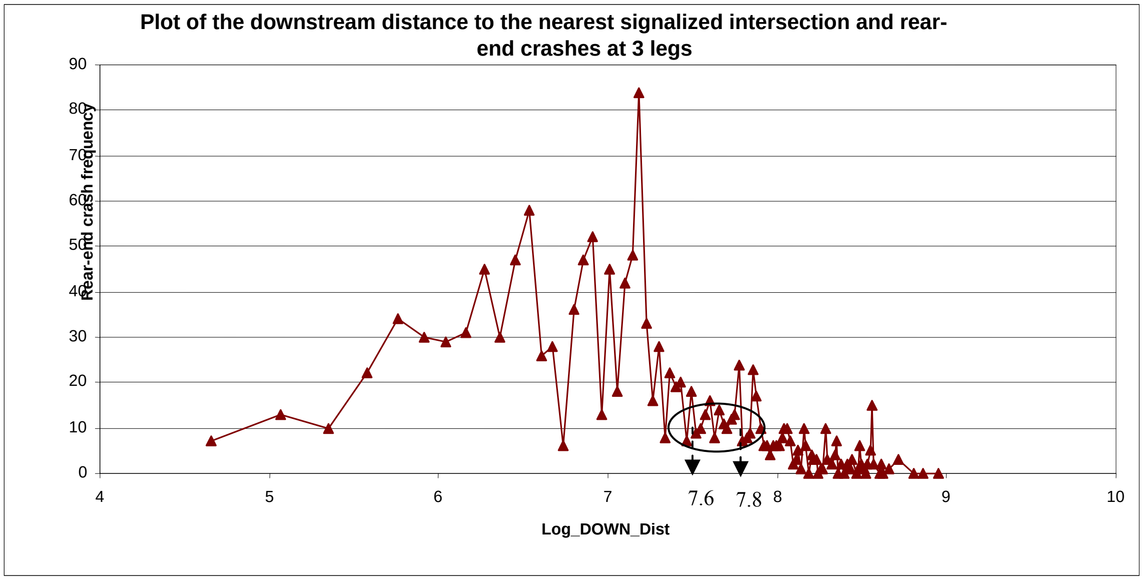


Figure 7-3: Plot of the Downstream Distance to the Nearest Signalized Intersection and Rear-end Crashes at 3-Legged Intersections

From Figure 7-3, it is noticed that rear-end crashes at 3 legs tend to decrease after a range of 7.6 to 7.8 for the log downstream distance (i.e., 0.38 to 0.46 miles), and there is no more trend fluctuation after this cut-off range. Also, it can be deduced that rear-end crashes decrease with relatively large downstream distance at 3 legs.

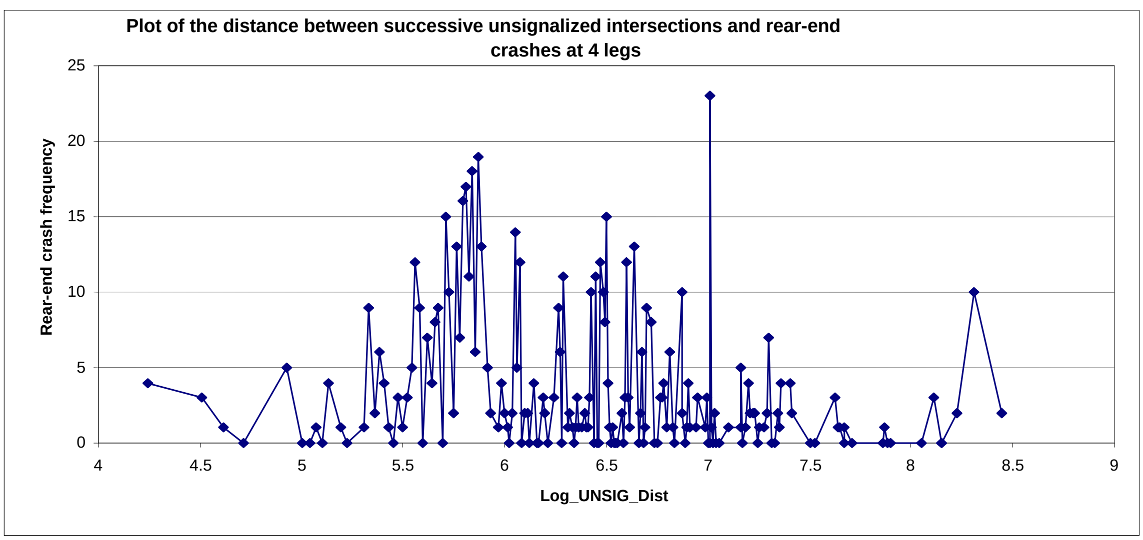


Figure 7-4: Plot of the Distance between Successive Unsignalized Intersections and Rear-end Crashes at 4-Legged Intersections

From Figure 7-4, it is noticed that there is a fluctuation in the trend, and it is difficult to determine the effect of the distance between successive unsignalized intersections on rear-end crashes at 4 legs from this plot.

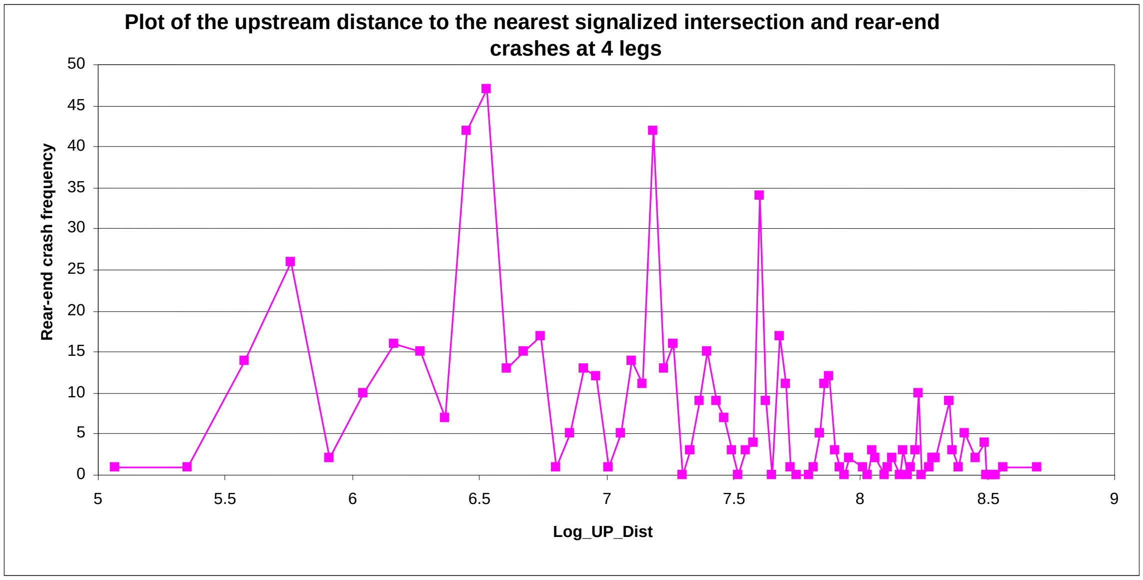


Figure 7-5: Plot of the Upstream Distance to the Nearest Signalized Intersection and Rear-end Crashes at 4-Legged Intersections

From Figure 7-5, it is noticed that rear-end crashes at 4 legs tend to decrease with relatively large upstream distance. Roughly, the cut-off range for the clear reduction starts from 7.6 to 7.8 (i.e., 0.38 to 0.46 miles). Also, the highest rear-end crash frequency nearly occurs at a log upstream distance of around 6.5 (0.13 miles).

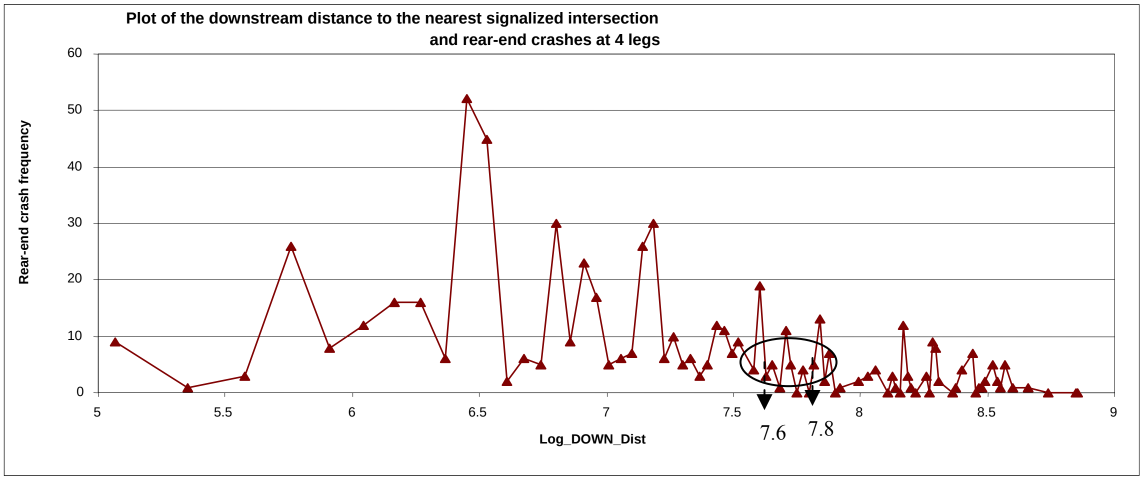


Figure 7-6: Plot of the Downstream Distance to the Nearest Signalized Intersection and Rear-end Crashes at 4-Legged Intersections

From Figure 7-6, it is noticed that the least magnitude of fluctuation occurs after a log downstream range distance of 7.6 to 7.8 (i.e., 0.38 to 0.46 miles), and generally, rear-end crashes decrease with relatively large downstream distance at 4 legs. Also, the highest rear-end crash frequency nearly occurs at a log downstream distance of around 6.5 (0.13 miles).

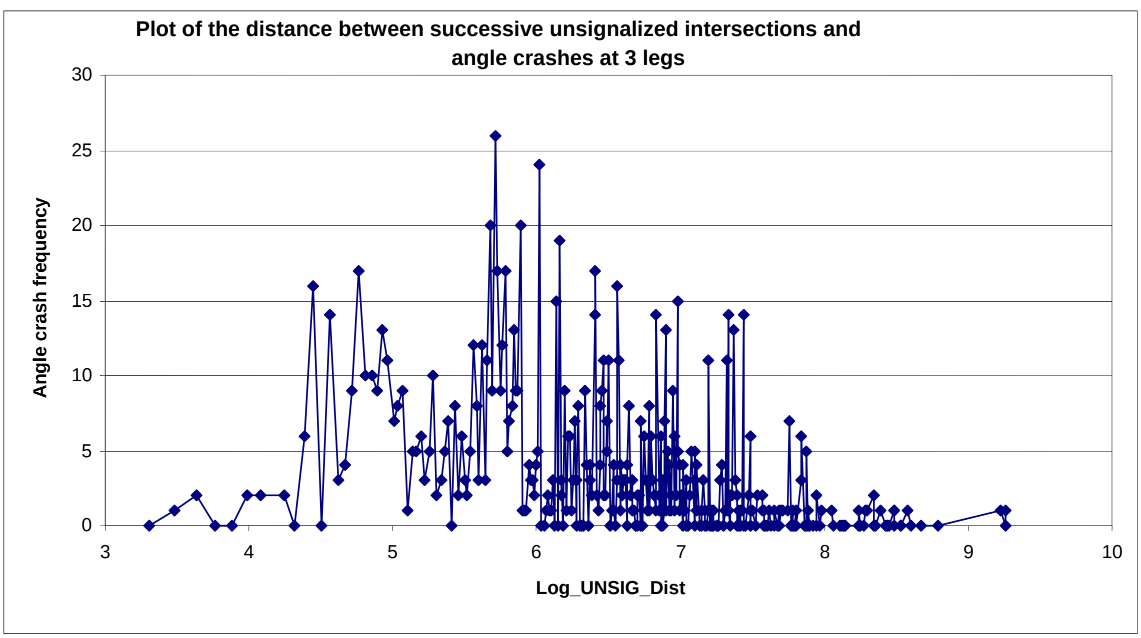


Figure 7-7: Plot of the Distance between Successive Unsignalized Intersections and Angle Crashes at 3-Legged Intersections

From Figure 7-7, it is noticed that there is a fluctuation in the trend, and it is difficult to determine the effect of the distance between successive unsignalized intersections on angle crashes at 3 legs from this plot.

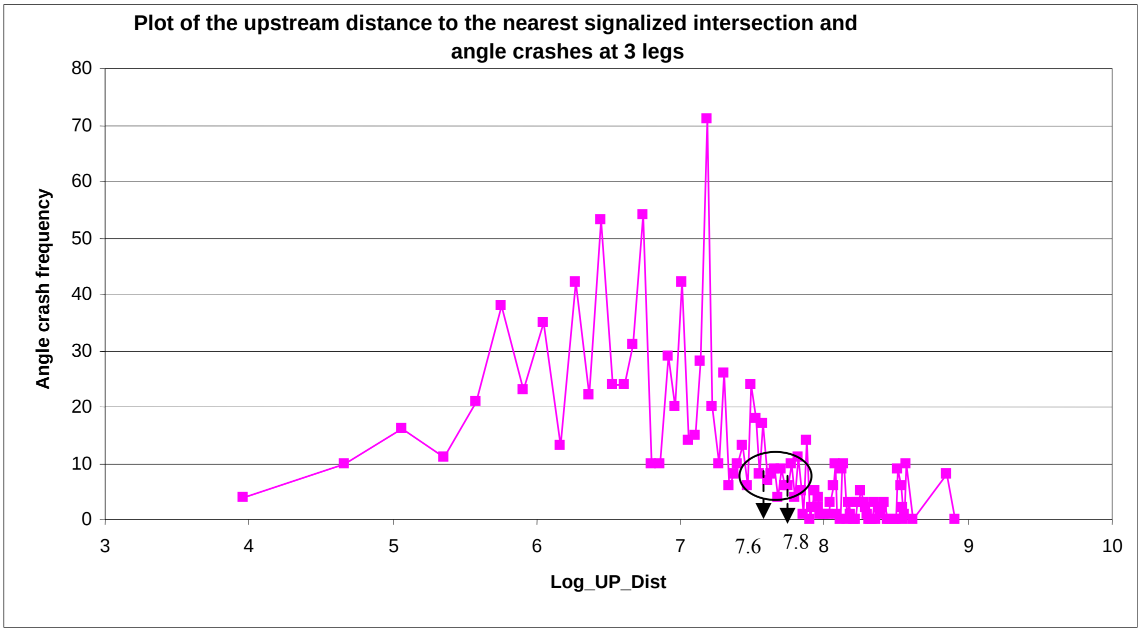


Figure 7-8: Plot of the Upstream Distance to the Nearest Signalized Intersection and Angle Crashes at 3-Legged Intersections

From Figure 7-8, it is noticed that angle crashes at 3 legs tend to decrease after a range of 7.6 to 7.8 for the log upstream distance (i.e., 0.38 to 0.46 miles), and there is no more trend fluctuation after this cut-off range. Also, it can be deduced that angle crashes decrease with relatively large upstream distance at 3 legs.

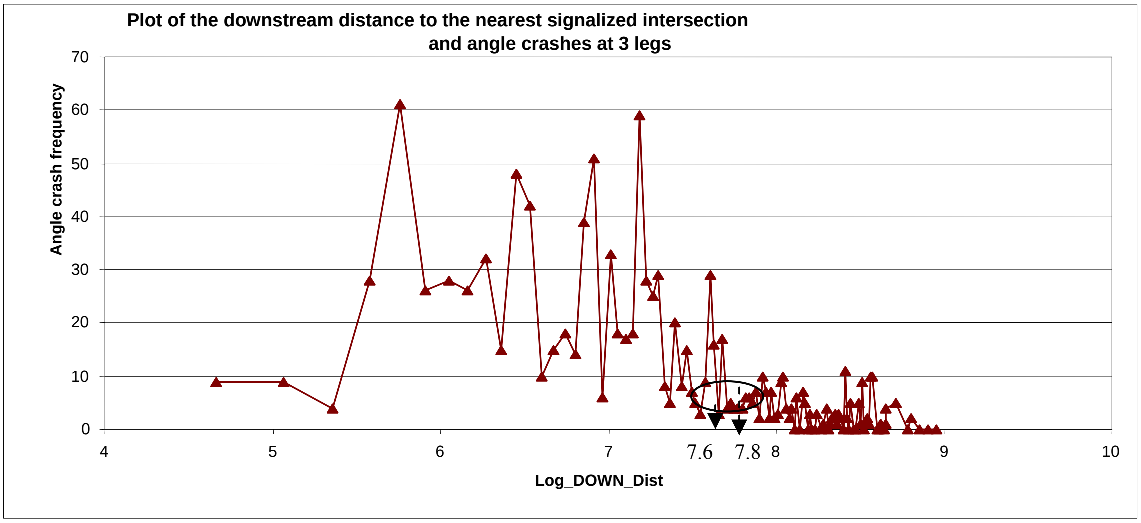


Figure 7-9: Plot of the Downstream Distance to the Nearest Signalized Intersection and Angle Crashes at 3-Legged Intersections

From Figure 7-9, it is noticed that angle crashes at 3 legs tend to decrease after a range of 7.6 to 7.8 for the log downstream distance (i.e., 0.38 to 0.46 miles), and there is no more trend fluctuation after this cut-off range. Also, it can be deduced that angle crashes decrease with relatively large downstream distance at 3 legs.

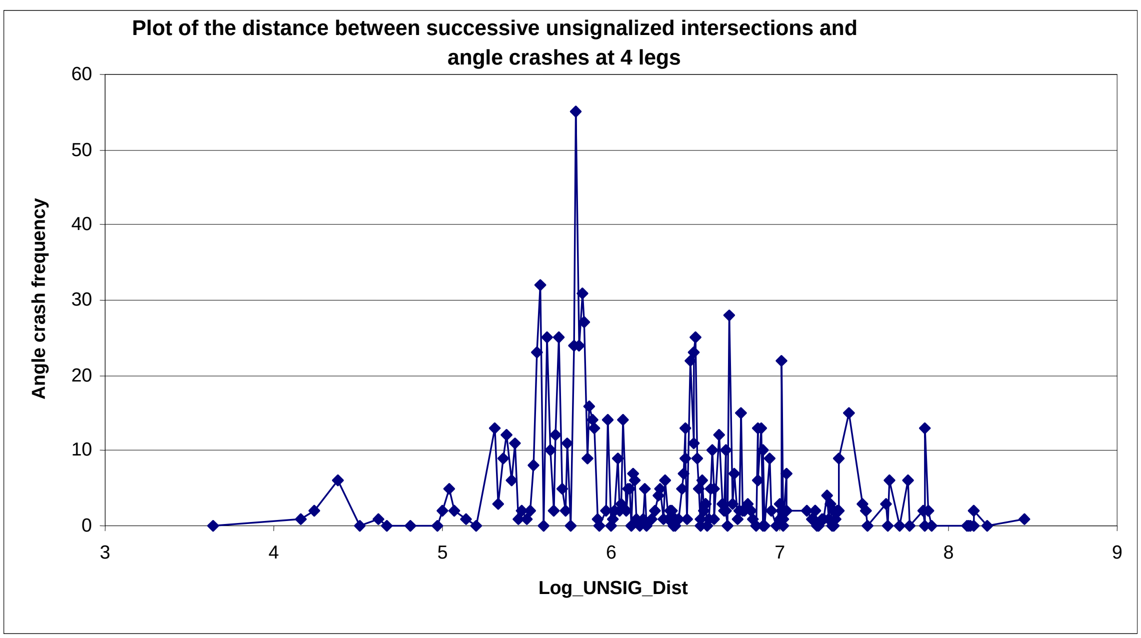


Figure 7-10: Plot of the Distance between Successive Unsignalized Intersections and Angle Crashes at 4-Legged Intersections

From Figure 7-10, it is noticed that there is a fluctuation in the trend, and it is difficult to determine the effect of the distance between successive unsignalized intersections on angle crashes at 4 legs from this plot.

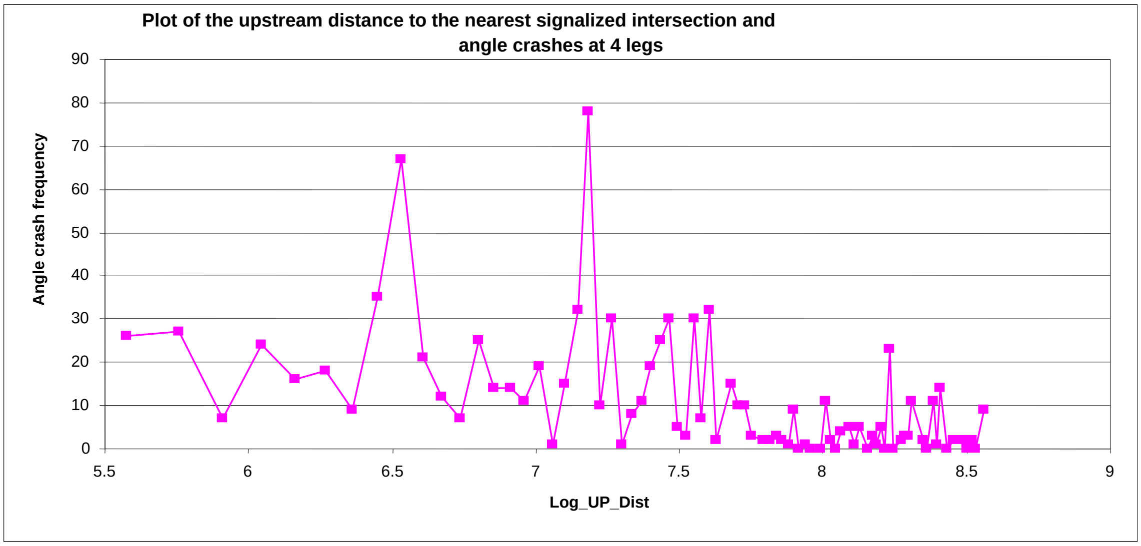


Figure 7-11: Plot of the Upstream Distance to the Nearest Signalized Intersection and Angle Crashes at 4-Legged Intersections

From Figure 7-11, it is noticed that angle crashes at 4 legs tend to decrease with relatively large upstream distance. Roughly, the cut-off range for the clear reduction starts from 7.6 to 7.8 (i.e., 0.38 to 0.46 miles). Also, the second highest angle crash frequency nearly occurs at a log upstream distance of around 6.5 (0.13 miles).

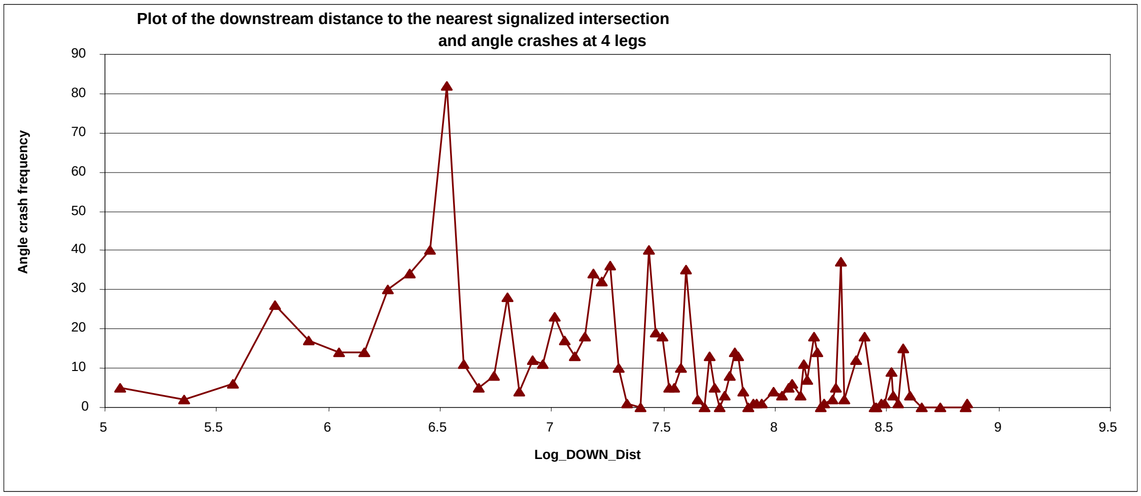


Figure 7-12: Plot of the Downstream Distance to the Nearest Signalized Intersection and Angle Crashes at 4-Legged Intersections

From Figure 7-12, it is noticed that angle crashes at 4 legs tend to decrease with relatively large downstream distance.

