Final Report

A Model-Based Risk Map for Roadway Traffic Crashes

Project: MBTC 2098

Submitted to:

Mack-Blackwell Rural Transportation Center

Date: July 29, 2008

Contact:

Chang S. Nam, Ph.D., CHFP
Assistant Professor
Department of Industrial Engineering
University of Arkansas
4207 Bell Engineering Center
Fayetteville, AR 72701
Phone: (479) 575 - 2563
Fax: (479) 575 - 8431
Email: cnam@uark.edu

Joon J. Song, Ph.D.
Assistant Professor of Statistics
Department of Mathematical Sciences
University of Arkansas
SCEN 309B
Fayetteville, AR 72701
Phone: (479) 575-6319
Fax: (479) 575-8630
Email: jjsong@uark.edu

ACKNOWLEGEMENTS

We wish to thank the following people for their assistance in providing access to documents needed to perform this research.

- Mark K. Bradley, Staff Research Engineer, Arkansas State Highway & Transportation Department
- Marc A. Mauer, Civil Engineer, Arkansas State Highway & Transportation Department
- Karen Bonds, Traffic Records Program Manager / FARS Supervisor, Highway Safety Office
- Jared Wiley, Planning & Research Division, Arkansas State Highway & Transportation Department
- Greg Nation, Planning & Research Division, Arkansas State Highway & Transportation Department
- Jon Waldrip, Head of Traffic Safety Section, Highway Safety Office, Arkansas State Police

We also wish to acknowledge the work of the following students that comprises parts of this study.

- Pingjian Yu, Graduate Student of Industrial Engineering, University of Arkansas
- Yohannes T. Tekle, Graduate Student of Mathematical Sciences, University of Arkansas
- Jacob Mercer, Graduate Student of Industrial Engineering, University of Arkansas

The support of the Mack-Blackwell National Rural Transportation Study Center at the University of Arkansas and assistance from the Department of Industrial Engineering also made this research possible.

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

A Model-Based Risk Map for Roadway Traffic Crashes

Project Objectives

Visualization of traffic safety data that transforms spatial data into a visual form can help highway engineers and traffic safety officials to effectively analyze the data and make decisions on which roadways and road side features to improve by providing the spatial distribution of the data. However, research efforts in the visualization of traffic safety data, which are usually stored in a large and complex database, are quite limited because of methodological constraints (Miaou and Song, 2005b; Miaou, Tandon, and Song, 2005; Smith, Harkey, and Harris, 2001). For example, there are only a few model-based maps that can account for the high variance of traffic crash estimates in low population areas, and at the same time clarify overall geographic trends and patterns. In addition, designers of roadways historically did not take into account the full range of driver characteristics, such as driver perception-response time, age differences, etc. (Dewar and Olson, 2002). One of the most important components of the roadway transportation is the human driver whose error is a factor in about 90% of traffic crashes (Treat et al., 1977). Therefore, it is very important for highway engineers and traffic safety officials to identify and understand the basics of human factors as relevant to driving and traffic safety.

To address these two issues, we conducted two studies:

1. The objectives of Study 1 were to generate a data set based on several data sets including traffic safety data, highway inventory data, and GIS data, to conduct exploratory data analysis, and to perform spatial analysis, with exploratory spatial data analysis and Bayesian spatial modeling to estimate and map crash risk.
2. The objective of Study 2 was to evaluate closely the issues involved with road traffic safety in the state of Arkansas.

Study 1: Statistical Analysis of Arkansas Traffic Crash Data

Introduction

In recent years, many methodologies for mapping quantity of interest have been developed in a wide range of fields, such public health, social sciences, and engineering. An example is diseasing mapping in public health policy studies, which has been used to describe the spatial variation in disease incidence, to identify areas or locations with potentially elevated disease risk, and provide an informative map for risk assessment, prioritization, and resource allocation in order to improve disease risk (Carline and Louis, 1996; Xia et al., 1997; Ghosh et al., 1999; Lawson et al., 1999; Dey et al., 2000; Sun et al., 2001; Miaou et al., 2003). In traffic safety planning, mapping traffic crash risk corresponds to disease mapping in public health. For traffic crash risk mapping, Miaou et al. (2003) built model-based risk maps for county-level traffic crashes in Texas using hierarchical Bayes models. Aguero-Valverde and Jovanis (2006) also developed full Bayes hierarchical spatial models of county-level road crash frequency, with socioeconomic, transportation-related, and environmental factors. Several multivariate approaches were proposed to deal with crash counts by severity or location in which a traffic crash occurs (Miaou and Song, 2005; Song et al., 2006; Park and Lord, 2007; Ma et al., 2008). For identifying sites on a road network, which potentially present high traffic crash risks, ranking sites based on some criteria is a popular approach. A simple and straightforward way is to rank them according to the number of crashes. However, this method has several drawbacks (Miranda-Moreno et al., 2005). In order to overcome the shortcomings, several different ranking criteria have been proposed to identify hazardous sites for further engineering evaluation and safety improvement (Persaud et al., 1999; Heydecker and Wu, 2001; Hauer et al., 2002; Kononov, 2002; Hallmark and Basavaraju, 20002; Midwest Research Institute, 2002; Agent et al., 2003; Miaou and Song, 2005a, 2005b; Brijs et al., 2007; Li and Zhang, 2007).

The objectives of the statistical analysis are to generate a data set based on several data sets including traffic safety data, highway inventory data, and GIS data, to conduct exploratory data analysis, and to perform spatial analysis, with exploratory spatial data analysis and Bayesian spatial modeling to estimate and map crash risk.

As mentioned earlier, this project focuses on KAB crashes on rural, two-lane, low volume, state-maintained highway in Arkansas in 2004. Hence, we retrieve the crashes and related information from the data sets addressed earlier.

Data

The state of Arkansas consists of 75 counties (Figure 1). Annual KAB crash frequencies for rural, two-lane, low volume, state-maintained highway at the county level in 2004 were used for subsequent statistical analyses.

Figure 1. 75 Counties in Arkansas

This type of roadway is 10.3% miles of Arkansas roads, and about 17.5% of the crashes are fatal, incapacitating injury, and non-incapacitating injury (or KAB) crashes in 2004. 22.7% of the crashes are KAB crashes on the roadway in 2004.White County has the highest number (=51) of KAB crashes in 2004, and three counties, Clay County, Greene

County, and Little River County, have the lowest number of KAB crashes (=3) in the year. Figure 2 displays annual KAB crash frequencies for the roadway at the county level in 2004, exhibiting a spatial pattern.

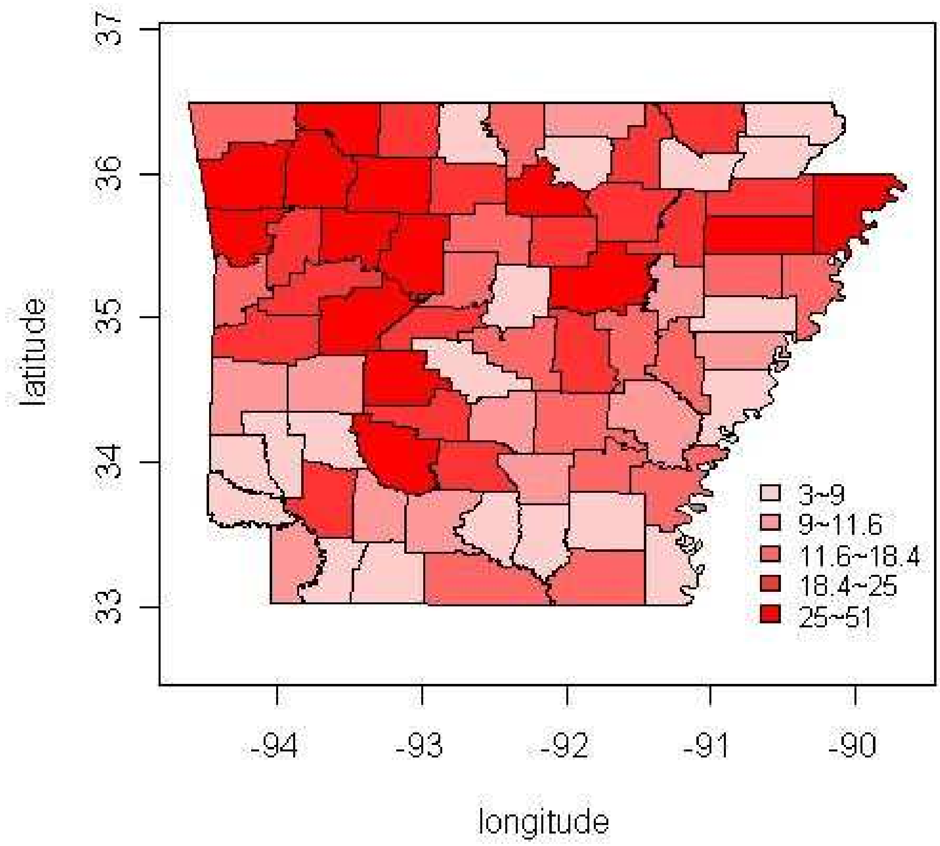


Figure 2. KAB Crashes on Rural, Two-Lan, Low-Volume, State-Maintained Roads in Each Arkansas County: 2004 (Darker area indicates more crashes)

A band of dark red counties stretches horizontally from the northwest region and thinning into lighter hues toward the south and toward the northeast corner. Such a pattern is a good reason to suspect spatial autocorrelation in the data. The observed total vehicle-miles traveled (VMT) in 2004 is shown in Figure 3, representing the size of the population at risk.

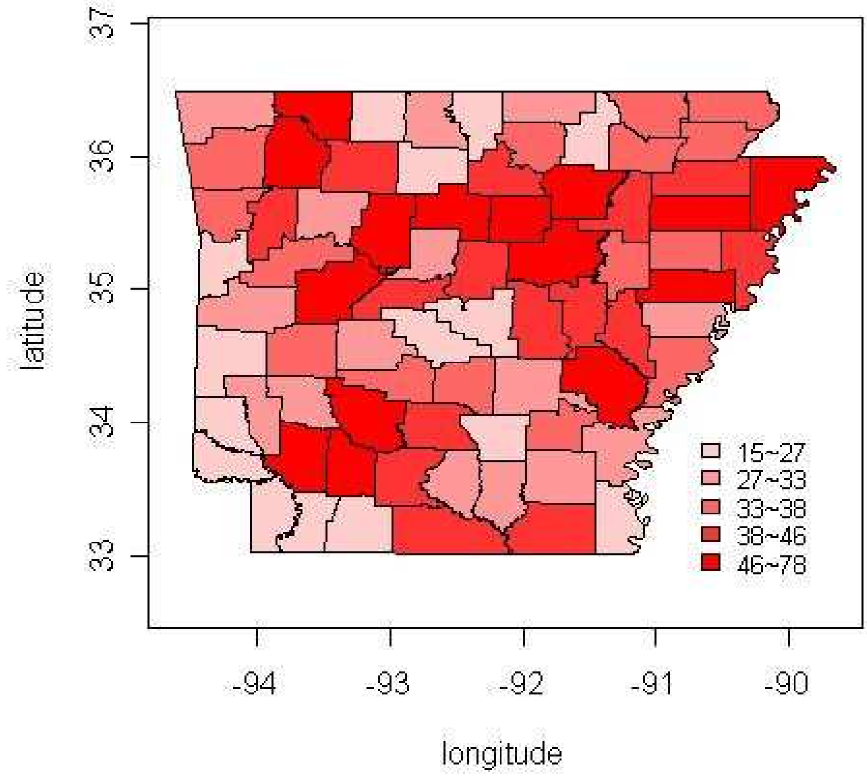
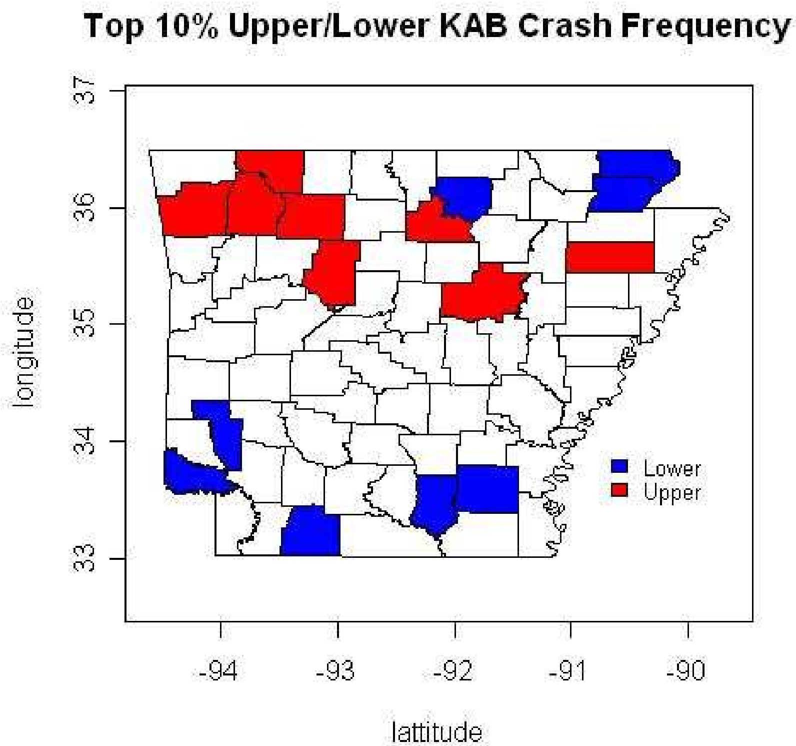


Figure 3. Vehicle-Miles Traveled on Rural, Two-Lane, State-Maintained Roads in Each Arkansas County (2004)

Figure 4 displays counties whose number of KAB crashes and rates in 2004 are in the top ten percent (red) and in the bottom ten percent (blue). The frequencies and rates are summarized by descriptive statistics (Table 1)) and boxplots (Figure 5). White county is identified as an outlier in the boxplot for crash frequency, with crash frequency 51 and crash rate 0.78, while Pulaski county has the highest crash rate (0.85) with crash frequency 17.

Table 1. Descriptive Statistics for Crash Frequency and Rate

	Mean	SD	Median	MAD	Min	Max	Range	Skew
Rate	0.46	0.21	0.45	0.21	0.08	0.85	0.77	0.27
Crash	16.69	9.76	14	8.9	3	51	48	0.96



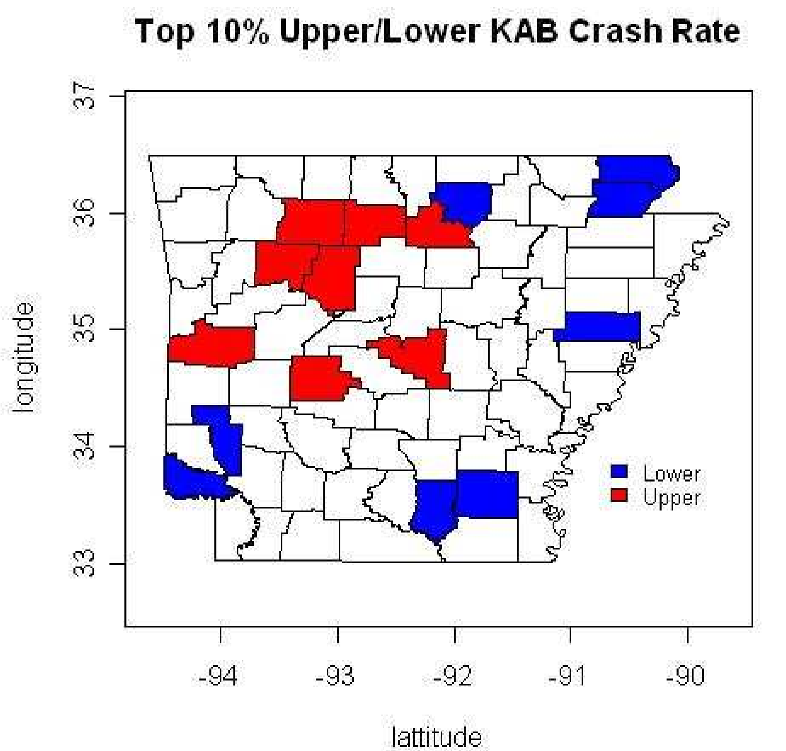


Figure 4. Top 10 Upper/Lower Frequencies and Rate in Arkansas (2004)

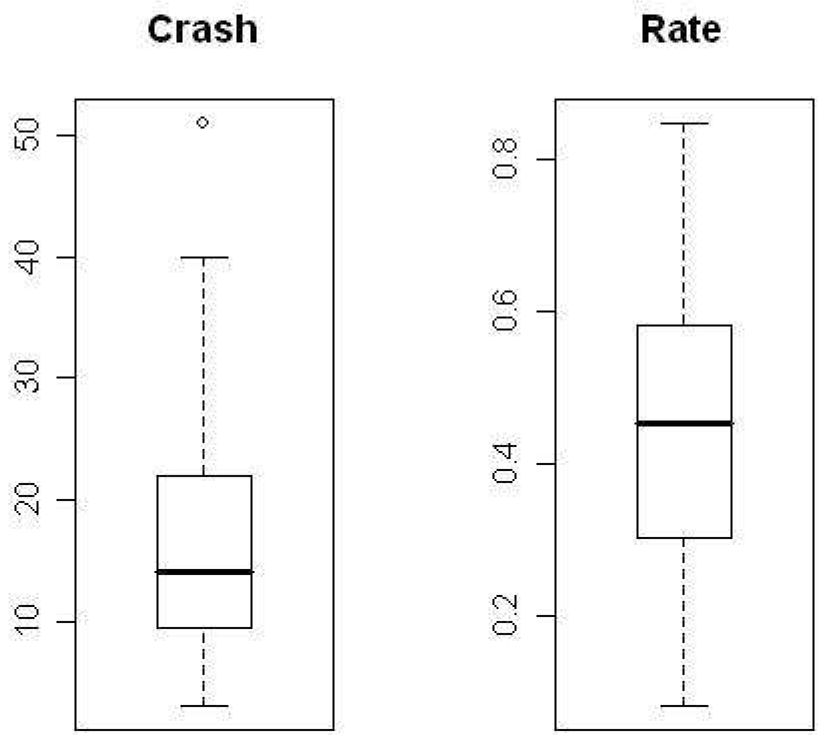


Figure 5. Boxplots for Crash Frequency and Rate

Exploratory Spatial Data Analysis

Similar to Exploratory Data Analysis (EDA) addressed in the previous paragraph, Exploratory Spatial Data Analysis (ESDA) is carried out for measuring spatial dependence, detecting spatial patterns in the spatial data, identifying "hotspots", and visualizing and mapping spatial data and related statistics. ESDA is an extension of EDA and a set of statistical tools for exploring and understanding spatial properties of the data. It is important to account for spatial dependence in modeling and the corresponding statistical inference because it can lead to unstable parameter estimates and yield unreliable significance test. There are two broad types of techniques for ESDA, graphical techniques and numerical techniques. The former includes thematic map, Moran's scatter plot, and significance map. For the latter, there are generally two quantities of interest, global spatial autocorrelation and local spatial autocorrelation. The rationale behind spatial analysis is based on Tobler's First Law of Geography, "Everything is related to everything else, but near things are more related than distant things".

Global spatial autocorrelation is a global measurement of spatial autocorrelation over the entire observations over an area of interest, used for testing spatial autocorrelation to detect departures from spatial randomness. The global Moran's I is the most popular statistics for measuring global spatial autocorrelation, and is used in testing the significance of the correlation with the null hypothesis of completely spatial independence. The statistic is given by

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)},$$

where x_i is the measurement at location i, $i = 1, \cdots n$, w_{ij} is an spatial weight between location i and j, and $\bar{x}$ is the sample mean. The expected value and variance are $E_N(I) = -\frac{1}{n-1}$ and $Var(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_3^2}{(n-1)(n+1)S_3^2}$, where $S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij} + w_{ji}\right)^2$, $S_2 = \sum_{i=1}^{n}\left(w_{i.} + w_{.i}\right)^2$, $S_3 = \sum_{i}\sum_{j} w_{ij}$ and $w_{i.} = \sum_{j} w_{ij}$. In a spatially random data set, the observation is unrelated to the location on the map where that observation is taken. If the test statistic I is too far from the its expected value under spatial randomness, we conclude that the data set exhibits spatial autocorrelation. The test can performed using a normality approximation or using a pseudo distribution. Anselin (1995) recommends testing using a pseudo-distribution, and that is the approach used in this paper. For an in-depth discussion of the use of the normal approximation and pseudo-distributions to test spatial randomness, refer to Schabenberger and Gotway (2004). Global Moran's I has some similar properties to those of classical correlation coefficient. For instance, this statistics ranges from -1 to1, and high positive value of the statistics represents that neighboring values tend to cluster together. That the statistic is zero means that there is no spatial autocorrelation and the observations are randomly distributed over the space.

Global Geary's C is another global spatial autocorrelation measure, ranging from 0 to

2. The statistic is equal to 1 for no spatial autocorrelation, and equal to 0 and 2 for strong positive and negative spatial autocorrelation, respectively. This statistic is given by

$$C = \frac{n-1}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - x_j\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)}.$$

For Arkansas crash data, we found significant global autocorrelation by rejecting the null hypothesis, with very small p-values for both statistics (I=0.31 and C=0.74), which indicates that the data set are significantly spatially correlated. The positive Moran's I indicates that the overall spatial pattern is positive and counties tend to be surrounded by counties with a similar number of KAB crash rate, which is confirmed in the following maps. The significant autocorrelation is visualized by Moran's I scatter plot and choropleth Moran scatterplot. Figure 6 displays Moran's I scatter plot for Arkansas traffic crash rate in 2004, plotting KAB crash rates against weighted average of the neighboring values. Each point on the graph corresponds two one location from the data set. The plot is also useful for identifying locations whose surrounding spatial autocorrelation pattern is different from the overall data set. The plot is divided into four quadrants. For instance, High-High (upper-right) quadrant indicates that high crash rates are surrounded by high rates and High-Low (lower-right) quadrant indicates that high crash rates are surrounded by low crash rates.

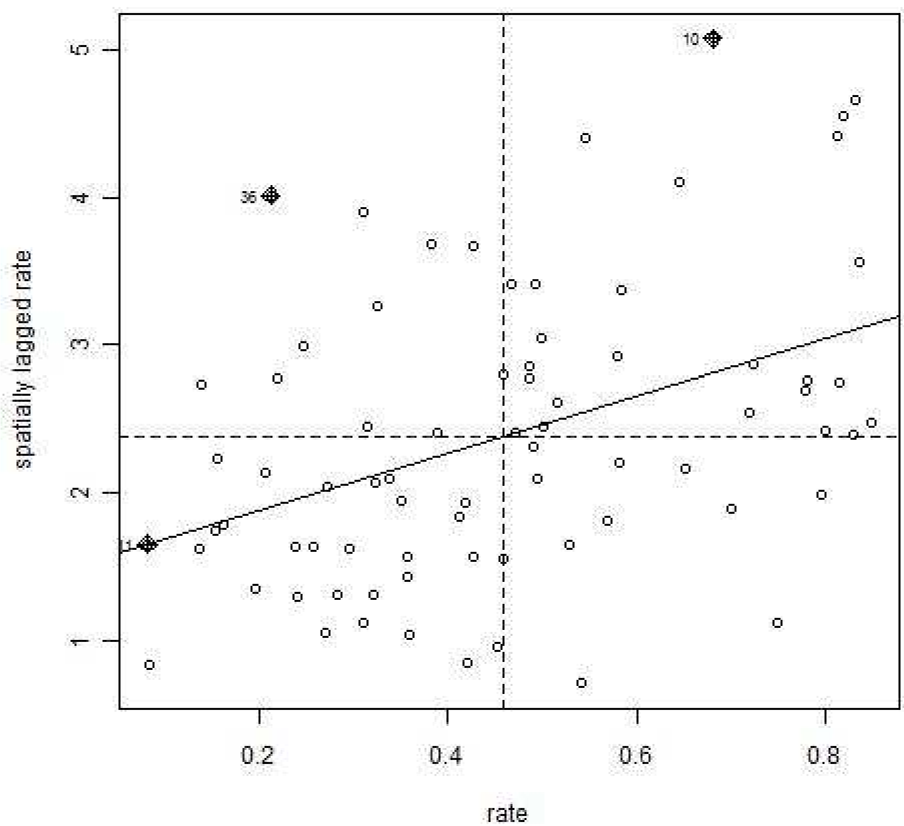


Figure 6. Moran's I Scatter Plot for Arkansas Crash rate in 2004. Upper-left quadrant: Low-High (negative spatial autocorrelation), upper-right quadrant: High-High (positive spatial autocorrelation), Lower-left quadrant: Low-Low (positive spatial autocorrelation), lower-right quadrant: High-Low (negative spatial autocorrelation)

Hence, the two quadrants represent positive and negative spatial autocorrelation, respectively. The scatterplot also allows us to investigate the proportion of counties that deviate from the overall pattern of positive spatial autocorrelation. Points lying in the upper left quadrant correspond to counties with a low crash rate, whose neighbors experienced a high number of crashes. Points lying in the lower right quandrant correspond to counties with a high crash rate, whose neighbors experienced a dissimilarly

low number. These counties exhibited a negative spatial autocorrelation pattern, and therefore deviate from the overall pattern of a clustering of similar values.
Note that three observations 10 (Madison), 11 (Greene county), and 36 (Faulkner county) are identified as leverage points, which strongly influence the spatial autocorrelation.. The choropleth Moran scatterplot is a mapping of the Moran scatterplot onto the actual map, shown in Figure 7.