7.5 Modeling Rear-end Crash Frequency at 3 and 4-Legged Unsignalized Intersections Using the NB Formulation

After using SAS (2002) with the "proc genmod" procedure, the NB rear-end crash frequency model for both 3 and 4-legged unsignalized intersections is shown in Table 7-4. This table includes the generalized R-square criterion as a goodness-of-fit statistic.

Table 7-3: Variables Description for 3 and 4-Legged Unsignalized Intersections

Variable Description	Variable Levels for 3 Legs	Variable Levels for 4 Legs
Crash location in any of the 6 counties	Orange, Brevard, Hillsborough, Miami-Dade, Leon and Seminole	Orange, Brevard, Hillsborough, Miami-Dade, Leon and Seminole
Existence of stop sign on the minor approach	= 0; if no stop sign exists; = 1; if stop sign exists	= 0; if no stop sign exists; = 1; if only one stop sign exists on one of the minor approaches; = 2; if one stop sign exists on each minor approach
Existence of stop line on the minor approach	= 0; if no stop line exists; = 1; if stop line exists	= 0; if no stop line exists; = 1; if only one stop line exists on one of the minor approaches; = 2; if one stop line exists on each minor approach
Existence of crosswalk on the minor approach	= 0; if no crosswalk exists; = 1; if crosswalk exists	= 0; if no crosswalk exists; = 1; if only one crosswalk exists on one of the minor approaches; = 2; if one crosswalk exists on each minor approach
Existence of crosswalk on the major approach	= 0; if no crosswalk exists; = 1; if one crosswalk exists on one of the major approaches; = 2; if one crosswalk exists on each major approach	= 0; if no crosswalk exists; = 1; if one crosswalk exists on one of the major approaches; = 2; if one crosswalk exists on each major approach
Control type on the minor approach	= 1; if stop sign exists (1-way stop); = 3; if no control exists; = 5; if yield sign exists	= 2; if stop sign exists on each minor approach (2-way stop); = 3; if no control exists on both minor approaches; = 4; if stop sign exists on the first minor approach, and no control on the other
Size of the intersection [a]	= 1; for “1x2”, “1x3” and “1x4” intersections; = 2; for “2x2” and “2x3” intersections; = 3; for “2x4”, “2x5” and “2x6” intersections; = 4; for “2x7” and “2x8” intersections; = 5; for “3x2”, “3x3”, “3x4”, “3x5”, “3x6” and “3x8” intersections; = 6; for “4x2”, “4x4”, “4x6” and “4x8” intersections	= 2; for “2x2” and “2x3” intersections; = 3; for “2x4”, “2x5” and “2x6” intersections; = 4; for “2x7” and “2x8” intersections; = 5; for “3x2”, “3x3”, “3x4”, “3x5”, “3x6” and “3x8” intersections; = 6; for “4x2”, “4x4”, “4x6” and “4x8” intersections
Type of unsignalized intersection	= 1; for access point (driveway) intersections; = 2; for ramp junctions; = 3; for regular intersections; = 4; for intersections close to railroad crossings [b]	= 1; for access point (driveway) intersections; = 3; for regular intersections; = 4; for intersections close to railroad crossings [b]
Number of right turn lanes on the major approach	= 0; if no right turn lane exists; = 1; if one right turn lane exists on only one direction; = 2; if one right turn lane exists on each direction [c]	= 0; if no right turn lane exists; = 1; if one right turn lane exists on only one direction; = 2; if one right turn lane exists on each direction
Number of left turn lanes on the major approach	= 0; if no left turn lane exists; = 1; if one left turn lane exists on only one direction; = 2; if one left turn lane exists on each direction [d]	= 0; if no left turn lane exists; = 1; if one left turn lane exists on only one direction; = 2; if one left turn lane exists on each direction
Number of left turn movements on the minor approach	= 0; if no left turn movement exists; = 1; if one left turn movement exists	= 0; if no left turn movement exists; = 1; if one left turn movement exists on one minor approach only; = 2; if one left turn movement exists on each minor approach

Variable Description	Variable Levels for 3 Legs	Variable Levels for 4 Legs
Land use at the intersection area	= 1; for rural area; = 2; for urban/suburban areas	= 1; for rural area; = 2; for urban/suburban areas
Median type on the major approach	= 1; for open median; = 2; for directional median; = 3; for closed median; = 4; for two-way left turn lane; = 5; for markings;= 6; for undivided median; = 7; for mixed median [e]	= 1; for open median; = 4; for two-way left turn lane; = 6; for undivided median
Median type on the minor approach	= 1; for undivided median, two-way left turn lane and markings; = 2; for any type of divided median	= 1; for undivided median, two-way left turn lane and markings; = 2; for any type of divided median
Skewness level	= 1; if skewness angle <= 75 degrees; = 2; if skewness angle > 75 degrees	= 1; if skewness angle <= 75 degrees; = 2; if skewness angle > 75 degrees
Posted speed limit on the major road	= 1; if posted speed limit < 45 mph; = 2; if posted speed limit >= 45 mph	= 1; if posted speed limit < 45 mph; = 2; if posted speed limit >= 45 mph
Number of through lanes on the minor approach [f]	= 1; if one through lane exists; = 2; if two through lanes exist; = 3; if more than two through lanes exist	= 2; if two through lanes exist; = 3; if more than two through lanes exist
Natural logarithm of the section annual average daily traffic on the major road; Natural logarithm of the upstream and downstream distances (in feet) to the nearest signalized intersection from the unsignalized intersection of interest; Left shoulder width near the median on the major road (in feet); Right shoulder width on the major road (in feet); Percentage of trucks on the major road; Natural logarithm of the distance between 2 successive unsignalized intersections [g]		

[a] The first number represents total number of approach lanes for the minor approach, and the second number represents total number of through lanes for the major approach

[b] Railroad crossing can exist upstream or downstream the intersection of interest

[c] One right turn lane on each major road direction for 3-legged unsignalized intersections: Two close unsignalized intersections, one on each side of the roadway, and each has one right turn lane. The extended right turn lane of the first is in the influence area of the second.

[d] One left turn lane on each major road direction for 3-legged unsignalized intersections: One of these left turn lanes is only used as U-turn.

[e] Mixed median is directional from one side, and closed from the other side (i.e., allows access from one side only)

[f] Surrogate measure for AADT on the minor approach

[g] Continuous variables

Table 7-4: Rear-end Crash Frequency Model at 3 and 4-Legged Unsignalized Intersections

	Three-Legged Model		Four-Legged Model	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
Intercept	-6.6300 (0.9229)	<0.0001	-12.7601 (1.7815)	<0.0001
Natural logarithm of AADT on the major road	0.5830 (0.0811)	<0.0001	1.2288 (0.1519)	<0.0001
Natural logarithm of the upstream distance to the nearest signalized intersection	-0.1376 (0.0406)	0.0007	N/S [b]	
Natural logarithm of the downstream distance to the nearest signalized intersection	N/S		-0.1244 (0.0681)	0.0678
Natural logarithm of the distance between 2 successive unsignalized intersections	N/S		0.0966 (0.0552)	0.0800
Unsignalized intersections in urban/suburban areas	0.6919 (0.2399)	0.0039	N/S	
Unsignalized intersections on rural areas	--- [c]		N/S	
Posted speed limit on major road >= 45 mph	0.2183 (0.0948)	0.0212	N/S	
Posted speed limit on major road < 45 mph	--- [c]		N/S	
Divided median on the minor approach	-0.2308 (0.1431)	0.1068	N/S	
Undivided median on the minor approach	--- [c]		N/S	
Undivided median exists on the major approach	N/S		0.4209 (0.1638)	0.0102
Two-way left turn lane exists on the major approach	N/S		0.3267 (0.1677)	0.0514
Open median exists on the major approach	N/S		--- [c]	
Left shoulder width near the median on the major road	0.0831 (0.0338)	0.0138	N/S	
Unsignalized intersections close to railroad crossings	0.5062 (0.4247)	0.2333	N/S	
Regular unsignalized intersections	0.4313 (0.1044)	<0.0001	N/S	
Unsignalized ramp junctions	0.6043 (0.2414)	0.0123	N/A [d]	
Access point unsignalized intersections (Driveways)	--- [c]		N/S	
One right turn lane exists on each major road direction	-0.2822 (0.2843)	0.3208	N/S	
One right turn lane exists on only one major road direction	0.1932 (0.1113)	0.0826	N/S	
No right turn lane exists on the major approach	--- [c]		N/S	

	Three-Legged Model		Four-Legged Model	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
Dummy variable for Seminole County	0.2595 (0.1856)	0.1622	0.2199 (0.2681)	0.4121
Dummy variable for Orange County	0.3032 (0.1587)	0.0561	0.1694 (0.2596)	0.5141
Dummy variable for Miami-Dade County	0.7018 (0.1597)	<0.0001	0.6764 (0.2596)	0.0092
Dummy variable for Leon County	1.2358 (0.1550)	<0.0001	0.8147 (0.2730)	0.0028
Dummy variable for Hillsborough County	0.7221 (0.1545)	<0.0001	1.1996 (0.2390)	<0.0001
Dummy variable for Brevard County	--- [c]		--- [c]	
Dispersion	0.9376 (0.0828)		0.4463 (0.0870)	
Generalized R-square [e]	0.178		0.313	

[a] Standard error in parentheses
[b] N/S means not significant
[c] Base case
[d] N/A means not applicable
[e] Generalized R-square = 1 – (Residual deviance/Null deviance). The residual deviance is equivalent to the residual sum of squares in linear regression, and the null deviance is equivalent to the total sum of squares (Zuur et al., 2007)

7.5.1 Three-Legged Model Interpretation

From Table 7-4, there is a statistical significant increase in rear-end crashes with the increase in the logarithm of AADT, as rear-end crashes always occur at high traffic volumes. This is consistent with that concluded by Wang and Abdel-Aty (2006), who found that the logarithm of AADT per lane increases rear-end crash frequency at signalized intersections.

There is a reduction in rear-end crashes with the increase in the logarithm of the upstream distance to the nearest signalized intersection. This is expected since there is enough spacing for vehicles to accommodate high AADT and frequent stops in rush hours, and thus rear-end crash risk decreases.

There is an increase in rear-end crashes in urban/suburban areas, when compared to rural areas. This is anticipated since there are higher volume and more intersections in urban (and suburban) areas, hence a higher rear-end crash risk.

Compared to access points, regular unsignalized intersections have longer stretches on the minor approach, thus rear-end crashes increase, and as shown in Table 7-4, the increase is statistically significant. As expected, rear-end crashes are high at unsignalized intersections next to railroads due to sudden unexpected stops that can propagate to intersections nearby. Also, ramp junctions have high probability of rear-end crashes due to sudden stops in merging areas.

The existence of one right turn lane from one major direction only increases rear-end crashes, compared to no right turn lanes. This shows that separating right and through maneuvers near unsignalized intersections might not be beneficial in some cases.

The highest significant increase in rear-end crashes occurs at Leon County (when compared to Brevard County). This might be explained that Leon County has the capital of Florida, thus having more central governmental agencies which generate more trips. It is mostly rural, and that is why it might have more unsignalized intersections. It can be also noticed that compared to the eastern part of Florida (represented by Brevard County), the highest increase in rear-end crashes occurs in the northern part (represented by Leon County), followed by the western part (represented by Hillsborough County), then the southern part (represented by Miami-Dade County), and finally the central part (represented by Orange and Seminole Counties).

7.5.2 Four-Legged Model Interpretation

From Table 7-4, as found in the 3-legged model, increasing the logarithm of AADT significantly increases rear-end crashes.

There is a reduction in rear-end crashes with the increase in the logarithm of the downstream distance to the nearest signalized intersection. The estimated coefficient is statistically significant at the 90% confidence.

The finding that there is an increase in rear-end crashes with the increase in the logarithm of the distance between successive unsignalized intersections should not be deceiving, as this could be masked by the variable “logarithm of the downstream distance to the nearest signalized intersection”. The relatively short downstream distance can cause a backward shockwave, resulting in turbulence at nearby unsignalized intersections, thus rear-end crash risk can be high.

Two-way left turn lanes as well as undivided medians on the major approach significantly increase rear-end crashes, when compared to having an open median. This shows the hazardous effect of having two-way left turn lanes for 4-legged intersections. This conforms to the study done by Phillips (2004) who found that two-way left turn lanes experience more crashes than raised medians.

Similar to the 3-legged model, the central part in Florida (represented by Orange and Seminole Counties) experience the least rear-end crash increase when compared to the eastern part (represented by Brevard County).

To show the result of the MARS model and the coefficients of different basis functions, the MARS model for 4-legged rear-end crash frequency is presented in Table 7-5.

Table 7-5: Rear-end Crash Frequency Model at 4-Legged Unsignalized Intersections Using MARS

Basis Function	Basis Function Description	Estimate *	P-value
Intercept	Intercept	23.0519 (7.2920)	0.0016
Log_AADT	Natural logarithm of AADT on the major road	-2.2023 (0.6749)	0.0012
Hills_County	Hillsborough County	-24.7873 (4.9271)	<0.0001
Undivided_Median	Undivided median on the major approach	-1.1506 (0.7342)	0.1179
Hills_County * Undivided_Median	An interaction term	1.3625 (0.5361)	0.0114
Log_AADT * Hills_County	An interaction term	2.4150 (0.4542)	<0.0001
$(\text{Log\_AADT} - 10.27505)_+$	A truncated power basis function for "Log_AADT" at "10.27505"	2.8190 (0.6533)	<0.0001
	Generalized R-square	0.55	

[*] Standard error in parentheses

For the shown MARS model, it is noticed that MARS selects only those significant levels of categorical variables, and it does not show all possible levels as the NB model. Also, it is noticed that there are two interaction terms. Thus, the two variables in each interaction term should be interpreted together. The first interaction term is between Hillsborough County and undivided median, while the second is between the logarithm of AADT and Hillsborough County. The equation representing the first interaction term is: "-24.7873 * Hills_County – 1.1506 * Undivided_Median + 1.3625 * Hills_County * Undivided_Median".

The interpretation for the shown equation is as follows: for the existence of undivided median on the major approach (i.e., Undivided_Median = 1), the equation becomes "(-24.7873 + 1.3625) * Hills_County – 1.1506", which can be simplified as "-23.4248 * Hills_County – 1.1506". Thus, the individual coefficient of "Hills_County" is "-23.4248". This means that, for the existence of undivided median on the major approach, the frequency of rear-end crashes decreases for Hillsborough County, when compared to the other five counties used in the analysis.

The equation representing the second interaction term is: "-2.2023 * Log_AADT – 24.7873 * Hills_County + 2.4150 * Log_AADT * Hills_County + 2.8190 * $(\text{Log\_AADT} - 10.27505)_+$". The interpretation for the shown equation is as follows: for Hillsborough County (i.e., Hills_County = 1) and Log_AADT > 10.27505 (i.e., AADT > 29,000), the equation becomes "(-2.2023 + 2.4150) * Log_AADT + 2.8190 * (Log_AADT – 10.27505) – 24.7873", which can be simplified as "3.0317 * Log_AADT – 53.7527". Thus, the individual coefficient of "Log_AADT" is "3.0317". This means that, for Hillsborough County, the frequency of rear-end crashes increases as long as AADT is greater than 29,000 vehicles per day.

From Table 7-5, it is noted that there is a nonlinear performance for the continuous variable "Log_AADT", as shown in its truncated basis function at "10.27505". In order to better understand the nonlinear function of "Log_AADT", a plot for its basis function is shown in Figure 7-13. The basis function "f(Log_AADT)" according to the fitted MARS model is "-2.2023 * Log_AADT + 2.8190 * (Log_AADT – 10.27505)$_+$".

As previously shown in Equation (7.3), the term "(Log_AADT – 10.27505)$_+$" equals "Log_AADT – 10.27505" when Log_AADT > 10.27505, and zero, otherwise. By this, the plot in Figure 7-13 can be formed, where the basis function "f(Log_AADT)" is plotted against all the values of "Log_AADT". From this figure, it can be noticed that there is only one knot (10.27505), when there is a sudden break in the straight line. This demonstrates the nonlinear performance of the variable "Log_AADT" with rear-end crash frequency.

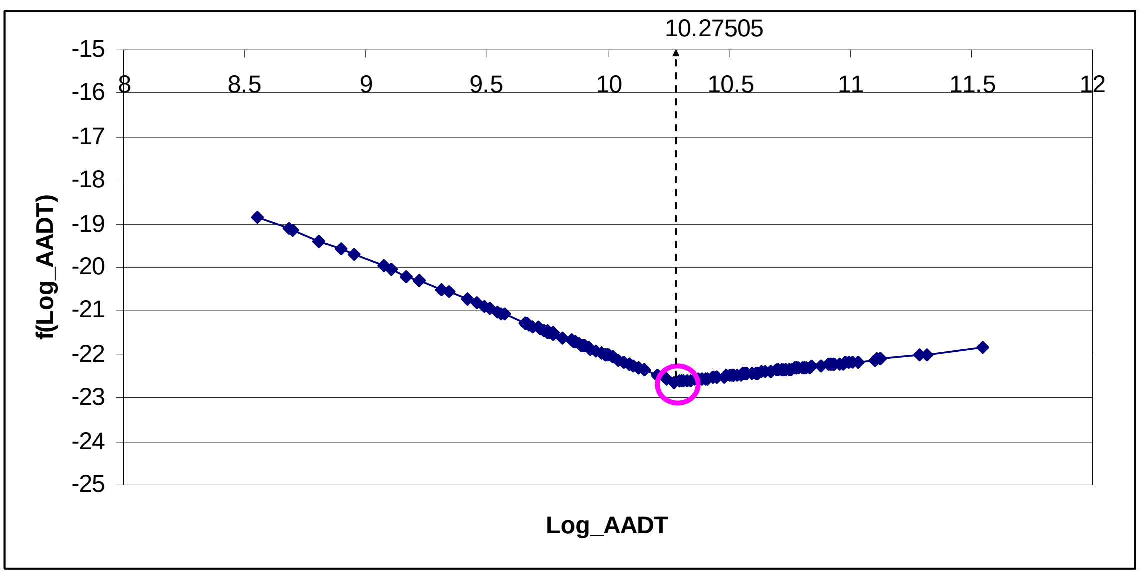


Figure 7-13: Plot of the Basis Function for "Log_AADT"

7.6 Comparing MARS and NB Models

For the first application of MARS in this study, a comparison between the two fitted MARS models and the corresponding NB models, while treating the response in each as a discrete one (i.e., crash frequency), is shown in Table 7-6. The R package (51) was utilized to estimate the MARS models via the library "polspline". The MARS models were generated using the default *GCV* value "3" in R. From this table, it is noticed that the MSPE for MARS in the 3-legged model is slightly lower than the corresponding NB model, and the MAD values are the same. As for the 4-legged model, the MSPE value for MARS is lower than the NB model, while the MAD is higher. This indicates that the MARS technique has a promising prediction capability. Also, the generalized R-square is much higher for the MARS models.

Table 7-6: Comparison between the Fitted MARS and NB Models in terms of Prediction and Fitting

		Rear-end three-legged model		Rear-end four-legged model	
		MARS	**NB**	**MARS**	**NB**
Prediction	**MAD ***	1.01	1.01	0.96	0.82
	MSPE *	2.54	2.55	1.98	2.62
Fitting	**Generalized R-square**	0.42	0.17	0.55	0.31

* MAD and MSPE values are normalized by the average of the response variable

7.7 Examining Fitting MARS Model with Continuous Response

To examine the higher prediction capability of MARS while dealing with continuous responses (Friedman, 1991), the two MARS models using the same important NB covariates were fitted while considering the response as the crash frequency normalized by the natural logarithm of AADT. It is worth mentioning that the natural logarithm of AADT was only used as the denominator of the response variable, i.e., not an explanatory variable as in the previous case. A default *GCV* value of "3" was used while fitting the models. The assessment criteria for the generated MARS models are shown in Table 7-7.

By comparing the MAD and MSPE values from this table with those from the previously fitted MARS models in Table 7-6, it is noticed that the MAD and MSPE values shown in Table 7-7 are lower. The estimated MSPE values are very close to "zero", indicating a very high prediction capability. This demonstrates the higher prediction performance of MARS while dealing with continuous responses. Also, the generalized R-square values in Tables 7-6 and 7-7 are very close.

Table 7-7: Prediction and Fitting Performance of the Two MARS Models Using a Continuous Response Formulation

		Rear-end three-legged model	Rear-end four-legged model
		MARS [1]	MARS [1]
Prediction	MAD [2]	1.07	0.95
	MSPE [2]	0.27	0.31
Fitting	Generalized R-square	0.39	0.46

[1] Response is the crash frequency normalized by the natural logarithm of AADT

[2] MAD and MSPE values are normalized by the average of the response variable

7.8 Using MARS in Conjunction with Random Forest

Since the MARS technique showed similar efficient prediction performance to the NB framework (with higher prediction capability while dealing with continuous responses), an additional effort to examine screening all possible covariates before fitting a MARS model, was attempted. This leads to utilizing the random forest technique (Breiman, 2001) before fitting a MARS model for variable screening and ranking important covariates. Using the R package, all possible covariates in the two attempted models were screened via the library "randomForest". The random forest technique was performed with 50 trees grown in the two training datasets. To examine whether this number can lead to stable results, the plot of the OOB error rate against different tree numbers for the three-legged training dataset (just as an example for illustration purposes) is shown in Figure 7-14. From this figure, it can be noticed that after 38 trees, the OOB error rate starts to stabilize. Hence, the attempted number of trees "50" was deemed large enough to obtain stable results. This was also concluded for the four-legged training dataset.

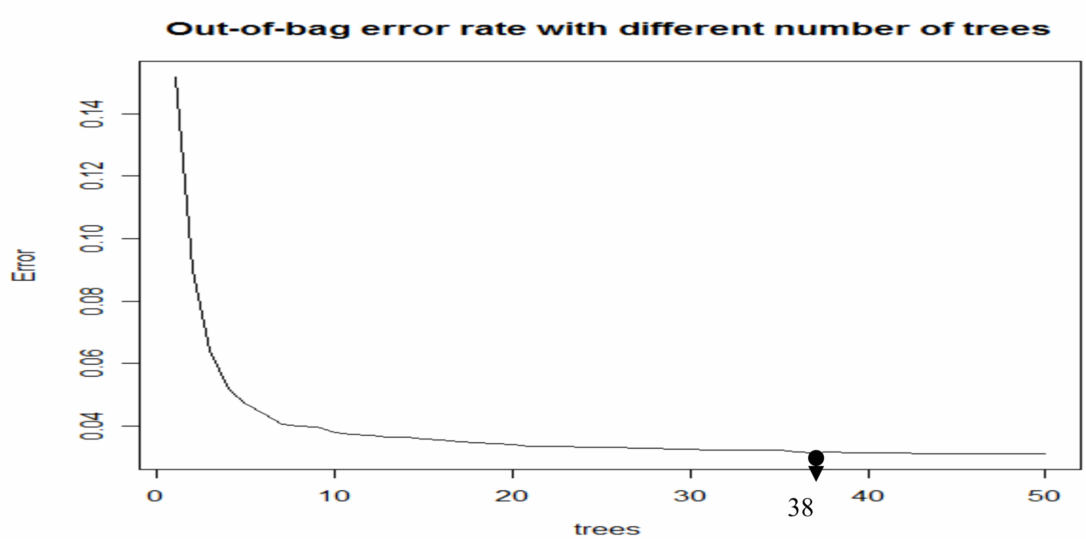


Figure 7-14: Plot of the OOB Error Rate against Different Number of Trees

Figure 7-15 shows the purity values for every covariate. The highest variable importance ranking is the percentage of trucks, followed by the natural logarithm of the distance between two unsignalized intersections, etc., until ending up with the existence of crosswalk on the major approach. The resulted variable importance ranking demonstrates the significant effect of the spatial covariates on rear-end crashes, with the distance between successive unsignalized intersections being the most significant. The second significant spatial variable is the upstream distance to the nearest signalized intersection, followed by the downstream distance. The upstream distance was also found significant in the fitted three-legged NB model. To screen the covariates, a cut-off purity value of "1.5" was used. This leads to selecting eight covariates (labeled from "1" till "8" in Figure 7-15). Those eight covariates were then fitted using MARS, with the response being the crash frequency normalized by the natural logarithm of AADT, since it revealed the best promising prediction performance.