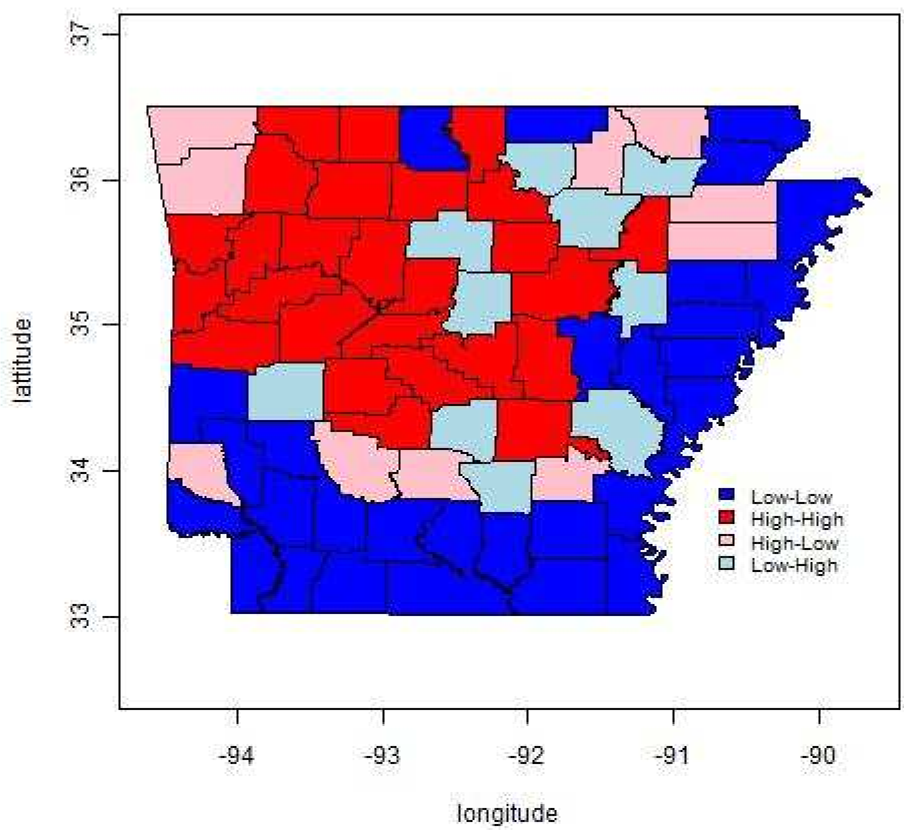


Figure 7. Choropleth Moran Scatterplot for Arkansas Crash Rate in 2004

These plots allow for a quick comparison of KAB crash rate at county level compared to that of its neighbors. Counties shaded a dark red correspond to points in the upper right quadrant. These are counties with high values, surrounded by high values (high-high). Counties shaded a dark blue correspond to points in the lower left quadrant (low-low). Thus, a dark color indicates positive spatial autocorrelation. Counties shaded a light red correspond to points in the lower right quadrant (high-low), while points for counties shaded a light blue can be found in the upper left quadrant (low-high). A light color indicates negative spatial autocorrelation. It is found that the KAB crash rates exhibit positive spatial autocorrelation in most of the state. A band of dark red stretches across the top of the map and the bottom portion is mostly covered by dark blue. We can also pick out counties with negative spatial autocorrelation. For example, the light red county surrounded by dark blue in the bottom left corner is Sevier County. Future research might address the question of why Sevier County is located in low KAB crash region.

The global spatial autocorrelation addresses whether the data set as a whole exhibits spatial autocorrelation, while we may want to determine if there is spatial autocorrelation around a specific county. Local spatial autocorrelation answers its purpose. The spatial autocorrelation is generally used for identifying local spatial cluster and outlier, assessing stationarity, and assessing influence of individual observations to the global spatial autocorrelation. Local indicator of spatial autocorrelation (LISA) is the most popular local spatial autocorrelation measure, introduced by Anselin (1995). This statistic is a means of determining whether a particular location is surrounded by locations with attributes that have markedly similar (positive spatial correlation) or dissimilar (negative spatial correlation) values. Such locations are referred to as "hotspots". There are several commonly used statistics which measure this clustering of similar values around a particular point, including local Moran's I , G and G*. In this study, we utilize local Moran's I in the subsequent analysis. Local Moran's I statistic is given by

$$I_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \sum_{j=1}^{n} w_{ij} \left(x_j - \bar{x}\right),\ i = 1, \cdots, n.$$

The interpretation of this statistic is identical to that of the global Moran's I addressed above. That the statistic at a location equals to zero indicates a lack of spatial autocorrelation at the location. A large positive value of the statistic indicates positive spatial autocorrelation. Like the global Moran's I, the local Moran's I can be used to perform a statistical test for significance of spatial autocorrelation (permutation test). The global Moran's I tests for spatial autocorrelation of the entire data set; the local Moran's I tests for spatial autocorrelation around a particular point. However, the local Moran's I to test hypotheses of non-spatial autocorrelation raises what is known in statistics as the multiplicity problem. When utilizing LISA's to detect hotspots, a statistical test is performed at each location in the data set. A spatial data set with an attribute at each of m locations will require a hypothesis test at each of the m locations. A total of n hypotheses are required, so the multiplicity problem arises.

The multiplicity problem refers to the fact that when performing multiple statistical tests, the likelihood of making a mistake increases. If the probability of incorrectly detecting spatial autocorrelation at each location, the significant level, is set at 0.05, then the probability of incorrectly detecting spatial autocorrelation for at least one of the n locations becomes $1-0.95^n$. Thus the probability of making a mistake goes up as the number of locations increases. If no adjustment is made, it is highly likely that some locations will be identified as hotspots even though they are actually not hotspots, called as false positives. The Bonferroni method is the traditional technique for dealing with the multiplicity problem. The Bonferroni method controls Type I error (α) by testing each of the hypotheses at level α / n, where n is the number of hypotheses being tested. In local autocorrelation testing, n is the number of locations. The drawback to the Bonferroni method is that it is too conservative. While successfully controlling for the possibility of making false positives, this method raises the probability of making the mistake of overlooking locations which truly are significant for spatial autocorrelation; that is, making false negatives

The False Discovery Rate (FDR) controlling method presents a compromise between the overly conservative Bonferroni method and the pitfalls of multiple testing with no adjustment. The false discovery rate refers to the proportion of false discoveries over the total number of discoveries made.

$$FDR = \frac{\text{Number of INCORRECTLY rejected Null hypotheses}}{\text{Number of rejected Null hypotheses}}$$

The FDR controlling method sets the expected value of the FDR, usually at 0.05. By allowing for some false discoveries, this method allows for more true discoveries while still accounting for multiple testing. The FDR controlling method is appropriate in situations where making some false discoveries is acceptable, and is suitable for the case of Arkansas crash data. If the multiplicity problem is not adjusted for, many false positives are detected, leading to waste resources allocated to them. Overly conservative Bonferroni method might miss most, possibly all, of hotspots, and thus it results in failing to address any problems indicated by the presence of those hotpots. The FDR controlling method provides a desirable compromise between these two extremes. Additionally, Caldas de Castro and Singer (2006) concluded through the use of simulations that FDR is the best method to apply when testing for local spatial autocorrelation. In this paper, we perform LISA testing using FDR, Bonferroni and unadjusted, and compare the results. We then proceed with our analysis using the results of the FDR controlling method. Mapping the local Moran's I statistics is used to identify locations whose spatial autocorrelation pattern deviate from that of the overall data set, and to further understand how the spatial autocorrelation pattern changes across space. The LISA is calculated for Arkansas crash data, and the significant values under different p-value adjustments are depicted in Figure 8.

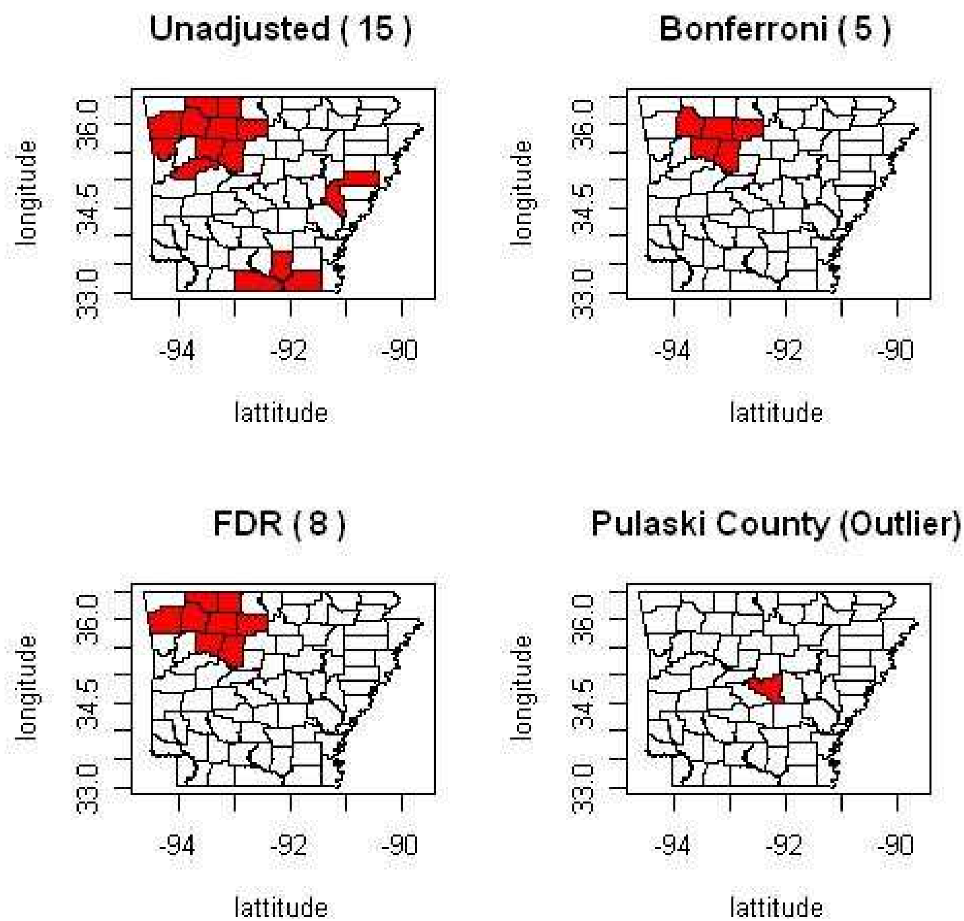


Figure 8. Significant LISA Values under Different p-value Adjustments

As expected, the two extreme adjustments detect the smallest (Bonferroni) and largest (Unadjusted) significant counties, and FDR points out eight significant counties. A noteworthy result is that Pulaski county is not identified as a hotspot in the analysis even though it is identified an extreme outlier in EDA. According to local Moran's I test, Pulaski County is not significant for spatial autocorrelation; there is no significant clustering of high KAB crash rates around Pulaski county. Note that all of the counties identified as hotspots, under all three multiple testing procedures, had a positive Moran's I statistic; that is, the counties with high KAB crash rates identified as hotspots are surrounded by counties with a similarly high crash rate. Likewise, low crash rate hotspots are surrounded by counties with low crash rates. For the purpose of potential management zones for traffic safety, Figure 9 maps global and local Moran's I statistics. The map for the global statistics can be partitioned into three zones: zone 1 (red), zone 2 (blue), and zone 3 (light blue and light red). Based on the interpretation of autocorrelation, these zones can be interpreted as high crash rate zone, low crash rate zone, and unstable zone, respectively. Since zone 1 is primarily of interest in traffic safety, the zone can be divided into two subdivisions based on the local statistics: subdivision 1 (red) and subdivision 2 (the rest of high crash rate). This can give some insight into planning traffic safety management zone.

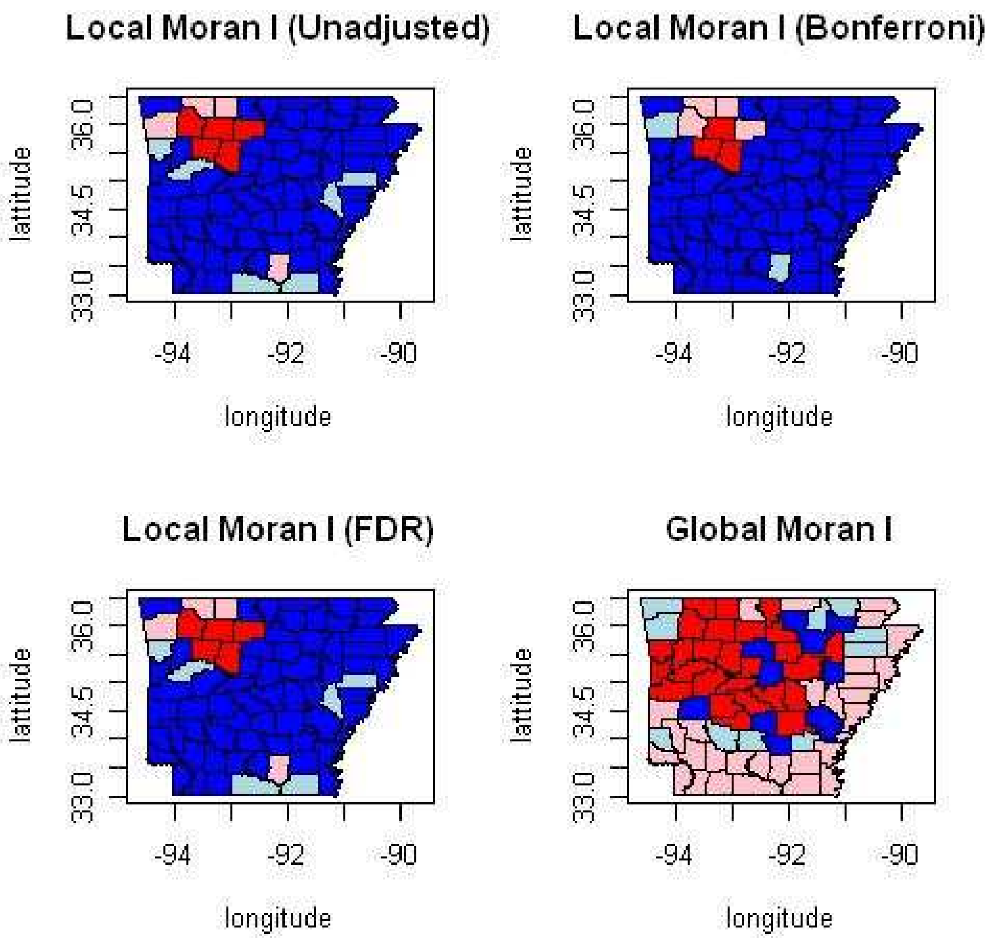


Figure 9. Significance Maps and Potential Management Zones: for three LISA maps, Blue (p-value>0.05), Light Blue (0.01<p-value<0.05), Light Red (0.001<p-value<0.01), and Red (p-value<0.001). For Global Moran's I map, High-High (Red), Low-Low (Blue), High-Low (Light Red), and Low-High (Light Blue)

Model-based risk mapping

As part of modeling efforts, the Poisson hierarchical Bayes model is employed. It is shown that risk estimates based on the hierarchical Bayesian modeling have several advantages over raw estimates. In particular, raw estimates are typically unreliable in the areas where incidences of traffic crashes are relatively rare. The hierarchical Bayes models with spatial random effect can address such drawbacks (Banerjee et al., 2004). We define variable Y_i and v_i as the total number of reported KAB crashes and the observed total VMT on the rural roads of interest in county i in 2004, respectively. We also consider four covariates. The first covariate is a surrogate variable for taking accounting of road surface condition. In particular, time that the road surface is wet due to rain, snow and ice is of interest. Hence, we used the proportion of KAB crashes that occurred under wet pavement condition due to rain and snow as a surrogate variable. We define a KAB crash related to wet pavement whose road surface in crash data is recorded as wet or ice. The second one is a surrogate variable intended to represent spatial difference in the number of sharp horizontal curves in different counties. We chose the proportion of KAB crashes that occurred on sharp horizontal curves in each county as a surrogate variable, and we define a KAB crash related to sharp horizontal curve whose crash roadway alignment in crash data is recorded as curve. The third covariate is intended to represent degree of road hazards. We chose the proportion of KAB crashes related to fix objects, such utility pole, guard rail, and sign, in each county as a surrogate variable. These variables are summarized by boxplot in Figure 10 and by descriptive statistics in Table 2.

Table 2. Descriptive Statistics for the Covariates

Statistics	Wet	Curve	Object
Mean	0.22	0.55	0.54
SD	0.13	0.21	0.19
Median	0.22	0.55	0.54
MAD	0.12	0.27	0.17
Min	0	0.11	0.19
Max	0.67	1	1
Range	0.67	0.89	0.81
Skew	0.54	-0.07	0.41

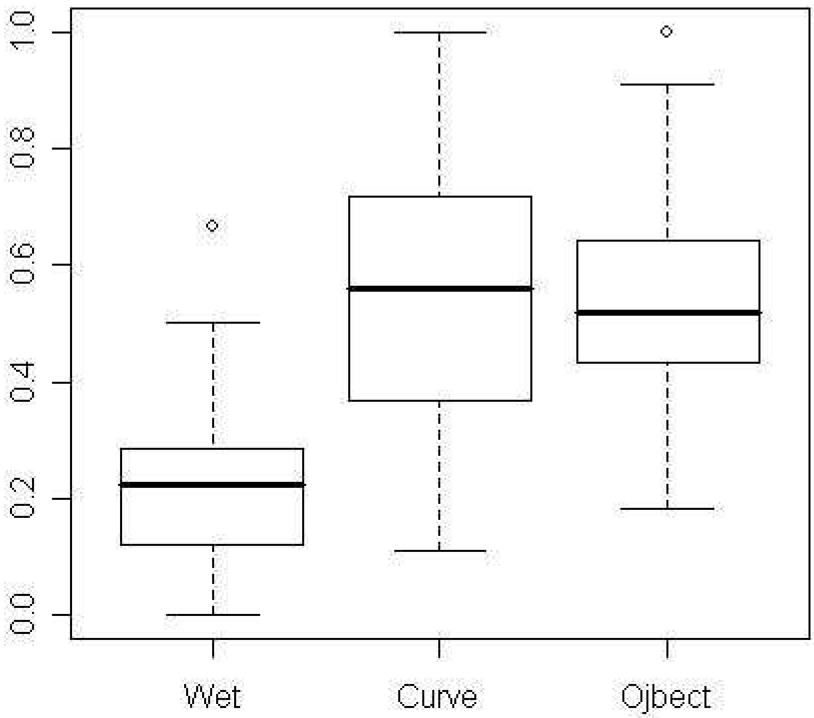


Figure 10. Boxplots for Three Covariates

Spatial distributions of these variables are shown in Figure 11, Figure 12, and Figure 13, respectively. Two outliers are detected in the two boxplots (Wet and Object), and both outliers correspond to Greene County.

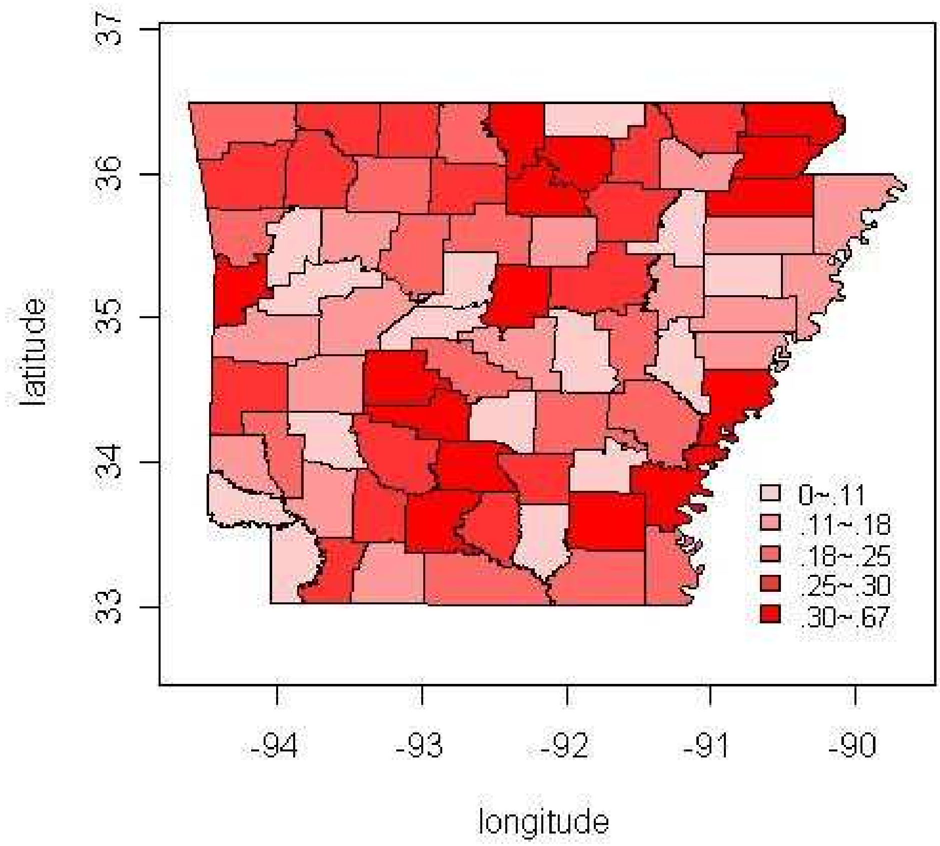


Figure 11. Proportion of KAB Crashes that Occurred Under Wet Pavement Conditions for Each County in 2004

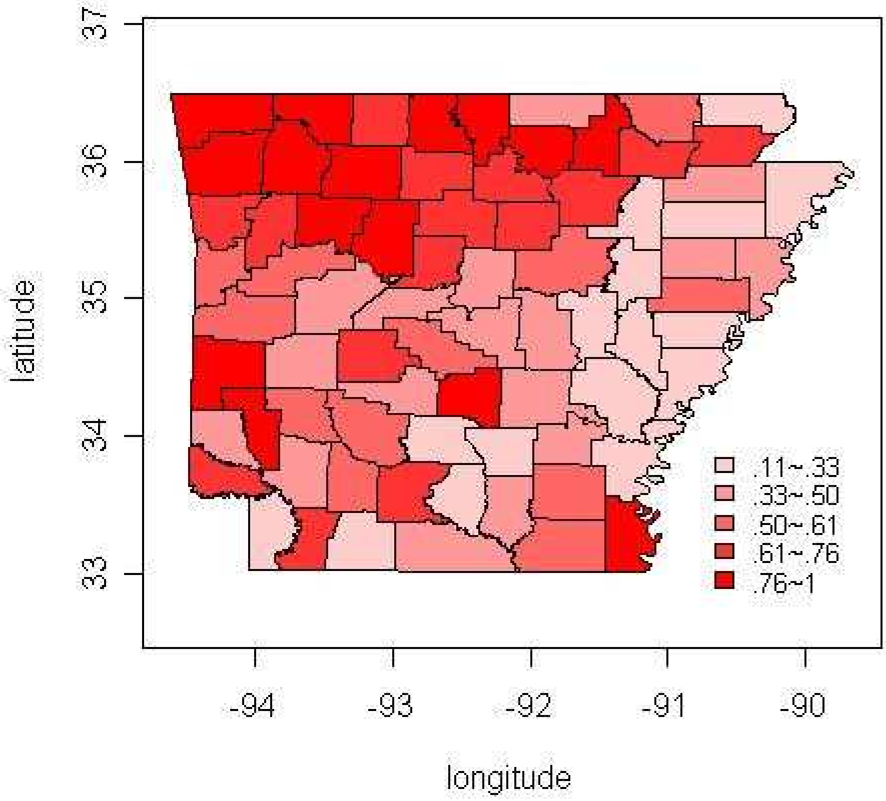


Figure 12. Proportion of KAB Crashes that Occurred on Sharp Horizontal Curves in Each County in 2004

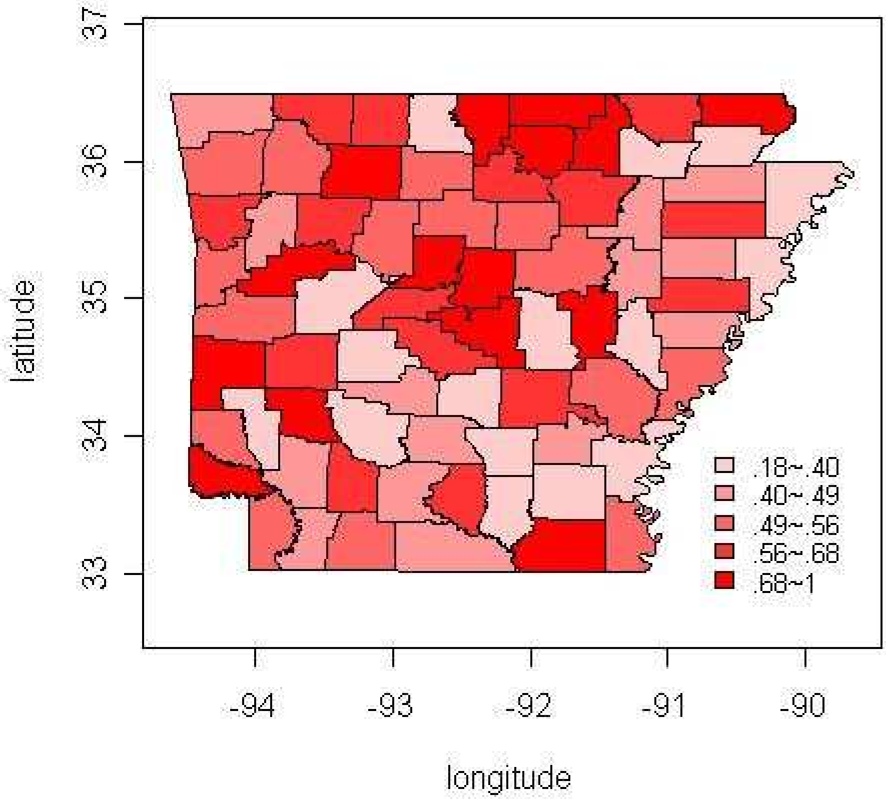


Figure 13. Proportion of KAB Crashes Involving Vehicles that Hit Fixed Objects on the Roadside for Each County in 2004

In the modeling, conditional on mean μ_i, the total number of KAB crashes Y_i is assumed to be mutually independent and Poisson distributed as $Y_i \sim Poisson(\mu_i)$, where $\mu_i = \nu_i \lambda_i$ and λ_i is the KAB crash rate. The rate is modeled in generalized linear mixed model (GLMM) framework,

$$\log(\lambda_i) = \sum_k \beta_k x_{ik} + \phi_i + \varepsilon_i\ , \quad i = 1, \cdots, n,$$

where x_{ik} are covariates discussed earlier, β_k is a regression coefficient, ϕ_i is a spatial random effect, and ε_i is a unstructured random effect.

Before Bayesian analysis, we fit the data to a classical Poisson regression without the two random effects, ϕ_i and ε_i, and calculate the resulting residuals for investigating spatial variation in control of the covariates. Figure 14 maps the residuals, showing that they are still not spatially independent. This is confirmed by Moran's I test of the residuals. The null hypothesis is rejected with very small p-value. These facts envision spatial modeling for the data.

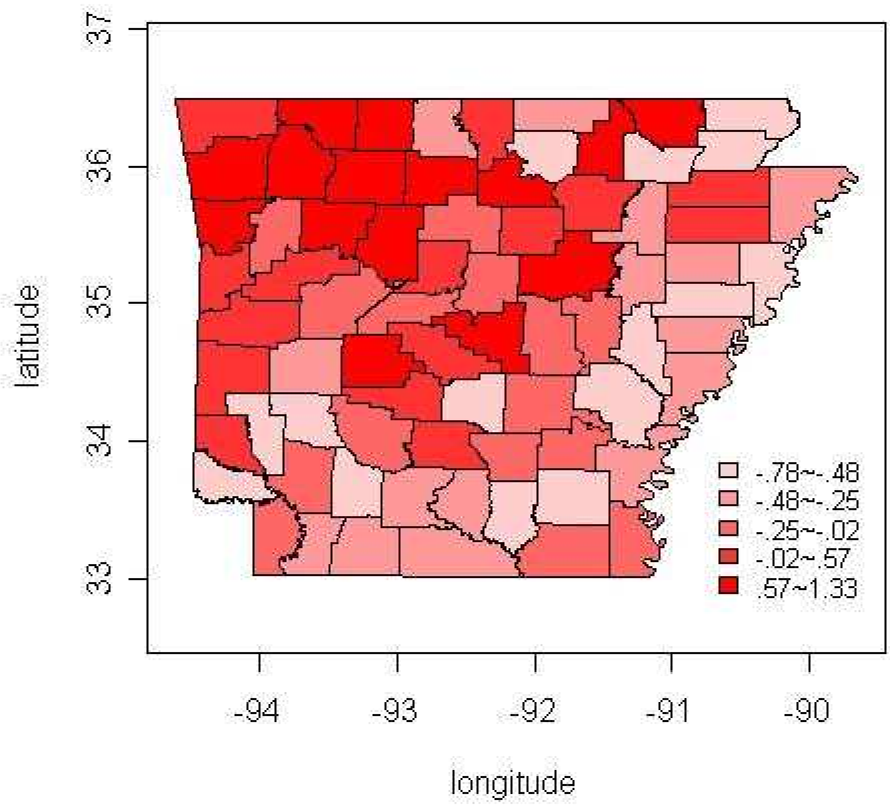


Figure 14. Residuals from a Poisson Regression Model without spatial and residual random effects

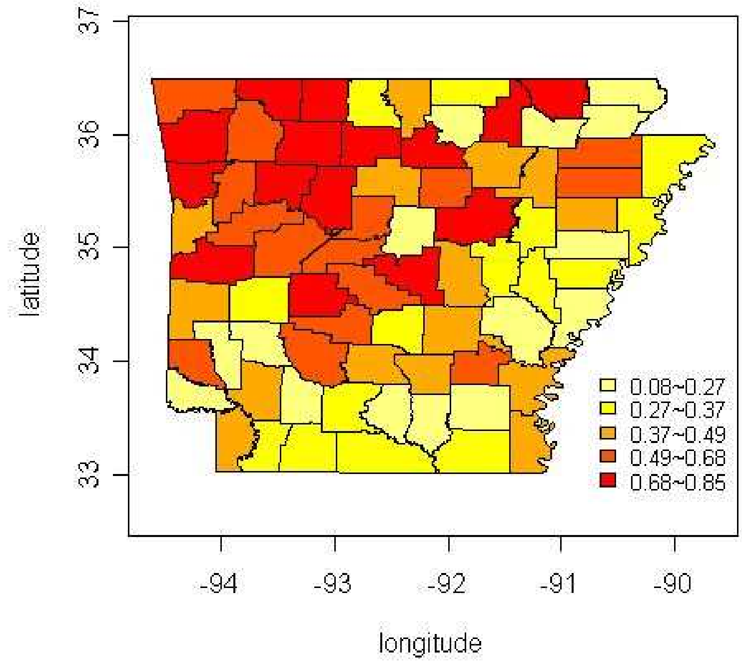
In order to complete Bayesian modeling, we need to specify prior distributions for the model parameters. We place independent and noninformative normal prior on β_k and ε_i . With Regard with spatial modeling in the GLMM, we adopt conditional autoregressive (CAR) model (Besag 1974 and 1975) for the spatial random effect ϕ_i, and the joint distribution of ϕ_i is given by

$$p(\phi_i,\cdots,\phi_n) \propto \exp\left(-\frac{1}{2\tau^2}\sum_{i\neq i^*_i} w_{ii^*}(\phi_i-\phi_{i^*})^2\right),$$

where τ^2 is a variance parameter and w_{ii^*} is a spatial weight associated with county i and i^*. Markov Chain Monte Carlo (MCMC) techniques are implemented to sample the posterior distributions. The MCMC simulation reached convergence quite quickly, so 10,000 iterations are performed, with discarding the first 5,000 iterations as burn-in. The simulation is summarized in Table 3, and the estimated KAB crash rates per MVMT by county from the model is shown in Figure 15, with raw annual KAB crash rates per MVMT by county for 2004. The wet pavement and horizontal curve variable are found as the significant variables to explain the crash rate, but none of them are statistical significant.

Table 3. Posterior Summary of the regression coefficients

Coefficient	Mean	2.5%	Median	97.5%
β_1 (Wet)	-0.6876	-2.673	-0.6695	1.169
β_2 (Curve)	0.5710	-0.7090	0.5825	1.819
β_3 (Object)	-0.1434	-1.497	-0.1337	1.121



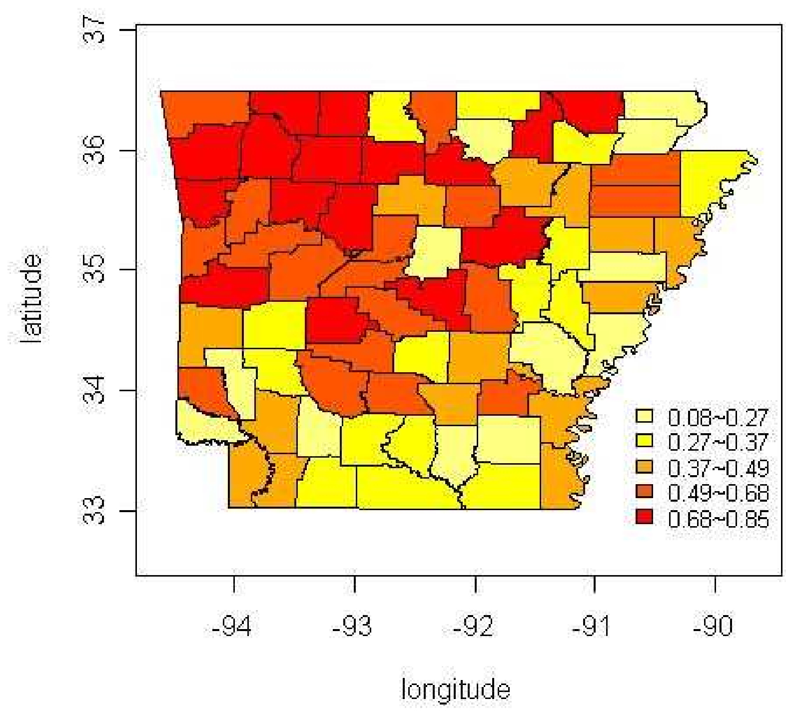


Figure 15. Raw Annual KAB Crash rates (Upper) and Estimate KAB Crash Rates from the model (Lower) per MVMT by County in 2004

Discussion

In this project, we demonstrate the use of some of statistical methodologies for spatial analysis for roadway traffic crashes at the county level. ESDA is a simple and powerful tool for exploring spatial data. In the beginning of spatial analysis, this approach can give some insight into further analysis. For the data set analyzed in this project, global and local spatial autocorrelation are measured and tested using the approach. The results motivate to consider spatial effect in modeling and mapping crash risk. In the modeling, we create three covariates in that there is not available any direct measure representing crash characteristics of interest. The estimation suggests us to consider more extensive characteristics. In this analysis, only one-year crash data is used in the analysis. An obvious extension of the current study is to develop risk maps for traffic crashes over several years. This can provide temporal trend of crash risk. Finally, traffic crash data are fundamentally network-based data rather than county-based data. To our knowledge, there are few statistical methods for network-based data. We expect that these network-based crash risk map will be more useful for roadway safety planners and engineers to 1) identify potentially risky road segment or intersection in network, 2) to optimally allocate resource to the roadway for improving and reducing the number and severity of traffic crashes, and 3) monitor and evaluate the safety performance on the roadway after improvement projects.

References

1. Agent, K.R., L. O'Connell, E.R. Green, D. Kreis, J.G. Pigman, N. Tollner, and E. Thompson. 2003. Development of Procedures for Identifying High-Crash Locations and Prioritizing Safety Improvements. Research Report KTC-03-15/SPR250-02-1F. Kentucky Transportation Center, University of Kentucky.
2. Aguero-Valverde, J. and P.P. Jovanis. 2006. Spatial Analysis of Fatal and Injury Crashes in Pennsylvania. *Accident Analysis & Prevention* 38:618-625.
3. Anselin, L. 1995. Local Indicators of Spatial Association-LISA. *Geographical Analysis* 27:93-115.
4. Banerjee, S., B.P. Carlin, and A.E. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman and Hall/CRC.
5. Besag, J. 1974. Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion). *Journal of the Royal Statistical Society, Series B* 36:192-236.
6. Besag, J. 1975. Statistical Analysis of Non-Lattice Data. *Statistician* 24:179-195.
7. Brijs, T., D. Karlis, F. Van den Bossche, and G. Wets. 2007. A Bayesian Model for Ranking Hazardous Road Sites. *Journal of the Royal Statistical Society, Series C* 170:1001-1017.
8. Caldas de Castro, M. and Singer, B. H. 2006.Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association. *Geographical Analysis* 38:180-208
9. Carlin, B.P. and T.A. Louis 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. London, England: Chapman and Hall/CRC.
10. Dey, D., S. Ghosh, and B.K. Mallick. 2000. *Bayesian Generalized Linear Model*. New York, NY: Marcel Dekker.
11. Ghosh, M., K. Natarajan, L.A. Waller, and D. Kim. 1999. Hierarchical Bayes for the Analysis of Spatial Data: An Application to Disease Mapping. *Journal of Statistical Planning and Inference* 75:305–318.
12. Hallmark, S. and R. Basavaraju. 2002. Evaluation of the Iowa DOT's Safety Improvement Candidate List Process. CTRE Project 00-74. Final Report. Center for Transportation Research and Education, Iowa State University.

13. Hauer, E., J. Kononov, B. Allery, and M.S. Griffith. 2002. Screening the Road Network for Sites with Promise. *Transportation Research Record* 1784:27-32.
14. Heydecker, B.G. and J. Wu. 2001. Identification of Sites for Road Accident Remedial Work by Bayesian Statistical Methods: An Example of Uncertain Inference. *Advances in Engineering Software* 32:859-869.
15. Kononov, J. 2002. Identifying Locations with Potential for Accident Reductions: Use of Direct Diagnostics and Pattern Recognition Methodologies. *Transportation Research Record* 1784: 153-158.
16. Lawson, A.B. 2001. Tutorial in Biostatistics: Disease Map Reconstruction. *Statistics in Medicine* 20:2183–2204.
17. Lawson, A.B., A. Biggeri, D. Bohning, E. Lessafre, J.F. Viel, and R. Bertollini, eds. 1999. *Disease Mapping and Risk Assessment for Public Health*. Chichester, UK: Wiley.
18. Li, L. and Y. Zhang. 2007. Bayesian Approach Based on Geographic Information System to Identifying Hazardous Roadway Segments for Traffic Crashes. *Transportation Research Record* 2024:63-72.
19. Ma J., K. M. Kockelman, and P. Damien. 2008. A Multivariate Poisson-Lognormal Regression Model for Prediction of Crash Counts by Severity, Using Bayesian Methods. *Accident Analysis & Prevention* 40: 964-975.
20. Miaou, S.-P. and J. J. Song. 2005a. Bayesian Ranking and Selection of Sites for Engineering Safety Improvements: Statistical Criterion, Treatability Assessment, and Cost Function. *Accident Analysis & Prevention* 37: 699-720.
21. Miaou, S.-P. and J. J. Song. 2005b. Application of Geocoded Traffic Crash Records and Crash-Risk Mapping Technology in Roadway Safety Improvement Projects, Southwest Region University Transportation Center.
22. Miaou, S.-P., J. J. Song, and B. Mallick. 2003. Roadway Traffic-Crash Mapping: a Space-Time Modeling Approach. *Journal of Transportation and Statistics* 6: 33-57.
23. Midwest Research Institute. 2002. SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites: Task K. White Paper for Module 1—Network Screening. Prepared for Federal Highway Administration.

24. Miranda-Moreno, L.F., L. Fu, F.F. Saccomanno, and A. Labbe. 2005. Alternative Risk Models for Ranking Locations for Safety Improvement. *Transportation Research Record* 1908:1-8.
25. Park, E. S., and D. Lord. 2007. Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record* 2019:1-6.
26. Persaud, B., C. Lyon, and T. Nguyen. 1999. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. *Transportation Research Record* 1665:7-12.
27. 1999.Schabenberger, O. and C. A. Gotway. 2004. *Statistical Methods for Spatial Data Analysis,* Chapman & Hall/CRC.
28. Song, J. J. 2004. Bayesian multivariate spatial analysis and their applications. Statistics. College Station, Texas A&M University. Ph.D.
29. Song, J. J., M. Ghosh, S.P. Miaou, and B. Mallick. 2006. Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis*. 97: 246-273.
30. Sun, D., R.K. Tsutakawa, H. Kim, and Z. He. 2000. Spatio-Temporal Interaction with Disease Mapping. *Statistics in Medicine* 19:2015–2035.
31. Xia, H., B.P. Carlin, and L.A. Waller. 1997. Hierarchical Models for Mapping Ohio Lung Cancer Rates. E*nvironmetrics* 8:107–120.

Study 2: Evaluating Arkansas Roadway Intersection Accidents Using Traffic Safety Analysis Methods - Generalized Estimating Equations and Roadway Observation

ABSTRACT

Considerable research has been made in recent years to evaluate road traffic safety. This is especially true with the United States, whose international rank in public safety is rapidly declining. In 2004, Arkansas ranked as the third highest in traffic fatalities among all the states. These are striking numbers that reflect the lack of attention that the state has received in terms of evaluating road traffic safety. Historically, this safety is measure by one of two methods: Statistical analyses of historical data or hands-on, observational analyses of present safety conditions. Rarely in research are both methods used within the same study. With this in hand, the objective of this research was to evaluate closely the issues involved with road traffic safety in the state of Arkansas. A database of all road traffic accidents within Arkansas between 2002 and 2004 was used in order to perform statistical testing and analyses. The study focused on intersection related crashes occurring on road segments within US highways, State highways, and Interstates with medium to heavy traffic volumes. In conjunction with these analyses, several hands-on observations of intersection locations were made to compare actual road safety with the statistical results, as well as to provide additional information that was not represented within any collected data. After carefully choosing key road segment locations throughout Arkansas, the intersections were surveyed for potential crash hazards. With the combination of these two approaches the leading factors for collisions in Arkansas were evaluated and preventative measures were recommended. Of all the potential factors, substantial attention was given to the human factors involved with road collisions. Historically, these factors have been found to be the most common of all factors, easiest to prevent, and therefore needing the most immediate attention.

The statistical models developed for Arkansas roadways were the Poisson, Negative Binomial, and Logistic regression models. Among the significant contributors to crash frequency and severity were road width, number of lanes, pavement condition, horizontal and vertical curvature of the road design ($p < 0.01$). Also, weather and light conditions, seat belt usage, age, alcohol consumption, and number of passengers were shown to be significant to predicting crash frequencies and/or severities ($p < 0.01$).

The observational analysis provided many insights on how road infrastructure and road surroundings can affect driving patterns and driver behavior. Poor signage, lane markings, traffic signals, and obstacles such as medians all can potentially decrement the driver's experience and increase the risk of collision.

The unique aspect of combining these two methods showed a vast improvement on the understanding of road traffic accidents and safety within the state of Arkansas. Their results give great insights and highlight potential issues of the driver behaviors and roadway characteristics that effect road traffic safety.

CHAPTER I - INTRODUCTION

1.1 Background

Road traffic safety has been a major issue across the world for decades, and there has been an increasing demand for it in recent years due to numerous factors. First, the human population is growing at exponential rates, putting more and more drivers on the road each year. In addition, travel distances for these drivers and vehicles have also recently been showing increases. Consequently, automobiles are spending more time active on the roads than ever before (Federal Highway Administration, 1992-2006). The result of these and other aspects equates to the heightened vulnerability of vehicles being involved in road traffic accidents. Despite the current efforts of addressing road safety, the number of people affected by each accident grows every day (Peden, 2004). Regardless of these increasing trends and what may be written about them, the majority of road traffic accidents are entirely preventable, given the proper attention. To that end, researchers have expressed the need to observe and analyze past and present accidents in order to find the significant factors that are associated with increases in collision risks. With the understanding of those factors, their research can then lead to new preventative measures.

Within each crash there are several characteristic levels in which it can be detailed and explained. These levels can contain a number of elements, or factors, that describe the accident from all different angles. In general these levels include environmental, geographical, infrastructural, vehicular, temporal, and human conditions, which all play an integral part in a explaining a collision (Evans, 2004). These range from the time of day, weather conditions, road design, age, type, and make of the vehicles, to driver cell phone or seat belt use. Hundreds of factors can potentially play a part in every road traffic accident. Thus, it becomes important to determine which factors are critical, leading causes of each wreck. With access to crash data, researchers can begin to analyze several factors using statistical modeling to accurately predict and measure crash outcomes. With new statistical packages available, more and more complex methods can be applied to fit almost any data into meaningful results. Kim et al. (2007) has discussed the analysis of crash outcome probabilities by using a hierarchical logistic model as the base of study for their data. In another study by Milton et al. (2008), crash severity was under observation, using a mixed logit model. Other studies use statistical modeling to determine crash counts (Abdel-Aty and Radwan, 2000), crash rates (Anastasopoulos, Tarko, and Mannering, 2007), or overall road safety indexes (de Leur and Sayed, 2002). The amount of unique models and applications for accident data is enormous, which means incredible care must be taken when choosing the proper model.

Results from these statistical analyses can shed some light on numerous methods for improving road traffic safety. Geographical/infrastructural factors can be affected by the redesigning, engineering, and maintenance of existing roadways, giving drivers more mobility, awareness, and control. Vehicular factors can be affected by the designing of newer, safer vehicles. Law enforcement and driver education can lead to a better control and understanding of the human factors involved in road traffic accidents. The key is to pinpoint which factors are the most important factors, and then to apply the necessary provisions (Janson and Karimkhani, 2001).