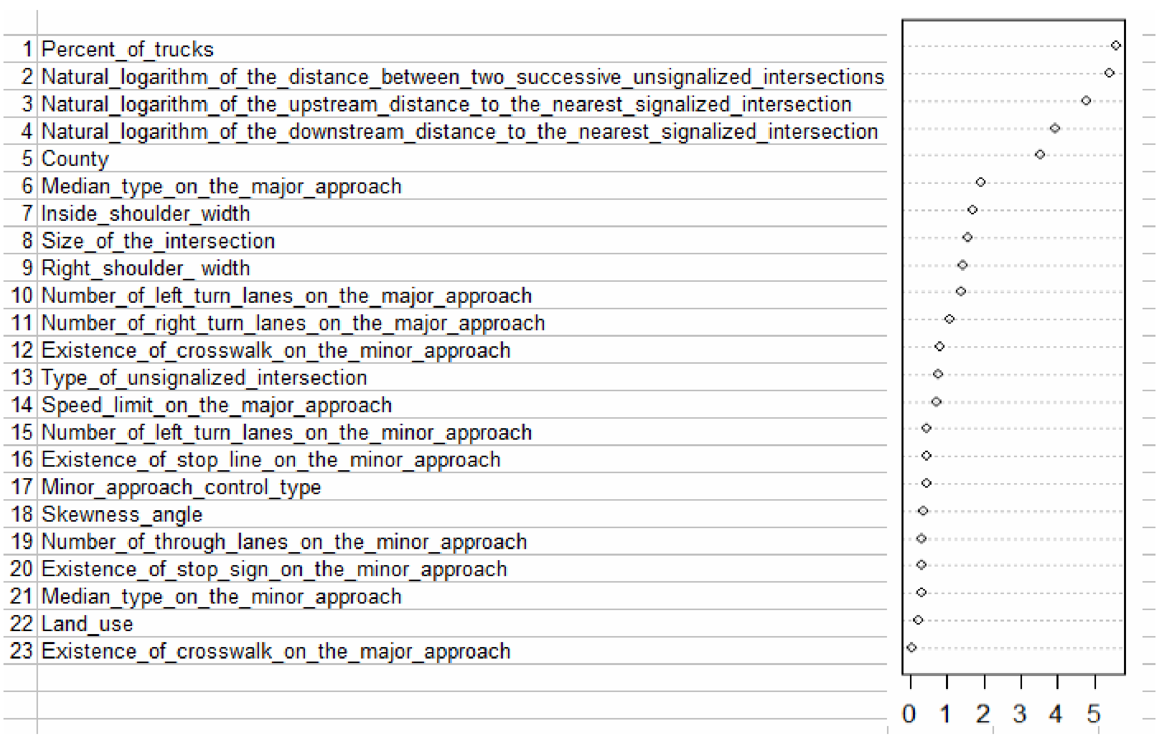


Figure 7-15: Variable Importance Ranking Using Node Purity Measure

The final fitted MARS model using the eight selected covariates at 3-legged unsignalized intersections is presented in Table 7-8, where the response is the crash frequency normalized by the logarithm of AADT. From this table, it is noticed that the negative coefficient for the logarithm of the upstream distance concurs with that deduced from Table 7-4.

Table 7-8: MARS Model at 3-Legged Unsignalized Intersections after Screening the Variables Using Random Forest

Basis Function	Basis Function Description	Estimate *	P-value
Intercept	Intercept	0.1360 (0.0521)	0.0091
Log_Up_Dist	Natural logarithm of the upstream distance to the nearest signalized intersection	-0.0263 (0.0045)	<0.0001
Leon_County	Leon County	0.0938 (0.0141)	<0.0001
Miami_County	Miami-Dade County	0.0017 (0.0175)	0.9196
Hills_County	Hillsborough County	0.0421 (0.0134)	0.0017
ISLDWDTH	Inside shoulder width (in feet)	-0.0737 (0.0139)	<0.0001
ISLDWDTH * Miami_County	An interaction term	0.0754 (0.0115)	<0.0001
	Generalized R-square	0.35	

[*] Standard error in parentheses

To assess whether there is an improvement over the two generated MARS models using the important variables from the NB model, the same evaluation criteria were used, as shown in Table 7-9. Comparing the MAD and MSPE values in Tables 7-7 and 7-9, it is noticed that there is always a reduction (even if it is small) in the MAD and MSPE values in Table 7-9, hence higher prediction accuracy. The resulted generalized R-square values are relatively high, hence encouraging model fit. This demonstrates that using MARS after screening the variables using random forest is quite promising.

Table 7-9: Prediction and Fitting Assessment Criteria for the Two MARS Models after Screening the Variables Using Random Forest

		Rear-end three-legged model	Rear-end four-legged model
		MARS	**MARS**
Prediction	**MAD ***	1.03	0.87
	MSPE *	0.25	0.28
Fitting	**Generalized R-square**	0.35	0.50

* MAD and MSPE values are normalized by the average of the response variable

7.9 Predicting Angle Crashes Using the MARS Technique

After exploring rear-end crashes in the previous sections of this chapter using MARS, another frequent crash type at unsignalized intersections (which is angle crash) was investigated in the following sections. The same unsignalized intersections sample was also used (2475 intersections).

7.9.1 Modeling Angle Crash Frequency at 3 and 4-Legged Unsignalized Intersections Using the NB Technique

The NB angle crash frequency model for both 3 and 4-legged unsignalized intersections is shown in Table 7-10. This table includes the generalized R-square criterion as a goodness-of-fit statistic.

Table 7-10: Angle Crash Frequency Model at 3 and 4-Legged Unsignalized Intersections

	Three-Legged Model		Four-Legged Model	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
Intercept	-7.1703 (1.3369)	<0.0001	-9.0650 (1.6736)	<0.0001
Natural logarithm of AADT on the major road	0.6741 (0.1120)	<0.0001	0.7151 (0.1662)	<0.0001
Natural logarithm of the upstream distance to the nearest signalized intersection	-0.0878 (0.0493)	0.0747	N/S [b]	
Natural logarithm of the distance between 2 successive unsignalized intersections	N/S		0.1200 (0.0604)	0.0471
Percentage of trucks on the major road	0.0272 (0.0168)	0.1049	N/S	
Unsignalized intersections close to railroad crossings	0.4368 (0.5317)	0.4114	1.0322 (0.3608)	0.0042
Regular unsignalized intersections	0.4069 (0.1193)	0.0007	0.4959 (0.1341)	0.0002
Unsignalized ramp junctions	0.5238 (0.3137)	0.0949	N/A [d]	
Access point unsignalized intersections (Driveways)	--- [c]		--- [c]	
One left turn lane exists on each major road direction	0.3495 (0.1754)	0.0463	0.4647 (0.2067)	0.0246
One left turn lane exists on only one major road direction	0.1642 (0.1324)	0.2149	0.6440 (0.2420)	0.0078
No left turn lane exists on the major approach	--- [c]		--- [c]	
One right turn lane exists on each major road direction	N/S		0.5842 (0.2678)	0.0292
One right turn lane exists on only one major road direction	N/S		0.0869 (0.2149)	0.6860
No right turn lane exists on the major approach	N/S		--- [c]	
One left turn exists on any of the minor approaches	-0.6274 (0.2112)	0.0030	N/S	
No left turn lane exists on the minor approach	--- [c]		N/S	
Mixed median exists on the major approach	-0.7215 (0.2795)	0.0099	N/A	
Undivided median exists on the major approach	-0.4342 (0.1504)	0.0039	0.3488 (0.2144)	0.1038
Marking exists on the major approach	-0.3797 (0.3128)	0.2248	N/A	
Two-way left turn lane exists on the major approach	-0.3779 (0.1891)	0.0457	0.0059 (0.1828)	0.9743
Closed median exists on the major approach	-0.5805 (0.2529)	0.0217	N/A	

	Three-Legged Model		Four-Legged Model	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
Directional median exists on the major approach	-0.6773 (0.2874)	0.0184	N/A	
Open median exists on the major approach	--- [c]		--- [c]	
Posted speed limit on major road >= 45 mph	0.2201 (0.1156)	0.0568	N/S	
Posted speed limit on major road < 45 mph	--- [c]		N/S	
"4x2", "4x4", "4x6" and "4x8" intersections	N/S		0.0443 (0.5968)	0.9408
"3x2", "3x3", "3x4", "3x5", "3x6" and "3x8" intersections	N/S		0.9531 (0.3527)	0.0069
"2x7" and "2x8" intersections	N/S		0.8813 (0.7924)	0.2660
"2x4", "2x5" and "2x6" intersections	N/S		0.2661 (0.2806)	0.3430
"2x2" and "2x3" intersections	N/S		--- [c]	
Dummy variable for Seminole County	0.1889 (0.2394)	0.4302	-0.0427 (0.2795)	0.8786
Dummy variable for Orange County	0.6930 (0.1911)	0.0003	0.0604 (0.2669)	0.8211
Dummy variable for Miami-Dade County	0.7522 (0.2104)	0.0004	1.0695 (0.2575)	<0.0001
Dummy variable for Leon County	0.8489 (0.1985)	<0.0001	0.5336 (0.2786)	0.0555
Dummy variable for Hillsborough County	1.0528 (0.1988)	<0.0001	1.1046 (0.2304)	<0.0001
Dummy variable for Brevard County	--- [c]		--- [c]	
Dispersion	1.1442 (0.1113)		0.8379 (0.1043)	
Generalized R-square	0.19		0.31	

[a] Standard error in parentheses   [b] N/S means not significant   [c] Base case   [d] N/A means not applicable

7.9.1.1 Three-Legged Model Interpretation

From Table 7-10, there is a statistical significant increase in angle crashes with the increase in the logarithm of AADT (which inherently means an increase in traffic volume). As AADT relatively increases, vehicles coming from the minor approach find it difficult to cross the major road due to congestion, hence angle crash risk might increase.

There is a reduction in angle crashes with the increase in the logarithm of the upstream distance to the nearest signalized intersection. This is expected since there is enough spacing for vehicles on the minor approach to cross the major road, and thus angle crash risk decreases.

There is an increase in angle crashes with the increase in truck percentage. This is anticipated due to possible vision blockage caused by trucks, thus angle crash risk could increase .

Compared to access points, regular unsignalized intersections have longer stretches on the minor approach, thus angle crashes increase, and as shown in Table 7-10, the increase is statistically significant. Also, ramp junctions have high angle crashes due to traffic turbulence in merging areas.

The existence of one left turn lane on each major road direction significantly increases angle crashes, compared to no left turn lanes. This is due to a high possible conflict pattern between vehicles crossing from both minor and major approaches.

Compared to open medians, undivided medians have the least significant decrease in angle crashes due to the reduction in conflict points.

Compared to the eastern part of Florida (represented by Brevard County), the highest increase in angle crashes occurs in the western part (represented by Hillsborough

County), followed by the northern part (represented by Leon County), then the southern part (represented by Miami-Dade County), and finally the central part (represented by Orange and Seminole Counties).

7.9.1.2 Four-Legged Model Interpretation

From Table 7-10, as found in the 3-legged model, increasing the logarithm of AADT significantly increases angle crashes.

The finding that there is an increase in angle crashes with the increase in the logarithm of the distance between successive unsignalized intersections could be masked by the variable “logarithm of the downstream distance to the nearest signalized intersection”. The relatively short downstream distance can cause a backward shockwave, resulting in turbulence at nearby unsignalized intersections, thus angle crash risk could be high.

Similar to the 3-legged model, compared to access points, regular unsignalized intersections as well as unsignalized intersections next to railroads experience a significant increase in angle crashes.

The existence of one left and right turn lane on each major road direction significantly increases angle crashes, compared to no left and right turn lanes, respectively. Once more, this is due to a high possible conflict pattern between vehicles crossing from both minor and major approaches.

Two-way left turn lanes as well as undivided medians on the major approach increase angle crashes, when compared to open medians, and the increase is statistically significant for undivided medians. This shows the hazardous effect of having two-way left turn lanes for 4-legged intersections. This conforms to the study done by Phillips

(2004) who found that two-way left turn lanes experience more crashes than raised medians.

As the size of intersections increase, angle crashes increase. This is anticipated due to the higher angle crash risk maneuver at relatively bigger intersections. Intersections with 3 total lanes on the minor approach have the only significant increase.

Similar to the 3-legged model, the highest increase in angle crashes occurs in the western part (represented by Hillsborough County), followed by the northern part (represented by Leon County), then the southern part (represented by Miami-Dade County) when compared to the eastern part (represented by Brevard County). The central part (represented by Orange and Seminole Counties) has no significant effect on angle crashes.

To show the result of the MARS model and the coefficients of different basis functions, the MARS model for 4-legged angle crash frequency is presented in Table 7-11.

Table 7-11: Angle Crash Frequency Model at 4-Legged Unsignalized Intersections Using MARS

Basis Function	Basis Function Description	Estimate *	P-value
Intercept	Intercept	2.1314 (5.3912)	0.6928
Log_AADT	Natural logarithm of AADT on the major road	0.6831 (0.5134)	0.1840
Hills_County	Hillsborough County	-5.5343 (1.9559)	0.0049
Orange_County	Orange County	-1.4406 (0.4560)	0.0017
Size_Lanes_3	“3x2”, “3x3”, “3x4”, “3x5”, “3x6” and “3x8” intersections	-6.3123 (2.2146)	0.0046
Acc_Point	Access points	-1.3737 (0.3382)	<0.0001
Hills_County * Size_Lanes_3	An interaction term	7.4050 (1.8259)	<0.0001
$(\text{Log\_AADT} - 10.389)_+$	A truncated power basis function for “Log_AADT” at “10.389”	6.4480 (1.3054)	<0.0001
$(\text{Log\_AADT} - 11.112)_+$	A truncated power basis function for “Log_AADT” at “11.112”	-25.5042 (7.3651)	0.0005
	Generalized R-square	0.52	

* Standard error in parentheses

From Table 7-11, it is noticed that MARS selects only those significant levels of categorical variables, and it does not show all possible levels as the NB model. Also, it is noticed that there is an interaction term. Hence, the two variables forming the interaction term should be interpreted together. The interaction term is between Hillsborough County and unsignalized intersections with three total lanes on the minor approach. The equation representing this interaction term is: “-5.5343 * Hills_County – 6.3123 * Size_Lanes_3 + 7.4050 * Hills_County * Size_Lanes_3”.

The interpretation for the formed equation is described as follows: for the case of Hillsborough (i.e., Hills_County = 1), the equation becomes “(-6.3123 + 7.4050) *

Size_Lanes_3 – 5.5343”, which can be simplified as “1.0927 * Size_Lanes_3 – 5.5343”. Thus, the individual coefficient of “Size_Lanes_3” is “1.0927”. This means that, in Hillsborough County, the angle crash frequency increases for intersections with three total lanes on the minor approach, when compared to other intersection sizes used in the analysis.

Also, from Table 7-11, it is noted that there is a nonlinear performance for the continuous variable “Log_AADT”, as shown in its truncated basis function at “10.389” and “11.112”. In order to understand the nonlinear function of “Log_AADT”, a plot for its basis function is shown in Figure 7-16. The basis function “f(Log_AADT)” according to the fitted MARS model is “0.6831 * Log_AADT + 6.4480 * $(\text{Log\_AADT} – 10.389)_+$ - 25.5042 * $(\text{Log\_AADT} – 11.112)_+$”.

As previously shown in Equation (7.3), the term “$(\text{Log\_AADT} – 10.389)_+$” equals “Log_AADT – 10.389” when Log_AADT > 10.389, and zero, otherwise. The same also applies for “$(\text{Log\_AADT} – 11.112)_+$”. By this, the plot in Figure 7-16 can be formed, where the basis function “f(Log_AADT)” is plotted against all the values of “Log_AADT”. From this figure, it can be noticed that there are two knots, “10.389 and 11.112”, when there is a sudden break in the straight line. This demonstrates the nonlinear performance of the variable “Log_AADT” with angle crash frequency.

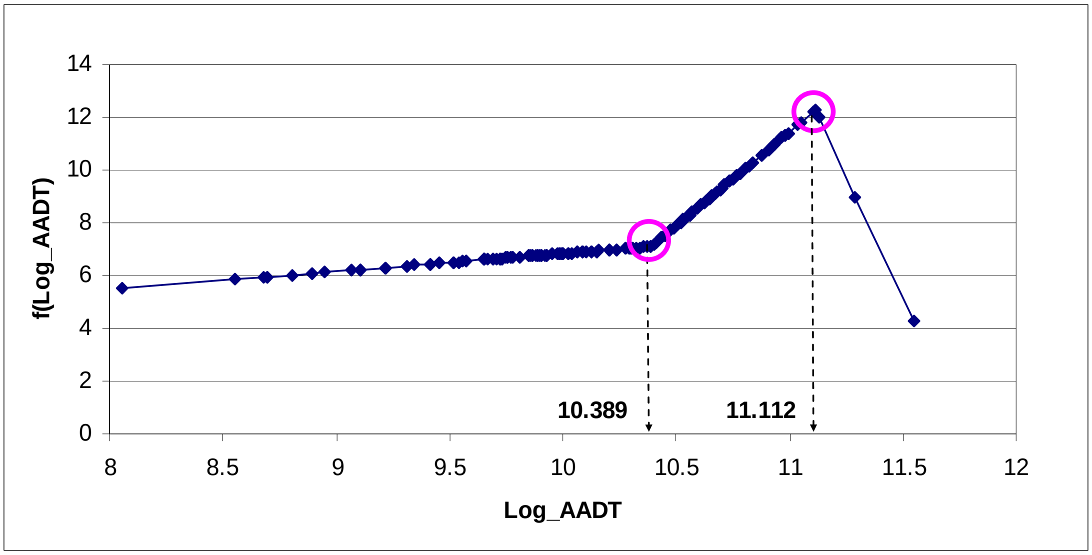


Figure 7-16: Plot of the Basis Function for "Log_AADT"

7.9.2 Comparing MARS and NB Models

For the first application of MARS in this study, a comparison between the two fitted MARS models and the corresponding NB models, while treating the response in each as a discrete one (i.e., crash frequency), is shown in Table 7-12. The R package was utilized to estimate the MARS models via the library "polspline". The MARS models were generated using the default *GCV* value "3" in R. From this table, it is noticed that the MSPE values for MARS in the 3 and 4-legged models are lower than the corresponding NB models. As for the MAD values, they are lower for the NB models. However, there is still a great potential of applying the MARS technique. The generalized R-square is much higher for the MARS models.

Table 7-12: Comparison between the Fitted MARS and NB Models in terms of Prediction and Fitting

		Angle three-legged model		Angle four-legged model	
		MARS	NB	MARS	NB
Prediction	MAD *	1.27	1.07	1.08	0.85
	MSPE *	3.08	3.96	2.95	3.30
Fitting	Generalized R-square	0.39	0.19	0.52	0.31

* MAD and MSPE values are normalized by the average of the response variable

7.9.3 Examining Fitting MARS Model with Continuous Response

To examine the higher prediction capability of MARS while dealing with continuous responses (Friedman, 1991), the two MARS models using the same important NB covariates were fitted while considering the natural logarithm of crash frequency. A default *GCV* value of “3” was used while fitting the models. The assessment criteria for the generated MARS models are shown in Table 7-13.

By comparing the MAD and MSPE values from this table with those from the previously fitted MARS models in Table 7-12, it is noticed that the MAD and MSPE values shown in Table 7-13 are much lower, hence higher prediction capability. Also, the generalized R-square values in Table 7-13 are higher than those in Table 7-12.

Table 7-13: Prediction and Fitting Performance of the Two MARS Models Using a Continuous Response Formulation

		Angle three-legged model	Angle four-legged model
		MARS [1]	MARS [1]
Prediction	MAD [2]	1.01	0.69
	MSPE [2]	0.74	0.61
Fitting	Generalized R-square	0.47	0.67

[1] Response is the natural logarithm of crash frequency

[2] MAD and MSPE values are normalized by the average of the response variable

7.9.4 Using MARS in Conjunction with Random Forest

Since the MARS technique showed promising prediction performance, especially while dealing with continuous responses, an additional effort to examine screening all possible covariates before fitting a MARS model, was explored. This leads to utilizing the random forest technique (Breiman, 2001) before fitting a MARS model for variable screening and ranking important covariates. Using the R package, all possible covariates in the two attempted models were screened via the library "randomForest". The random forest technique was performed with 50 trees grown in the two training datasets. To examine whether this number can lead to stable results, the plot of the OOB error rate against different tree numbers for the four-legged training dataset (an example for illustration purposes) is shown in Figure 7-17. From this figure, it is noticed that after 38 trees, the OOB error rate starts to stabilize. Hence, the attempted number of trees "50" was deemed large enough to obtain stable results. This was also concluded for the three-legged training dataset.

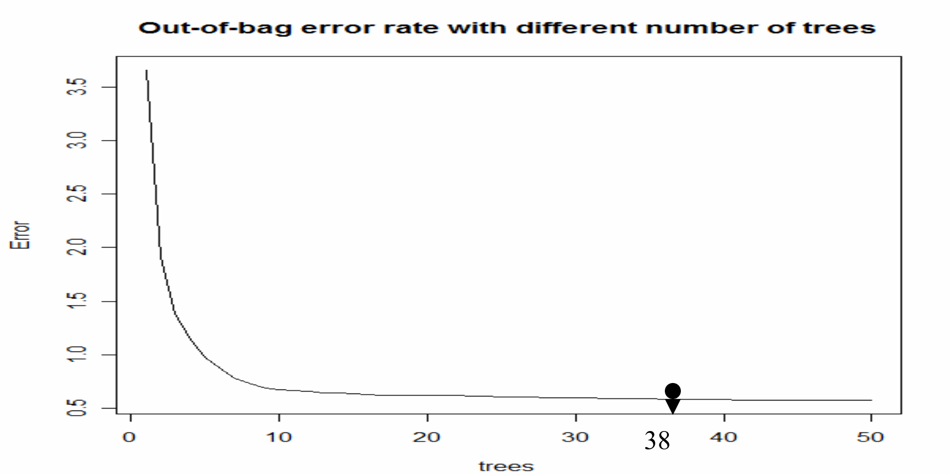


Figure 7-17: Plot of the OOB Error Rate against Different Number of Trees

Figure 7-18 shows the purity values for every covariate. The highest variable importance ranking is the natural logarithm of AADT, followed by the county location, then the natural logarithm of the distance between two unsignalized intersections, etc., until ending up with the existence of crosswalk on the major approach. The resulted variable importance ranking demonstrates the significant effect of the spatial covariates on angle crashes, with the distance between successive unsignalized intersections being the most significant. This variable was also found significant in the fitted four-legged NB model. To screen the covariates, a cut-off purity value of "10" was used. This leads to selecting seven covariates (labeled from "1" till "7" in Figure 7-18). Those seven covariates were then fitted using MARS, with the response being the natural logarithm of crash frequency, as it revealed the most promising prediction capability.