For any of this to work, the accident data used must be as reliable as possible. Richard Scurfield of the World Bank's Transport Department states that one of the biggest obstacles facing crash analysis today is the abundance of poor quality data (2002). Some studies have even shown that although this analysis is a beneficial method in determining the important relationships within crash data, statistical studies may not be enough in several cases where the data is not completely accurate. For example, an experiment performed by de Leur et al. showed the increased reliability of using a proactive, observational study of roadways (2002). This type of analysis gives first hand and real world views of the road system, showing the nature of traffic flows and trends that might not be fully describable in statistical crash data. The quality of data found in these experiments was shown to be vastly superior. However, due to the amount of time and high costs associated with observing all the necessary road systems across the globe, this method often times becomes infeasible. There is a natural trade-off between practicality and accuracy when dealing with these two approaches. Historically, road traffic safety is evaluated using one of these two methods. Rarely are the two researched in conjunction with each other, which is regrettable due to the amount of information gained from using both perspectives.

Although road traffic safety has been an area widely studied for years, there is an increasing need for more specialized studies. Trends and factors related to roadway accidents are highly useful to road designers and drivers alike, but trends and factors are known to vary in different settings. A 2004 study conducted by the US Census Bureau found that the average number of traffic fatalities for every 100,000 vehicle miles traveled ranged from 0.87 in Massachusetts to 2.28 in Mississippi. According to this ranking, Arkansas places as the third highest state having an average of 2.22 fatalities per 100,000 vehicle miles traveled (US Census Bureau, 2004). Despite these high numbers, Arkansas is a state that has yet to be fully evaluated in terms of road traffic safety. To the knowledge of the author, no study has yet been published that has considered the conditions involving roadway traffic accidents throughout Arkansas. A study based exclusively on Arkansas may be able to reveal the reasons, factors, or trends behind a traffic rating of over 2.5 times the safest state ratings; a rating that all states should be able to achieve.

1.2 Research Goals

With this information at hand, the objective of this research was to evaluate closely the issues involved with road traffic safety in the state of Arkansas. This overall objective served as the keystone effort accomplished by the following research goals.

1.2.1 Evaluating Road Safety in Arkansas Using Statistical Analyses

A database of all road traffic accidents and road segments within Arkansas between 2002 and 2004 was used in order to perform statistical testing for the analysis of road traffic safety issues. Models were developed to measure the effect of numerous potential crash factors associated with both crash frequency and crash severity. These models were used to determine the significance of each crash factor, which corresponded to several aspects of the crash, including the time, location, weather conditions, road features, vehicle, and driver. For this analysis, the decision was made to focus specifically on intersection related crashes due to the historically large proportion of road traffic crashes which occur at intersections. This is chiefly due to the increased vehicle contact and conflict (Abdel-Aty, Keller, and Brady; 2005).

1.2.2 Evaluating Road Safety in Arkansas Using Observational Analyses

Several intersection locations were chosen to be evaluated using hands-on observations. This required on-site examinations and surveys of road conditions, driver behaviors, and the effect that road conditions have on driver behaviors. This type of analysis allows for several crash hazards to be observed, analyzed, and described in a way that is not represented within any collected historical data; especially with regards to human factors.

1.2.3 Developing Implications of the Two Methods Used Together

Each method gives a different perspective of road traffic safety. The limitation of one study may be the strength of the other study. More importantly, using both of these methods for safety evaluation gives a combined insight that is vastly superior to either of the stand alone methods.

CHAPTER II - LITERATURE REVIEW

2.1 Growing Need for Improved Road Traffic Safety

Road traffic safety is not a new issue. It has been around since the first automobile moved onto the road, but its importance has grown drastically. In less than a century since its invention, the automobile became the leading cause of young adults' deaths in the United States (Mashaw and Harfst, 1991). Today, these trends have grown and spread all around the world. Road traffic accidents and injuries are quickly becoming the leading concerns in global health, especially in developing countries (Peden, 2004). As these risks have continued to escalate, the United States has failed to maintain its position as the world leader in safety, and continues to fall in the ranks (Evans, 2003). There is now a great need for safety attention in the United States, and in particular, road traffic safety. This need can be seen through numerous risks that are currently growing in impact. Population growth and technology are just a couple of these risks, while lack of litigation is another.

2.1.1 Increased Volume of Vehicles

World population is undoubtedly growing as it always has. By the year 2000, the global population officially exceeded 6 billion, and it is projected that it will jump to 7 billion early within the next decade (US Census Bureau, 2008). Generally, the population growth has been steady over the past 60 years, showing an annual increase of around 1% for the United States. Between 2000 and 2005, the United States had a total population growth of about 5.3% (NHTSA, 2000-2006). This trend also continues when considering the number of registered vehicles and licensed drivers within the United States. Throughout the past decade, a typical year produced nearly 2 million new licensed drivers and around 2.5 million new registered vehicles. This increase has dramatically increased the vehicle volume on today's roads, making travel all the more demanding for each driver (Pickering, 2004). Highway statistics from the US Department of Transportation have also shown an increase in the average total automobile kilometers travelled annually (FHA, 1992-2006). What this means is that not only are there more cars on the roads each year, but each vehicle is active on the roads longer. People are now travelling longer distances for work or for recreational travel than ever before, further increasing the volume on the United States roadways (Pickering, 2004). Leonard Evans, DPhil., who has been one of the lead researchers of traffic safety for well over 30 years, has suggested the two most important factors in traffic safety: the individual driver's behavior, and the behavior of all other road users (Evans, 2003). Therefore, with a substantial increase of road users, the workload demanded on each individual driver also increases. This gives rise to potentially more and more road traffic accidents if not prevented through road traffic safety measures.

2.1.2 Increased Driver Inattention

The ability to drive and to drive safe depends on the mitigation of a number of important tasks which often relate to driver focus and control (Salvucci, 2006). In terms of control tasks, the driver must have their hands on the wheel in order to steer and have their feet on the brake and acceleration pedals to drive. Focusing tasks not only include the driver keeping their eyes on the road, but processing what is going on in order to stay in the proper lane, maintain their speed, obey traffic signs and signals, and avoid any sudden hazards. The level of focus that the driver has at any one moment also affects their ability to make decisions while driving. According to their comprehensive study, Weirwille et al. found that all of these primary tasks require some amount of cognitive processing from the driver (Weirwille, Tijerina, Kiger, Rockwell, Lauber, and Bittner Jr., 1996).

The danger in road traffic safety is when drivers fail to perform these tasks, by taking their hands off the wheel and their eyes off the road (Pickering, 2004; Wogalter and Mayhorn, 2005). In most cases, this is a consequence of additional tasks performed by the driver that are not related to the primary task of controlling the vehicle and focusing on the road. The National Highway Traffic Safety Administration performed a study which surveyed drivers who admitted to performing tasks such as talking on cell phones, changing radio stations, eating, talking with passengers, fixing their hair, and even daydreaming while driving (Sundeen, 2007). Just as for the primary tasks, these additional tasks also require cognitive processing by the driver. However, several studies have shown that the processing capability of any single driver is limited (Weirwille et al., 1996). When any one task demands too much of the driver's total cognitive capacity, overall performance of that task may be degraded. This is also true when several tasks require more than the driver's total mental capability; one, many, or all of the tasks' performances can be degraded. For a driver, a task such as adjusting the radio station requires some of the attention that was being used to focus on the road, as well as a hand that is no longer on the wheel and both eyes which are no longer on the road. Distracting tasks like this, along with many others, create an enormous amount of mental workload for the driver, which could potentially lower the performance of the primary tasks. This can result in vehicles swerving in lanes, speeding, driving through stop signs, or running into objects. Therefore, it is not surprising to find out that the more distractions that are presented the more at risk drivers are at being involved in a collision or being injured (McCormick, 2003, Pickering, 2004).

Today, with the rise of technology there is no shortage of distractions, especially for drivers (Trbovich and Harbluk, 2003; Wogalter et al., 2006). At the helm of these distractions are cell phones and their growing usage in everyday life. Cellular telephones were introduced in the early 1980's with the intent of having a quick, convenient, and remote source of communication; ideal for emergency situations. During the first decade, the cell phone was thought of more or less as a novelty item, which due to its bulkiness and price often found very few users (Sundeen, 2007). Today they have evolved into a widespread commodity, cheap in price and with limitless functionalities; texting, e-mail, video and image recording are all examples of the cell phones use today (Wogalter et al., 2006). In 1995 there were a total of 28.1 million wireless subscribers, according to the Cellular Telecommunications and Internet Association. That total grew to 97 million in 2000, 194.4 million in 2005, and as of the beginning of 2008 it has reached over 254.6 million subscribers (CTIA, 2008). The CTIA also showed that over 80% of the US population owns some type of cellular phone, as opposed to only the 11% in 1995. The exponential rise of wireless subscribers has also shown a substantial increase in frequency of use (Wogalter et al., 2006). Cell phones are no longer used merely for the rare emergency, and because of their mobility, calls no longer have to wait for the office or at home. People are more accessible because of cell phones, which allow them to make calls at practically any time of the day and at low costs. In a 2001 study, researcher David Strayer found that 85% of all cell phone users admitted to using them while driving, and that nearly 60% of all cell phone conversations occur while in a vehicle (Strayer, Drews, Albert, and Johnston, 2001).

The use of cell phones presents several potentially distracting activities for a driver of an automobile. According to Goodman et al., these tasks include acquiring the phone, dialing, engaging in communication, and other associated tasks such as text messaging, or reading a map or calendar. In general the cell phone is not immediately in the hand of the driver, but rather it is somewhere where it must be found and grasped. Phones in pockets, purses, dash board consoles, or other areas require the driver to move one or both hands off the steering wheel to search for the phone. Dialing is also a task that requires at least one hand, and generally both eyes (Goodman, Tijerina, Bents, and Weirwille, 1999; Wogalter et al., 2006). The actual conversations can vary substantially with how much cognitive possessing actually occurs, depending on whether the driver is talking or listening. It also depends on how engaged the driver is with the conversation. In general, the more engaged a driver is in conversation, the less engaged they are in focusing on what is happening on the road (Lamble, Kauranen, Laakso, and Summala, 1999).

In 2004, General Motors released a public statement claiming that driver distractions contributed to more than 25 percent of automobile crashes (Pickering, 2004). However, recent studies have shown that up to 78 percent of crashes studied over a 12 month data collection period were due to driver inattention; 60 percent of near-crashes were also shown to be caused by distraction (Klauer, Neale, Dingus, Ramsey, and Sudweeks, 2006). Statistics from the 2006 study by the National Highway Traffic Safety Administration has even shown that driver inattention generated 4.9 million crashes, 34,000 fatalities, 2.1 million injuries, and up to $184 billion in economic damage. Therefore the move recently has been for state legislatures to pass laws and regulations of cellular phones to help lower these numbers. Every state and area in the US has at least proposed some form of cell phone regulation in the past five years, yet only New York, New Jersey, and District of Columbia have passed laws banning hand-held cell phones (Sundeen, 2007). Trends are now showing that there will be a rise in hands-free cell phones in the future. Whether or not this will make a significant contribution to crash safety is yet to be seen. Currently, many researchers are analyzing the effects of both hands-on and hands-free technology, and the differences between them. Some studies have already shown that hands-free cell phones do not make a significant improvement over handheld cell phones, despite the fact that they eliminate the distractions of searching for and manipulating the device (Redelmeier and Tibshirani, 1997; Tijerina, 2000). More research in this area is needed in the coming future.

Cellular phones are currently the leading source of in-vehicle distraction, but they are being closely followed by the steady rise of new auxiliary devices entering into the global marketplace; these include products like Personal Data Assistants (PDAs), Global Positioning and Navigation Systems (GPS), and MP3 players (Pickering, 2004; Salvucci, Markley, Zuber, and Brumby, 2007; Sundeen, 2007). New devices also create new forms of distraction for the driver, but the effects are the same. Navigating through maps and menus and the physical manipulation of these devices are putting the driver at risk of collision as their attention is drawn away from their primary task of driving (Salvucci et al., 2007). Intentionally, these devices were designed to assist or enhance the driver's performance in some way, as with the GPS and its ability to direct lost drivers. Cell phones and MP3 players have even been shown to increase performance of driver tasks such as lane keeping and speed maintenance in situations where fatigue is a factor (Goodman et al., 1999). But researchers argue that despite these benefits, they are still outweighed by their distracting effects (McCormick, 2003).

Jim Geschke, vice president and general manager of Johnson Controls has stated that it is inevitable that drivers will find more ways to bring excessive information into the vehicle. Drivers do not necessarily need cell phones and GPS devices to drive safer, but they believe that they do. When the information the drivers want is not already within the vehicle, they will bring it in themselves through the use of these devices. However, Geschke goes on to say that this typically is never done in a safe manner. This has led to a vast increase in human/machine interactions research between drivers and their vehicles (McCormick, 2003). Today, it has become an increasing responsibility for the automobile companies to understand the cognitive workload on their drivers, so that they can develop the safest ways of meeting their needs. If the automobile companies do not invest in researching these topics, drivers will continue to bring in new and more distracting devices, and potentially put everyone on the road at risk of injury.

2.1.3 Increased Road Traffic Accidents within the United States

Despite the fact that population and technology growth has spread worldwide, the United States has been one country that has failed to keep up with road traffic safety. The two decades between 1979 and 2000 have shown several countries such as Canada, Britain, and Australia of having an overall reduction in traffic fatalities of 50%, 46%, and 48% respectively (Evans, 2003; FHA, 1979-2000). The United States during that same period had only reduced its numbers by 18%. For 2002, the United States saw a total of 42,000 road fatalities; 15,000 or more of which could have been saved if the country kept up with the global trends. Internationally, the United States, which once led the ranks in traffic safety during the early 1980's has now fallen to 9th place (Hakim, 2003) and is currently still declining.

In a 2003 editorial, researcher Leonard Evans gave a comparison of air traffic and road traffic safety litigation. In his study, Evans pointed out the effects of the terrorist attacks on September 11th, 2001, where nearly 3,000 American peoples were killed. America's focus turned quickly to rid the country of such a tragedy from ever happening again and increased airline safety measures drastically (Evans, 2003). Yet, for virtually every month since these attacks, more Americans died on the roads due to preventable traffic accidents (NHTSA, 2002-06). Road traffic safety has not received nearly the amount of attention as it is deserved, comparably. It has been viewed that there is somewhat of an unbalanced litigation in the United States safety policies, which direct the focus away from the critical countermeasures needed for road traffic safety improvements. This inattention to prevention has even suggested an estimated 100,000 American lives lost over the last two decades (Evans, 2003). Evans also suggests that the focus that has been made on road traffic accidents has been more on the side of reduction in crash and injury severity, rather than the more critical aspect of accident prevention. The United States safety litigation generally implies that crashes are and will always be inevitable events, when in fact they are all, to some extent, preventable with some underlying understanding of the situations (Peden, 2004). In order for the United States to follow the trends of Canada, Britain, and the like, they must first emphasize the fact that road traffic accidents are a public health issue, which they currently do not emphasize. This would lead to greater support for scientific research and studies on crashes and their countermeasures (Evans, 2004; Nantulya and Reich, 2002; Peden, 2001).

In 1968, a researcher by the name of William Haddon Jr. illustrated the possible opportunities for road traffic safety intervention (Peden, 2004). He summarized the interactions of the human, vehicle, and environmental factors throughout three phases of a crash: pre-crash, crash, and post crash. His work produced what became the Haddon Matrix (Figure 1), which displayed several opportunities for reducing the risk of accidents and reducing the risk of injury or consequences of a crash.

		Factors		
Phase		Human	Vehicles and Equipment	Environment
Pre-Crash	Crash Prevention	Information Attitudes Impairment Police Enforcement	Roadworthiness Lighting Braking Handling Speed Management	Road Design and Road Layout Speed Limits Pedestrian Facilities
Crash	Injury Prevention during the crash	Use of Restraints Impairment	Occupant Restraints Other Safety Devices Crash-Protective Design	Crash-Protective Roadside Objects
Post-Crash	Life Sustaining	First-Aid Skill Access to Medics	Ease to Access Fire Risk	Rescue Facilities Congestion

Figure 1: The Haddon Matrix

Although trends in the United States are worse in many aspects, global road traffic safety is just as big of a concern. According to a study performed by the World Health Organization in 2004, road traffic injuries ranked 9th on a worldwide compilation of leading causes of the global burden of disease and injury in 1990; it was just under tuberculosis and measles. It was projected that road traffic injuries would rise up to be 3rd in the year 2020, just under heart disease and uni-polar major depression; war was projected to be 8^{th}. These projections showed that road traffic deaths will increase substantially in low-income countries, even though there will be an overall 30% decrease for high-income countries like the United States and Britain (Peden, 2004). Another study has shown that 85% of all deaths from road traffic accidents occurred in developing nations, as well as 90% of all disability adjusted life years (DALYs) lost (Nantulya et al., 2002). Among the reasons for this rising burden (in addition to the rising populations) were also poor enforcement and regulation of safety laws, poor public health infrastructure, and poor access to health services. The countries considered by Nantulya and Reich in their study to be developing nations were China, India, South America, the Western Pacific, and South East Asian countries. Africa, the Western Pacific, and South East Asian countries are currently the countries with the highest number of deaths per 100,000 in population (Peden, 2004). These rates range from 19 to 30 deaths per 100,000 in population.

2.2 Characteristic Factor Levels of a Collision

Within any road traffic accident, whether a single-vehicle or a multiple-vehicle crash, there exist several different levels of characteristic factors that make up the details of the crash. These levels consist of the environmental, temporal, geographic, infrastructural, vehicular, and human aspects of an accident. In essence, each road accident consists of a road, its surroundings, and its victims. Any detail that describes these things, both before and after the collision, is considered to be a characteristic factor of a collision. Throughout the research community, each category has been shown to be of great importance, however greater emphasis today has been on the infrastructural and human factors involved (Janson et al., 2001; Noy, 1997; Rasmussen, Nixon, and Warner, 1990).

2.2.1 Environmental Factors

Studies usually differ when it comes to what details of a crash site should go into each category. This is especially true for environmental factors. Road attributes are occasionally included in the environmental category, as in the research of Janson et al. (2001) and Shankar et al. (2004). More commonly however, researchers narrow these road attributes into another category; infrastructural factors (shown later in this literature review). The environment, in its simplest form, is regarded as the uncontrollable elements of a location that affect the road and its surroundings, but that are not an actual part of the road or its surroundings; weather and atmosphere are the prime examples of environmental factors. These factors are exogenous in nature, due to the fact that they are beyond the control of any person or policy (Chang and Graham, 1993). Atmospheric conditions of a particular segment of road, such as whether it was clear, raining, snowing, sleeting, or extremely windy can all have a major impact on the road's overall safety. Other examples include the state of the road surface (icy, dry, wet), lighting conditions (daylight, dark, dark but lighted, cloudy), or other uncontrollable environmental issues (fog, smog) that can affect the vision of the driver or the drive of the vehicle (Kim, Lee, Washington, and Choi, 2007; Yau, Lo, and Fung, 2006).

Weather and atmospheric conditions have always been and will continue to be a part of nature, which cannot be prevented. Yet their effects can. The presence of street lights, dark or cloudy conditions, salt-based chemicals for slick surfaces, roadway coverings, as well as warning systems are all ways to reduce the effects of the environment (Ahmad and Rahman, 2003; Carson and Mannering, 2000).

As with many of the other characteristic factors, environmental factors should not be treated independently. In general, many environmental factors depend highly on the time of day, season of the year, and other temporal factors (Carson et al., 2000; Lord and Persaud, 2000).

2.2.2 Temporal Factors

As was mentioned in the previous section, most characteristic factors involved with a road traffic accident are dependent on other characteristic factors. One of the larger interdependencies is between environmental and temporal factors (Carson et al., 2000). Temporal factors are those which specify or reference a collision with respect to one instance in time. Along with weather, Carson et al. mention that traffic volumes are highly dependent on the time of day. For example, the rush hours in which individuals drive to work in the morning and from work in the afternoon are known for their increased road congestion. Lunch-hour traffic is another example. Therefore, the time of day that a collision occurs can play a large part in the analysis of traffic safety. Late night and early morning times can also be attributed to human factors such as fatigue and sleepiness (Baulk, Biggs, Reid, van den Heuvel, and Dawson, 2007). Another important temporal factor is the day of the week. In some areas, Fridays and Saturdays may experience higher traffic volumes do to vacations and recreational trips as an example. Seasonal information may also prove to be an important factor. Colder seasons of the months between December and February may lead to greater or more extreme environmental factors (Carson et al., 2000). To fully understand the risks involved through temporal factors, it is most beneficial to have all aspects of the time of a crash known: time of day, day of the week, month, and year. The importance of these factors cannot be overstated, as they are the key to discovering the trends within road traffic accidents. Predictability is a leading feature for accident prevention and cannot be completed without known references in time (Lord et al., 2000).

2.2.3 Geographic Factors

In the context of these subcategories, geographic factors should not be confused with infrastructural factors. Geographic, for the purposes of road traffic safety, is in reference to a physical location, and not the characteristic shapes and curves of the road system (Van Beeck, Mackenbach, Looman, and Krunst, 1991). A simple example of a geographic factor would be the state, county, or city that a particular collision occurred in. These factors can be as broad as the country where the collision occurred, to as detailed as the name of the street, section, and mile number of a particular segment of road. In case a particular road does not have a specific section or mile associated to it, a geographic factor could consist of a simple reference point. For example, a collision that occurred a few blocks away from a major intersection could be referenced as such to that intersection, given a proper distance and directional heading. Therefore, directions can also serve as a geographical factor. One of the most important geographic factors considered today is the distinction between urban and rural roads (Gårder, 2005). Urban and rural distinctions change, however, from county to county and from city to city. A general rule from the US Census Bureau used by policy makers is to classify urban or metropolitan areas if they contain a total metro area population of at least 100,000 residents or if they are economically tied to those core metro areas. Nonmetropolitan or rural regions are those outside a metro's boundaries that do not include cities with any more than 50,000 residents (Ricketts, Johnson-Webb, and Taylor, 1998).

2.2.4 Infrastructural Factors

This category describes the physical layout and design of a particular road system. Within the infrastructural factors, there lie two important areas: the road itself, and its immediate surroundings (de Leur et al., 2003). A road can be designed based on its composition and its shape. Examples of these factors include the type of road surface (concrete, asphalt, dirt, gravel) used as well as the physical grade and curvature of the road itself (straight, curved, level, uphill, downhill). Traffic lines are also a key to the infrastructure. These lines help designate right-of-way policies by directing traffic into their designated lanes, showing where a vehicle can pass other vehicles, or where the vehicle can safely make a turn (Flahaut, 2003). Surface infrastructure can refer to the original designed conditions of the road (as the above examples), but more importantly it can refer to unintended conditions such as potholes or worn out traffic lines. Potholes can increase damage to the vehicle, which may in turn cause the vehicle to lose control and wreck with other objects, whereas worn traffic lines can lead to driver confusion (Karlaftis and Golias, 2002). The type of road (US highway, interstate, city road, on ramp, off ramp) and its relation to other roads (intersection, merging lane, alley, driveway) are other important infrastructural factors of a road segment (Van Beeck et al., 1991).

Apart from the road itself, the road's immediate surroundings are also factors when considering road traffic safety (Peden, 2004). If collisions occur at the edge of a road, then the infrastructures of these areas are important as well. For example, the side of a road may consist of a ditch, trench, sidewalk, median strip, or a fixed object; all of which play a critical part when considering the impact of a collision (Yamamoto and Shankar, 2004). Road signs and traffic controls are often considered to be significant factors of a location (Peden, 2004). Road signs give drivers the ability to predict the physical infrastructure or important events ahead of them, which allow them to better prepare their actions. Traffic controls help to direct traffic, whether by a sign (STOP, YEILD, CAUTION) or if there is a light that changes accordingly with traffic. Although these controls often help traffic flow, they may be misused if people choose to ignore them or if the controls are not functioning properly (Escalera, Moreno, Salichs, Armingol, 1997). Signs and controls that are placed in poor areas, not functioning properly, or hidden from view can lead a driver to misinterpret road and traffic conditions ahead of them, which can greatly affect the overall safety of the driver and other vehicles on the road.