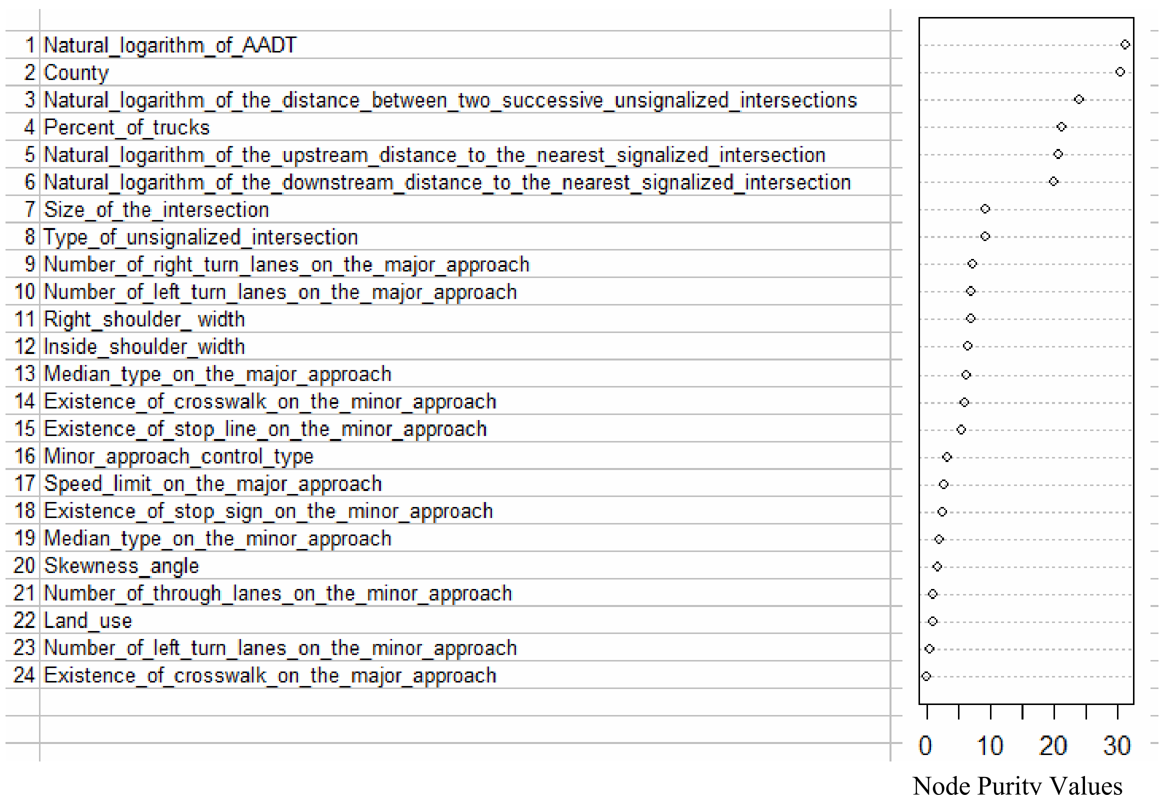


Figure 7-18: Variable Importance Ranking Using Node Purity Measure

The final fitted MARS model using the seven selected covariates at 4-legged unsignalized intersections is presented in Table 7-14, where the response is the logarithm of angle crash frequency. From this table, it is noticed that the positive coefficient for the logarithm of AADT concurs with that deduced from Table 7-10. Also, there is a nonlinear performance for the continuous variable "Log_AADT" with the logarithm of angle crashes , as shown in its truncated basis function at "10.778" and "11.112".

Table 7-14: MARS Model at 4-Legged Unsignalized Intersections after Screening the Variables Using Random Forest

Basis Function	Basis Function Description	Estimate *	P-value
Intercept	Intercept	-2.9252 (0.8759)	0.0009
Log_AADT	Natural logarithm of AADT on the major road	0.2376 (0.0852)	0.0055
Hills_County	Hillsborough County	0.5529 (0.0922)	<0.0001
Miami_County	Miami-Dade County	0.5362 (0.1031)	<0.0001
$(\text{Log\_AADT} - 11.112)_{+}$	A truncated power basis function for "Log_AADT" at "11.112"	-8.3871 (1.9002)	<0.0001
$(\text{Log\_AADT} - 10.778)_{+}$	A truncated power basis function for "Log_AADT" at "10.778"	2.6198 (0.6390)	<0.0001
	Generalized R-square	0.65	

* Standard error in parentheses

To assess whether there is an improvement over the two generated MARS models using the important variables from the NB model, the same evaluation criteria were used, as shown in Table 7-15. Comparing the MAD and MSPE values in Tables 7-13 and 7-15, it is noticed that there is a reduction (even if it is small) in the MAD and MSPE values in Table 7-15, hence better prediction accuracy. The resulted generalized R-square values are relatively high, hence encouraging model fit. This demonstrates that using MARS after screening the variables using random forest is quite promising.

Table 7-15: Prediction and Fitting Assessment Criteria for the Two MARS Models after Screening the Variables Using Random Forest

		Angle three-legged model	Angle four-legged model
		MARS	**MARS**
Prediction	**MAD ***	0.99	0.69
	MSPE *	0.74	0.58
Fitting	**Generalized R-square**	0.47	0.65

* MAD and MSPE values are normalized by the average of the response variable

7.10 General Conclusions from the MARS Analysis

This chapter investigated multiple applications of a new methodology "MARS" for analyzing motor vehicle crashes, which is capable of yielding high prediction accuracy. This was the motivation of this study by applying it to extensive data collected at unsignalized intersections. Rear-end and angle crashes were selected for the scope of the analysis and assessment.

The fitted NB rear-end regression models showed several important variables affecting safety at unsignalized intersections. These include traffic volume on the major road, the upstream and downstream distances to the nearest signalized intersection, median type on the major approach, land use at the intersection's influence area, and the geographic location within the state.

For the NB angle crash models, the important factors include traffic volume on the major road, the upstream distance to the nearest signalized intersection, the distance between successive unsignalized intersections, median type on the major approach, percentage of trucks on the major approach, size of the intersection and the geographic location within the state.

While comparing the MARS and NB models using a discrete response for both fitted rear-end and angle crash models, it was concluded that both MARS and NB models yielded efficient prediction performance, hence MARS can be used as an effective method for prediction purposes.

Treating crashes as continuous response while fitting MARS models was explored. It was concluded that the fitted MARS models always yielded better prediction performance than MARS models with the discrete response.

Finally, a smarter technique of fitting MARS models using the screened variables from the random forest technique was attempted. It was concluded that applying MARS in conjunction with the random forest technique showed better results than fitting MARS model using the important variables from the NB model.

The findings of this study point to that the MARS technique is recommended as a robust method for effectively predicting crashes at unsignalized intersections if prediction is the sole objective. Hence, for achieving the most promising prediction accuracy, important variables should be initially selected using random forest before fitting a MARS model. Still, NB regression models are recommended as a valuable tool for understanding those geometric, roadway and traffic factors affecting safety at unsignalized intersections, as they are easy to interpret.

CHAPTER 8. ACCESS MANAGEMENT ANALYSIS

8.1 Introduction

This chapter is mainly concerned with access management analysis related to unsignalized intersections. This is performed with respect to the six median types specified in this research. The need to address the safety effects of different median types reflects an increased attention to access management analysis. As previously mentioned in Chapter 3, the six median types identified are closed, directional, open, undivided, two-way left turn lane and marking medians. An additional median type was identified in Chapter 4, which is the mixed median (directional from one side, and closed from the other). The first two types, as well as mixed medians are considered restricted medians (i.e., no vehicle can cross from the side streets or driveways "access points"), whereas the last four types are unrestricted medians (i.e., vehicles can cross from the side streets or driveways through each median). Restricted medians always exist at 3-legged intersections, as they restrict the full major street crossing, thus, even if two driveways exist on both sides of any of these medians, they are treated as two separate 3-legged intersections. On the other hand, unrestricted medians could exist on either 3 or 4-legged intersections. They could exist on 4-legged intersections, since from the geometry aspect, they can not restrict vehicles crossing the full major street's width.

An extensive literature review regarding access management analysis was previously presented in Chapter 3 of this dissertation.

8.2 Preliminary Analysis: Comparing Crashes at Different Median Types

After identifying the seven median types at unsignalized intersections, it is essential to give insight to the number of intersections falling within each median type (based on the collected data in this study), as well as the frequency of crashes within each type. This will formulate a preliminary perspective for the safest and most hazardous median types at unsignalized intersections. The total number of intersections used in this analysis is 2498 intersections. The number of intersections associated with each identified median type is shown in Figure 8-1. From this figure, it is noticed that intersections associated with open medians were the most dominant in the dataset, followed by undivided medians, then closed medians, then two-way left turn lanes, then directional medians, then mixed medians, and finally marking medians (since they rarely exist at intersections' approach).

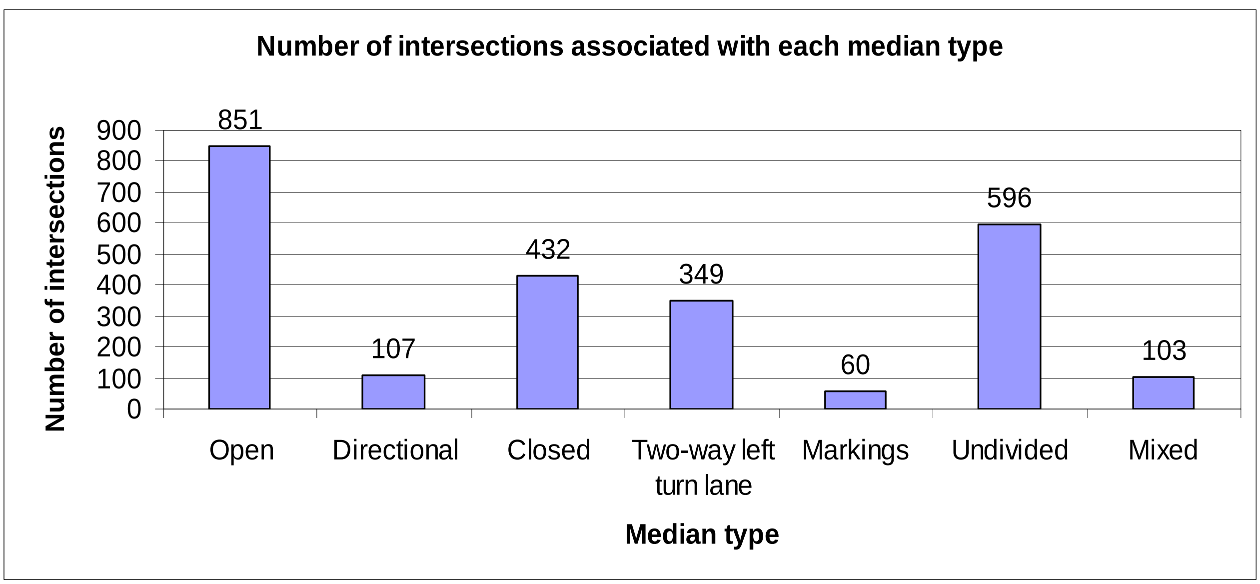


Figure 8-1: Plot of the Number of Intersections Associated with Each Median Type (Based on the Collected Data)

To provide an insight to the distribution of crashes at each median type, the plot of the average total crash per intersection in 4 years "from 2003 until 2006" associated with each median type is presented in Figure 8-2. The average total crash per intersection associated with each median type was presented - and not the total crashes - to account for the actual intersection sample at each median type (i.e., the normalization by the number of intersections was beneficial in this case).

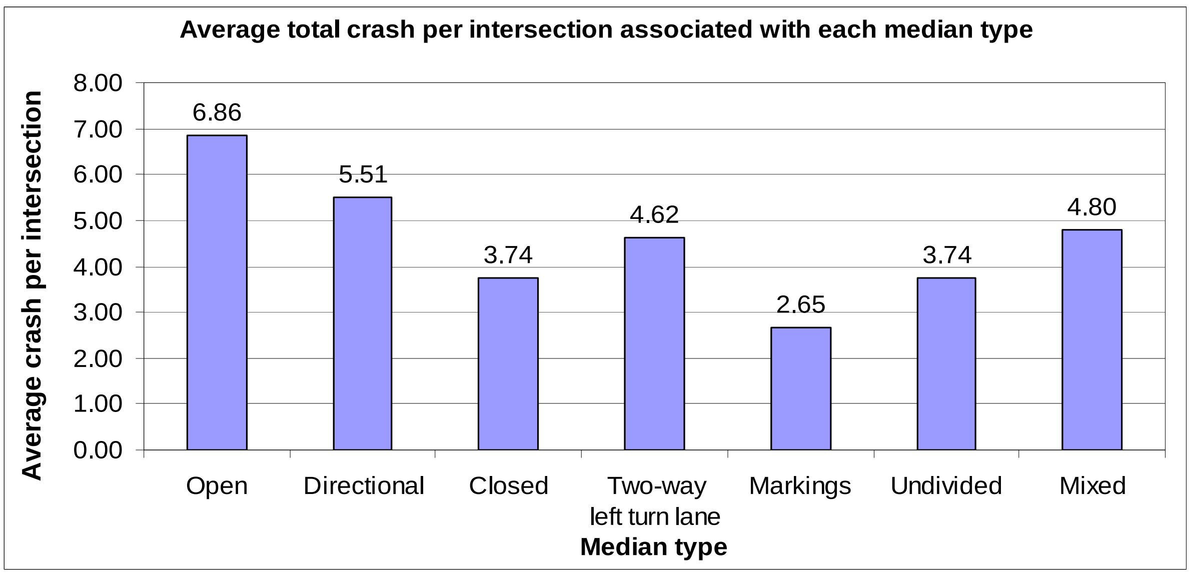


Figure 8-2: Plot of the Average Total Crash per Intersection in Four Years Associated with Each Median Type (Based on the Collected Data)

From Figure 8-2, the highest average number of crashes per intersection occurs at intersections associated with open medians, followed by directional medians, mixed medians, two-way left turn lanes, undivided and closed medians, and finally markings. Thus, it can be concluded that open medians are preliminarily considered as the most hazardous median type. This is attributed to the large number of conflict patterns at open medians, when compared to other median types.

To break down the most frequent types of crashes at unsignalized intersections in the 4-year analysis period "from 2003 until 2006" (based on the collected data in this study), the plot of the average total crash per intersection associated with each median type for each of the five most frequent crash types "rear-end, head-on, angle, left-turn and side-swipe" for each median type is presented in Figure 8-3.

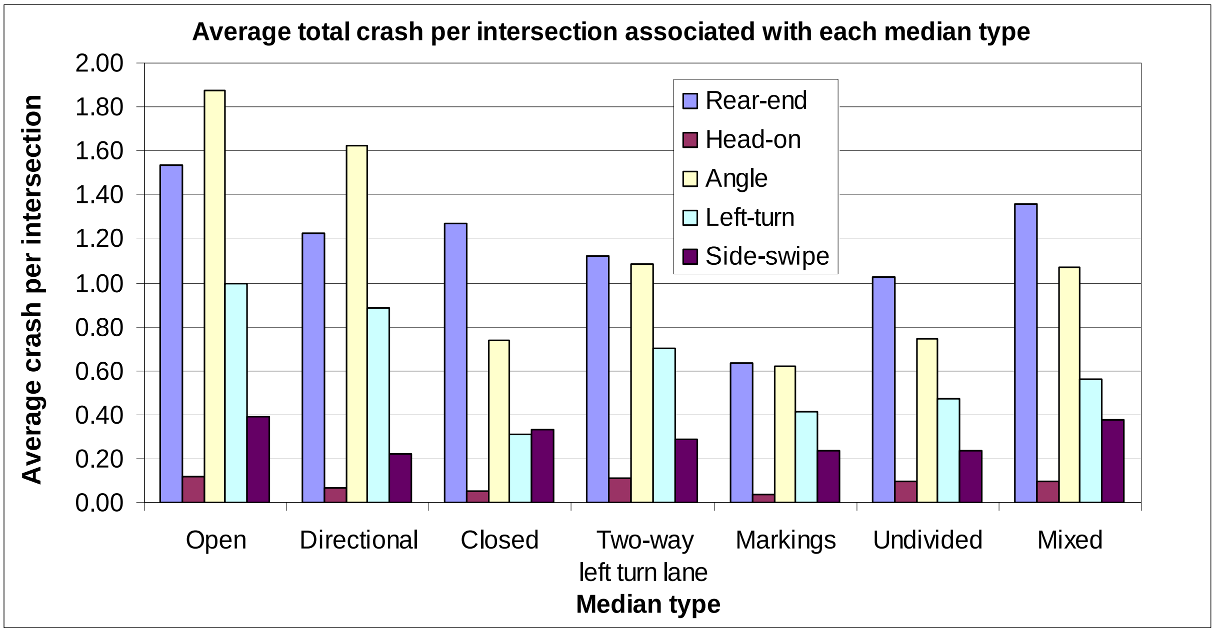


Figure 8-3: Plot of the Average Total Crash per Intersection in Four Years for the Five Most Frequent Crash Types Associated with Each Median Type (Based on the Collected Data)

From Figure 8-3, open medians have the highest average value for all the five most frequent crash types. This result is consistent with that from Figure 8-2. Marking medians have the lowest averages, except for left-turn and side-swipe crashes. Closed medians have the lowest average left-turn crash, since no left-turn maneuver is allowed at both major and minor intersection approaches. The explanation of having left-turn crashes at closed medians might be due to the existence of a nearby median opening at the intersections' influence area, but not at the approach itself (i.e., the separation median between the two major directions in front of the intersection is relatively small in length, thus allowing for left-turn maneuvers at a relatively small distance from the intersection of interest, but still in the influence area).

Directional medians have the lowest average side-swipe crashes, since there is a separation raised median-structure between the two left-turn vehicles from each major road direction. However, the existence of some side-swipe crashes could be explained by two main reasons. The first one is the officer's mistake in documenting the resulted crash pattern, and the second is the tiny thickness for the separation raised median (can act as if it is a painted marking), allowing some vehicles to go over it, hence, side-swipe crash is probable.

The two highest crash averages at each median type are rear-end and angle crashes. This result conforms to previous studies (e.g., Summersgill and Kennedy, 1996; Layfield, 1996; Pickering and Hall, 1986; Agent, 1988 and Hanna et al., 1976). Since marking medians have a relatively low crash average per intersection, as well as low intersection sample representation (as shown in Figure 8-3, and aided by Figures 8-1 and 8-2), they were excluded from further analysis in this chapter.

8.3 Possible Median-related Crashes at Different Median Types

Most of the safety research documents the safety performance of the intersection as a whole, and does not evaluate the safety performance of the median area by itself (e.g., Gluck et al., 1999). Thus, the main objective of the analysis done in this chapter is to identify various crash patterns that could occur at each of the identified median types, i.e., identify median-related crashes at the collected unsignalized intersections in the six counties. Thus, median-related crashes were isolated from other crash patterns that could occur at intersections. Hence, a clearer understanding (after removing unrelated median crashes) can be done to investigate the relationship between median-related crash occurrence and those geometric, traffic and driver features. This will provide a precise mechanism to identify the safest and most hazardous medians at unsignalized intersections, thus, identification of the significant countermeasures as a remedy for any safety deficiency at different median types would be beneficial.

Different median-related crash conflicts existing at open, closed, undivided, two-way left turn lane, directional and mixed medians are shown in Tables 8-1 and 8-2 for 4 and 3-legged intersections, respectively, where each possible conflict represents a certain crash pattern. Each possible crash pattern is sketched at 4-legged intersections for different median types in Figures 8-4 to 8-6. It is noted that for 3-legged intersections, patterns 4 till 9 do not exist at unrestricted medians (i.e., open, undivided and two-way left turn lane medians). Possible crash patterns at 3-legged intersections for directional and mixed medians are sketched in Figures 8-7 and 8-8, respectively.

Table 8-1: Possible Median-related Crash Conflicts at 4-legged Unsignalized Intersections

	Unrestricted medians			Restricted medians		
	Open median	Undivided median	Two-way left turn lane median	Directional median	Mixed median	Closed median
Pattern	**Crash type**	**Crash type**	**Crash type**	**Crash type**	**Crash type**	**Crash type**
1	U-turn (Rear-end)	N/A*	U-turn (Rear-end)	N/A	N/A	N/A
2	Left-turn (Angle)	Left-turn (Angle)	Left-turn (Angle)	N/A	N/A	N/A
3	Left-turn (Angle)	Left-turn (Angle)	Left-turn (Angle)	N/A	N/A	N/A
4	Side-swipe (Left-turn)	N/A	Side-swipe (Left-turn) or head-on	N/A	N/A	N/A
5	Right-angle (Angle)	Right-angle (Angle)	Right-angle (Angle)	N/A	N/A	N/A
6	Right-angle (Angle)	Right-angle (Angle)	Right-angle (Angle)	N/A	N/A	N/A
7	Left-turn (Angle)	Left-turn (Angle)	Left-turn (Angle)	N/A	N/A	N/A
8	Left-turn (Angle) (Head-on)	Left-turn (Angle) (Head-on)	Left-turn (Angle) (Head-on)	N/A	N/A	N/A
9	Rear-end	Rear-end	Rear-end	N/A	N/A	N/A

* N/A means not applicable

Table 8-2: Possible Median-related Crash Conflicts at 3-legged Unsignalized Intersections

	Unrestricted medians			Restricted medians		
	Open median	Undivided median	Two-way left turn lane median	Directional median	Mixed median	Closed median
Pattern	Crash type	Crash type	Crash type	Crash type	Crash type	Crash type
1	U-turn (Rear-end)	N/A*	U-turn (Rear-end)	U-turn (Rear-end)	U-turn (Rear-end)	N/A
2	Left-turn (Angle)	Left-turn (Angle)	Left-turn (Angle)	Left-turn (Angle)	Left-turn (Angle)	N/A
3	Left-turn (Angle)	Left-turn (Angle)	Left-turn (Angle)	N/A	N/A	N/A
4	N/A	N/A	N/A	N/A	N/A	N/A
5	N/A	N/A	N/A	N/A	N/A	N/A
6	N/A	N/A	N/A	N/A	N/A	N/A
7	N/A	N/A	N/A	N/A	N/A	N/A
8	N/A	N/A	N/A	N/A	N/A	N/A
9	Rear-end	Rear-end	Rear-end	Rear-end	Rear-end	N/A

* N/A means not applicable

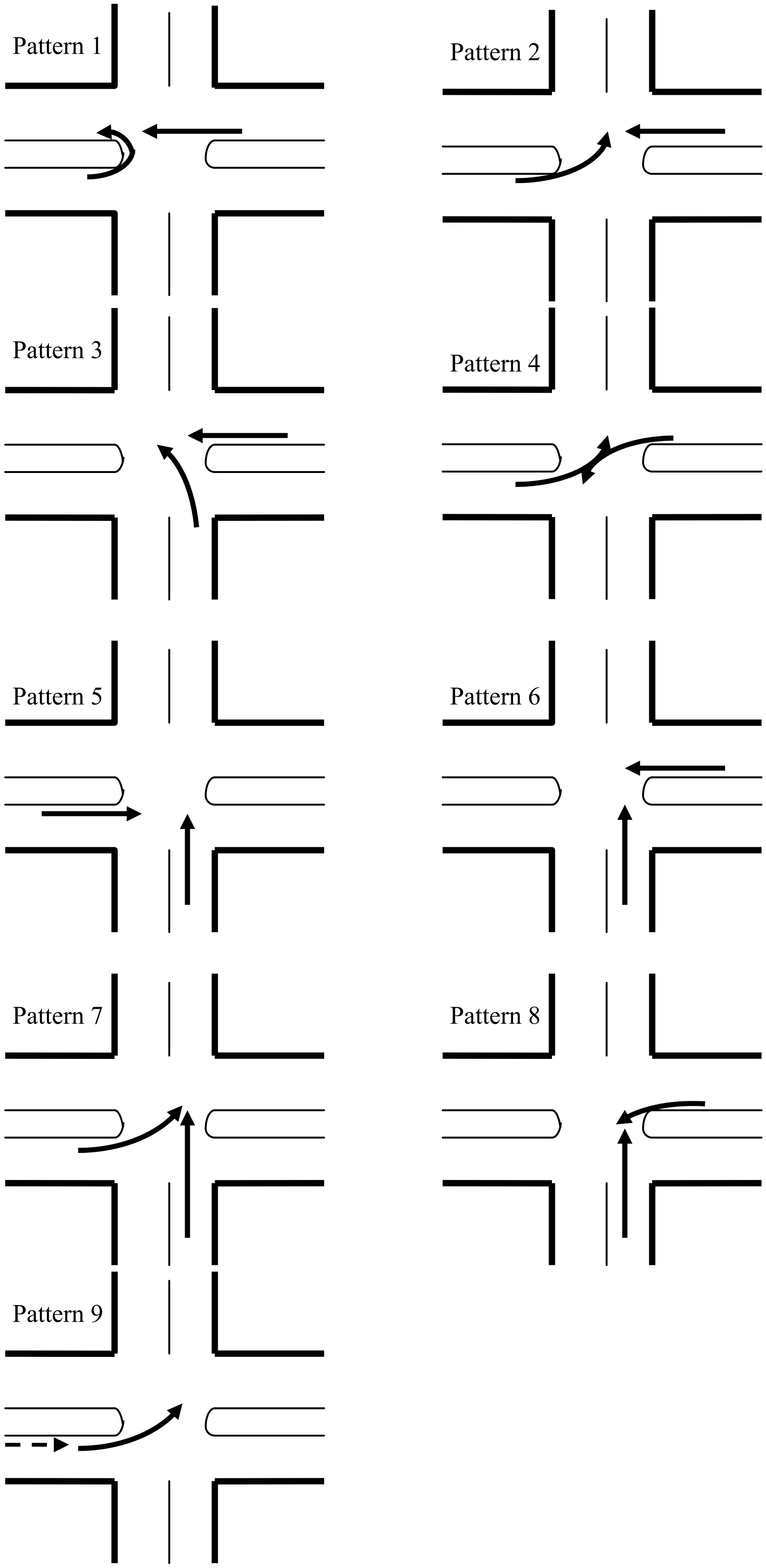


Figure 8-4: Possible Median-Related Crash Patterns at Open Medians at 4-legged Intersections

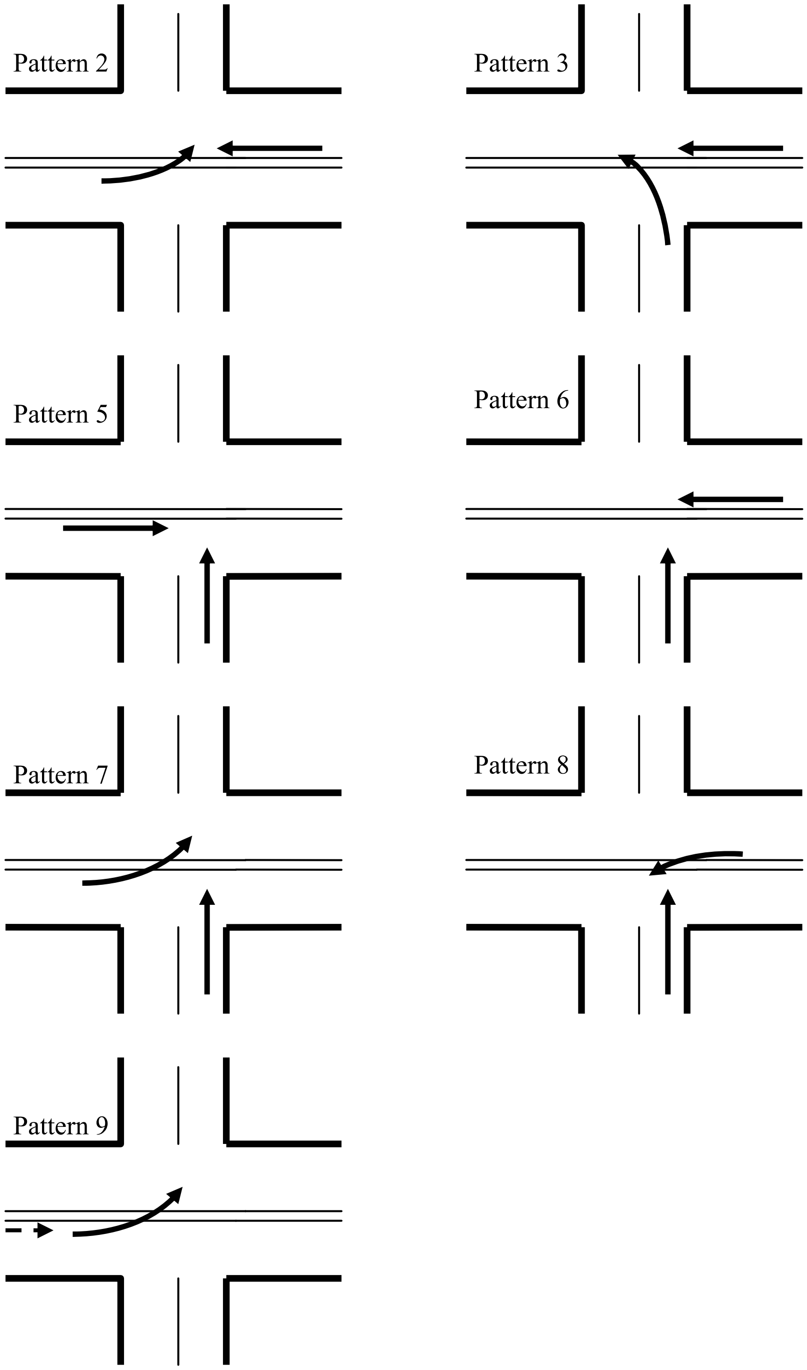


Figure 8-5: Possible Median-Related Crash Patterns at Undivided Medians at 4-legged Intersections

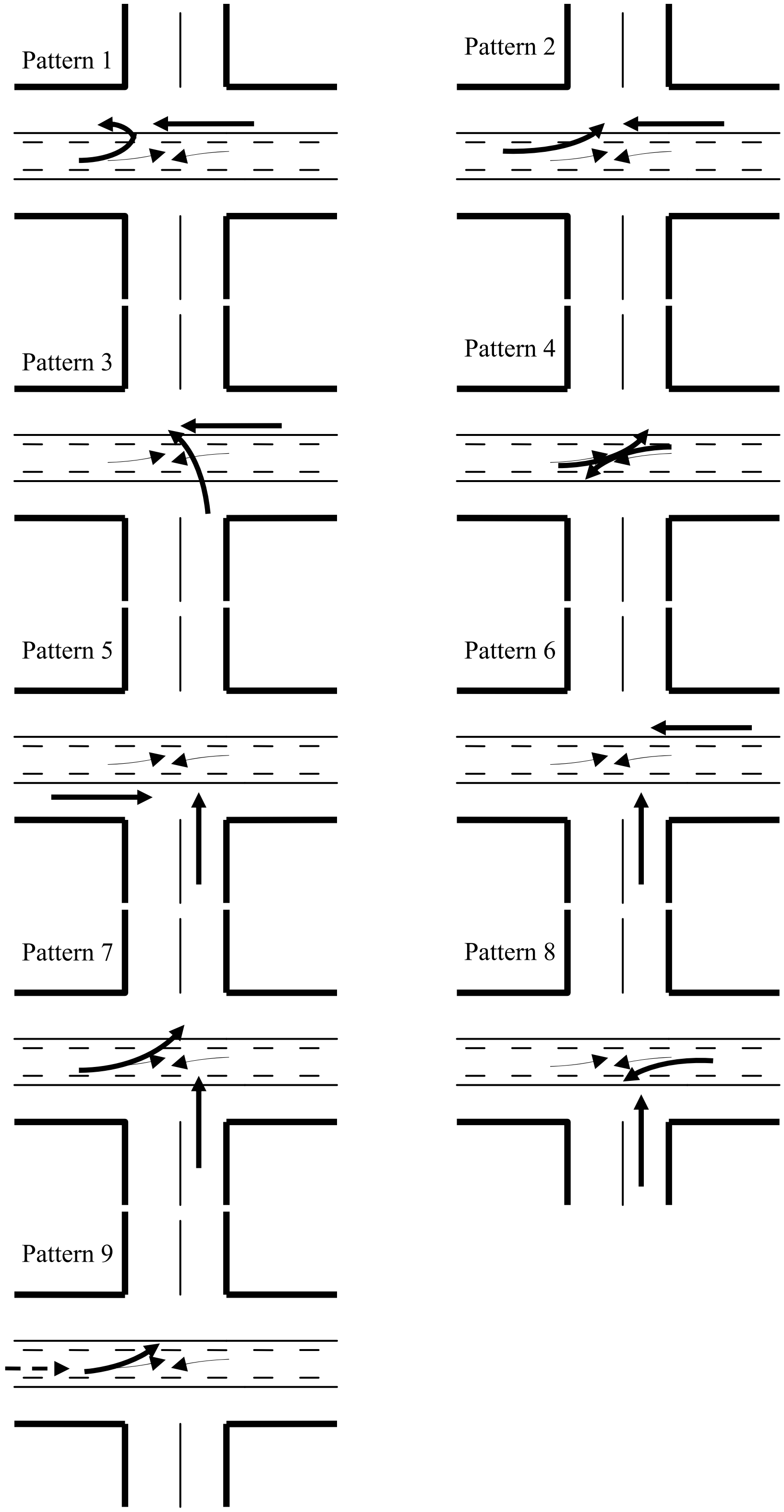


Figure 8-6: Possible Median-Related Crash Patterns at Two-Way Left Turn Medians at 4-legged Intersections

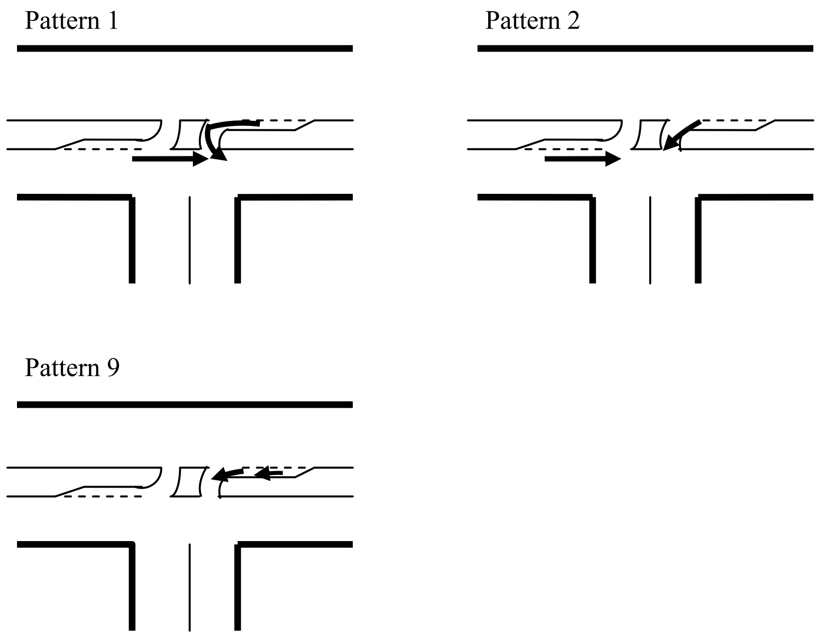


Figure 8-7: Possible Median-Related Crash Patterns at Directional Medians at 3-legged Intersections

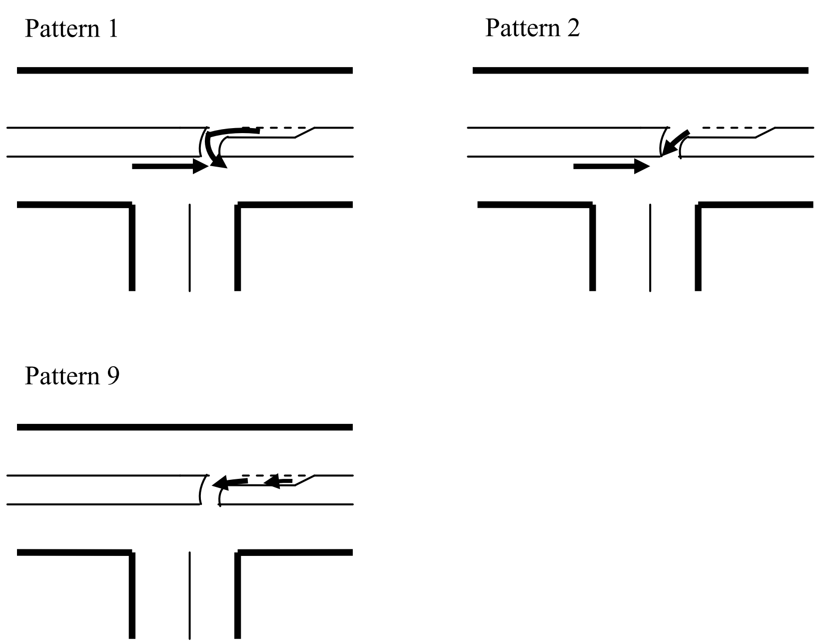


Figure 8-8: Possible Median-Related Crash Patterns at Mixed Medians at 3-legged Intersections

8.4 Screening for Median-related Crashes in the Dataset

After identifying all possible median-related crash patterns, all crashes in the 4-year analysis period were screened, so as to account for those crash patterns only. The variables used for screening is "ACCSIDRD", which is defined as the location of the crash (accident) on the roadway. The used code for screening is "M" (i.e., crashes occurring on the median side). This was the only variable that could be relied on for separating median-related and intersection-related crashes.

After screening for median-related crashes, the final number of crashes was 300. Afterwards, it was decided to select a representative sample to make sure that median-related crashes (and not intersection-related crashes) exist in those identified crashes (i.e., the analysis dataset truly represents median-related crashes). The selected random crash sample was 30 crashes (10%). Long-form crash reports for those randomly selected crashes was extracted from the "Hummingbird" Web-based service released by FDOT. A sketched diagram from a sample crash report illustrating the existence of pattern 8 for two-way left turn lane medians is shown in Figure 8-9.

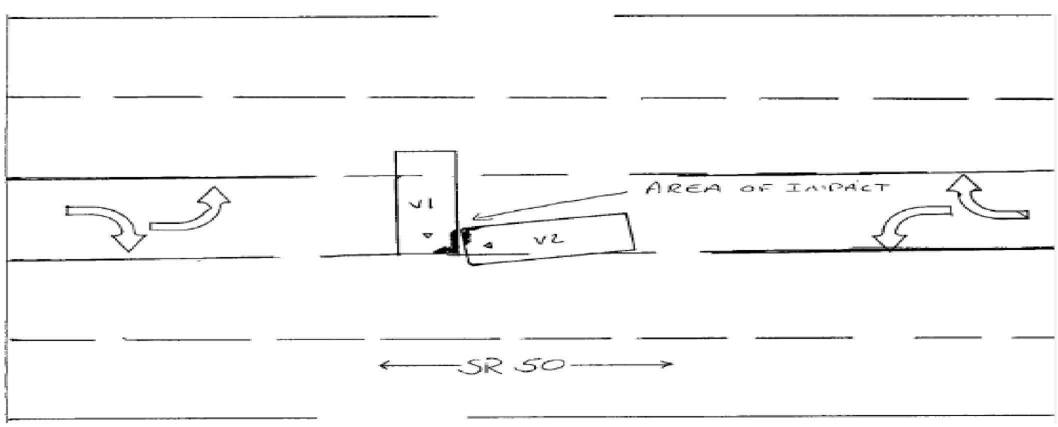


Figure 8-9: A Sketched Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 8 at Two-way Left Turn Lanes (Retrieved from "Hummingbird" Intranet Website)

For the crash report presented in Figure 8-9, the officer reported the crash pattern as a left-turn crash, as shown in Figure 8-10. The code "04" is for collision with motor vehicle in transport (Left-turn).

FIRST / SUBSEQUENT HARMFUL EVENT(S)

01 Collision With MV in Transport(Rear End)	15 Collision With Animal	29 MV Ran Into Ditch/Culvert	1	2	3
02 Collision With MV in Transport(Head On)	16 MV Hit Sign / Sign Post	30 Ran Off Road Into Water	04	04	/
03 Collision With MV in Transport(Angle)	17 MV Hit Utility Pole / Light Pole	31 Overturned	/	/	/
04 Collision With MV in Transport(Left Turn)	18 MV Hit Guardrail	32 Occupant Fell From Vehicle	/	/	/
05 Collision With MV in Transport(Right Turn)	19 MV Hit Fence	33 Tractor/Trailer Jackknifed	/	/	/
06 Collision With MV in Transport(Sideswipe)	20 MV Hit Concrete Barrier Wall	34 Fire			
07 Collision With MV in Transport(Backed Into)	21 MV Hit Bridge/Pier/Abutment/Rail	35 Explosion			
08 Collision With Parked Car	22 MV Hit Tree /Shrubbery	36 Downhill Runaway			
09 Collision With MV on Roadway	23 Collision With Construction Barricade Sign	37 Cargo Loss or Shift			
10 Collision With Pedestrian	24 Collision With Traffic Gate	38 Separation of Units			
11 Collision With Bicycle	25 Collision With Crash Attenuators	39 Median Crossover			
12 Collision With Bicycle (Bike Lane)	26 Collision With Fixed Object Above Road	77 All Other (Explain in Narrative)			
13 Collision With Moped	27 MV Hit Other Fixed Object				
14 Collision With Train	28 Collision With Moveable Object On Road				

Figure 8-10: Reported Left-turn Crash by the Officer for the Crash in Figure 8-9

Another diagram from another sample crash report illustrating the existence of pattern 4 for open medians is shown in Figure 8-11**.** For this particular crash, the officer reported it as an angle crash, as shown in Figure 8-12**.** The code "03" is for collision with motor vehicle in transport (Angle).

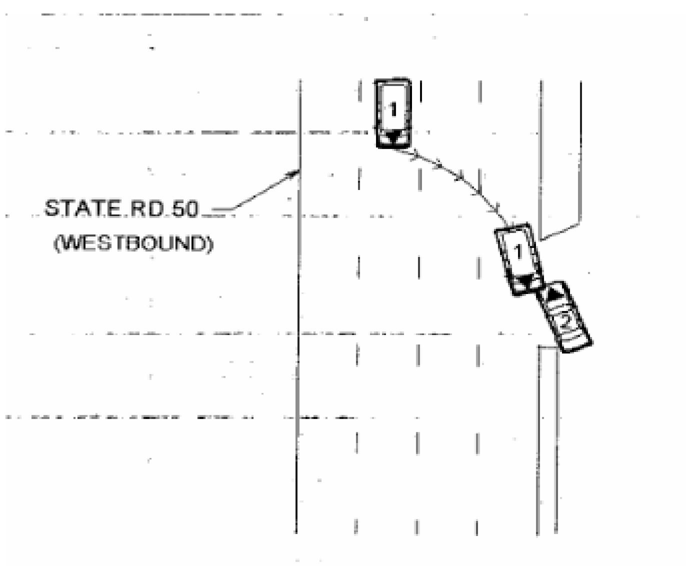


Figure 8-11: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 4 at Open Medians (Retrieved from "Hummingbird" Intranet Website)

FIRST / SUBSEQUENT HARMFUL EVENT(S)			
01 Collision With MV in Transport (Rear End)	15 Collision with Animal	29 MV Ran Into Ditch / Culvert	1 2 3
02 Collision With MV in Transport (Head-on)	16 MV Hit Sign / Sign Post	30 Ran Off Road / Into Water	03
03 Collision With MV in Transport (Angle)	17 MV Hit Utility Pole / Light Pole	31 Overturned	
04 Collision With MV in Transport (Left Turn)	18 MV Hit Guardrail	32 Occupant Fell From Vehicle	
05 Collision With MV in Transport (Right Turn)	19 MV Hit Fence	33 Tractor / Trailer Jackknifed	
06 Collision With MV in Transport (Sideswipe)	20 MV Hit Concrete Barrier Wall	34 Fire	
07 Collision With MV in Transport (Backed Into)	21 MV Hit Bridge / Pier / Abutment / Rail	35 Explosion	
08 Collision With Parked Car	22 MV Hit Tree / Shrubbery	35 Downhill Runaway	
09 Collision with MV on Other Roadway	23 Collision with Construction Barricade Sign	37 Cargo Loss or Shift	
10 Collision with Pedestrian	24 Collision with Traffic Gate	38 Separation of Units	
11 Collision with Bicycle	25 Collision with Crash Attenuators	39 Median Crossover	
12 Collision with Bicycle (Bike Lane)	26 Collision with Fixed Object Above Road	77 All Other (Explain in Narrative)	
13 Collision with Moped	27 MV Hit Other Fixed Object		
14 Collision with Train	28 Collision with Moveable Object on Road		

Figure 8-12: Reported Left-turn Crash by the Officer for the Crash in Figure 8-11

A third diagram from a sample crash report illustrating the existence of pattern 9 for two-way left turn lane medians is shown in Figure 8-13. For this particular crash, the officer reported it as a rear-end crash, as shown in Figure 8-14. The code “01” is for collision with motor vehicle in transport (Rear-end).

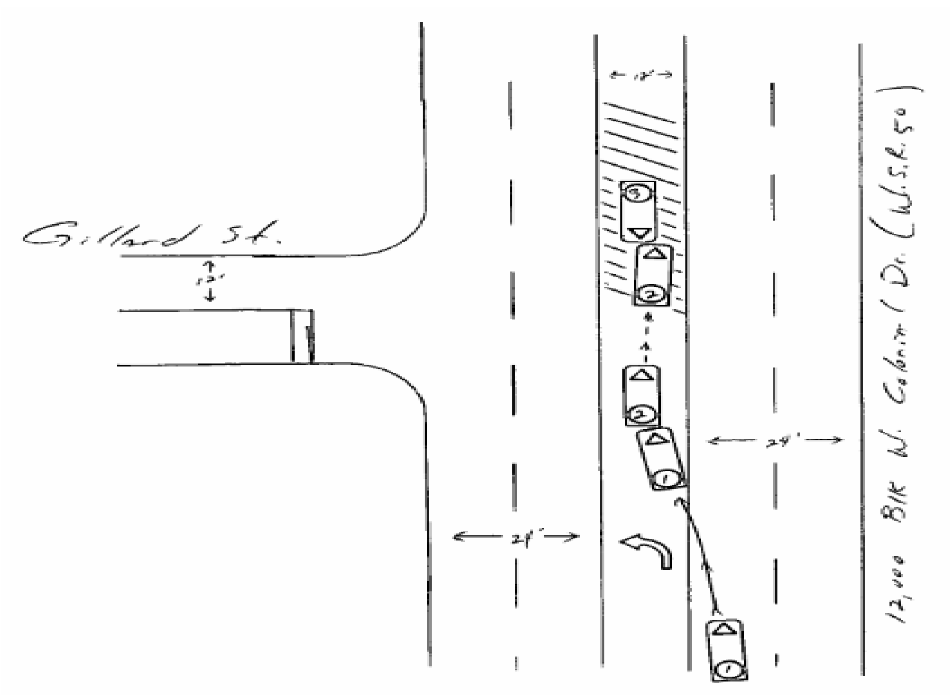
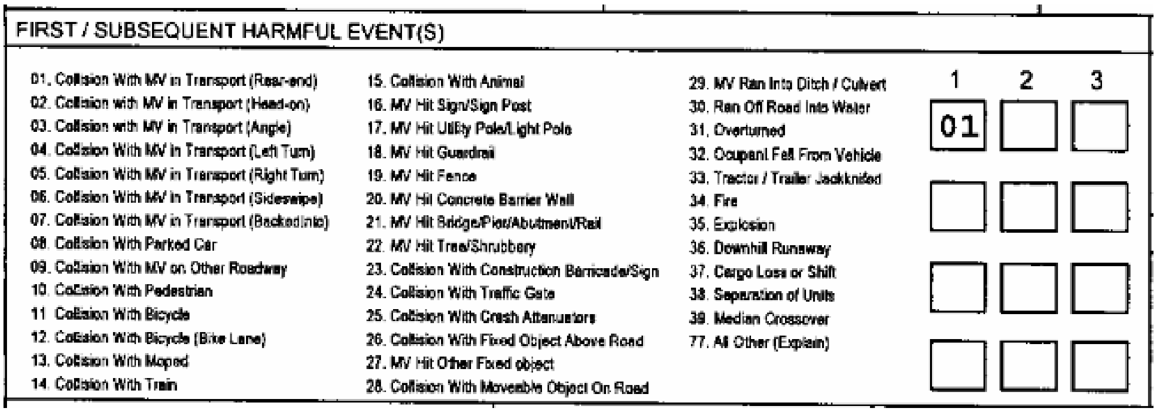


Figure 8-13: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 9 at Two-way Left Turn Lanes (Retrieved from “Hummingbird” Intranet Website)



FIRST / SUBSEQUENT HARMFUL EVENT(S)

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

1	2	3
01		

Figure 8-14: Reported Left-turn Crash by the Officer for the crash in Figure 8-13

From the randomly selected 30 crash reports, 22 were identified as a result of the patterns initially sketched. The remaining 8 were median-related crashes, but not as a result of the patterns sketched. They were rather single-vehicle crashes that occurred at the median (e.g., hitting a fixed object or a sign or a pole) or other two or multi-vehicle crashes apart from those nine identified crash patterns. Hence, there is enough evidence that the collected sample is a true representation of median-related crashes as a result of any of the patterns sketched at each median type.

Since there were some other crash patterns outside the scope of the identified nine crash patterns, it was decided to identify two new crash patterns (pattern 10 and pattern 11). Pattern 10 accounts for two or multi-vehicle median-related crashes other than those nine crash patterns. Pattern 11 accounts for any single-vehicle crash (such as hitting a fixed object or a sign or a pole on the median).

Two sketched diagrams from two sample crash reports illustrating the existence of pattern 10 for two or multi-vehicle crashes other than those nine identified crash patterns are shown in Figures 8-15 and 8-16.