One of the reasons that infrastructural factors garner so much attention in research today is that they are factors that can be altered (de Leur et al., 2003). Geographically, locations cannot change; a road in Johnson County will always be in Johnson County, unless, of course, the name changes. Time is a function that is constantly changing, but the way in which it changes cannot be altered; a person can avoid a certain road at a particular time, but they cannot avoid that particular time. Roads are always being influenced by their environment, and although they can reduce the effects of if, they cannot alter the environmental factors. The infrastructure of a road, however, can be altered. It may not always be the most cost effective method, but preventative measures can be made by using road maintenance to fix pot holes and lane markings or by adding or changing road signs and traffic control units to better direct traffic. Roads can even be widened or moved to include more lanes in the case where traffic flows become too great (Noland, 2002).

2.2.5 Vehicular Factors

For every roadway location there exists environmental, temporal, geographic, and infrastructural factors, regardless of whether or not an accident occurs at that location. Every location has a geographic reference point, physical characteristics, and is influenced both by time and the environment surrounding them. However, the vehicular and human factors are the characteristics that are externally brought into the location influencing traffic safety. The vehicle plays a major role in road traffic accidents, and several crash-influential factors can be attributed to it. These factors may include the age, type, make, and body of the vehicle (Evans, 2004). Older vehicles may have engine issues that cause the car to die in the middle of a busy interstate, or worn tires may lead to a dangerous blowout, for example. Each vehicle on the road has its own unique physical limitations that may be exceeded due to severe environmental problems or bad infrastructures (Peden, 2004). A passenger car may not be able to function well in conditions where ice covers a stretch of road, whereas a sports utility vehicle that can withstand such conditions may have a tendency to roll over in steep, curvy road conditions. Another example where physical vehicle factors plays a part is in situations where a wreck occurs between large and small vehicles. In these situations, the smaller cars are at greater risks simply due to its size disadvantage (Evans, 2004). Therefore the size and current condition of each vehicle can turn out to be a major cause of a road traffic accident.

Additional vehicle factors that are of importance are the lighting and warning systems of each automobile (Zhang, Huang, Roetting, Wang, Wei, 2006). In dark settings, proper lighting is crucial for drivers to physically see the road and its surroundings. If headlights are not in working condition, not only is the driver's vision impaired, but other vehicles on the road may not see the vehicle as well or at all in dark conditions. Brake lights and turning signals are used to warn following vehicles that the vehicle will be making a sudden departure from their current speed or direction. Without these properly working devices, vehicles may fail to become aware of these changes and cause a rear end or other type of accident.

2.2.6 Human Factors

According to Evans in 2003, the two most important factors in road traffic safety are the individual driver's behavior and the behavior of every other vehicle on the road. Human factors, in the context of road traffic safety, are the factors that are in the direct control of the driver as well as the personal, physical, or psychological characteristics of the driver (NHTSA, 2008). A study performed in 1980 proved that 90% of road traffic accidents were attributed to human factors, either directly or indirectly through other factors (Sabey and Taylor, 1980). The most common personal characteristics of a person consists of their age, gender, race, weight and overall health conditions. Other personal characteristics can further describe the health and state of a person, such as any physical disabilities they may have; vision and hearing impairments are examples of these. The level of fatigue and sleepiness of the driver is also a major concern (Baulk et al., 2007). The actions performed (or not performed) by the driver are also key human factors. For example, obeying traffic laws, speeding, wearing a seat belt, or driving while under the influence of drugs or alcohol are all types of actions that the driver has direct control over, which may impact the occurrence or severity of a collision (Janson et al., 2001; Rasmussen et al., 1990).

In general, human factors can be the most difficult to measure or determine at any particular crash site (Sundeen, 2007). Personal characteristics aside, the actions that a driver was engaged in before the collision may be unclear and may depend on the driver's own interpretation of what happened. Yet, driver distraction and inattention are still considered as the root cause of many collisions (Strayer et al., 2001; Sundeen, 2007). Talking on cell phones, putting on makeup, eating, adjusting the radio, grabbing something from another seat, or looking at maps are all examples of driver distractions and are considered to be human factors. Numerous researchers, such as Sheridan (2004), Horrey et al. (2006), and Neyens et al. (2007), have shown driver distraction and inattention to be any action that diverts the driver's main attention from the road and it's surrounding causing a decrement in driver awareness and road traffic safety. Cell phones normally get the most attention from a human factors standpoint, simply because they are one of the easiest aspects to measure (Sundeen, 2007). According to his study, Sundeen explains that it is because of their visibility that cell phones get spotted and remarked as an important safety hazard. On the other hand, there are now devices that exist that are not as visible as cell phones that tend to distract drivers, such as navigation devices, PDAs, and MP3 players (Pickering, 2004; Salvucci et al., 2007).

Much like road infrastructure and vehicles, human factors gain a generous amount of attention due to their preventability. Human factors, more than any other factor, are under the control of the driver. By simply altering their behavior, drivers can easily avoid a number of instances where they might have found themselves in danger of collision. The issue with these measures, on the other hand, is the willingness of drivers to actually alter their behaviors (Rumar, 1988).

2.3 Statistical Analysis of Road Traffic Safety

The use of statistical studies of historical, numeric data has become increasingly popular in many areas of study over the past few decades, including not only crash analysis, but also economic, biological, and sociological applications (Anastasopoulos et al., 2007). The main attraction to the analysis of past data is in its ability to unlock potential methods of predicting the future or inferring the past based on historical trends. By quantitatively determining these trends, it becomes easy to understand the relationships between one or more factors. In the world of statistics, the main method for determining these trends is the use of regression modeling (Al-Ghamdi, 2002; Berhanu, 2004). Regression, in its broadest sense, is a way of developing a “best fit” model that encompasses a number of independent, explanatory variables and a single, dependent response (Lewis-Beck, 1980).

The decision about what data should be considered to be either explanatory or response depends heavily on what the experimenter wants to analyze. The answer is not always apparent. In the case of crash analysis, the response might be the number of crashes that occurred within a certain area and the explanatory variables could be the time of day, road traffic volume, age of the driver, or any other numeric data gathered from the crash site (Anastasopoulos et al., 2007). Perhaps the model would show that an increase in traffic volume leads to an increase in crash frequency. Using regression, it becomes simple to determine which numeric variables in a process significantly affect the numeric response being observed. However, there exists a vast number of unique models that can be applied to historical data. Choosing the best model is the key to reliable results (Lord, Washington, and Ivan, 2005). Each regression model should have a method for evaluating the goodness-of-fit to the data, which will in turn determine whether or not the model is feasible. A poorly fitted model has little or no predictive capabilities, and could be considered scientifically useless (Saccomanno, Nassar, and Shortreed, 1996). Deciding which regression model is best for any specific data depends on many underlying assumptions about the data (Lord et al., 2005). The first of which is the nature of the response. The dependent variable, or response, of the model is assumed to be a random variable. The nature of the response, and the model itself, can then be defined by the type of random variable the response is and the probability distribution assumed by the model. The most common type of regression model used is that of a linear regression model where the independent and dependent variables are assumed to be continuous random variables (Berhanu, 2004). Within each model there are two values for each response: the expected value of the line and the true value of the data. The errors, which are the differences between the two, are then minimized using the method of least squares. Regression uses this method to change the parameters of the linear model in such a way that the error terms are as small as possible. The result of minimizing the errors is what ends up being the "best fit" model (Lewis-Beck, 1980). Also, it is assumed in all models that these error terms are normally distributed; a.k.a. each error term is independent of any other error term (Jones and Jørgensen, 2001; Kim et al., 2007).

However, the assumption of continuous variables is often times inaccurate, especially in crash analysis. This has led researchers to find better models for their analysis. In some studies, the assumption of continuous variables is addressed (Al-Ghamdi, 2002; Bernahu, 2004; Kim et al., 2007). For these studies, the assumptions were relaxed so that independent variables could be discrete or even binary. In 2001, Jones et al. proposed a model for crash data in the Norway. For their analysis, the response variable was binary, determining whether a crash was fatal or not. Also, their model consisted of several other binary and continuous independent variables. For this to work, Jones et al. developed a logistic regression model, which relaxes the model assumptions and allows the response to be binary (Jones et al., 2001).

Kim et al. preformed a similar study in 2007, modeling the types of crashes occurring in Georgia. Crash types such as angle, sideswipe, rear-end, or single vehicle crashes were analyzed using the logistic model (Kim et al., 2007). The logistic model is very similar to that of the simple linear model, but the change in assumptions also leads to a change in model parameter estimation. One key feature that makes logistic regression attractive is its ability to calculate an odds ratio, which allows the experimenter to interpret the change of an event's likeliness to occur given a change in the independent variable described by that odds ratio (Al-Ghamdi, 2002).

Another aspect of these two logistic regression studies is that the data used for the models were hierarchical in nature. This refers to the fact that there are people within each vehicle, within each crash. It is safe to assume that the responses between passengers in the same vehicle are correlated with each other, as are responses between vehicles that are within the same crash. This would mean that there is a violation of the normally distributed error terms in these experiments, because they are not fully independent. Therefore, these models were adjusted in such a way that data was clustered among passengers in the same vehicle and vehicles in the same crash. Then, each cluster of data was treated as independent. This is called a hierarchical logistic regression, and works around the assumption of normally distributed errors (Jones et al., 2001; Kim et al., 2007).

Another case where the continuous random variable may not be the best choice for a regression model would be involved with responses that represent a count or a frequency. A response that represents a count or a frequency, such as the number of crashes in a particular area per year, is necessarily a positive and discrete number (Anastasopoulos et al., 2007). Also, many studies have shown that crash occurrence can be more realistically described as a Poisson process. Poisson is a discrete probability distribution that represents the probability of a number of events occurring during a fixed period of time, such as customer arrivals in a store per hour (Abdel-Aty et al., 2000; Bernahu, 2004; Lord et al., 2005). When regression takes on responses that are Poisson distributed, the model must be adjusted. This is done by transforming the responses using the logarithm of each response. The explanatory variables are left alone, as only the response is transformed. From here each independent variable is treated as usual, where the regression technique attempts to find the best fit linear trend of the explanatory variables and log transformed response (Abdel-Aty et al., 2000). Other studies have used other unique regression models, changing the type of response and underlined probability distribution. Negative Binomial regression, which is used to data similar to Poisson regression, is used often when the underlying assumptions of the Poisson are violated (Bernahu, 2004). A 2007 study showed the use of a Tobit regression model for crash rate analysis, similar to crash frequency analysis but with different underlining assumptions to the model. Founded by James Tobin, Tobit regression was originally used for economic analysis, but was later applied to crash rate analysis. Its model contains a method of censoring the range of the dependent variable by clustering of data, rather than data truncation (Anastasopoulos et al., 2007).

2.4 Roadway Safety Audits

The techniques to evaluate road traffic safety do not have to be limited to the evaluation of road traffic crashes. The prerequisite for this type of technique is that a crash actually occurs. But since the biggest issue with road traffic safety today is the avoidance and reduction of crashes it is best to find another method that does not consider crashes having already occurred (Evans, 2003). The proactive approach would be to evaluate the roadways before a potential crash even occurs. A roadway audit has the ability to catch troublesome aspects of the road, such as potholes, infrastructure, signage, and other aspects that could potentially lead to road traffic un-safety (Allsop, 1997). In their 2003 study, de Leur et al. proposed a method to formally evaluate not only currently existing roads, but also roads that have not yet been built. The ability to look at and evaluate a road system in its planning process can lead to huge cost savings in the future (de Leur et al., 2003). Putting in the effort ahead of time prevents changes to have to be made later on when the infrastructure turns out to be weak. Dwight Horne, director of the Office of Highway Safety Infrastructure states that the redesigning of a road is much more cost-effective than the reconstruction of a road (Horne, 1999).

The process of a roadway audit is performed by a well trained, multi-disciplinary team of auditors. The auditors work independently of the road engineers and project managers. They can evaluate road systems in any of five stages of design: the feasibility study, draft design, detailed design, pre-opening, and post-opening of the road system (Allsop, 1997). The auditors then document their finding, which include the potential safety hazards, in a documented form that goes to the roadway project managers. At this point, the project managers evaluate the findings and make any necessary changes they feel applicable (Horne, 1999). Although it is not a guarantee that all changes will be made, the roadway audit allows each of the safety measures to at least be fully considered.

CHAPTER III - DATA DESCRIPTION

This study uses data collected from the state of Arkansas during the period of 2002 through 2004. The Arkansas State Highway and Transportation Department (AHTD) has allowed the use of two essential databases in order to study and analyze roadway traffic accidents throughout the state. The first database is referred to as the crash database and contains a log of all vehicle crashes reported or collected by the Arkansas State Police during this time period. Each crash is described in detail within each log, containing information about all persons, vehicles, and conditions involved with the accident. The second database is known as the Arkansas roadway inventory database. The details contained in this database pertain to the intricate road systems within Arkansas, listing every major and minor road, along with the geography of each road segment. Together, these two databases contain all the necessary information needed to support the objective of the study.

3.1 Collecting Crash Data

Crash data consists of a number of descriptive characteristics associated with any particular road traffic accident that has been recorded in some fashion. The level of detail can vary substantially, depending on the situation and who is collecting the data. In most general cases, crash data is collected by city or federal officials who are present at a crash scene. Whenever a crash is reported and the proper authorities are notified, it is generally required by law to document and log pertinent information about the accident. Documentation is usually performed by filling out forms or inquiry sheets, allowing the information to be further recorded and archived later. Traffic accidents can vary substantially in size and severity, which causes some accidents to require more or less information. Some smaller, single vehicle accidents may not appear to need a largely detailed report to explain their cause. Other, larger accidents require enough information about the crash in order to determine the cause of the wreck, perhaps for insurance or legal reasons: who was involved, whether it was due to driver error, road issues, weather, or any number of factors, etc. Over time, most agencies have developed a standard amount of information to be documented for each traffic accident.

For the state of Arkansas, all city and state officials are required to record several pieces of information at each crash site, called variables in this study. The information recorded includes several elements of temporal, environmental, geographical, infrastructural, vehicular, and human factors as described in the literature review. All officials are required to fill out as much information for every traffic accident as possible, regardless of the magnitude of the accident. After documentation is complete, the crash information is sent to the AHTD to be logged into the crash database.

At the AHTD, all the information from the crash report is uploaded into the database and checked. Then, using that crash data, the department generates several other important pieces of information. Many important crash variables within the database are not gathered directly at the time of the crash's initial investigation. This is usually due to the fact that some information may not be readily available to the city or state official when filling out a crash report. Road details such as the average daily traffic (ADT) or whether or not it is located in an urban or rural city can be determined after the initial crash report, as long as a specific street name and reference point are listed. Passenger ages can also be generated back at the AHTD, as long as their dates of birth are recorded. The details of this database, and the variables included within it is covered more thoroughly in the following section.

3.2 Crash Database

For every roadway traffic accident in the state of Arkansas that is reported and investigated, its details are entered into the crash database. There are 82 characteristic pieces of information for each entry that is entered into the database. Instead of having only one entry per crash, the data takes into account that each crash contains a particular number of vehicles and that each vehicle contains a particular number of passengers. Because of this hierarchical like form, every individual person involved with the accident gets an entry in the database. The information variables in each entry reflect elements, or factors, describing the details of the crash site in terms of temporal, environmental, geographical, infrastructural, vehicular, and personal attributes. Along with these factors, the database also contains information about the outcomes of the accident. The database is divided into three sections of data, relating to the levels of the crash hierarchy: Crash, Vehicle, and Person levels. Vehicle and Person levels pertain specifically to the vehicular and personal factors, respectively, as described in section 2.2 of the literature review. The Crash level, however, is a much broader category. The temporal, environmental, geographical, and infrastructural related factors are all contained within the Crash attribute.

The crash database includes a total of 136,164 data entries over the three year span among all of Arkansas' 75 counties and 920 cities. Table 1 shows a general breakdown of the number of data entries throughout the three years.

Table 1: Arkansas State Highway and Transportation Department Database

	2002	**2003**	**2004**	**Total**
Crashes	70,903	70,912	74,059	215,874
Vehicles	128,727	127,216	133,204	389,147
People	190,296	187,225	196,428	573,949
Average Number of Vehicles per Crash				1.803
Average Number of Passengers per Vehicle				1.475

3.2.1 Crash Level

As was mentioned earlier, every road traffic accident can be considered as a single crash involving vehicles involving people. These three things form a natural ordering hierarchy, with the Crash level being the broadest level. Because of this, the Crash level contains the bulk of the information recorded within the crash database. The entire database covers over 80 distinct characteristics (called variables) that detail the events and conditions of the crash site and the crash itself. 43 of these categories are considered to be within the Crash level. Crash level, in terms of this database, refers to any element of the accident that describes the crash as a whole, including its outcomes. For the most part this includes the temporal, environmental, geographical, and infrastructural details of the crash scene as described in the literature review. Examples include the time of day, weekday, road surface conditions, weather, city name, road type, junction type, etc. All of these variables can be used to describe the factors that may have potentially led to or even caused the accident. However, there are a few Crash level variables that are not considered to be any of these four 'contributing' factors, because they detail the specific outcomes of the crash. An outcome refers to the type of collision that occurred, the severity of the crash, or even the number of fatalities. These variables are not contributing factors to the crash; instead they are resulting circumstances of the crash. A complete table of Crash level information variables is shown in Table 2, along with a brief description of the variable labeled as *temporal, environmental, geographical, infrastructural*, or *outcome*. Two variables are labeled as *reference*, and their only purpose is identifying a specific crash, assigning the crash and the form used by the city or state official a specific code.

Table 2: Crash Level Information (Variables)

Variable Name	Description	Example Values	Type
CRASHDATE	Date of Crash (encoded)	37260	Temporal
WEEKDAY	Day of the Week	FRI	Temporal
CRASHTIME	Time of Crash (encoded)	1.520833333	Temporal
ATMOSPHERICCONDITIONS1	Atmospheric Conditions	Clear	Environmental
LIGHTCONDITIONS	Light Conditions	Daylight	Environmental
ALCOHOLINVOLVED	Was Alcohol Involved?	N	Environmental
RURALURBAN	Accident Locale	Rural	Geographical
COUNTY	County	Union	Geographical
COUNTYNO	County Number	70	Geographical
INCITY	Crash in City?	N	Geographical
CITY	City	Hamburg	Geographical
DISTANCEFROMNEARESTCITY	Distance from Nearest City	5.4 Mi	Geographical
DIRECTIONFROMNEARESTCITY	Direction from Nearest City	S	Geographical
ROUTE	Road Route Number	275	Geographical
SECTON	Road Section Number	1	Geographical
LOGMILE	Road Logmile Location	180	Geographical
ATINTERSECTINGSTREET	At Intersecting Street?	N	Geographical
REFERENCEPOINT	Reference Point (Any Text)	Camp Road	Geographical
DISTFROMNEARESTINTERSECT	Distance from Nearest Intersection	1.8 Mi	Geographical
DIRFROMNEARESTINTERSECT	Direction from Nearest Intersection	N	Geographical
RAILROADIDNUMBER	Railroad Identification Number	434457U	Geographical
ROADSURFACECONDITION	Road Surface Condition	Dry	Infrastructural
ROADSYSTEM	Road System Type	State Highway	Infrastructural
ROADSURFACTTYPE	Roadway Surface Type	Asphalt	Infrastructural
ROADWAYALIGHMENT	Roadway Alignment	Curve	Infrastructural
ROADWAYPROFILE	Roadway Profile	Grade	Infrastructural
CRASHINCONSTZONE	In Construction/Maintenance Zone?	N	Infrastructural

TRAFFICFLOW	Traffic Flow	Not Divided	Infrastructural
NUMBEROFLANES	Number of Lanes	2	Infrastructural
RELATIONTOJUNCTION	Relation to Junction	Driveway	Infrastructural
TYPEOFTRAFFICCONTROL	Type of Traffic Control	Stop Sign	Infrastructural
CONTROLFUNCTIONING	Control Functioning Properly	Device Functioning Properly	Infrastructural
TYPEOFCOLLISION	Type of Collision	Rear End	Outcome
FIREOCCURRENCE	Occurrence of Fire?	N	Outcome
HITRUNCRASH	Hit and Run?	Y	Outcome
CRASHSEVERITY	Crash Severity (1-5)	5	Outcome
NUMBEROFFATALITIES	Number of Fatalities (Severity 1)	0	Outcome
NUMBERIFINJURIES	Number of Injured Persons (Severity 2-4)	0	Outcome
NUMBERINVOLVED	Number of Persons Involved	1	Outcome
NUMBEROFVEHICLES	Number of Vehicles Involved	1	Outcome
INVESTIGATINGAGENCY	Investigating Agency	Arkansas State Police	Outcome
CRASHNUMBER	Crash Number (Year + reference #)	200200001	Reference
FORMCODE	Form Code	07/3/021:47:06PM,Station11	Reference

The variable ALCOHOLINVOLVED is labeled as an *environmental* factor. It may seem intuitive that this variable be considered as a *human* factor, due to the fact that it is the driver's choice whether or not to drive while intoxicated. Although this is true, the database contains a similar, more descriptive variable within the Person level data. The difference is that the variable ALCOHOLDRUGIMPAIRMENT pertains only to the person driving the vehicle. The reason ALCOHOLINVOLVED is considered as a Crash level characteristic is because it does not pertain to a single individual. A drunk driver on the road affects everybody else on that road. As far as a sober driver is concerned, he has no control over the drunk driver in the other lane. Therefore, if alcohol was involved with any of the persons or vehicles within a crash, it is treated as just another *environmental* factor or obstacle.

In order to avoid a convoluted database, the AHTD has developed a method to standardize several of the variables in terms of their values. This makes sorting and searching through the database much easier. This entails that most of the variables have a certain range of values that they can be, limiting the variation of data that could be entered. For example, the values for the variable ROADSURFACECONDITION can only be Wet, Dry, Ice, Sand, Dirt, Oil, Other, and Unknown. This helps eliminate the variation between the terms Ice, Icy, Slick, Frozen, and Slippery, which all mean essentially the same thing. Other variables, like ROUTE or REFERENCEPOINT may have to be entered in as any text, just because there are so many different possibilities for those values.

3.2.2 Vehicle Level

The next level in the AHTD crash database is the Vehicle level. The data within this level refer to the attributes assigned to each vehicle that was involved with a certain accident. It includes 17 of the total 80 crash variables within the database. The details described within this data include many factors about the type and condition of each vehicle, as well as the actions that the vehicle was engaged in prior to the accident. A list of all 17 variables within the Vehicle level is shown in Table 3, along with a brief description and example entry. The variable VEHICLENUMBER is a number that references each vehicle within a crash. It has no purpose other than as a reference.

The variable DRIVERSCONDITION appears to be misplaced in the Vehicle level. This variable describes the conditions of the driver of each vehicle as reported by a state or city official in the official crash report. Conditions such as bad eyesight, bad hearing, or drowsiness are documented in this variable. These are clearly human factors, regardless of whether they are controllable by the human or not. However, for the purpose of staying consistent with the database, this variable was left in the Vehicle level.