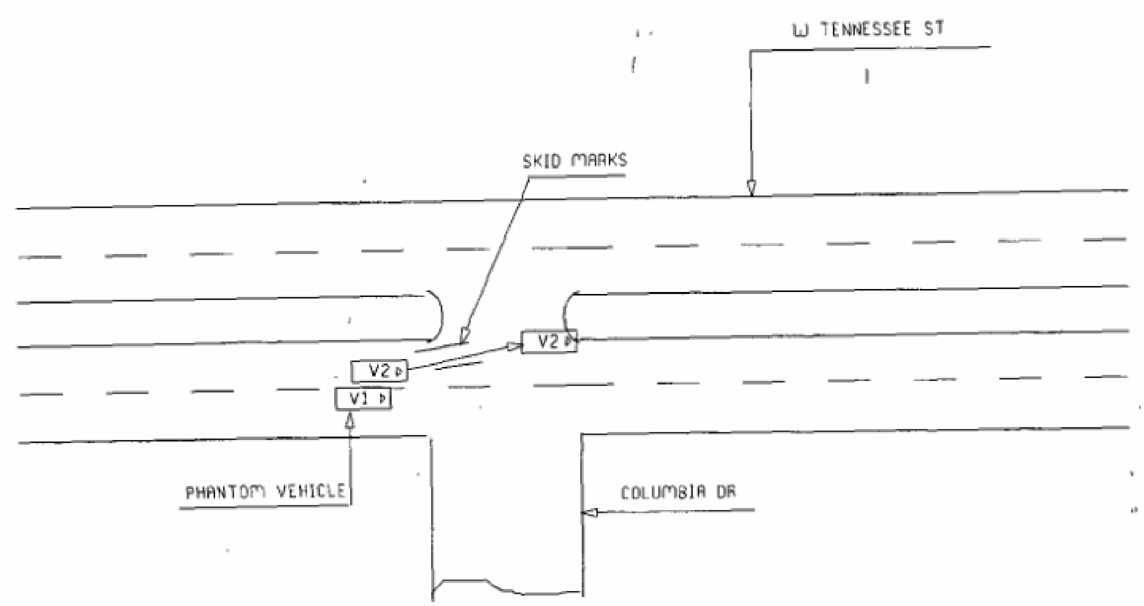


Figure 8-15: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 10 for Two-vehicle Crashes other than the Nine Identified Crash Patterns (Retrieved from "Hummingbird" Intranet Website)

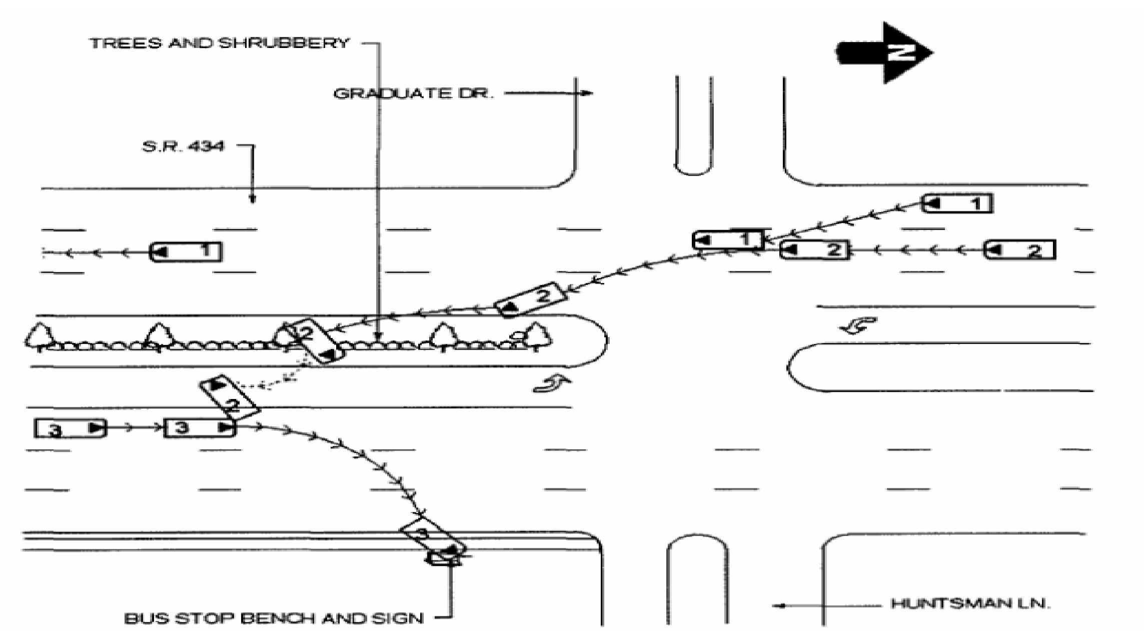


Figure 8-16: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 10 for Multi-vehicle Crashes other than the Nine Identified Crash Patterns (Retrieved from "Hummingbird" Intranet Website)

From Figure 8-15, vehicle 1 “v1” tried to change its lane, then it hit vehicle 2 “v2”, causing “v2” to skid towards the median and “v2” finally hit the median. As for Figure 8-16, vehicle 1 tried to change its lane, and vehicle 2 was running at high speed. Vehicle 2 tried to avoid hitting vehicle 1, but it could not. Hence, vehicle 2 lost control and crossed over the tree and shrubbery median. Additionally, vehicle 2 – because of the high collision reaction – went on the other direction and hit vehicle 3 on the lane just beside the median, causing vehicle 3 to lose control and hit the bus stop sign on the very right side of the roadway. These two crashes are very uncommon, hence, they were not introduced in the nine identified patterns.

Other two diagrams from two sample crash reports illustrating the existence of pattern 11 for single-vehicle crashes are shown in Figures 8-17 and 8-18.

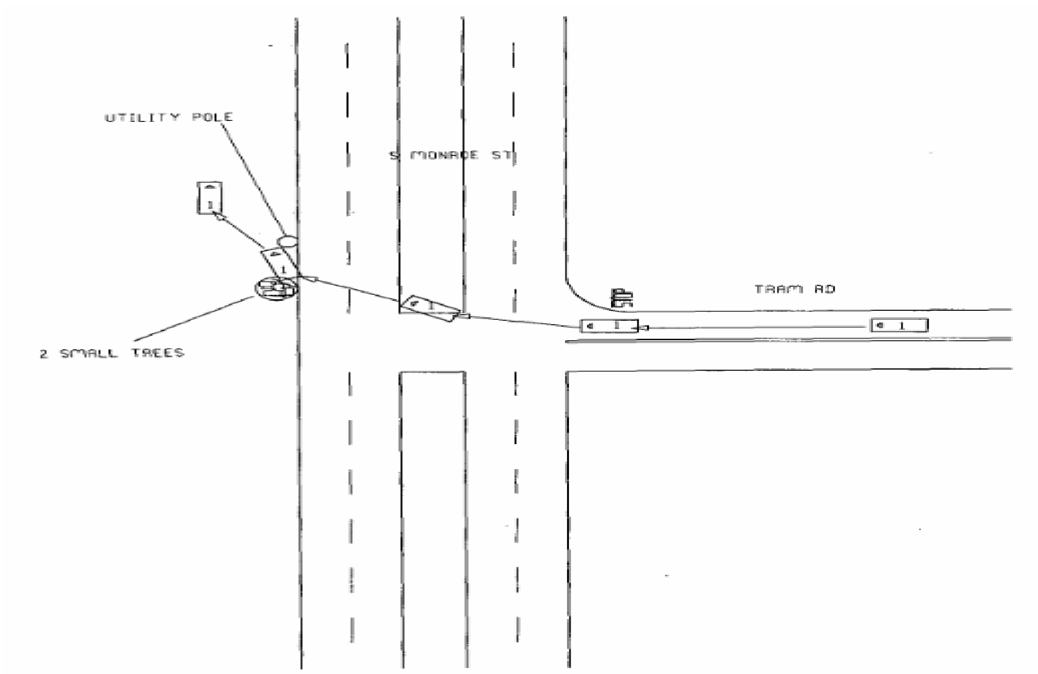


Figure 8-17: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 11 for Single-vehicle Crashes (Retrieved from “Hummingbird” Intranet Website)

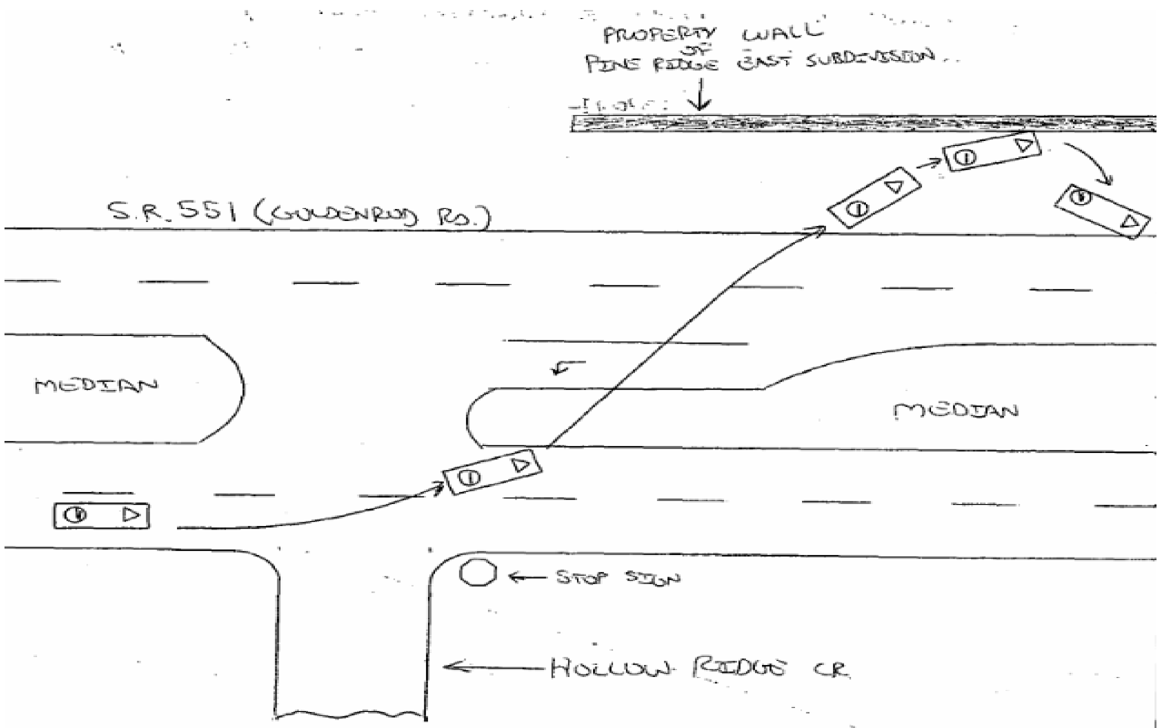


Figure 8-18: A Diagram from a Sample Crash Report Demonstrating the Existence of Pattern 11 for Single-vehicle Crashes (Retrieved from "Hummingbird" Intranet Website)

From Figure 8-17, a vehicle was coming out from the minor approach at a high speed and could not see the stop sign. Thus, the vehicle crossed over the median, and ended up with hitting a utility pole on the further direction. As for Figure 8-18, the driver of vehicle 1 lost control, resulting in crossing over the median, and hitting both a utility pole and a property wall.

Since closed medians were considered as the base case (as no median-related crash could exist in the ideal condition, except for some limited two or single-vehicle crashes such as vehicle crossing over the median), any crash occurring at closed medians is assigned a pattern zero (pattern 0). Thus, pattern 0 is always associated with closed median crashes.

In the median-related crash dataset, there were 300 observations (300 crashes), and 6 of those crashes have some missing values for some important variables, and the associated crash patterns for those crashes were difficult to identify. Hence, they were excluded, and the final dataset contains 294 observations.

Additionally, due to data limitations, some of the identified crash patterns were extremely difficult to be differentiated from each other. For example, patterns 5 and 6 are very similar, as the vehicle's movement on the minor approach is the same. The only difference is the vehicle's movement on the major approach (on the lane just next to the median), and in the used dataset, the direction of travel on the major and minor approaches was not available. Hence, any crash associated with patterns 5 or 6 is assigned a pattern 5. Similarly, patterns 2, 3 and 7 are left-turn crashes, and they are hard to be differentiated, hence, any crash associated with patterns 2 or 3 or 7 is assigned a pattern 2. Additionally, patterns 1 and 9 could be rear-end crashes, and they are hard to be differentiated as well, hence, any crash associated with patterns 1 or 9 is assigned a pattern 1.

Thus, the possible existing patterns in the identified median-related crashes are patterns 0, 1, 2, 4, 5, 8, 10 and 11. A cross-tabulation (2x2 contingency table for each median type by the crash pattern) is shown in Table 8-3.

Table 8-3: A 2x2 Contingency Table for Median Type by Crash Pattern

		Pattern								Total
		0	1	2	4	5	8	10	11	
Median type	Closed	37 (100.00)*	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	37
	Open	0 (0.00)	15 (17.05)	4 (4.55)	3 (3.41)	10 (11.36)	2 (2.27)	1 (1.14)	53 (60.23)	88
	Directional	0 (0.00)	1 (10.00)	1 (10.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (10.00)	7 (70.00)	10
	Two-way left turn lane	0 (0.00)	9 (7.03)	35 (27.34)	14 (10.94)	39 (30.47)	6 (4.69)	3 (2.34)	22 (17.19)	128
	Undivided	0 (0.00)	3 (13.04)	8 (34.78)	0 (0.00)	4 (17.39)	3 (13.04)	1 (4.35)	4 (17.39)	23
	Mixed	0 (0.00)	2 (25.00)	3 (37.50)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	3 (37.50)	8
Total		37	30	51	17	53	11	6	89	294

* Row percentages in parentheses

From Table 8-3, it is noticed that the most frequent crash pattern in the dataset is pattern 11 (single-vehicle median-related crashes), followed by pattern 5 (right-angle crashes), then pattern 2 (left-turn or angle crashes), then pattern 0 (for any closed median crashes), then pattern 1 (mostly rear-end crashes), then pattern 4 (mostly side-swipe crashes), then pattern 8 (mostly head-on crashes), and finally pattern 10 (two or multi-vehicle crashes other than the identified patterns).

Also, it can be noticed that single-vehicle crashes are the most frequent crashes for open and directional medians, accounting for 60.23% and 70%, respectively of crashes at those median types. An important finding is that 54.5% of head-on median-related crashes (pattern 8) occur at two-way left turn lanes. This is a relatively high percentage, and indicates the hazardous effect of two-way left turn lanes on head-on median-related crashes.

For two-way left turn lane medians, pattern 5 (right-angle crashes) is the most frequent crash pattern, accounting for 30.47% of crashes at these medians.

For undivided medians, pattern 2 (left-turn or angle crashes) is the most frequent crash pattern, accounting for 34.78% of crashes at these medians.

For mixed medians, patterns 2 and 11 (single-vehicle crashes) are the most frequent crash patterns, accounting for 75% (together) of crashes at these medians.

8.5 Preliminary Methodological Approach: Multinomial Logit Framework

According to Agresti (2007), logistic regression model is usually used to model binary response variables. A generalization of it models categorical responses with more than two categories (levels). This model is named multinomial logit, where the counts in the categories of the response variable follows a multinomial distribution. It is used to model nominal responses, where the order of the categories is not of concern. The multinomial logit model was described by Haberman (1982) and Press (1972).

With j = 1, 2, 3, ---, J, let J denote the number of categories for the response y. Also, let $\{\pi_1, ---, \pi_J\}$ denote the response probabilities, satisfying the condition that $\sum_j \pi_j = 1$. Multinomial logit models simultaneously use all pairs of categories by specifying the odds "likelihood" of an outcome in a category relative to another.

Multinomial logit models for nominal response variables pair each category with a baseline category. Assuming that the last category "J" is the baseline, the possible "J-1" logit models are:

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \qquad j = 1, 2, ---, J-1 \qquad (8.1)$$

where: α is the intercept to be estimated for each of the "J-1" models, β is the vector of parameter estimates for each of the "J-1" models and x is the vector of fitted covariates.

This means that the possible number of equations is "J-1" and the number of parameters to be estimated is "(J-1) * (p+1)", by assuming p covariates (excluding the intercept). The parameters of this model are estimable by maximization of the multinomial likelihood.

The probability of all categories except for the baseline category within the response y is estimated as:

$$\pi_j = \frac{\exp(\beta x)}{1+\sum_{j=1}^{J-1}\exp(\beta x)}, \qquad j = 1, 2, ---, J\text{-}1 \qquad (8.2)$$

The probability of the baseline category "J" within the response y is estimated as:

$$\pi_J = \frac{1}{1+\sum_{j=1}^{J-1}\exp(\beta x)} \qquad (8.3)$$

A special case of the multinomial logit model exists when J=2, i.e., the response has only two categories. Hence, the multinomial logit model converges to the binomial logit one.

8.6 Multinomial Logit Model Estimation

In this chapter there were six median types identified, hence the multinomial logit model could be appropriate for possible interpretation of geometric and traffic factors leading to crashes at specific median types with respect to a base type. The base median type decided in the analysis procedure is closed median, since in the ideal condition, no median-related crash exists, except for some single-vehicle crashes.

The multinomial logit model was fitted for the five types “open, directional, two-way left turn lane, undivided and mixed”, and the baseline category was closed medians. The fitted multinomial logit model did not converge in the beginning, because as previously mentioned, there were 294 median-related crashes, and this sample is considered limited with those specific median types. Hence, the best way was to combine some median types. The most relevant way for doing so is having two main median types, restricted and unrestricted medians.

From the traffic perspective, restricted medians include closed, directional and mixed medians, since no vehicle from the minor approach could cross to the further major direction. Also, based on Table 8-3, the most frequent crash patterns at directional and mixed medians are single-vehicle crashes, as they almost have the same construction characteristics. For this, closed, directional and mixed medians were assigned as restricted medians. On the other hand, unrestricted medians include open, two-way left turn lane and undivided medians, as there is no restriction to prevent vehicles from crossing to the further major direction from the minor approach. Hence, the multinomial logit model was converged to the binomial one. It is worth mentioning that a binomial logit model was attempted with the specified crash patterns as dummy covariates, but the model did not converge properly. Thus, crash patterns were classified as single and non-single vehicle crashes.

The fitted binomial logit model is shown in Table 8-4. This model is fitted for restricted medians with respect to unrestricted medians. The goodness-of-fit statistics are shown at the end of the table.

Table 8-4: Binomial Logit Model for Restricted Medians (Baseline is Unrestricted Medians)

Variable Description	Estimate	Standard Error	P-value
Intercept	26.2132	7.9672	0.0010
Natural logarithm of AADT on the major road	-1.4842	0.5954	0.0127
Natural logarithm of the upstream distance to the nearest signalized intersection	-0.6596	0.2372	0.0054
Natural logarithm of the downstream distance to the nearest signalized intersection	-1.1056	0.2625	<0.0001
Posted speed limit on major road >= 45 mph	0.9245	0.2901	0.0014
Posted speed limit on major road < 45 mph	--- [a]		
Single-vehicle crashes	0.9235	0.2451	0.0002
Non-single vehicle crashes	--- [a]		
One left turn lane exists on each major road direction	-1.4263	0.3406	<0.0001
One left turn lane exists on only one major road direction	0.2463	0.3073	0.4228
No left turn lane exists on the major approach	--- [a]		
Number of observations	***294***		
Log-likelihood at convergence	***-77.55***		
AIC [b]	***171.11***		
Pseudo R-square	***0.45***		

[a] Base case [b] Akaike Information Criterion (= -2 * log-likelihood + 2 * number of parameters)

From Table 8-4, the likeliness of having a median-related crash at restricted medians increases as the logarithm of AADT decreases (i.e., inherently decreasing traffic volume). This means a higher probability of single-vehicle crashes or lower chance of two or multi-vehicle crashes. This result is assessed by the positive coefficient of single-vehicle crashes in the model. Hence, the probability of having single-vehicle median-related crashes at restricted medians is exp(0.9235) “2.52” higher than that for non-single vehicle crashes. Also, the AADT interpretation indicates that median-related crashes at restricted medians increase at higher speeds. This is assessed as well in the model, where the probability of having median-related crashes at restricted medians at speeds equal to or above 45 mph is exp(0.9245) “2.52” higher than that at lower speeds. This is logic, since single-vehicle median-related crashes always occur at relatively higher speeds.

As the upstream and downstream distance to the nearest signalized intersection increases, the likeliness of having a median-related crash at restricted medians decreases. This indicates the importance of setting back restricted medians (closed or directional or mixed) from nearby signalized intersections to avoid conflict with intersection queues (backward shock waves). A similar finding related to median openings installation was concluded by Koepke and Levinson (1992).

The likeliness of having a median-related crash at restricted medians while having one left turn on each major direction is exp(-1.4263) “0.24” times that while having no left turn lane at all. This indicates the importance of having an exclusive left turn lane on each major approach for separating left turning vehicles from through vehicles, hence median-related crashes are reduced.

8.7 Second Methodological Approach: Bivariate Probit Framework

After examining the multinomial (binomial) logit approach in the previous two sections, this section emphasizes another methodological approach for analyzing median-related crashes, the bivariate probit framework. According to Greene (2003), the bivariate probit is a natural extension of the probit model that allows two equations with correlated disturbances. This is similar to the seemingly unrelated models. The general equation for the two-equation model is:

$$y_1^* = x_1{}'\beta_1 + \varepsilon_1, \qquad y_1 = 1 \text{ if } y_1^* > 0; 0 \text{ otherwise} \qquad (8.4)$$

$$y_2^* = x_2{}'\beta_2 + \varepsilon_2, \qquad y_2 = 1 \text{ if } y_2^* > 0; 0 \text{ otherwise} \qquad (8.5)$$

The characteristics of the error terms "ε_1 and ε_2" are specified according to:

$$E\left[\varepsilon_1 \mid x_1, x_2\right] = E\left[\varepsilon_2 \mid x_1, x_2\right] = 0 \tag{8.6}$$

$$Var\left[\varepsilon_1 \mid x_1, x_2\right] = Var\left[\varepsilon_2 \mid x_1, x_2\right] = 1 \tag{8.7}$$

$$Cov\left[\varepsilon_1, \varepsilon_2 \mid x_1, x_2\right] = \rho \tag{8.8}$$

where ρ is the correlation coefficient between the two error terms. The bivariate probit model converges to two separate binomial probit models when ρ equals zero (i.e., when there is no correlation between the two error terms in both equations).

The model parameters of the two probit equations are estimated simultaneously using the maximum likelihood estimation. A detailed explanation of the parameters' estimation is found in Greene (2003).

8.8 Bivariate Probit Model Estimation

For estimating the bivariate probit model, the first dependent variable for the first probit equation was the median type (restricted or unrestricted), and the second dependent variable for the second equation was the median crash pattern (single vs. non-single crashes). The fitted bivariate probit model is shown in Table 8-5. The first probit model has unrestricted medians as the baseline for the dependent variable, while the second probit model has non-single vehicle crashes as the baseline. The goodness-of-fit statistics are shown at the end of the table. Also, the correlation coefficient "rho" between the two error terms in both equations is presented.

Table 8-5: Bivariate Probit Model Estimates

Variable Description	Estimate	Standard Error	P-value
First probit model (Baseline is unrestricted medians)			
Intercept	15.1250	3.9052	0.0001
Natural logarithm of AADT on the major road	-1.0831	0.2831	0.0001
Natural logarithm of the upstream distance to the nearest signalized intersection	-0.3905	0.1210	0.0013
Natural logarithm of the downstream distance to the nearest signalized intersection	-0.5164	0.1233	0.0000
Posted speed limit on major road >= 45 mph	1.0396	0.2718	0.0001
Posted speed limit on major road < 45 mph	--- [a]		
Single-vehicle crashes	2.3885	0.2637	0.0000
Non-single vehicle crashes	--- [a]		
Second probit model (Baseline is non-single vehicle median crashes)			
Intercept	-0.6685	0.1735	0.0001
Width of the median on the major road (in feet)	0.0438	0.0081	0.0000
Posted speed limit on major road >= 45 mph	-0.4726	0.1850	0.0106
Posted speed limit on major road < 45 mph	--- [a]		
Error terms correlation coefficient (ρ)	***-0.8775***	***0.1368***	***0.0000***
Number of observations	***294***		
Log-likelihood at convergence	***-271.04***		
AIC [b]	***560.08***		
Pseudo R-square	***0.15***		

[a] Base case [b] Akaike Information Criterion

The signs of the parameters in the first probit model look identical to those from Table 8-4. This demonstrates a validation of using the binomial logit and bivariate probit frameworks for analyzing median-crashes.

From the second probit model, as the median width on the major road increases, the likeliness of having a median-related crash at restricted medians increases as well. Since single-vehicle median-related crashes are more likely to occur at restricted medians, thus hitting a wide median is considered more hazardous than hitting a narrow median. This is due to the denser physical nature of wide medians in collision.

The probability of having single-vehicle median crashes at speeds equal to or above 45 mph is less than that at lower speeds, i.e., single-vehicle median crashes are more probable at lower speeds. Although this is unexpected, this might be explained as drivers are more likely to experience risky maneuvers nearby medians at relatively high congestion to escape from a traffic jam. Hence, drivers could end up hitting a median as a result of traffic rage.

The coefficient of correlation "ρ" between the two error terms in both models is statistically different from zero, hence illustrating the validity of using the bivariate probit framework.

8.9 General Conclusions from the Access Management Analysis

The access management analysis performed in this chapter dealt with analyzing six main median types associated with unsignalized intersections/access points. These six median types were open, closed, directional, two-way left turn lane, undivided and mixed medians. Also, crash conflict patterns at each of these six medians were identified and applied to a dataset including median-related crashes. In this case, separating median-related and intersection-related crashes was deemed significant in this analysis. From the preliminary analysis, open medians were considered the most hazardous median type, and closed and undivided medians were the safest.

It was concluded that single-vehicle crashes were the most probable crash patterns from a sample of around 300 median-related crashes in six counties in Florida. The second most frequent crashes were right-angle crashes. Of the least probable crashes were head-on crashes. For open, directional and mixed medians, single-vehicle crashes were the most frequent, accounting for 60%, 70% and 38% of total crashes at those medians, respectively. For two-way left turn lane medians, right-angle crashes were the most frequent, accounting for 30%. As for undivided medians, left-turn and angle crashes were the most frequent, accounting for 35%. Since single-vehicle crashes were the most frequent at directional and mixed medians, these two medians - in addition to closed medians – were classified as restricted medians. This is also supported by the traffic perspective that they restrict minor vehicles' path to the further major direction. In the same manner, open medians, two-way left turn lanes and undivided medians were classified as unrestricted medians.

Using restricted and unrestricted medians showed better results than using the six median types individually. Similarly, using single and non-single median crash patterns was deemed significant for the modeling approach. The binomial logit and bivariate probit models demonstrated the importance of median-related variables affecting median-related crashes. Examples of these variables are median width, speed limit on the major road, logarithm of AADT, logarithm of the upstream and downstream distances to the nearest signalized intersection and crash pattern.

CHAPTER 9. CONCLUSIONS

9.1 Summary and Contributions

This study attempted to provide insight into the safety analysis of unsignalized intersections. Few studies have addressed the safety of these intersection types. One important reason is the inadequacy and difficulty to obtain data at these intersections, as well as the limited crash counts. Another reason is that authorities mainly focus on signalized intersections, since they have more crashes and are relatively larger in size.

Massive data collection effort has been conducted for the scope this study. There were 2500 unsignalized intersections collected from six counties in the state of Florida. These six counties were Orange, Seminole, Hillsborough, Brevard, Leon and Miami-Dade. These selected counties are major counties representing the central, western, eastern, northern and southern parts in Florida, respectively. Hence, a geographic representation of the state of Florida was achieved. Important intersections' geometric and roadway features, minor approach traffic control, major approach traffic flow and crashes were obtained. The analyzed years of crashes were four years (from 2003 till 2006).

In this study, traffic volume (or AADT) on the major approach was included as an explanatory variable in various crash models (i.e., total crashes, crash types such as rear-end and angle crashes and crash severity). This covariate was usually found to be the most significant variable affecting intersection safety.

The AADT on the minor approaches was not available for most of the cases, since they are mostly non-state roads. However, for the scope of this study, this was explored by a surrogate measure, which was represented by the number of through lanes on this approach. This surrogate measure was investigated while analyzing crash injury severity as well as rear-end and angle crashes. However, this covariate was not usually found to be significant.

This study explored new important roadway and traffic covariates that were not examined before. Examples of those new roadway covariates are the existence of crosswalks on the minor and major approaches, number of left and right turn lanes on the major approaches, effect of various minor approach control types (e.g., stop sign, no control and yield sign), various sizes of intersections, intersection type (whether it is a regular unsignalized intersection, access point or ramp junction), various median types on the major approach (open, closed, two-way left turn lane, etc.), distance between unsignalized intersections and signalized ones (from both the upstream and downstream aspects), distance between successive unsignalized intersections, and left (or median) shoulder width.

The analysis conducted in the fifth chapter of this dissertation used a coordinated method of the NB model, as well as the reliability method (in terms of the full Bayesian updating framework) for reducing uncertainty in predicting crash frequency at 3 and 4-legged unsignalized intersections. A broad exploration of both non-informative and informative priors was conducted using both the NB and the log-gamma likelihood functions.

It was concluded that the log-gamma likelihood function is strongly recommended as a robust distribution for updating the parameters of the NB probabilistic models. Also, results from this study show that the full Bayesian updating framework for updating parameter estimates of probabilistic models is promising. However, the use of the estimates from the NB regression models (without updating) still led to favorable results, where the prediction accuracy was 78% for the 3-legged model, and 68% for the 4-legged model..

The analysis conducted in the sixth chapter attempted to provide deep insight into factors affecting crash injury severity at 3 and 4-legged unsignalized intersections using the most comprehensive data collected at those locations by using the ordered probit, binary probit and nested logit frameworks. The common factors found in the fitted probit models are the logarithm of AADT on the major road, and the speed limit on the major road. It was found that higher severity (and fatality) probability is always associated with a reduction in AADT, as well as an increase in speed limit.

The fitted probit models showed several important traffic, geometric and driver-related factors affecting safety at unsignalized intersections. Traffic factors include AADT on the major approach, and the number of through lanes on the minor approach (surrogate measure for AADT on the minor approach). Geometric factors include the upstream and downstream distance to the nearest signalized intersection, existence of stop lines, left and right shoulder width, number of left turn movements on the minor approach, and number of right and left turn lanes on the major approach. As for driver factors, young and very young at-fault drivers were always associated with the least fatal/severe probability compared to other age groups. Also, heavily-populated and highly-urbanized areas experience lower fatal/severe injury.

Comparing the aggregated binary probit model and the disaggregated ordered probit model showed that the aggregate probit model produces comparable if not better results, thus for its simplicity the binary probit models could be used to model crash injury severity at unsignalized intersections if the objective is to identify the factors contributing to severe injuries in general rather than the specific injury category. The nested logit models did not show any improvement over the probit models.

The seventh chapter investigated multiple applications of a new methodology "MARS" for analyzing motor vehicle crashes, which is capable of yielding high prediction accuracy. Rear-end and angle crashes were selected for the scope of the analysis and assessment.

The fitted NB rear-end regression models showed several important variables affecting safety at unsignalized intersections. These include traffic volume on the major road, the upstream and downstream distances to the nearest signalized intersection, median type on the major approach, land use at the intersection's influence area, and the geographic location within the state. For the NB angle crash models, the important factors include traffic volume on the major road, the upstream distance to the nearest signalized intersection, the distance between successive unsignalized intersections, median type on the major approach, percentage of trucks on the major approach, size of the intersection and the geographic location within the state.

MARS yielded the best prediction performance while dealing with continuous responses (either crash frequency normalized by the logarithm of AADT or the logarithm of crash frequency). Additionally, screening the covariates using random forest before fitting MARS model showed the best results. Hence, the MARS technique is recommended as a robust method for effectively predicting crashes at unsignalized intersections if prediction is the sole objective.

Finally, an access management analysis was performed with respect to six main median types associated with unsignalized intersections/access points. These six median types were open, closed, directional, two-way left turn lane, undivided and mixed medians. Also, crash conflict patterns at each of these six medians were identified and applied to a dataset including median-related crashes. In this case, separating median-related and intersection-related crashes was deemed significant in the analysis. From the preliminary analysis, open medians were considered the most hazardous median type, and closed and undivided medians were the safest.

It was concluded that single-vehicle crashes were the most probable median-related crash patterns, followed by right-angle crashes. Of the least probable crashes were head-on crashes. For open, directional and mixed medians, single-vehicle crashes were the most frequent, accounting for 60%, 70% and 38% of total crashes at those medians, respectively. For two-way left turn lane medians, right-angle crashes were the most frequent, accounting for 30%. As for undivided medians, left-turn and angle crashes were the most frequent, accounting for 35%.

The binomial logit and bivariate probit models demonstrated the importance of median-related variables affecting median-related crashes, such as median width, speed limit on the major road, logarithm of AADT, logarithm of the upstream and downstream distances to the nearest signalized intersection and crash pattern.

9.2 Research Applications

The results from this study from the different methodological approaches for analyzing safety at unsignalized intersections can be applicable to diagnose some safety deficiencies identified.

As a traffic application for alleviating crashes at intersections with only one stop sign, installing another stop sign on the left side of the minor road at those stop-controlled intersections might be useful. This countermeasure was examined by Polaris (1992), who found it to be effective in some cases.

Also, in order to increase drivers' awareness of the existence of stop signs, rumble strips can be installed at intersection approaches in order to call their attention. Rumble strips are usually recommended for application when measures such as pavement markings or flashers were tried and showed failure to alleviate high crash occurrence. Moreover, rumble strips can be coordinated with a "STOP AHEAD" device, i.e. when the driver crosses the rumble strip, this control device starts flashing.

Additionally, maintenance of stop signs should be performed at a high standard to ensure that their effectiveness is obtained. According to MUTCD, stop signs should be kept clean, and visible at all times (at day and night). Improper signs should be replaced without delay. Special care should be taken to make sure that trees, shrubs, and other vegetations do not block stop signs.

From the identification of various factors contributing to crash severity at unsignalized intersections using the probit modeling analysis, since it was found that prohibiting left turn maneuvers from the minor approaches reduces crash severity, hence, as an alternative, encouraging right turns from the minor approaches, followed by U-turns from the major road is very essential. This is consistent with the study done by Liu et al. (2007) who found that there is a reduction in total crashes and fatality for right turns followed by U-turns, as an alternative to direct left turn maneuvers from driveways. Prohibiting left turns from the minor approaches could be enforced by designing closed medians at the intersection's approach. This was also concluded from the access management analysis that closed medians are the safest median types.

Also, some countermeasures that can be dealt with to reduce injury severity at unsignalized intersections could be done by designing safety awareness campaigns encouraging speed control, and enforcement on speeding. Also, having a 90-degree intersection design is the most appropriate safety design for reducing severity. Moreover, making sure of marking stop lines at unsignalized intersections is essential.

From analyzing rear-end and angle crashes, since the increase in the upstream distance to the nearest signalized intersection from the unsignalized intersection of interest decreases both crash types, it is recommended to have a relatively large spacing between signalized and unsignalized intersections. The minimum spacing between both intersections (based on the analysis) from both the upstream and downstream sides is recommended to be around 0.38 to 0.46 miles. It was observed that a clear reduction in both crash types was gained after this distance range, and this effect was more obvious on rear-end crashes. Moreover, the least magnitude of crash fluctuation was observed after this specified range.

Since two-way left turn lanes were always associated with higher rear-end crashes at 4-legged unsignalized intersections, it is strongly recommended to avoid installing this median type at 4-legged intersections. A similar conclusion was reached from the access management analysis for another crash pattern, where two-way left turn lanes have right-angle crashes as the most dominant. As a remedy, installing closed medians could be useful, hence both intersections on both sides of the major road will be treated as at two separate 3-legged intersections. Another possible remedy is to install two-way left turn lanes at 3-legged unsignalized intersections only.

From the access management analysis, since open medians were always associated with the highest average crashes when compared to other median types (i.e., the most hazard median type), and closed median was the safest median, it is recommended to close median openings at most intersections. This indeed will help reduce traffic conflict points, hence, safety could be increased.

Since left-turn and angle crash patterns were the most dominant at undivided medians, it is recommended to avoid left turn maneuvers at unsignalized intersections having undivided medians at their approach. This could be enforced by installing a left-turn prohibition sign on both major and minor approaches. In this case, vehicles are only allowed to make a right turn maneuver.

Also, it is recommended to set back signalized intersections from restricted medians (i.e., closed, directional and mixed) across from driveways and unsignalized intersections to reduce median-related crash risk. Additionally, it is essential to separate left turning vehicles from through vehicles for a suitable deduction in restricted median-related crashes.

9.3 Further Research

From the reliability analysis in terms of the Bayesian updating concept, further research could be conducted to extend this work. This can be done by examining other distributions that can be used as likelihood functions for updating the parameter estimates of the NB model, such as the log-normal and beta distributions. Moreover, validating the updating procedure can be performed at some other locations rather than unsignalized intersections, such as signalized intersections, toll plazas and roadway segments.

From the crash severity analysis, although the work carried out provided useful information about various geometric, traffic and driver factors affecting crash injuries at 3 and 4-legged unsignalized intersections, further research could be conducted to extend this work. Since the probit models illustrated the significance of the spatial effect of the spacing between signalized and unsignalized intersections, analyzing unsignalized intersections along with the stretches linking them as one entity can be an encouraging prospect. This result suggests that spatial correlation between intersections exists, and unsignalized intersections should not be treated as isolated locations.

From the MARS analysis, even though the application of MARS models showed promising results, validating this method can be performed at some other locations rather than unsignalized intersections, such as signalized intersections and roadway segments. Additionally, using some other techniques for variables' screening (such as classification and regression trees "CART") before fitting a MARS model can be explored.

From the access management analysis, exploring other covariates such as the length of the median opening across from the driveway or unsignalized intersection might be useful. This could help formulate a broad perspective for the effect of wide and narrow median openings on traffic safety.

LIST OF REFERENCES

1. A Policy on Geometric Design of Highways and Streets, 4th Edition. AASHTO, Washington, D.C., 2001.
2. Abbess, C., Jarett, D. and Wright, C. Accidents at Blackspots: estimating the Effectiveness of Remedial Treatment, With Special Reference to the "Regression-to-Mean" Effect. Traffic Engineering and Control, Vol. 22, No. 10, pp. 535-542, 1981.
3. Abdel-Aty, M., Chen, C. and Schott, J. An assessment of the effect of driver age on traffic accident involvement using log-linear models. Accident Analysis and Prevention, 30(6), 851–861, 1998.
4. Abdel-Aty, M. and Radwan, E. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention, 32 (5), pp. 633-642, 2000.
5. Abdel-Aty, M. Analysis of driver injury severity levels at multiple locations using probit models. Journal of Safety Research, Vol. 34, No. 5, pp. 597-603, 2003.
6. Abdel-Aty, M. and Abdelwahab, H. Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. Accident Analysis and Prevention, 36, pp. 447-456, 2004.
7. Abdelwahab, H. and Abdel-Aty, M. Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. Transportation Research Record: Journal of the Transportation Research Board, No. 1784, TRB, National Research Council, Washington, D.C., pp. 115-125, 2002.
8. Abraham, A., Steinberg, D. and Philip, N. Rainfall Forecasting Using Soft Computing Models and Multivariate Adaptive Regression Splines. IEEE SMC Transactions:

Special Issue on Fusion of Soft Computing and Hard Computing in Industrial Applications, February 2001.

9. Agent, K. Traffic Control and Accidents at Rural High-Speed Intersections. Transportation Research Record 1160, Transportation Research Board, pp. 14-21, Washington D.C., 1988.
10. Agresti, A. An Introduction to Categorical Data Analysis. Wiley Series, 2nd Edition, 2007.
11. Aitchison, J. and Silvey, S. The Generalization of Probit Analysis to the Case of Multiple Responses. Biometrika, 44:131-40, 1957.
12. Akaike, H. A new look at the statistical model identification, IEEE Transactions on Automatic Control 19 (6): 716–723, 1974.
13. Anastasopoulos, P. and Mannering, F. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention, Vol. 41, pp. 153-159, 2009.
14. Attoh-Okine, N., Mensah, S. and Nawaiseh, M. A new technique for using multivariate adaptive regression splines (MARS) in pavement roughness prediction. Proceedings of the Institution of Civil Engineers, 156 (1), pp. 51-55, February 2003.
15. Bauer, K., and Harwood, D. Statistical models of at-grade intersection accidents. Report No. FHWA-RD-96-125, Federal Highway Administration, 1996.
16. Ben-Akiva, M. and Lerman, S. Discrete Choice Analysis: Theory Application to 29 Travel Dem. MIT Press, Cambridge, MA, 1985.
17. Bliss, C. The calculation of the dosage-mortality curve. Annals of Applied Biology, (22)134-167, 1935.

18. Bonneson, J. and McCoy, P. Effect of Median Treatment on Urban Arterial Safety. An Accident Prediction Model. Transportation Research Record 1581, pp.27-36, Transportation Research Board, National Research Council, Washington D.C., USA, 1997.
19. Bowman, B. and Vecellio, R. Investigation of the impact of medians on road users. U.S. Department of Transportation, Federal Highway Administration (McLean, Va. and Springfield, VA), 1994.
20. Box, G. and Tiao, G. Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Mass, 1992.
21. Breiman, L. Random Forests, Machine Learning, 45 (1), pp. 5-32, October 2001.
22. Briglia Jr., P. An evaluation of 4-way stop sign control. ITE Journal, 52, 16–19, 1982.
23. Brude, U. Traffic Safety at junctions, 3rd European Workshop on Recent Developments in Road Safety Research, VTI Rapport 366A, Statens Vaeg - Och Trafikinstitut, Linkoeping, Sweden, pp. 55-61, 1991.
24. Byrd, M. and Stafford, D. Analysis of delay and user costs at unwarranted four-way stop sign controlled intersections. Transportation Research Record, vol. 956, pp. 30–32, Washington, D.C., 1984.
25. Caliendo, C., Guida, M. and Parisi, A. Crash-prediction models for multilane roads. Accident Analysis and Prevention, Vol. 39, No. 4, pp. 657-670, 2007.
26. California Department of Public Works. Evaluation of Minor Improvements: Flashing Beacons, Safety Lighting, Left-Turn Channelization, Traffic Department, 1967.

27. Chang, L. and Mannering, F. Analysis of Injury Severity and Vehicle Occupancy in Truck- and Non-truck-Involved Accidents. Accident Analysis and Prevention, 31, pp. 579–592, 1999.

28. Chin, H. and Quddus, M. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis and Prevention, 35, pp. 253-259, 2003.

29. Chovan, J., Tijerina, L., Pierowicz, J. and Hendricks, D. Examination of Unsignalized Intersection, Straight Crossing Path Crashes and Potential IVHS Countermeasures. Report No. DOT-HS-808 152, Washington, D.C., National Highway Traffic Safety Administration, 1994.

30. Chovan, J., Tijerina, L., Everson, J., Pierowicz, J. and Hendricks, D. Examination of Intersection, Left Turn Across Path Crashes and Potential IVHS Countermeasures. Report No. DOT-HS-808 154, Washington, D.C., National Highway Traffic Safety Administration, 1994.

31. Crash Analysis Reporting System. http://tlhost01.dot.state.fl.us/bluezone/FDOT_Session/default.htm.

32. Cribbins, P., Horn, J., Beeson, F. and Taylor, R. Median Openings on Divided Highways: Their Effect on Accident Rates and Level of Service. Highway Research Record 188, Highway Research Board, 1967.

33. Datta, K. Head-On, Left-Turn Accidents at Intersections with Newly Installed Traffic Signals. Transportation Research Record No. 1318, TRB, The National Academies, Washington, D.C., pp. 58–63, 1991.

34. Datta, K. and Dutta, U. Traffic Signal Installation and Accident Experience. ITE Journal, Washington, D.C., pp. 39–42, September 1990.

35. David, N. and Norman, J. Motor vehicle accidents in relation to geometric and traffic features of highway intersections: Vol. II - Research report. Report No. FHWA-RD-76- 129, Federal Highway Administration, 1976.

36. David, N. and Norman, J. Motor Vehicle Accidents in Relation to Geometric and Traffic Features of Highway Intersections, Report Number FHWA-RD-76-128, Federal Highway Administration, Washington D.C., USA, p. 236, 1975.

37. De Veaux, D., Psichogios, R. and Ungar, L. Comput. Chem. Eng. 17, 819, 1993.

38. Dean, C. Modified Pseudo-Likelihood Estimator of the Over-dispersion Parameter in Poisson Mixture Models. Journal of Applied Statistics, 21 (6), pp. 523-532, 1994.

39. Del Mistro, R. Accidents at Urban Intersections Tee Versus Cross Intersections, Technical Report Number RF/4/79, National Institute for Transport and Road Research, Pretoria, South Africa, p. 18, 1979.

40. Del Mistro, R. Accidents at Urban IntersectionsA Second Study, Technical Report Number RF/5/81, National Institute for Transport and Road Research, Pretoria, South Africa, p. 59, 1981.

41. Dissanayake, S. and Lu, J. Access Management Techniques to Improve Traffic Operations and Safety: A Case Study of a Full vs. Directional Median Opening. Proceedings of the 2003 Mid-Continent Transportation Research Symposium, Ames, Iowa, August 2003.

42. Duncan, C., Khattak, A. and Council, F. Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions. Transportation Research

Record, Vol. 1635 (pp. 63–71), Washington, DC: Transportation Research Board, National Research Council, 1998.

43. Effects of Signalization on Intersection Safety. Safety Operations Unit, Publication 8, Traffic and Safety Division, New York State Department of Transportation, January 1982.
44. El-Basyouny, K. and Sayed, T. Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models. Paper Presented at the Transportation Research Board 85th annual meeting, Washington, D.C., 2006.
45. Elvik, R. Predictive Validity of Empirical Bayes Estimates of Road Safety. Transportation Research Board 87th annual meeting, No. 0220, Washington, D.C., 2008.
46. Fambro, D. Geometric design consideration for rural roads. Report 1 125- 1 F, Texas Transportation Institute, 1989.
47. Fatality Analysis Reporting System. http://www.fars.k6ya.org/fd1999/.
48. Florida Department of Highway Safety and Motor Vehicles. Traffic Crash Statistics Report, 2003, http://www.hsmv.state.fl.us.
49. Florida Department of Highway Safety and Motor Vehicles. Traffic Crash Statistics Report, 2005, http://www.hsmv.state.fl.us.
50. Florida Department of Highway Safety and Motor Vehicles. Traffic Crash Statistics Report, 2006, http://www.hsmv.state.fl.us.
51. Florida Department of Highway Safety and Motor Vehicles. Traffic Crash Statistics Report, 2007, http://www.hsmv.state.fl.us.

52. Florida Department of Highway Safety and Motor Vehicles. http://www.flhsmv.gov/html/safety.html.

53. Florida Department of Transportation Map (2007). http://www.dot.state.fl.us/safety/SRTS_files/SRTSCoordinators1_18_07Map2a.gif.

54. Florida Department of Transportation. http://www.dot.state.fl.us/.

55. Florida Department of Transportation Median Handbook, Tallahassee, Florida, 1997.

56. Florida Department of Transportation Median Handbook, Interim Version, February 2006.

57. Florida Department of Transportation. Florida Strategic Highway Safety Plan, Florida Department of Transportation, 2006.

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

59. Francis, L. Martial chronicles. Is MARS better than neural networks. Casualty Actuarial Society Forum, Winter 2003, 27-54, 2003.

60. Friedman, J. Multivariate adaptive regression splines. Ann Stat 19:1–141, 1991.

61. Gardoni, P., Der Kiureghian, A. and Mosalam, K. Probabilistic Models and Fragility Estimates for Bridge Components and Systems. PEER Report No. 2002/13. Pacific Earthquake Engineering Research Center, Berkeley, CA, 2002.

62. Gattis, J. and Low, S. Intersection Angle Geometry and the Driver's Field of View. Transportation Research Board 77th annual meeting, No. 1612, Washington, D.C., 1998.

63. Geedipally, S. and Lord, D. Effects of Varying Dispersion Parameter of Poisson-gamma Models on Estimation of Confidence Intervals of Crash Prediction Models.

Transportation Research Board 87th annual meeting, No. 1563, Washington, D.C., 2008.

64. Gluck, J., Levinson, H. and Stover, V. Impacts of Access Management Techniques. NCHRP Report 420, TRB, National Research Council, Washington, D.C., 1999.
65. Google Earth. http://earth.google.com/.
66. Greene, W. Econometric Analysis. Pearson Education, Inc., 5th Edition, 2003.
67. Grimm, R., Behrens, T., Marker, M. and Elsenbeer, H. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma, Vol. 146, No. 1-2, pp. 102-113, July 2008.
68. Haberman, J. Analysis of dispersion of multinomial responses. Journal of American Statistical Association, 77, pp. 568–580, 1982.
69. Hall, J. Economic Benefit of Accident Reductions. Prepared for the 1998 Annual Meeting of the Institute of Transportation Engineers, Toronto, Canada, 1998.
70. Hanna, J., Flynn, T. and Webb, L. Characteristics of intersection accidents in rural municipalities. Transportation Research Record 601, Transportation Research Board, 1976.
71. Harb, R., Yan, X., Radwan, E. and Su, X. Exploring precrash maneuvers using classification trees and random forests. Accident Analysis and Prevention, 41 (1), pp. 98-107, January 2009.
72. Harwood, D. NCHRP Synthesis of Highway Practice 191: Use of Rumble Strips to Enhance Safety, Transportation Research Board of the National Academies, 1993.