Table 3: Vehicle Level Information (Variables)

Variable Name	Description	Example Values
VEHCILENUMBER	Vehicle Number	1
CONTRIBUTINGFACTOR1	Contributing Factor	Careless / Prohibited Driving
VEHICLEACTION	Vehicle Action	Going Straight
HARMFULEVENT	Harmful Event	Motor Vehicle in Transport
COLLISIONWITHFIXEDOBJECT	Collision with Fixed Object	Fence
VISIONOBSCUREMENT	Vision Obscured?	Not Obscured
DRIVERSCONDTIION*	Driver's Condition	Appeared Normal
VEHICLEDIRECTIONOFTRAVEL	Vehicle Direction of Travel	N
VEHMAKE	Vehicle Make	Nissan
VEHICLETYPE	Vehicle Type	Passenger Car
VEHICLEMODEL	Vehicle Model	Altima
VEHICLEBODY	Vehicle Body	4 Door
NUMBEROFTRAILERS	Number of Trailers	0
DAMAGECLOCKPOINT	Vehicle Damage Clockpoint	Front
NUMBEROFOCCUPANTS	Number of Occupants	1
PRIORVEHICLEDEFECTS	Prior Vehicle Defects	No Defects
FIRSTHARMFULEVENTOCCURRED	First Harmful Event Locale	On Roadway

*the condition of the driver is a human factor, but will remain in the Vehicle Level for consistency

3.2.3 Person Level

The last level within the ASHD crash database consists of the Person level. Here, all the data that is recorded can be related to each individual person that was involved with a particular crash. The exceptions to this are the cases in which women who are pregnant are only recorded as one individual, which happens on occasion. The information and details related to each individual can be used to detail their personal attributes, such as age, gender, race, name, and their home state. Other pieces of information that are considered to be in the Person level correspond to the location and action of each person within the car, such as determining who was driving and where the passengers sat. Another important human factor located in this data describes each passenger's restraint type, which is to say whether or not they were wearing a seat belt. Driver's may also be drug or alcohol tested after the accident, in which case the results are also documented as a Person level factor. The entire list of 20 variables within the Person level is shown in Table 4, along with a brief description and example for each variable.

Some variables, such as INJURYSEVERITY, CITATIONNUMBER1, CITATIONNUMBER2, and AIRBAG may be considered as an outcome. It is true that all of these variables are important outcomes of a roadway traffic accident. However, for the sake of this study, and staying consistent with the database, these variables will remain as personal factors. Although they are indeed outcomes, each of these variables describes the condition of every person involved in the crash, and is therefore a personal attribute.

Table 4: Person Level Information (Variables)

Variable Name	Description	Example Values
PERSONNUMBER	Person Number	1
PERSONTYPE	Person Type	1
SEATPOSITION	Seat Position	X
RESTRAINTCODE	Type of Restraint	Lap & Shoulder Belt
AIRBAG	Airbag Details	Non-Deployed Airbag
EJECTIONCODE	Ejection Code	Not Ejected
RACE	Race	B
SEX	Sex	M
AGE	Age	34
INJURYSEVERITY	Injury Severity (1-5)	5
PEDESTRIANLOCATIONACTION	Pedestrian Location Action	0
NAME	Person Name	(blank for privacy issues)
CITATIONNUMBER1	Citation Type	Reckless/Careless Driving
CITATIONNUMBER2	Citation Type	Suspended License
DATEOFBIRTH	Date of Birth (encoded)	25500
LICENSESTATE	License State	AR
LICENSETYPE	License Type	DL
BACTESTED	BAC Tested?	N
BACCRESULT	BAC Results	0
ALCOHOLDRUGIMPAIRMENT	Alcohol/Drug Impairment	N

3.3 Road Inventory Database

Like most states, Arkansas is home to a complex network of road systems. The variety of roads in this network is vast, as it includes many hundreds of miles of interstate, State highways, U.S. highways, county roads, and city streets. Figure 2 shows an aerial map of the state of Arkansas which shows all the major road systems within the state; including interstates, State highways, and U.S. highways.

Due to the complexity of this network, a roadway inventory database was created to keep a record of all the different road segments. More importantly, this database keeps a record of the smaller subsections of each road. Many roads span from one end of the state to the other, changing in size, shape, condition, and jurisdiction. To overcome this, road surveys have been conducted by the AHTD to break down all major interstates, U.S., and State highways into a Route, Section, and Logmile. Every highway and interstate is first broken into several large route segments that are individually numbered; then that route is further broken into several smaller sections. Finally, each section of road is broken into a logmile, which is in reference to the posted mileage that surrounds these road systems. For an even more detailed road segment, the logmile reference is reported in hundredths of a mile. Each set of Route, Section, and Logmile references can be viewed as a unique address for the location of these road segments.

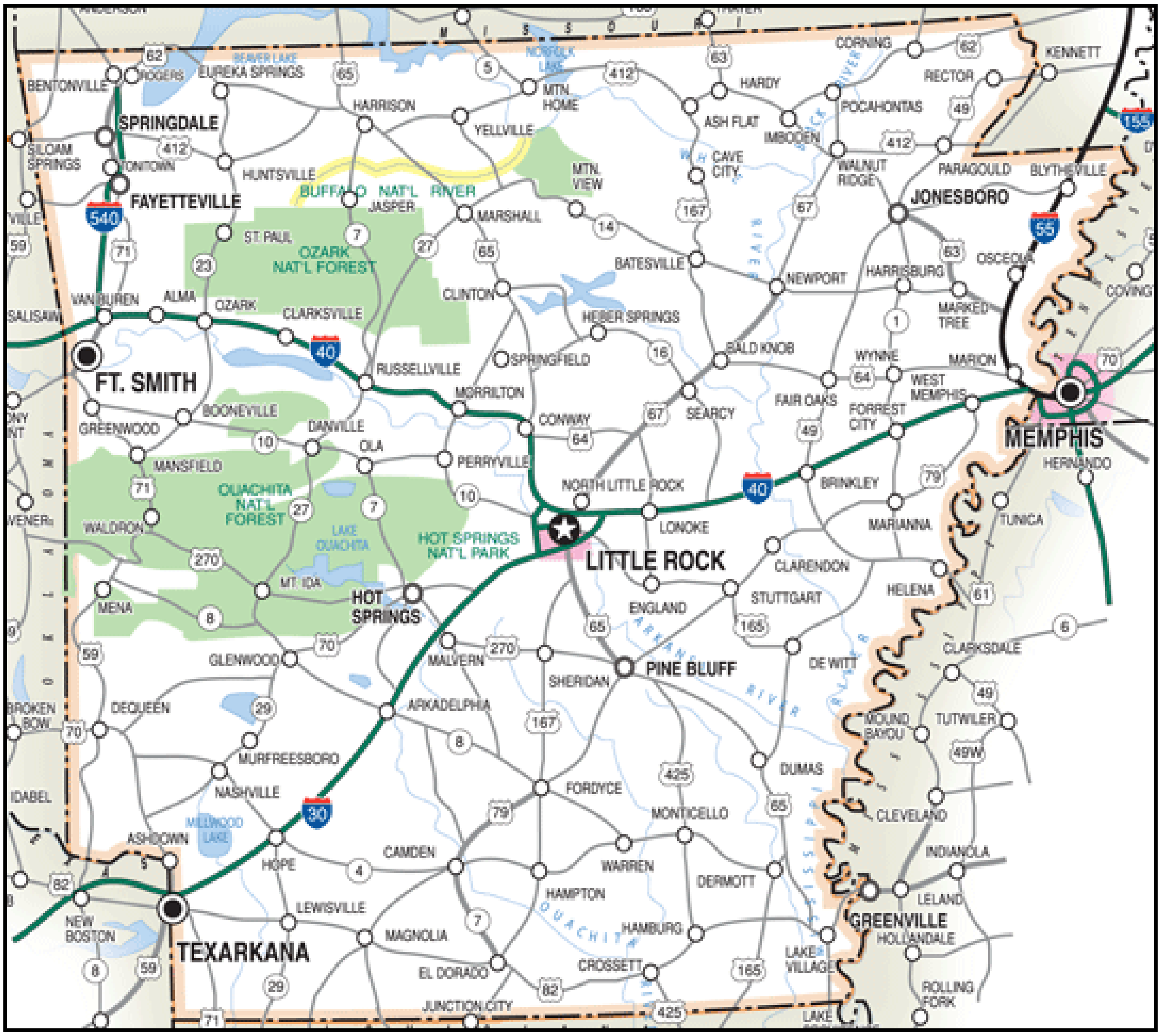


Figure 2: Arkansas State Highways, U.S. Highways, and Interstate Systems

Unfortunately, mile markers are only a commodity used for major road systems like these highways and interstates. For roads within cities and counties, including back roads, defining a unique address can be difficult. County roads and city streets must be identified first by the street name or county road number. Because of the lack of mileage markers, these roads are generally not broken into any smaller sub-sections. They can, however, be identified using a direction as a reference, such as North Main Street and South Main Street.

In total, the AHTD roadway inventory database contains over 115,000 identifiable road segments between all road types for the years 2002 through 2004. These road sections can range anywhere from 0.01 miles in length, to well over 400 miles in length. Not all of these 115,000 road segments are unique, however. One large road segment that spans 100 miles in length can be one entry, whereas that same segment can be broken into 100 smaller one mile sections that serve as 100 separate entries. The point here is that many road sections are duplicated two, three, or more times. The data from 2002 actually contains over 136,000 road segments, merely because several larger road segments were broken down into several other smaller road segments. Both 2003 and 2004 contained a little over 115,000. A breakdown of the number of segments within each type of road system is displayed in Table 5.

Table 5: Arkansas Road Inventory Breakdown of Road Segments

Road Segment Type	**2002***	**2003***	**2004***
Interstate	2,256	1,001	1,000
U.S. Highway	7,130	5,083	5,098
State Highway	16,218	11,261	11,309
County Road	57,725	48,365	48,428
City Street	52,200	49,524	49,472
Frontage/Other	633	553	553
Total	136,162	115,787	115,860
Average Length	0.869	0.857	0.857
Standard Deviation	5.613	5.905	5.902

*Table values in units of road segments

Each entry within the roadway inventory database refers to one specific segment of road within Arkansas. Within each entry, the database contains 50 descriptive pieces of information about the road segment. Much of this information is categorical in nature, such as the type of the road system, its functional class, population group, and surface type. The total list of descriptive variables for road segments within this database is shown in Table 6. Along with the variable name, there is also a short description of what the variable describes and an example for each one.

Table 6: Arkansas Road Inventory Database Variables

Variable	Full Name/Description	Example Values
DSTNO	District Number	12
CONTY	County Number	75
ROUTE	Route	65
SECTN	Section	13B
LMPTR	Beginning Logmile	0.17
ENDLM	Ending Logmile	2.1
RDLEN	Segment Length	1.93 Miles
SEQCN	Sequence Number	Z
RECON	Record Control	Mileage
RTFIX	Route Prefix	Interstate
GOVCO	Government Control	Municipal/City
DOMAN	Domain	State Agencies
POPGR	Population Group	2,499 or less
URBAN	Rural/Urban Area Code	Rural
URBAC	Urbanized Area Code	Fort Smith
PLACE	Place Code	Texarkana
FNCLA	Functional Classification	Interstate-Rural
NHSYS	National Highway System/Funding Eligibility	National Highway System
SYSTA	System Status	Open to public travel
SPSYS	Special System	Airport Road
ADT	Average Daily Traffic Volume	2000
ACCES	Control of Access	Full control of access
FROAD	Frontage Road	Frontage left of main lanes
TYOPR	Type of Operation	One way
NOLAN	Number of Lanes	2
SURTY	Surface Type Code	Bituminous Concrete
BUILT	Year Built	1956

RECONS	Year Reconstructed	1976
MEDWD	Median Width	8 Feet
TYDEV	Type of Development	Urban, Fringe
LNWID	Lane Width	7 Feet
SURWD	Surface Width	12 Feet
RSHOS	Right Surface Shoulder	Bituminous Concrete
LSHOS	Left Surface Shoulder	Bituminous Concrete
RSHOW	Right Width Shoulder	12 Feet
LSHOW	Left Width Shoulder	8 Feet
CURBS	Curbs	No Curbs
ROWWD	Right of Way Width	8 Feet
TERAN	Terrain	Flat
NAMES	Railroad Information	Union Pacific
RDWID	Roadway Width	40 Feet
EXLAN	Extra Lanes	Turn Lanes
TROAD	Type of Road	Main Lane
RAMPD	Ramp Designator	S
YRADT	Year ADT was last measured	2002
ROUGH	Roughness (IRI)	142
PAVCO	Pavement Condition	4
CONNC	Intermodal Connectors	Major Airport
TFILE	Type of File	County Road
APHN	Arkansas Primary Highway Network	National Highway System

CHAPTER IV – STATISTICAL ANALYSIS

4.1 Methodology of Statistical Analysis

For this study, statistical modeling was used to estimate crash occurrence, frequency, and severity. For the estimations of crash occurrence and frequency, Poisson and Negative Binomial regression models were used to fit roadway and crash data. Crash severity was analyzed through a Binary Logistic regression model. Each model takes into account a number of potential factors from both the Arkansas roadway inventory database and crash database. In the following sections these models are looked at in detail, including model assumptions, parameters, and estimation processes. The section concludes with an analysis of model limitations.

4.1.1 Crash Occurrence and Frequency

The first task for this study was to define and develop a mathematical model that manages to predict crash occurrence. In doing this, certain potential crash and road factors are built into the model and verified for significance after the model has been tested for goodness of fit. The methodology behind this test revolves around certain assumptions in which the crash and road data is based. In the case of crash occurrence, the assumption is that the response is either a binary or a count variable. As was mentioned earlier, the type of response is a key element in defining a model. Typically, a crash occurrence model that is binary revolves around the fact that there was a specific driver and vehicle situation (with corresponding environmental, temporal, geographical, infrastructural, vehicular, and human factors), which resulted either with or without a traffic accident; basically, and event occurred or it did not. However, this requires data on every single vehicle on any road at any one time. A more realistic model, and the model used for this study, revolves around crash frequency on road segments. This approach requires a response that is of count type, meaning the response is a positive, discrete random variable that corresponds to the number of crashes on any segment of road during a specified time period.

To this end, the main assumption made for this portion of the study was that crash frequency follows a Poisson distribution. Crash frequency is a number count, which means that it needs to be modeled as a discrete random variable. The Poisson distribution is one type of discrete probability distribution, and is used to describe the probability of a specific number of events occurring within a particular frame of time. It is also assumed that each event in the Poisson distribution is independent of any other event. The model's assumptions match the nature of crash occurrence quite well, and therefore Poisson is regarded as a popular method of analysis. Also, computation and solution inference is made quite simple by this model.

Because the response variable under study is assumed to follow the Poisson distribution, it is only natural to first apply the data in a Poisson regression model. Crash frequency acts as the discrete response variable, whereas any number of variables can make up the independent and explanatory variables. The independent variables are not limited to being continuous in nature, such as length of the road segment, or age of the driver; the variables could also be binary or categorical in nature. Depending on the type of variables set into the Poisson regression model, estimations may vary. The following paragraphs step through the general methodology behind the Poisson regression model.

It is important to note that the Poisson regression model is an extension of the family of models called Generalized Linear Models. This is because the model is trying to adapt a linear relationship between the factors and the response, as was simple linear regression. Simple linear regression takes the form

$$Y_i = \beta_o + \beta_1 x_i + \varepsilon_i \tag{1}$$

where x_i represents the ith explanatory variable, given the response of Y_i, and β_0 and β_1 are the estimated parameters of the intercept and slope for the best fit line. Ideally, if there was a true relationship between explanatory variables and the response, the model would be exact and there would be no need for an error term. Realistically, however, there is always variation from the "true" relationship, and therefore must be compensated through an error term. The final term in the model represents this error, which is the mathematical difference between the expected value of the response and the actual value of the response. The best fit model is the one that minimizes the sum total of the squared errors, which is done by manipulating the two parameter values. The value of β_1 is of particular importance, because it describes the effect that the explanatory variable has on the response. A positive β_1 would mean that an increase in x leads to an increase in the response, whereas a negative β_1 would mean that a increase in x would lead to an decrease in the response. If the overall goodness-of-fit for the model is decent, then it can be inferred that the explanatory variable is in fact a good predictor of the response Y.

Multiple linear regression takes this model one step further by adding in more explanatory variables to be considered; simple linear regression can only consider one explanatory variable. This model for this case is

$$Y_i = \beta_o + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \tag{2}$$

which includes several independent variables (x_1, x_2, …, x_k) and corresponding slope parameters (β_1, β_2, … β_k). Again, the model is chosen on a least squared errors method to minimize ε_i; the error term associated with the ith element within the model. As with all regression models, the error terms must be normally distributed, or the assumptions of the model are not valid. Error terms can be tested for normality through the use of a normal probability plot of the errors, in which it should form a relatively straight line.

Poisson regression extends these basic ideas by adjusting for the fact that the response Y_i is a discrete random variable following Poisson assumptions. The probability density function associated with the Poisson random variable is as follows:

$$\Pr(Y_i | x_i) = \frac{\exp(-\lambda_i)\lambda_i^{Y_i}}{Y_i!} \tag{3}$$

This equation describes the probability associated with Y events occurring given the expected occurrence rate of λ. The equation is conditional on the x_i factors that are being tested within the model. The occurrence rate λ is actually a function of the x_i variables and can be expressed through the following function:

$$\lambda_i = \beta_o e^{\sum \beta_j x_{ij}} \tag{4}$$

This equation is the basic model for the Poisson regression, which at first sight looks nothing like the simple or multiple regression models. However, what follows is a logarithmic transformation of the response variable λ. This is a useful transformation that takes the discrete random variable λ and makes it into a continuous random variable in the form of ln(λ). The new model becomes

$$\ln(\lambda_i) = \beta_o + \sum_{j=1}^{k} \beta_j x_{ij} \tag{5}$$

This new model now looks exactly like the multiple linear regression model, except for the response, which is now a logarithmic transformation.

Because Y_i is assumed to be Poisson distributed, it is important to note a few aspects of the distribution. One key feature of the Poisson variable is the fact that the mean and the variance of the distribution are said to be equal. In other terms:

$$Var\{Y_i | x_i\} = E\{Y_i | x_i\} = \lambda_i \tag{6}$$

Both the variance and mean are equal to the expected number of occurrences, λ_i. The equal mean and variance is often times a roadblock, as it may not always be true for the model, but this will be relaxed later.

The next step for the model is to find the estimates for the parameters to find the best fit equation to the data. Simple and multiple linear regression models estimate their parameters $\beta_o + \beta_i$ through the use of the least squared errors technique. However, to get the best fit parameters when the distribution is assumed to be Poisson, the estimation technique needs to be based off of a different method. One popular method for this is referred to as the Maximum Likelihood function. In essence, this function is a function of the probability mass function shown in equation (3). To derive the best estimate for λ_I, the Likelihood function for Poisson data is:

$$L(\lambda) = \prod_i \frac{\exp(-\lambda_i)\lambda_i^{Y_i}}{Y_i!} \tag{7}$$

or the joint densities associated with the *i* values of *Y*. To maximize this function, the derivative of is taken with respect to λ and set equal to zero. The result is the Maximum Likelihood estimator for λ:

$$\hat{\lambda}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{8}$$

Once this estimation of λ is made, the model then calculates the parameters of $\beta_o + \beta_i$ that produces the best fit results. The slope terms have similar interpretations as they did for multiple linear regression; positive values lead to positive correlations between a specific factor and the response, and negative values lead to negative correlations.

Sometimes, however, the historical data may not fit all of the assumptions of the Poisson distribution; namely the fact that the variance is equal to the mean. In many real world processes, especially crash data, the variance is larger than the mean. This causes the problem of overdispersion. This is a large issue, because if the response variable is overdispersed, the estimations may not be statistically valid. Luckily, models have been developed to handle this issue. This is accomplished by allowing the variance to be greater than the mean, which is represented in mathematical terms as:

$$E\{Y_i\} = \lambda_i \tag{8}$$

$$Var\{Y_i\} = \lambda_i[1 + (1/\phi)\lambda_i] \tag{9}$$

Now, the variance is altered by a factor ϕ which represents the overdispersion factor. The smaller this value becomes, the larger the overdispersion.

Next, the extra variation is accounted for in the formulation of the occurrence rate (donated here as μ_i rather than λ_i for distinction between models). The new model becomes:

$$\ln(\mu_i) = \beta_o + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \tag{10}$$

The probability density function is then:

$$\Pr(Y_i | x_i, \mu_i) = \frac{\exp(-\lambda_i \mu_i)(\lambda_i \mu_i)^{Y_i}}{Y_i!} \tag{11}$$

However, this new function is not conditional on the explanatory factors alone. In order to have this density function unconditional of the additional error term, it must be integrated out. The error term here is assumed to be gamma distributed. Once the expression is integrated over μ_i, the density function becomes

$$\Pr(Y_i | x_i) = \frac{\Gamma(\theta + Y_i)}{\Gamma(\theta) Y_i!} r^{\theta} (1 - r)^{Y_i} \tag{12}$$

where

$$r = \frac{\theta}{\theta + \lambda_i} \tag{13}$$

and

$$\theta = \frac{1}{\phi} \tag{14}$$

These resulting equations form the Negative Binomial regression model. Again, the Negative Binomial is a model that is an extension on the Poisson model, which allows for the data to be overdispersed. Whenever the expression for θ becomes equal to zero, the Negative Binomial model reduces back to the Poisson model.

Parameter estimates for the Negative Binomial regression model can be made in the same fashion as the Poisson's parameters. Maximum Likelihood is again the most common method. For the Negative Binomial model, the Likelihood function is the joint product of densities, or

$$L(\lambda) = \prod \frac{\Gamma(\theta + Y_i)}{\Gamma(\theta) Y_i!} r^{\theta} (1-r)^{Y_i} \tag{15}$$

Again, once this function is maximized for λ, the parameters for $\beta_o + \beta_i$ can be found that produce the best fit model.

The goodness-of-fit for either the Poisson or Negative Binomial distribution can be determined by the value of the final likelihood value. The greater this value becomes the better fit the model. The value may change depending on what explanatory values goes into the model. Insignificant predictor variables will increase errors and lower the likelihood. Variables can be evaluated through the use of p-values calculating their significance in predicting the response. Variables with a p-value of 0.01 or less can be considered significant contributors to the response, based on a 99% or greater certainty.

Other common goodness-of-fit measures include the Akaike information criterion (*AIC*) and the Bayesian information criterion (*BIC*). Both methods take into account that adding parameters to a model increases its complexity. Both are functions of the logarithmic value of the maximum likelihood value and the number of parameters within the model. The *AIC* takes the form

$$AIC = 2k - 2\ln(ML) \tag{16}$$

where k is the number of unknown parameters and ML is the maximum likelihood value. *BIC* takes the form

$$BIC = k\ln(n) - 2\ln(ML) \tag{17}$$

where n is the number of observations used in the model. Essentially, the smaller these values are the better fit the entire model becomes.

4.1.2 Crash Severity

For responses that represent counts, such as crash frequency, Poisson and Negative Binomial models are well suited. However, sometimes the response can be discrete and binomial. A binomial response can only be one of two possible choices; generally this is a value of 1 or 0. Crash occurrence can be modeled by a binary response, such as whether a crash occurred or did not occur. Crash or injury severity can also be modeled as binary; the response could be 1 if an injury or crash was severe or 0 otherwise.

The basic regression model for dichotomous responses (meaning only two values) is the logistic regression model. The logistic regression model is a generalized linear model, an extension of the general linear models. It is able to handle discrete output data, similar to Poisson and Negative Binomial models. Extensions of the logistic regression model allow the responses to take on non-dichotomous responses that are categorically based. Ordinal logistic regression allows the response to take on discrete values that have a common rank or order, such as survey results with answers ranked on satisfaction (dissatisfied, neutral, satisfied, etc.). Nominal logistic regression takes into account categorical data that does not have an obvious ranking, such as the county number of a location. The basic logistic regression, on the other hand, deals with binary responses. Unlike simple and multiple linear regression where the expected value for the response takes the form of equations (1) and (2), the expected value of the dichotomous response Y is given by the formula

$$\pi(x_i) = E\{Y_i | x_i\} = \frac{\exp(\beta_o + \beta_i' x_i)}{1 + \exp(\beta_o + \beta_i' x_i)} \tag{18}$$

where $\pi(x_i)$ represents the probability of a 1 occurring (or the proportion of 1's). The formula is conditional on the vector of explanatory factors x_i. The parameters are once again represented by $\beta_o + \beta_i'$, where the betas are treated as a vector corresponding to the vector of x_i factors.

The logistic model $\pi(x_i)$ can then be altered using a logarithmic transformation, usually denoted as a logit transformation. To do this, equation (18) is altered as follows:

$$\pi(x_i) = \frac{1}{1 + \exp(-\beta_o - \beta_i' x_i)} \tag{19}$$

Then solving for $\exp(\beta_o + \beta_i' x_i)$,

$$\frac{\pi(x_i)}{1 - \pi(x_i)} = \exp(\beta_o + \beta_i' x_i) \tag{20}$$

And finally,

$$\ln\left[\frac{\pi(x_i)}{1 - \pi(x_i)}\right] = \beta_o + \beta_i' x_i \tag{21}$$

The left hand side of equation (21) represents the logit function. Now, the right hand side is similar to that of multiple regression.

A useful aspect of the binary logistic regression model is the development of the odds ratio. The odds ratio is essentially the left hand side of equation (20). This ratio is the ratio of the odds of an event occurring in one group to the odds of the event occurring in another group. Using the example of crash severity where 1 represents a severe crash and 0 represents a non-severe crash, an odds ratio for an explanatory variable such as sex of the driver (Male =1, Female =0) may be 1.2. Since this ratio is greater than 1, it is interpreted that crashes involving male drivers are more likely to have been severe than for female drivers. To find out exactly how much more likely, the natural log of the odds ratio is taken and 1 is added to the number. Therefore in this example, crashes involving male drivers are 1.18 times as likely to be severe than for female drivers.

However, before accurate odds ratios can be made, the parameters need to be found. This is done using the maximum likelihood. The likelihood function for logistic regression models are of the form,

$$L(\beta) = \prod \left[\pi(x_i)^{Y_i} [1 - \pi(x_i)]^{1-Y_i} \right] \tag{22}$$

By maximizing this formula, the corresponding β coefficients can be determined and analyzed as in the previous models.

The goodness-of-fit can be interpreted in a number of ways for the logistic regression model. One way is to view the value of -2 times the natural log of the likelihood value. This is interpreted in many ways like the AIC and BIC, in that the lower this value, the better fit the model is. Other methods include the p-values of the explanatory variables. Variables with too high of a p-value can lower the goodness-of-fit of the model and may be considered insignificant.

4.1.3 Issues with Statistical Evaluation

Any statistical model that is chosen to describe a set of data must be based on the underlying assumptions associated with that model. If the data does not follow those assumptions, the resulting model fits may not be reliable. One of the major assumptions to the regression models mentioned for this study is that of independence within the data. Data entries that are dependent on other entries may have a confounding effect on the fitted model. Significant explanatory factors may be found to be insignificant, or vice versa. Independence can be checked through the use of the error terms. As was mentioned earlier, if the error terms follow a normal distribution, then there is a good chance that the data is relatively independent.

The reason that this becomes an issue with crash data is because of the natural breakdown of a road traffic accident. Earlier, a crash was described as a single event that encompasses one or many vehicles, which includes one or many passengers. This violates the assumption of independence. Consider two separate crashes that occur at two different locations and times. It is likely to think that a passenger involved in the first crash would have injuries independent of a passenger in a completely different crash at a completely different location. Now, consider one crash that happens between two vehicles, both of which containing two passengers. It can be assumed that the two passengers within the same vehicle will have very similar injuries; a direct violation of independence.

To overcome this, researchers can do one of two things: use only the data that is independent of all other data, or account for the inter-dependence within the model. The first way is the easiest method, as it does not involve more intricate and complicated statistical software, and is easier to interpret the results. This is what is done in the current study. To avoid the hierarchical relationship between crashes, vehicles, and passengers, the analysis on crash frequency was performed only based on roadside features and the total number of crashes involved on each particular road segment. The crash frequency is a count of total crash incidences, which does include anything about the number of individual vehicles or individual passengers that would lead to the inter-dependence within the data. The second analysis, which focuses on crash severity, deals heavily with human factors. To avoid dependence within the data, only single vehicle crashes were studied. This ruled out the dependence between vehicles. Also, only the driver's personal characteristics and human factors were considered in order to remove the dependence involved with any passengers.

The second method a researcher could use is to account for the inter-dependence within the model. This is achieved by redeveloping the Poisson, Negative Binomial, and Logistic Regressions to account for entries that have a hierarchical relationship. For each level of the hierarchy, such as crash or person, the model calculates a unique set of parameters. Although this method would produce very valid results, its models become complex very fast with increasing levels of hierarchy. Because of this, the analysis of these models was left outside the scope of this analysis.

Even if inter-dependence within a set of data is not a concern, there are other issues that may cause the need for model reconsideration; especially for the Poisson regression model. As was mentioned earlier, the Poisson regression model assumes that the response's mean is equal to its variance. If the variance is actually greater than the mean, then the Negative Binomial model may be a better fit. However, this may not be the only issue with the Poisson data. Occasionally, data that is said to be Poisson distributed (as crash data often is) can find that the response has several zeros. Because crash frequency often gets modeled, it is not hard to find that many road segments have no crashes throughout a given time period. In fact, it is common that there are more road segments without any crashes than there are road segments with one or more crashes. The model may run into estimation problems, or end up with inaccurate estimates if there become too many excess zeros within the data. This too can be handled through the use of a new model. Again, the current models can be modified to become a Zero-Inflated Poisson model. However, this is a highly complex model that will not be covered within this paper.

There is rarely such a thing as a perfectly fit model, but many of these more complex models get closer to best fit model than other simpler models. Yet, the Poisson, Negative Binomial, and Logistic models should not be completely omitted for consideration. More than not, these models are sufficient in crash prediction models, and are able to show significant goodness-of-fit. Only when these models fail to predict efficiently should more complex models be used to evaluate data.

4.2 Statistical Model Results and Discussion

Using the methodology behind the Poisson, Negative Binomial, and Logistic regression techniques, three models were developed and evaluated. The software used to run and evaluate each model was the statistical package SPSS. In this section the input parameters as well as the output from each model are given. The results from these models are then discussed and potential implications are drawn.

4.2.1 Poisson and Negative Binomial Regression

Using both the crash database and the roadway inventory database, road segment crash frequency was modeled through the use of Poisson and Negative Binomial Regression in SPSS. The entries for this model consisted of road segments within Arkansas that were part of a US highway, State highway, or interstate road. Unfortunately, the database is set up in a way that only these roads are specified with a unique location via route, section, and logmile entries. County roads and city roads had to be excluded because of the inconsistency within the database. County and city roads contain the large majority of crashes, and so this exclusion was one drawback of this analysis. The data was further reduced to contain only road segments of a length of one mile or less, and average daily traffic (ADT) volumes of 2000 or greater. This was done in order to eliminate some road segments that had large segment lengths and unrealistically small ADT values, which affected the overall crash rate. According to the Arkansas Highway Department, roads with an ADT value of 2000 or greater are considered to be medium to high volume roads. Therefore, low volume roads were also left out of the study; mainly due to the lack of data on these road segments.

Each road segment contained a range of logmile values. The crash occurrences within the crash database were then separated out into the logmile ranges from which they are addressed. This was the basis of calculating the total crash frequency. Also, because this analysis focuses on intersection crashes, only the crashes that had a ‘Yes’ value for the variable ATINTERSECTINGSTREET were considered. ‘Yes’ refers to the fact that the crash did occur at two or more intersecting roads. There is another variable that could have potentially been used, JUNCTIONTYPE, which defines the junction of roads if one exists (Intersection, Intersection related, No Junction, etc). When a crash is reported, both of these entries are supposed to be recorded, but rarely are actually recorded together. The AHTD has verified that ATINTERSECTINGSTREET is the most commonly recorded and most accurate of the two entries.

Also, the analysis was run on each of the three years of data. The size of the roadway inventory data changed considerably from year to year, therefore causing the need to run a new model for each year. Ideally, there would be an additional variable denoting the year of occurrence, but the road segments vary too much each year to make this feasible. The first year of roadway data had significantly more road segments than did the other years, possibly due to segment duplication. To avoid any problems with this issue, the years were evaluated separately.

The explanatory variables used initially in the models are given in Table 7. There were 14 variables in all, 7 of which were considered continuous. The other 7 variables were considered as discrete and categorical, and therefore either ordinal or nominal. Ordinal describes a discrete categorical variable that has a natural ranking. Nominal variables include categories that have no natural ranking. Descriptions of these variables are also shown in Table 7.

Table 7: Variables Included Initially in Poisson and Negative Binomial Models

Variable Name	Description of Variable	Value	Type
Crash Frequency	Total number of crashes on road segment i	0, 1, 2, …	Discrete (Scale)
District	The highway district in which the crash is located	1-12	Discrete (Nominal)
County	The county in which the crash is located	1-75	Discrete (Nominal)
Length	The length of the segment of road	0-1	Continuous
Population	The Population Group of the crash location's surroundings	0-9	Discrete (Ordinal)
Urban	The Urban/Rural Code of the crash location's surroundings	1-5	Discrete (Nominal)
FNCLA	Code for the Functional Class of the road segment	0-19	Discrete (Nominal)
ADT	Average Daily Traffic volume	> 1999	Continuous
Lanes	Number of lanes	2, 4	Discrete (Nominal)
Surface	Coded description of the type of surface material	0-90	Discrete (Nominal)
Lane Width	Width of the most narrow lane on the road	0-99	Continuous
Reconstruction	Year of the last reconstruction on segment	Year	Continuous
Terrain	Coded value for the physical surroundings of the road	0-4	Discrete (Nominal)
Road Width	Width of road, excluding medians	0-99	Continuous
Pav Condition	Score denoting the condition of the pavement	0-5	Continuous

The data was analyzed first using the Poisson model. The statistical package used was SPSS, which contains the Poisson model through the use of the GENLIN function. The Poisson is an extension of the generalized linear equations under this function. All 14 dependent variables were placed in the original model. A more detailed layout of the input parameters and SPSS coding is shown in APPENDIX A. The outputs of the Poisson model for each year are displayed in Tables 8-10.

The first table shows the results of the test of model effects for each of the three models (Table 8). Among the three years of data, nearly all of the explanatory variables were shown to be significant based off of their p-values (in bold). Ideally, each of the three years would have matched with their significant variables. In this case, only one or two variables were shown to be insignificant. For 2002 and 2004, the county location of crashes was shown to be insignificant based on a 99% confidence interval, yet it was significant during 2003. Similarly, the year of the last reconstruction was insignificant for 2003 and 2004, but was significant for 2002.

The model results, or the coefficients to the best fit line, are shown in Table 9. Assuming the model is fitted well, these values can be interpreted in terms of trends. For continuous variables like ADT, Road Width, Length, Pavement Condition, etc., the interpretation is straight forward. A positive coefficient means an increase in the variable causes an increase in the response. Length has a positive coefficient, which means that it tends to have a positive correlation with the response variable, crash frequency; the longer a road segment's length, the more potential crashes it can have on that segment. Traffic volume (ADT) also has a significant positive coefficient, which makes sense intuitively; the more vehicles on a single road at one time, the more potentially dangerous the road becomes. This can be compared with the recent claims that road traffic accidents are more frequent on roads with increasing traffic volumes (Pickering, 2004).

Table 8: Poisson Regression Test of Model Effects

Source	2002			2003			2004		
	Wald Chi-Square	**df**	**p-value**	**Wald Chi-Square**	**df**	**p-value**	**Wald Chi-Square**	**df**	**p-value**
(Intercept)	64.161	1	**0.000**	3.925	1	0.048	2.400	1	0.121
District	44.840	1	**0.000**	22.990	1	**0.000**	68.954	1	**0.000**
County	0.300	1	0.584	29.158	1	**0.000**	0.524	1	0.469
Length	2,971.487	1	**0.000**	2,840.746	1	**0.000**	2,681.359	1	**0.000**
Population	103.352	1	**0.000**	106.298	1	**0.000**	271.250	1	**0.000**
Urban	20.233	1	**0.000**	21.583	1	**0.000**	50.742	1	**0.000**
FNCLA	363.357	1	**0.000**	268.742	1	**0.000**	218.274	1	**0.000**
ADT	1,930.336	1	**0.000**	1,959.993	1	**0.000**	1,623.364	1	**0.000**
Lanes	1,569.870	1	**0.000**	1,396.879	1	**0.000**	1,507.755	1	**0.000**
Surface	48.551	1	**0.000**	45.816	1	**0.000**	51.583	1	**0.000**
Lane Width	578.831	1	**0.000**	332.851	1	**0.000**	277.697	1	**0.000**
Reconstruction	26.471	1	**0.000**	0.293	1	0.588	1.122	1	0.289
Terrain	111.393	1	**0.000**	140.584	1	**0.000**	164.228	1	**0.000**
Road Width	575.849	1	**0.000**	441.437	1	**0.000**	494.717	1	**0.000**
Pav Condition	44.130	1	**0.000**	272.401	1	**0.000**	177.126	1	**0.000**

Road width actually has a negative coefficient, which can be interpreted in the opposite fashion; an increased road width creates fewer crash occurrences. Wider roads lead to less potential contact between vehicles traveling parallel to each other within the lanes. Therefore, these values for road width seem very realistic. For variables that are nominal in nature, such as District, County, Urban, FNCLA, Surface, and Terrain, interpretation of these coefficients may be difficult. However, these coefficient values are often times miniscule enough that it does not affect the response in a major way, even though the variables themselves can be significant predictors. County was shown to be significant in 2003 with a coefficient of -0.002. Because this coefficient is so small, it cannot be truly interpreted that County 75 (Yell) was more dangerous than County 2 (Ashley). In fact, the differences in the response between counties will only be a fraction of a car accident with this small of a coefficient. Regardless of the interpretation, the model still shows many of these nominal variables to be significant.

Table 9: Poisson Regression Parameter Estimates

	2002				2003				2004			
		Std.	95% Wald Confidence Interval			Std.	95% Wald Confidence Interval			Std.	95% Wald Confidence Interval	
Parameter	**B**	**Error**	**Lower**	**Upper**	**B**	**Error**	**Lower**	**Upper**	**B**	**Error**	**Lower**	**Upper**
(Intercept)	-2.384	0.298	-2.968	-1.801	-0.568	0.287	-1.130	-0.006	-0.450	0.290	-1.019	0.119
District	0.027	0.004	0.019	0.035	0.020	0.004	0.012	0.028	0.033	0.004	0.025	0.041
County	0.000	0.000	-0.001	0.001	-0.002	0.001	-0.003	-0.002	0.000	0.000	-0.001	0.001
Length	1.837	0.034	1.771	1.903	1.862	0.035	1.793	1.930	1.776	0.034	1.709	1.844
Population	0.092	0.009	0.074	0.110	0.096	0.009	0.077	0.114	0.145	0.009	0.128	0.162
Urban	-0.134	0.030	-0.192	-0.075	-0.141	0.030	-0.201	-0.082	-0.210	0.030	-0.268	-0.152
FNCLA	0.125	0.007	0.112	0.137	0.107	0.007	0.094	0.120	0.094	0.006	0.082	0.107
ADT	5E-05	1E-06	4E-05	5E-05	4E-05	1E-06	4E-05	5E-05	4E-05	1E-06	4E-05	4E-05
Lanes	0.681	0.017	0.647	0.715	0.641	0.017	0.608	0.675	0.646	0.017	0.613	0.678
Surface	-0.031	0.004	-0.039	-0.022	-0.029	0.004	-0.037	-0.020	-0.031	0.004	-0.039	-0.022
Lane Width	0.106	0.004	0.098	0.115	0.085	0.005	0.076	0.094	0.082	0.005	0.073	0.092
Reconstruction	0.002	0.000	0.001	0.002	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001
Terrain	-0.374	0.035	-0.443	-0.304	-0.423	0.036	-0.493	-0.353	-0.447	0.035	-0.515	-0.379
Road Width	-0.027	0.001	-0.030	-0.025	-0.024	0.001	-0.026	-0.022	-0.024	0.001	-0.026	-0.022
Pav Condition	0.011	0.002	0.008	0.015	-0.028	0.002	-0.032	-0.025	-0.023	0.002	-0.026	-0.019

Table 10 shows the goodness-of-fit for each year's model. All three seem to be relatively close for all the values and criteria. It may be difficult to interpret these results currently, because there has not yet been a model to compare the Poisson model to. Each criterion shown above is in a 'smaller is better' form. These values may be small or large. Until another model is run, this cannot be fully interpreted.

Table 10: Poisson Regression Model Goodness-of-Fit

	2002			2003			2004		
	Value	df	Value/ df	Value	df	Value/ df	Value	df	Value/ df
Deviance	1.90E+04	5,859	3.247	1.88E+04	6,052	3.107	2.02E+04	6,055	3.342
Scaled Deviance	1.90E+04	5,859		1.88E+04	6,052		2.02E+04	6,055	
Pearson Chi-Square	2.77E+04	5,859	4.72	2.72E+04	6,052	4.489	2.97E+04	6,055	4.902
Scaled Pearson Chi-Square	2.77E+04	5,859		2.72E+04	6,052		2.97E+04	6,055	
Log Likelihood[a]	-1.28E+04			-1.28E+04			-1.36E+04		
Akaike's Information Criterion (AIC)	2.57E+04			2.56E+04			2.71E+04		
Finite Sample Corrected AIC (AICC)	2.57E+04			2.56E+04			2.71E+04		
Bayesian Information Criterion (BIC)	2.58E+04			2.57E+04			2.72E+04		
Consistent AIC (CAIC)	2.58E+04			2.57E+04			2.73E+04		

*Table values are based on a smaller is better form

Before accepting the Poisson model, certain aspects of the data need to be verified. As was mentioned before, one assumption of the Poisson model is that the mean of the response is equal to its variance. However, when attempting to verify this with the crash data, it was found that the model actually encountered overdispersion. The response variance was in fact greater than its mean. This violation of the Poisson model may have a negative effect on the parameter estimates and the model's goodness-of-fit. In fact, it may also be said that this violation causes the estimations to be unreliable. To overcome this issue, the data was tested once again using SPSS, this time using a Negative Binomial model. The Negative Binomial regression model is another extension of the GENLIN function, but it accounts for overdispersion in the model. The setup was the same, with only minor changes in the input codes. Coding for the Negative Binomial tests are shown in Appendix A. The outputs of the new models are shown in Tables 11-13.

Table 11: Negative Binomial Regression Test of Model Effects

Source	2002			2003			2004		
	Wald Chi-Square	**df**	**p-value**	**Wald Chi-Square**	**df**	**p-value**	**Wald Chi-Square**	**df**	**p-value**
(Intercept)	26.617	1	**0.000**	4.751	1	0.029	4.715	1	0.030
District	27.800	1	**0.000**	13.702	1	**0.000**	41.875	1	**0.000**
County	1.353	1	0.245	6.833	1	**0.009**	0.069	1	0.793
Length	385.092	1	**0.000**	404.575	1	**0.000**	363.226	1	**0.000**
Population	44.935	1	**0.000**	49.892	1	**0.000**	109.996	1	**0.000**
Urban	0.437	1	0.509	1.776	1	0.183	5.868	1	0.015
FNCLA	55.371	1	**0.000**	44.183	1	**0.000**	27.558	1	**0.000**
ADT	417.261	1	**0.000**	416.820	1	**0.000**	339.181	1	**0.000**
Lanes	219.106	1	**0.000**	187.264	1	**0.000**	233.699	1	**0.000**
Surface	1.604	1	0.205	2.610	1	0.106	1.867	1	0.172
Lane Width	113.366	1	**0.000**	66.187	1	**0.000**	58.840	1	**0.000**
Reconstruction	26.236	1	**0.000**	8.372	1	**0.004**	3.997	1	0.046
Terrain	22.835	1	**0.000**	41.526	1	**0.000**	50.527	1	**0.000**
Road Width	136.358	1	**0.000**	103.173	1	**0.000**	127.611	1	**0.000**
Pav Condition	5.041	1	0.025	60.380	1	**0.000**	52.234	1	**0.000**

The end result of running the Negative Binomial regression shows that there are some slight differences with the model effects as compared to the Poisson model. Based on a 99% confidence interval, 2002 and 2004 both show 4 insignificant variables, whereas 2003 shows only 2 (significant p-values in bold text). The variables Urban and Surface are not significant during any of the three years. This suggests that the type of area in terms of the level of urban or rural surroundings in which the crash occurred is not a predictive measure for determining crash frequency. However, the population group, which gages the surrounding area in terms of increasing population, is significant. This may be interpreted in a manner that suggests that the population variable already has enough predicting power for the model and that the urban/rural variable is not even necessary. Although surface type is shown to be insignificant, it is important to note that this study only considered US highways, State highways, and interstates, which contain little variation in terms of the materials used for each road. Other variables that are shown to be insignificant using a 99% confidence interval are County (2002 and 2004), Pavement Condition (2002), and Reconstruction (2004). All other variables are shown to be significant with p-values of less than 0.01. It was expected that County would be highly significant do to the varying nature of the Arkansas landscape, but this was not the case. Again, this was probably due to the limitation of the study to use only highways and interstate roads, where there was much less variation in terms of roadways. County and city roads, which were left out of this study due to data limitations, contain much more varying attributes of surface type, shape, and size. For the sake of the data used, the interpretation should be that the county location is not significant for crashes specifically on these highways and interstates. The Highway District is significant, however, which suggests that the specific highway systems are significantly different in terms of crash frequency, but not from county to county.

Table 12 now shows the parameter estimates of these explanatory variables. Again, by looking at the coefficients of the variables shown to be significant, trends can be interpreted. Traffic volume, road segment length, lane width, and number of lanes all show a positive correlation with crash frequencies, whereas road width still has a negative correlation. Lane width’s results are interesting because of its positive correlation. It would seem to make sense that wider lanes would produce fewer crashes. However, the argument can be made that lane width is highly correlated with road width, and that road width’s negative coefficient may actually partially compensate for lane width’s positive coefficient.

Table 12: Negative Binomial Parameter Estimates

Parameter	2002				2003				2004			
			95% Wald Confidence Interval				95% Wald Confidence Interval				95% Wald Confidence Interval	
	B	Std. Error	Lower	Upper	B	Std. Error	Lower	Upper	B	Std. Error	Lower	Upper
(Intercept)	-3.315	0.643	-4.574	-2.056	-1.375	0.631	-2.611	-0.139	-1.386	0.638	-2.637	-0.135
District	0.04	0.008	0.025	0.055	0.028	0.008	0.013	0.042	0.048	0.007	0.033	0.062
County	-0.001	9E-04	-0.003	0.001	-0.002	9E-04	-0.004	0.000	0.000	8E-04	-0.001	0.002
Length	1.532	0.078	1.379	1.685	1.592	0.079	1.437	1.747	1.492	0.078	1.338	1.645
Population	0.134	0.02	0.095	0.173	0.141	0.02	0.102	0.18	0.196	0.019	0.159	0.233
Urban	-0.037	0.056	-0.147	0.073	-0.074	0.055	-0.182	0.035	-0.129	0.054	-0.234	-0.025
FNCLA	0.084	0.011	0.062	0.106	0.073	0.011	0.052	0.095	0.056	0.011	0.035	0.077
ADT	7.E-05	3.E-06	6.E-05	8.E-05	7.E-05	3.E-06	6.E-05	7.E-05	6.E-05	3.E-06	5.E-05	6.E-05
Lanes	0.519	0.035	0.45	0.587	0.469	0.034	0.402	0.537	0.506	0.033	0.442	0.571
Surface	-0.012	0.01	-0.031	0.007	-0.015	0.01	-0.034	0.003	-0.013	0.01	-0.032	0.006
Lane Width	0.093	0.009	0.076	0.11	0.073	0.009	0.055	0.09	0.069	0.009	0.052	0.087
Reconstruction	0.003	7E-04	0.002	0.005	0.002	6E-04	0.001	0.003	0.001	6E-04	2E-05	0.002
Terrain	-0.245	0.051	-0.345	-0.144	-0.32	0.05	-0.418	-0.223	-0.342	0.048	-0.436	-0.248
Road Width	-0.025	0.002	-0.029	-0.021	-0.021	0.002	-0.026	-0.017	-0.023	0.002	-0.027	-0.019
Pav Condition	0.008	0.003	0.001	0.014	-0.025	0.003	-0.032	-0.019	-0.023	0.003	-0.029	-0.017
(Negative binomial)	1				1				1			

Many of these significant variables and trends have been previously shown in studies within this subject. Abdel-Aty et al. found in 2000 that ADT volumes, road lengths, road widths, and urban/rural classification are all significant using a Negative Binomial model for crash frequency. Road width was also found to be significant in studies by Anastasopoulos et al. (2007) and Berhanu (2004). Wang et al. found that both the number of lanes and traffic volumes were significant in their 2006 study.