73. Harwood, D., Pietrucha, M., Wooldridge, M., Brydia, R. and Fitzpatrick, K. Median intersection design. NCHRP Report 375, National Cooperative Highway Research Program, 1995.

74. Harwood, D., Bauer, K., Potts, I., Torbic, D., Richard, K., Rabbani, E., Hauer, E. and Elefteriadou, L. Safety effectiveness of intersection left- and right-turn lanes. Report No. FHWARD-02-089, Federal Highway Administration, 2002.

75. Harwood, D., Council, F., Hauer, E., Hughes, W. and Vogt, A. Prediction of the expected safety performance of rural two-lane highways, Federal Highway Administration, Final Report, FHWA-RD-99-207, 2000.

76. Hauer, E. Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Elsevier Science Ltd, Oxford, 1997.

77. Hauer, E. The Safety of Older Persons at Intersections. Transportation in an Aging Society: Improving Mobility and Safety for Older Persons, Volume 2, Special Report 218, Transportation Research Board, pp. 194-252, 1988.

78. Hauer, E., Ng, JCN and Lovell, J. Estimation of safety at signalized intersections. Transportation Research Record, No. 1185, pp. 48–61, 1988.

79. Hauer, E. and Lovell, J. The safety effect of conversion to all-way stop control. Transportation Research Record, vol. 1068, pp. 103–107, Washington, D.C., 1986.

80. Hauer, E. Over-dispersion in Modeling Accidents on Road Sections and in Empirical Bayes Estimation. Accident Analysis and Prevention, 33 (6), pp. 799–808, 2001.

81. Heydecker, B. and Wu, J. Identification of sites for road accident remedial work by Bayesian statistical methods: an example of uncertain inference. Advances in Engineering Software 32, 859–869, 2001.

82. Huang, Y. and May, A. Accident Prediction Models and Applications for Unsignalized and Signalized Intersections, Intersections Without Traffic Signals II, Springer-Verlag, Berlin, Deutschland, Bochum, Germany, pp. 282-296, 1991.

83. Hummingbird Intranet Website. http://dotscoedms1.co.dot.state.fl.us/cyberdocs/cyberdocs.asp.

84. Hutchinson, T. Statistical Modeling of Injury Severity, with Special Reference to Driver and Front Seat Passenger in Single-Vehicle Crashes. Accident Analysis and Prevention, 18(2), 157– 167, 1986.

85. Imai, K. and Van Dyk, D. A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation. Journal of Econometrics, 124(2), pp. 311–334, 2005.

86. Imai, K. and Van Dyk, D. MNP: R Package for Fitting the Multinomial Probit Models. Journal of Statistical Software, 14(3), pp.1-31, 2005.

87. Ivan, J. and O'Mara, P. Prediction of Traffic Accident Rates Using Poisson Regression. Paper Presented at the 76th Annual Meeting of the Transportation Research Board, Washington, D.C., 1997.

88. Jonsson, T., Lyon, C., Ivan, J., Washington, S., Schalkwyk, I. and Lord, D. Investigating Differences in the Performance of Safety Performance Functions Estimated for Total Crash Count and for Crash Count by Crash Type. Paper Presented at the Transportation Research Board 88th annual meeting, Washington, D.C., 2009.

89. Kang, Y. Delay, Stop and Queue Estimation for Uniform and Random Traffic Arrivals at Fixed-Time Signalized Intersections. PhD Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, April 2000.

90. Khattak, A. Intersection safety, Nebraska Department of Roads. Research Project Number SPR-l (2) P544-SJ0105, March 2006.

91. King, G. and Goldblatt, R. Relationship of Accident Patterns to Type of Intersection Control. Transportation Research Record No. 540, TRB, The National Academies, Washington, D.C., pp. 1–12, 1975.

92. King, G. and Goldblatt, R. Urban Accident Patterns. Transportation Research Record 540, Transportation Research Board, Washington, D.C., pp. 1-12, 1975.

93. Kitto, H. Accident Rate at Urban Right-Angle Intersections. Road Research Unit Bulletin 48, National Roads Board, Wellington, NZ, p. 94, 1980.

94. Klop, J. and Khattak, A. Factors Influencing Bicycle Crash Severity on Two-Lane, Undivided Roadways in North Carolina. Transportation Research Record, Vol. 1674, pp. 78–85, Washington, D.C., 1999.

95. Kockelman, K. and Kweon, Y. Driver injury severity: an application of ordered probit models. Accident Analysis and Prevention, 34(4), 313–321, 2002.

96. Koepke, F. and Levinson, H. Access Management Guidelines for Activity Centers. NCHRP Report 348, National Cooperative Highway Research Program, Transportation Research Board, 1992.

97. Kuciemba, S. and Cirillo, J. Safety Effectiveness of Highway Design Features, Volume V: Intersections. FHWA-RD-91-048, Federal Highway Administration, Washington, D.C., 1992.

98. Kuhn, S., Egert, B., Neumann, S. and Steinbeck, C. Building blocks for automated elucidation of metabolites: Machine learning methods for NMR prediction. BMC Bioinformatics, 9 (400), September 2008.
99. Kulmala, R. Safety at Rural Three- and Four-Arm Junctions: Development and Applications of Accident Prediction Models. VTT Publications 233, Technical Research Centre of Finland, Espoo, 1995.
100. Kulmala, R. Safety at Highway Junctions Based on Predictive Accident Models, Third International Symposium on Intersections without Traffic signals, Portland, USA, pp. 151-157, 1997.
101. Lahsen, J., Schmidhammer, H. and Rode, B. Helv. Chim. Acta 84, 3299, 2001.
102. Lau, M., and May, A. Accident Prediction Model Development: Signalized Intersections, Research Report No. UCB-ITS-RR-88-7, Institute of Transportation Studies, University of California-Berkeley, December 1988.
103. Lau, M. and May, A. Accident Prediction Model Development: Unsignalized Intersections, Research Report No. UCB-ITS-RR-89-12, Institute of Transportation Studies, University of California-Berkeley, May 1989.
104. Layfield, R. Accidents at Urban Priority Crossroads and Staggered Junctions. TRL Report 185, Transport Research Laboratory, Crowthorne, UK, p. 120, 1996.
105. Leathwick, J., Rowe, D., Richardson, J . Elith, J. and Hastie, T. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Journal of Freshwater Biology, Vol. 50, 2034–2052, 2005.
106. Leong, H. Relationship between accidents and traffic volumes at urban intersections, Australian Road Research, Volume 5, Number 3, pp. 72-90, 1973.

107. Levinson, H., Potts, I., Harwood, D., Gluck, J. and Torbic, D. Safety of U-Turns at Unsignalized Median Openings: Some Research Findings. Transportation Research Record 1912, Transportation Research Board, Washington, D.C., pp. 72-81, 2005.

108. Li, X. , Lord, D., Zhang, Y. and Xie, Y. Predicting motor vehicle crashes using Support Vector Machine models. Accident Analysis and Prevention, 40 (4), pp. 1611-1618, 2008.

109. Liu, P., Lu, J. and Chen, H. Safety effects of the separation distances between driveway exits and downstream U-turn locations. Accident Analysis and Prevention, 40, 760–767, 2007.

110. Lord, D. The Prediction of Accidents on Digital Networks: Characteristics and Issues Related to the Application of Accident Prediction Models. Ph.D. Dissertation. Department of Civil Engineering, University of Toronto, Toronto, Ontario, 2000.

111. Lord, D. and Bonneson, J. Development of Accident Modification Factors for Rural Frontage Road Segments in Texas. Zachry Department of Civil Engineering, Texas A&M University, College Station, TX, 2006.

112. Lord, D. and Mahlawat, M. Examining the Application of Aggregated and Disaggregated Poisson-gamma Models Subjected to Low Sample Mean Bias. Paper Presented at the Transportation Research Board 88th annual meeting, No. 1290, Washington, D.C., 2009.

113. Lord, D. and Persaud, B. Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations (GEE) Procedure. Transportation Research Record, No. 1717, pp. 102-108, 2000.

114. Lord, D., Manar, A. and Vizioli, A. Modeling Crash-Flow-Density and Crash-Flow-V/C Ratio Relationships for Rural and Urban Freeway Segments. Accident Analysis and Prevention, 37 (1), pp. 185–199, 2005.

115. Lord, D., Washington, S. and Ivan, J. Poisson, Poisson-gamma and Zero Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. Accident Analysis and Prevention, Vol. 37, No. 1, pp. 35-14, 2005.

116. Lord, D. Modeling Motor Vehicle Crashes using Poisson-gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter. Accident Analysis and Prevention, 38 (4), pp. 751-766, 2006.

117. Lu, J., Liu, P. and Pirinccioglu, F. Determination of the Offset Distance between Driveway Exits and Downstream U-turn Locations for Vehicles making Right Turns Followed by U-turns, University of South Florida, November 2005.

118. Lu, J., Dissanayake, S. and Xu, L. Safety Evaluation of Right-Turns followed by U-turns as an Alternative to Direct Left Turns: Crash Data Analysis. Florida Department of Transportation, 2001 a.

119. Lu, J., Dissanayake, S. and Castillo, N. Safety Evaluation of Right-Turns followed by U-turns as an Alternative to Direct Left Turns: Conflict Data Analysis. Florida Department of Transportation, 2001 b.

120. Lu, J., Pirinccioglu, F. and Pernia, J. Safety Evaluation of Right-Turns followed by U-turns at Signalized Intersection (six or more lanes) as an Alternative to Direct Left Turns: Conflict Data Analysis. Florida Department of Transportation, 2004.

121. Lu, J., Pirinccioglu, F. and Pernia, J. Safety Evaluation of Right-Turns followed by U-turns (4-lane arterials) as Alternatives to Direct Left Turns: Conflict Data Analysis. Florida Department of Transportation, 2005.

122. Maher, M. and Summersgill, I. A comprehensive methodology for the fitting of predictive accident models. Accident Analysis and Prevention, 28, pp. 281-296, 1996.

123. Malyshkina, N. and Mannering, F. Analysis of the effect of speed limit increases on accident injury severities. Transportation Research Record, Vol. 2083, pp. 122–127, Washington, D.C., 2008.

124. Marek, J., Kyte, M., Zongzhong, T., Lall, K. and Voigt, K. Determining Intersection Traffic Control type Using the 1994 Highway Capacity Manual. ITE Journal, August 1997.

125. Margiotta, R. and Chatterjee, A. Accidents on Suburban Highways –Tennessee's Experience. Journal of Transportation Engineering, pp. 255-261, May/June 1995.

126. McCoy, P., Hoppe, W. and Dvorak, D. Benefit-Cost Evaluation of Left-Turn Lanes on Uncontrolled Approaches of Rural Intersections. Transportation Research Record 1026, Transportation Research Board, 1985.

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

128. McCoy, P., Tripi, E. and Bonneson, J. Guidelines for realignment of skewed intersections, Nebraska Department of Roads Research Project Number RES 1 (0099) P47 1, 1994.

129. McDonald, J. Relation between Number of Accidents and Traffic Volume at Divided-Highway Intersections. Bulletin 74, HRB, National Research Council, Washington, D.C., 1953.

130. McFadden, D. Modeling the Choice of Residential Location. Transportation Research Record 672, TRB, pp. 72-77, Washington, D.C., 1978.

131. McFadden, D. Econometric model of probabilistic choice. In Structural analysis of discrete data with econometric applications, Manski, McFadden, (Eds). Cambridge, MA: MIT Press, 1981.

132. McKelvey, R. and Zavoina, W. A Statistical Model for the Analysis of Ordinal Level Variables. Journal of Mathematical Sociology, 4:103-20, 1975.

133. Miaou, S. Measuring the Goodness-of-Fit of Accident Prediction Models, FHWA-RD-96-040, Federal Highway Administration, Washington, D.C., 1996.

134. Miaou, S. The relationship between truck accidents and geometric design of road sections: poisson versus negative binomial regressions. Accident Analysis and Prevention, 26, pp. 471–482, 1994.

135. Miaou, S. and Lord, D. Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes Methods. Transportation Research Record, No. 1840, pp. 31–40, 2003.

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

137. Miranda-Moreno, L., Fu, L., Saccomanno, F. and Labbe, A. Alternative Risk Models for Ranking Locations for Safety Improvement. Transportation Research Record, No. 1908, pp. 1–8, 2005.

138. Mitchell, R. Identzfiing and improving highway accident locations. PubIic Works, December 1972.

139. Mountain, L., Maher, M. and Fawaz, B. The influence of trend on estimates of accidents at junctions. Accident Analysis and Prevention, 30, pp. 641-649, 1998.

140. Mussone, L., Ferrari, A. and Oneta, M. An Analysis of Urban Collisions Using an Artificial Intelligence Model. Accident Analysis & Prevention, Vol. 31, No. 6, pp. 705-718, 1999.

141. Nagelkerke, N. Maximum Likelihood Estimation of Functional Relationships, Pays-Bas, Lecture Notes in Statistics, Volume 69, 110p ISBN 0-387-97721-X, 1992.

142. Najm, W., Mironer, M., Koziol, J., Wang, J.-S. and Knipling, R. Synthesis Report: Examination Of Target Vehicular Crashes And Potential ITS Countermeasures. Report No. DOT HS 808 263, Washington, D.C., National Highway Traffic Safety Administration, 1995.

143. Najm, W., Smith, J. and Smith, D. Analysis of Crossing Path Crashes (Report No. DOT HS 809 423), Washington, DC: National Highway Traffic Safety Administration, 2001.

144. Nassar, S., Saccomanno, F. and Shortreed, J. Road Accident Severity Analysis: A Micro Level Approach. Canadian Journal of Civil Engineering, 21, 847–855, 1994.

145. Nataf, A. Determination des Distribution don't les Marges sont Donnees, Comptes Rendus de l' Academie des Sciences, Paris, 225, 42-43, 1962.

146. Noland, R. and L. Oh. The Effect of Infrastructure and Demographic Change on Traffic-related Crashes: A Case Study of Illinois County-level Data. Accident Analysis and Prevention, 36, 525–532, 2004.

147. Nguyen-Cong, V., Van Dang, G., Rode, B. and Eur. J. Med. Chem. Eng. 31, 797, 1996.

148. O'Brien, A. Some Safety Considerations for Rural Intersections, 8th Australian Road Research Board Conference, Volume 8, Australian Road Research Board, Nunawading, Australia, Perth, Australia, pp. 1-9, 1976.

149. Oh, J., Lyon, C., Washington, S., Persaud, B. and Bared, J. Validation of the FHWA Crash Models for Rural Intersections: Lessons Learned. Transportation Research Record, No. 1840, pp. 41-49, 2003.

150. Pant, P. and Park, Y. Comparative Study of Rural Stop-controlled and Beacon-controlled Intersections, Transportation Research Record 1692, Transportation Research Board, Washington, D.C., USA, pp. 164-172, 1999.

151. Park, B. and Lord, D. Adjustment for Maximum Likelihood Estimate of Negative Binomial Dispersion Parameter. Transportation Research Board 87th annual meeting, No. 1701, Washington, D.C., 2008.

152. Park, E. and Lord, D. Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. Transportation Research Record, No. 2019, pp. 1–7, 2007.

153. Persaud. B. and Dzbik, L. Accident prediction models for freeways. Transportation Research Record, No. 1401, pp. 55–60, 1993.

154. Persaud, B, Lan, B., Lyon, C. and Bhim, R. Comparison of Empirical Bayes and Full Bayes Approaches for Before-after Road Safety Evaluations. Paper Presented at the Transportation Research Board 88th annual meeting, No. 2593, Washington, D.C., 2009.

155. Phillips, S. Empirical Collision Model for Four-Lane Median Divided and Five-Lane with TWLTL Segments. Ms Thesis, North Carolina State University, Raleigh, NC, 2004.

156. Pickering, D. and Hall, R. Accidents at Rural T-junctions. Research Report 65, Transport and Road Research Laboratory, Crowthorne, UK, p. 39, 1986.

157. Piegorsch, W. Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter. Biometrics, 14 (3), pp. 863-867, 1990.

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

159. Polanis, S. Reducing traffic accidents through traffic engineering. Transportation Quarterly, 46, 235– 242, 1992.

160. Potts, I., Harwood, D., Torbic, D., Richard, K., Gluck, J., Levinson, H., Garvey, P. and Ghebrial, R. Safety of U-Turns at Unsignalized Median Openings. NCHRP Report 524, TRB, National Research Council, Washington, D.C., 2004.

161. Powers, M. and Carson, J. Before-After Crash Analysis: A Primer for Using the Empirical Bayes Method. Report for the Montana Department of Transportation, 2004.

162. Press, S. Applied Multivariate Analysis. University of Chicago. New York: Holt, Rinehart and Winston, Inc., 1972.

163. Priest, R. Statistical Relationships between Traffic Volume, Median Width, and Accident Frequency on Divided Highway Grade Intersections. Highway Research News, No. 13, June 1964.

164. Put, R., Xu, Q., Massart, D. and Heyden, Y. Multivariate adaptive regression splines (MARS) in chromatographic quantitative structure–retention relationship studies. Journal of Chromatography A, 1055, pp. 11–19, 2004.

165. Quddus, M., Noland, R. and Chin, H. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research, 33(4), 445–462, 2002.

166. R Software. http://www.r-project.org/.

167. Renski, H., Khattak, A. and Council, F. Impact of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways. Paper Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C, 1998.

168. Retting, R., Weinsteinb, H. and Solomonb, M. Analysis of motor-vehicle crashes at stop signs in four U.S. cities. Journal of Safety Research, 34, 485– 489, 2003.

169. Riviere, C., Lauret, P., Ramsamy, J. and Page, Y. A Bayesian Neural Network Approach to Estimating the Energy Equivalent Speed. Accident Analysis & Prevention, Vol. 38, No. 2, pp. 248-259, 2006.

170. Roadway Characteristic Inventory. http://webapp01.dot.state.fl.us/Login/default.asp.

171. SAS Institute Inc. Version 9 of the SAS System for Windows. Cary, NC, 2002.

172. Saccomanno, F., Grossi, R., Greco, D. and Mehmood, A. Identifying black spots along Highway SS107 in southern Italy using two models. Journal of Transportation Engineering, 127 (6), 515–522, 2001.

173. Saha, K. and Paul, S. Bias-Corrected Maximum Likelihood Estimator of the Negative Binomial Dispersion Parameter. Biometrics, 61 (1), pp. 179-185, 2005.

174. Salman, N. and Al-Maita, K. Safety Evaluation at Three-Leg, Unsignalized Intersections by Traffic Conflict Technique. Transportation Research Board, Vol. 1485, pp.177-185, 1995.

175. Savolainen, P. and Mannering, F. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. Accident Analysis and Prevention, 39, pp. 955–963, 2007.

176. Sawalha, Z. and Sayed, T. Transferability of accident prediction models. Journal of Safety Science, 44 (3), pp. 209–219, 2006.

177. Sawalha, Z. and Sayed, T. Traffic accidents modeling: some statistical issues. Canadian Journal of Civil Engineering, 33 (9), pp. 1115–1124, 2006.

178. Sayed, T. and Rodriguez, F. Accident Prediction Models for Urban Unsignalized Intersections in British Columbia, Transportation Research Board, Vol. 1692, pp. 30-38, 1999.

179. Sayed, T. and Zein, S. Traffic Conflict Standards for Intersections, Transportation Planning and Technology, Vol. 22, pp. 309-323, 1999.

180. Sekulic, S. and Kowalski, B. MARS - A Tutorial J Chemometr 6 (4) 199-216, July-August 1992.

181. Sephton, P. Forecasting Recessions: Can We Do Better on MARS? Federal Reserve Bank of St. Louis Review, 83, 39–49, 2001.

182. Shankar, V., Mannering, F. and Barfield, W. Statistical Analysis of Accident Severity on Rural Freeways. Accident Analysis and Prevention, 28(3), pp. 391–401, 1996.

183. Shankar, V., Milton, J. and Mannering, F. Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical Inquiry. Accident Analysis & Prevention, Vol. 29, No. 6, pp. 829-837, 1997.

184. Stockton, W. and Bracckett, R. Stop, Yield, and No Control at Intersections, Report Number FHWARD-81/084, Federal Highway Administration, Washington, D.C., p. 102, 1981.

185. Strategies to Address Nighttime Crashes at Rural, Unsignalized Intersections. Center for Transportation Research and Education, Iowa State University, February 2008.

186. Summersgill, I. and Kennedy, J. Accidents at Three-Arm Priority functions on Urban Single Carriageway Roads. TRL Report 184, Transport Research Laboratory, Crowthorne, UK, p. 74, 1996.

187. Templer, J. Provisions for elderly and handicapped pedestrians, Volume 3: The development and evaluation of countermeasures. Report No. FHWA/RD-7913, Federal Highway Administration, 1980.

188. The New Web-based RCI Application. http://webservices.camsys.com/trbcomm/docs/presfla2005.htm.

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

190. Traffic Safety Toolbox. A Primer on Traffic Safety. ITE, Washington, D.C., 1987.

191. Transportation Engineering, Inc. Collision Analysis for District Wide Median Evaluation and Public Involvement, Florida Department of Transportation, 1995.

192. U.S. Census Bureau Profile of General Demographic Characteristics 2000. http://www.census.gov/statab/ranks/rank39.html.

193. U.S. Department of Transportation. Traffic safety facts, Report no. DOT HS-809-337, Washington, D.C., 2002.

194. Van Maren, P. Correlation of design and control characteristics with accidents at rural multilane highway intersections. Purdue University and Indiana State Highway Commission, July 1980.

195. Video Log Viewer Application. http://webapp01.dot.state.fl.us/videolog/.

196. Vogt, A. Crash Models for Rural Intersections: Four-Lane by Two-Lane Stop-Controlled and Two-Lane by Two-Lane Signalized. Report No. FHWA-RD-99-128, Federal Highway Administration, October 1999.

197. Vogt, A. and Bared, J. Accident Models for Two-Lane Rural Roads: Segments and Intersections, Report Number FHWA-RD-98-133, No. 3A5A, Final Report, Federal Highway Administration, McLean, USA, p. 179, 1998.

198. Walker, F. and Roberts, S. Influence of Lighting on Accident Frequency at Highway Intersections, Transportation Research Record 562, Transportation Research Board, Washington, D.C., pp. 73-78, 1976.

199. Wang, J. and Knipling, R. Intersection Crossing Path Crashes: Problem Size Assessment And Statistical Description. Report No. DOT-HS-808-190, Washington, D.C., National Highway Traffic Safety Administration, 1994.

200. Wang, X. Safety Analyses at Signalized Intersections Considering Spatial, Temporal and Site Correlation. PhD Dissertation, University of Central Florida, Orlando, Florida, 2006.

201. Wang, X. and Abdel-Aty, M. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention, 38, pp. 1137-1150, 2006.

202. Washington, S., Gibby, A., and Ferrara, T. Evaluation of high-speed isolated intersections in California, Report No. FHWAlCA/T0/9 1-2, Federal Highway Administration, June 1991.

203. Weerasuriya, S. and Pietrzyk, M. Development of Expected Conflict Value Tables for Unsignalized Three-Legged Intersections. Transportation Research Record No. 1635, TRB, pp. 121-126, 1998.

204. Wolfram Mathematica 6. http://www.wolfram.com/products/mathematica/index.html.

205. Xie, Y., Lord, D. and Zhang, Y. Predicting Motor Vehicle Collisions Using Bayesian Neural Network Models: An Empirical Analysis. Accident Analysis & Prevention, Vol. 39, No. 5, pp. 922-933, 2007.

206. Xiong, R. and Meullenet, J. Application of Multivariate Adaptive Regression Splines (MARS) to the Preference Mapping of Cheese Sticks. Journal of Food Science, Vol. 69, 2004.

207. Zhang, Y., Ye, Z. and Lord, D. Estimating the Dispersion Parameter of the Negative Binomial Distribution for Analyzing Crash Data Using a Bootstrapped Maximum Likelihood Method Paper Presented at the Transportation Research Board 86th annual meeting, Washington, D.C., 2007.

208. Zuur, A., Ieno, E. and Smith, G. Analyzing Ecological Data. Statistics for Biology and Health, New York, 2007.