Table 13: Negative Binomial Regression Goodness-of-Fit

	2002			2003			2004		
	Value	df	Value/df	Value	df	Value/df	Value	df	Value/df
Deviance	6.77E+03	5,859	1.155	6.90E+03	6,052	1.139	7.22E+03	6,055	1.192
Scaled Deviance	6.77E+03	5,859		6.90E+03	6,052		7.22E+03	6,055	
Pearson Chi-Square	1.04E+04	5,859	1.774	1.04E+04	6,052	1.713	1.09E+04	6,055	1.801
Scaled Pearson Chi-Square	1.04E+04	5,859		1.04E+04	6,052		1.09E+04	6,055	
Log Likelihood[a]	-8.40E+03			-8.58E+03			-8.83E+03		
Akaike's Information Criterion (AIC)	1.68E+04			1.72E+04			1.77E+04		
Finite Sample Corrected AIC (AICC)	1.68E+04			1.72E+04			1.77E+04		
Bayesian Information Criterion (BIC)	1.69E+04			1.73E+04			1.78E+04		
Consistent AIC (CAIC)	1.70E+04			1.73E+04			1.78E+04		

*Table values are based on a smaller is better form

Finally, it is important to determine whether or not the Negative Binomial model resulted in a better fit to the data. Looking at the goodness-of-fit values for each of the year's models in Table 13, this fact is verified. Every criterion calculated with SPSS shows a significant decrease compared to the results from the Poisson model. Because these criterion are based on a 'smaller is better' form, the Negative Binomial is concluded to be the better of the two models for the Arkansas crash data.

4.2.2 Binary Logistic Regression

For the second analysis of this study, a Binary Logistic regression model was built to describe the nature of crash severity. Crash severity is a binary response, where 1 represents a severe crash and 0 represents one that is not severe. Distinctions between the two are made based on a ranking scale similar to injury severity. For this analysis, only variables from the crash database were considered. Along with crash severity, 17 variables were initially included in the model. These variables are shown alongside their possible values in Table 14.

Many of the potential variables above are binary in nature, meaning they have only a value of 0 or 1. Injury severity is an ordinal value ranked from 5 to 1, where 1 is a fatal injury. Road system type is a nominal value from 1 to 5, because there is no natural ranking of these values. Other non-binary variables include the year of the accident, county location, number involved in each crash, and age of the driver. Instead of modeling road segments as in the previous analysis, this model is based on exclusive crash occurrences. To avoid any hierarchical nature within the data, only single-vehicle crashes were included in the model. For the human factor variables, the vehicle's driver's values were used.

Table 14: Variables Included in Logistic Regression Model

Variable	Description	Variable	Description
CRASHSEVERITY	Non-incapacitating or greater = 1	ROADWAYPROFILE	Level = 1
	Less than non-incapacitating = 0		Not Level = 0
YEAR	2002 = 1	WEEKDAY	Weekend = 1
	2003 = 2		Weekday = 0
	2004 = 3	NUMBERINVOLVED	1, 2, 3, …
ATMOSPHERICCONDITIONS	Clear = 1	ALCOHOLINVOLVED	No = 1
	Not Clear = 0		Yes = 0
LIGHTCONDITIONS	Daylight =1	RESTRAINTCODE	Safety belt = 1
	Not Daylight = 0		Other = 0
RURALURBAN	Rural = 1	SEX	M ale = 1
	Urban = 0		Female = 0
ROADSURFACECONDITION	Dry = 1	AGE	Actual Age of Driver
	Not Dry = 0	INJURYSEVERITY_ORD	1 = Fatal
ROADSYSTEM	Interstate = 1		2 = Incapacitating Injury
	US Highway = 2		3 = Non-incapacitating Injury
	State Highway = 3		4 = Possible Injury
	County Road = 4		5 = No Injury
	City Street = 5	LICENSESTATE	AR = 1
ROADWAYALIGNMENT	Straight = 1		Other state = 0
	Not Straight = 0	COUNTYNUMBER	1, 2, 3, …., 75

Again, SPSS was used to make model fits. The program simply uses its Logistic regression function to perform the analysis. Initially, the data was inputted and run for all 17 variables. To avoid correlation issues similar to Road Width and Lane Width in the previous analysis, the initial model was tested for correlated effects. Although this was not a significant issue before, it was believed that the data in the second analysis would have more correlation between some of the variables. To be sure, a correlation matrix was developed, which is displayed in Table 15.

From the correlation matrix, two variables were strongly related: Atmospheric Conditions and Road Surface Conditions. This makes sense, because when weather conditions are clear, the road surface tends to be dry. Also, when the weather is rainy, the road surface tends to be wet. Since the two were so related, one was left out. Because the correlation was so high, it did not matter which one was chosen to be removed, and so atmospheric conditions was taken out. Injury severity also has a natural correlation, although not as high as the previous correlation, to crash severity. Severe crashes tend to produce more severe injuries. Thus, injury severity was left out of the final model. Removing these two variables, the Binary Logistic regression was run once more. The results of this model calculation are shown in Tables 16 and 17.

Table 15: Correlation Matrix for Binary Logistic Regression Estimates

	Constant	YEAR	ATMOSPHERICCONDITIONS	LIGHTCONDITIONS	RURALURBAN	ROADSURFACECONDITION	ROADSYSTEM	ROADWAYALIGNMENT	ROADWAYPROFILE	WEEKDAY	NUMBERINVOLVED	ALCOHOLINVOLVED	RESTRAINTCODE	SEX	AGE	INJURYSEVERITY_ORD
Constant	1.0	-0.4	-0.1	0.0	-0.3	0.0	-0.3	-0.1	-0.1	-0.1	0.0	-0.2	-0.1	-0.3	-0.2	-0.5
YEAR	-0.4	1.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
ATMOSPHERICCONDITIONS	-0.1	0.0	1.0	0.0	0.0	-0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LIGHTCONDITIONS	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	-0.2	0.0	0.0	-0.1	0.0
RURALURBAN	-0.3	0.1	0.0	0.0	1.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1
ROADSURFACECONDITION	0.0	0.0	-0.8	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ROADSYSTEM	-0.3	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	-0.1
ROADWAYALIGNMENT	-0.1	0.0	0.0	0.0	0.1	0.0	0.0	1.0	-0.2	0.0	0.0	0.0	0.0	0.0	0.0	-0.1
ROADWAYPROFILE	-0.1	0.0	0.0	0.0	0.1	0.0	0.0	-0.2	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
WEEKDAY	-0.1	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	1.0	0.1	-0.1	0.0	0.0	0.0	0.0
NUMBERINVOLVED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	1.0	0.0	0.0	0.1	0.2	-0.3
ALCOHOLINVOLVED	-0.2	0.0	0.0	-0.2	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	1.0	-0.1	0.1	0.0	0.0
RESTRAINTCODE	-0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	-0.1	1.0	0.1	0.1	-0.1
SEX	-0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	1.0	0.1	0.0
AGE	-0.2	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.1	0.1	1.0	-0.2
INJURYSEVERITY ORD	-0.5	0.1	0.0	0.0	0.1	0.0	-0.1	-0.1	0.0	0.0	-0.3	0.0	-0.1	0.0	-0.2	1.0
LICENSESTATE	-0.3	0.0	0.0	-0.1	0.0	0.0	-0.1	0.0	0.0	0.0	0.1	0.0	-0.1	0.0	0.3	0.0
COUNTYNUMBER	-0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 16: Binary Logistic Regression Parameter Estimates and Effects

Variables	B	S.E.	Wald	df	p-value	Exp(B)
YEAR	0.032	0.026	1.497	1	0.221	1.032
LIGHTCONDITIONS	0.204	0.043	22.023	1	**0.000**	1.226
RURALURBAN	0.307	0.043	49.908	1	**0.000**	1.359
ROADSURFACECONDITION	0.51	0.05	103.161	1	**0.000**	1.665
ROADSYSTEM	-0.088	0.018	23.148	1	**0.000**	0.916
ROADWAYALIGNMENT	-0.155	0.048	10.35	1	**0.001**	0.857
ROADWAYPROFILE	-0.129	0.047	7.444	1	**0.006**	0.879
WEEKDAY	0.051	0.042	1.474	1	0.225	1.052
NUMBERINVOLVED	0.406	0.023	301.804	1	**0.000**	1.500
ALCOHOLINVOLVED	-0.678	0.058	136.105	1	**0.000**	0.508
RESTRAINTCODE	-0.896	0.045	404.526	1	**0.000**	0.408
SEX	0.092	0.045	4.214	1	0.040	1.096
AGE	-0.004	0.001	15.744	1	**0.000**	0.996
LICENSESTATE	0.613	0.06	102.981	1	**0.000**	1.846
COUNTYNUMBER	0	0.001	0.002	1	0.963	1.000
Constant	-1.051	0.151	48.503	1	**0.000**	0.349

Table 17: Logistic Regression Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	14,036.025[a]	.094	.136

From Table 16, all variables are considered significant based on a 99% confidence interval except for Year, Weekday, Sex, and County Number. The interpretation of this is simple; based on the Arkansas crash data, the results do not show significant change in crash severity from year to year, county to county, day to day, or between male and female drivers. These results were mostly as expected, mainly due to the nature of the response. Again, the response variable being modeled is crash severity, which was not expected to be dependent on day or location. This is opposite from the expected results about crash frequency, where day and location were expected to be significant. The risk of a crash during the weekend or in a specific county may be higher than another county on another day, but the severity of those crashes should be consistent around the state. Sex was actually found to have a p-value of 0.04, which is significant on a 95% confidence interval, but not on a 99% confidence interval.

The significant variables, on the other hand, are more difficult to interpret. Assuming this is a well fit model, the slope coefficients for the binary logistic model cannot be interpreted the same as before. Here, the trend is not specifically linear. In fact, most of the explanatory variables are binary values of 0 or 1, which would make an interpretable relationship between the coefficients and response nearly impossible. However, as was mentioned before, one positive aspect of using the logistic regression model is its calculation of the odds ratio. For the output given in this table, the odds ratio corresponds to the *Exp(B)* term. Road Surface Condition was determined significant with this model and has an odds ratio of 1.665. This odds ratio is calculated as the odds of equaling a 1 (or having a dry road condition) divided by the odds of equaling a 0 (or having a road condition that is not dry). Because this ratio is greater than 1, it can be interpreted as saying that severe crashes have a higher probability of occurring on dry roads as they do on roads that are not dry; all other things being equal. Although that may not seem intuitive, this is a very feasible situation. In fact, if road conditions are poor, such as wet or icy, drivers may be more alert and drive slower. Crashes may be more abundant during these conditions, but severe crashes may not be if drivers are driving slowly and cautiously. It is when conditions are clear that drivers tend to speed and drive more recklessly, causing more severe crashes.

Roadway curvature is shown to be significant both vertically and horizontally. The odds ratios for Roadway Profile and Roadway Grade are 0.857 and 0.879, respectively. Because these values are less than 1, it can be suggested that severe crashes are more prominent along curved roads than straight and level roads. This result is as expected, because of the increased risk involved when driver visibility is decremented by blind spots caused by curves.

The restraint code variable is also significant with an odds ratio of 0.408. Since 1 refers to the situation where the driver is wearing a seat belt, this odds ratio is interpreted as conveying that the probability of a severe crash is actually decreased when wearing a seat belt. This is important in other terms, because crash severity was shown to be correlated with injury severity. This means that the data has shown significant evidence that seat belts reduce crash and injury severity. The variable associated with alcohol shows similar results. Its odds ratio states that crash severity is lessened when ALCHOLINVOLVEMENT is equal to 1, meaning the driver has not had alcohol.

An odds ratio close to 1 for any of the variables suggests that there is no real difference in the odds of the specific values of the variable. For example, Sex was almost shown to be significant using a 99% confidence interval. However, even if it was concluded that the sex of the driver was significant, the odds ratio of the variable is 1.096. This suggests that even with a significant predicting relationship to crash severity, the risk is barely increased when the driver is male, rather than female. But this difference in risk is small in comparison to other variables significance.

Age is the only true continuous variable within this data, and therefore the odds ratio cannot be interpreted for this variable. The coefficient is -0.004, which shows that there is no major difference between crash severity between 18 year olds and 64 year olds, for example. However, the trend is still significant and can be somewhat interpreted as an increase in crash severity for older drivers.

In terms of previous studies, many of these findings are comparable to past research. Some results that are emphasized in this study as being significant both here and in previous research are road surface conditions (Kim et al., 2007; Shankar et al., 2004), lighting conditions (Al-Ghamdi, 2002; Anastasopoulos et al., 2007; Jones et al., 2003; Yau et al., 2006), roadway curvature (Abdel-Aty et al., 2000; Anastasopoulos et al., 2007; Kim et al., 2007), seat belt usage (Hutchins et al., 2003), driver age, gender, and alcohol usage (Abdel-Aty et al., 2000; Jones et al., 2003).

CHAPTER V – OBSERVATIONAL ANALYSIS

5.1 Methodology of Observational Analysis

Observational road studies in this report serve two important functions that statistical analysis fail to provide. First, as was mentioned earlier, is the observational study's ability to describe problem areas within a road system before a crash occurs. The crash data only focuses on the aspects of crashes that have already occurred. By observing several types of roads and locations, these problem areas can be analyzed and solutions can be recommended. This proactive approach disregards any past data and focuses only on the current roadway issues that can affect crashes in the future.

The second function that the observational analyses serve is their ability to measure potential factors or hazards that cannot be, or have not been, recorded in either the roadway inventory or crash database for Arkansas. This gives a much better representation of driver behavior as compared to the subjective measurements documented in the crash database. Also, aspects of the road that are not documented explicitly in the roadway inventory file can be evaluated. Overall, the observations can be used to bridge the informational gaps that the historical data may have had.

5.1.1 Choice of Locations

Although the quality of observational analyses is generally more accurate, as compared with the statistical analyses, the biggest drawback of the observational technique is the amount of time and money that it takes to evaluate every single road system. For an ideally proactive approach to road safety, observations should be made for every road type, segment, and location that is available. When this is not a feasible solution, it is common that a sample of all the roads is chosen to be evaluated. Because of the 115,000 potential road segments to be observed within the state of Arkansas, a method was developed to sample only a few of these segments.

The choice of road segments to be observed can be made using several methods, which depend heavily on the scope of the study. For this particular study it was decided to sample roads that have been historically more dangerous than other roads. These roads are determined by a ranking system that measures the crash rate among all potential road segments, with respect to the length and average daily traffic volume of each particular road segment (Anastasopoulos et al., 2007). The following equation shows the calculation method for each road segment's relative crash rate:

$$AccidentRate*_{Year,i} = \frac{Accidents_{Year,i}}{(ADT_{Year,i} \times Length_i \times 365)} \tag{23}$$

* Anastasopoulos et al., 2007

This expression represents accident rates for a specific year on a road segment *i*. $Accidents_{Year,i}$ is the total crash count for road segment *i* in a specific year, and $ADT_{Year,i}$ represents the Average Daily Traffic volume measured for that year on the *i*th road segment. For this analysis, the crash count consists of only those crashes that are denoted as being intersection related or at an intersection. This is done in order to stay consistent with the results of the statistical analysis which only takes into account the intersection crashes. To transform the ADT values into Average Yearly Traffic, this average is multiplied by 365. It is natural to assume a positive correlation between the length of a road segment and the number of crashes that occur on that segment; therefore the length is also factored into the rate function. Rates are usually given units in terms of the number of crashes per 100,000,000 vehicles traveled per road length, which would require the above rate to be multiplied by 100,000,000. However, this scaled factor does not affect the ranking order of road segments, and was thus left out of the expression.

For this particular study, it was important to focus on roads that are regarded as more dangerous, because these roads need the most attention. The reason they need the most attention is the fact that road segments with the highest crash rates are the locations that can stand for the most improvement. It is important to note that the ranking crash rate method is only used to sample roads that are assumed to be more dangerous than most other road segments. This does not mean that these choices are the most dangerous road segments in Arkansas.

Within the roadway inventory file, road segments occasionally change from year to year. To overcome this, a separate crash rate ranking was made for each of the three years. The choice of road segments for this study will then be based on those road segments that are consistently at the top of each year's ranking.

5.1.2 Location Procedures

Once all possible road segments have been ranked by their respective crash rates, the locations with the highest rankings will be visited. The ranking is based on intersection crashes because it is easier to encounter hazards when cars are in direct contact. Therefore, each road segment chosen will be analyzed based on the intersecting streets throughout the road segment. It is important to notice that there is a restraint on the Arkansas data, which only allows this portion of the study to be focused on State highways, US highways, and interstates. This is due to the lack of data for county roads and city streets. Crash frequency cannot be collected for segments of these roads, because they are not broken up into segments like the larger roads are. Therefore, the locations chosen for analysis will be those on the highways and interstates. This will also be in conjunction with the statistical analysis.

At each location, several observations will be made about the road and traffic flow. There are many aspects of the road itself, as well as driver behavior, which is not fully describable in the historical data. These are the most important aspects to record and survey. The crash database and roadway inventory database take into account general details about the shape of the road, such as the grade and curvature of the road, but this is not always enough. Road layout is also important. When considering the intersections of major highways and interstates, it is important to think about signage and lane markings. The data will say whether or not a crash occurred at a traffic light and whether that light was functioning, but it does not mention anything about the signage and how clear it was. Perhaps there were no signs or lane markings to guide the traffic. In cases like this, it is important to observe the flow of traffic and the behavior of the drivers through the entire intersection. If right-of-way and traffic flow is not properly displayed, it might be visible by the actions taken by the driver. Improper turns could be a sign that the driver did not know what action to take. Therefore, for each location, it is important to observe all signage, lane markings, layout of the road, and other surrounding factors such as buildings that may affect the way a person drives in that area.

Another aspect of the observational study is the analysis of driver behavior; in particular, the behavior that may not be attributed to bad signage or markings along the road. These aspects may be in terms of human factors; things that the driver is doing that may distract them from their driving. This may include whether or not they have any passengers, if they are talking on cell phones, or if they are doing any other distracting task that keeps their eyes off the road. Different locations may have different populations with different behaviors, so it is important to note these differences. This part of the study may be extremely subjective, but it allows some insights on the issues involved on the road today. It has already been shown that these human factors are nearly impossible to measure quantitatively, but studying driver behavior may be able to highlight important topics that need to be addressed.

5.2 Observational Approach Results and Discussion

Locations were chosen by means of a ranking system of crash rates among all possible locations. The results of the ranking were evaluated and several locations were chosen to be visited for a hands-on observational study. The findings made at each location, including any infrastructural problems, driver behaviors, signage issues, etc. were documented and discussed.

5.2.1 Choice of Locations Results

Crash rates for all three years (2002-2004) of historical crash data were analyzed. Road segments were chosen based on how high the ranking was in each year. Due to the lack of the physical ability to observe each of the ranked locations, many segments were chosen based on their relatively high ranks for each of the three years. Segments that were consistently high in each year were given more emphasis than a location that had only one year of high rank. This was done to avoid potential outliers, which may have been the cause of some extraneous factors involved with any one road segment during any one of the three years. The original rankings are shown in Tables 18-20. The final choices of locations are shown in Table 21 and again in Figure 3.

This figure shows the actual locations as an overview map throughout the state of Arkansas. Each one of these locations is shown in greater detail in Appendix B including the aerial screenshots of the intersection.

Table 18: 2002 Crash Rate Ranking

Rank	City	County	Route	Section	BegLogmile	Length*	ADT*	Frequency*	Crash Rate
1	North Little Rock	60	67	10	1.16	0.01	45,589	60	3.606E-04
2	Marion	18	64	17	19.26	0.01	6,100	8	3.593E-04
3	Little Rock	60	30	23	135.17	0.1	2,673	35	3.587E-04
4	Texarkana	46	30	11	0	0.01	21,000	26	3.392E-04
5	Blytheville	47	55	12	67.33	0.01	18,000	18	2.740E-04
6	White Hall	35	270	11	6.84	0.01	8,500	7	2.256E-04
7	Marion	18	64	17	19.26	0.02	6,100	10	2.246E-04
8	Blytheville	47	61	3	13.58	0.01	6,200	5	2.209E-04
9	Pine Bluff	35	79	09B	0	0.01	15,000	12	2.192E-04
10	Jacksonville	60	67	10	10.89	0.28	2,124	44	2.027E-04
11	Fayetteville	72	112	0	1.41	0.01	6,800	5	2.015E-04
12	Little Rock	60	365	12	0.69	0.01	15,000	11	2.009E-04
13	Little Rock	60	430	21	7.68	0.01	61,000	43	1.931E-04
14	Alma	17	71	15	0	0.01	13,000	9	1.897E-04
15	Sherwood	60	67	10	3.09	0.27	3,241	60	1.879E-04
16	Jonesboro	16	63	7	1.64	0.01	12,000	8	1.826E-04
17	El Dorado	70	82	05B	2.4	0.02	11,000	14	1.743E-04
18	Fort Smith	65	71	14B	3.53	0.01	21,000	13	1.696E-04
19	Van Buren	17	40	11	7.38	0.01	33,000	18	1.494E-04
20	Van Buren	17	59	6	0.94	0.01	9,300	5	1.473E-04
21	Dumas	21	54	2	0.68	0.01	5,600	3	1.468E-04
22	Fort Smith	65	271	1	0	0.1	2,300	12	1.429E-04
23	North Little Rock	60	67	10	0.84	0.36	3,341	61	1.390E-04
24	Van Buren	17	59	5	25.14	0.01	23,676	12	1.389E-04
25	Fort Smith	65	22	1	3.72	0.01	40,000	20	1.370E-04

*Length and ADT have units of miles, while Frequency has units of cashes. All other numbers are references.

Table 19: 2003 Crash Rate Ranking

Rank	City	County	Route	Section	BegLogmile	Length*	ADT*	Frequency*	Crash Rate
1	Jacksonville	60	67	10	10.89	0.28	2,124	51	2.349E-04
2	Gateway	4	62	2	18.88	0.02	2,400	3	1.712E-04
3	Dumas	21	54	2	0.68	0.01	5,600	3	1.468E-04
4	Clarksville	36	103	0	1.6	0.1	2,100	11	1.435E-04
5	Marion	18	64	17	19.26	0.02	6,000	6	1.370E-04
6	Gentry	4	59	01B	0.57	0.03	2,200	3	1.245E-04
7	North Little Rock	60	70	13	0.98	0.02	17,800	16	1.231E-04
8	Fort Smith	65	255	3	5.28	0.04	21,000	37	1.207E-04
9	Blytheville	47	18	6	15.99	0.02	5,700	5	1.202E-04
10	Fort Smith	65	22	1	0.06	0.01	12,000	5	1.142E-04
11	El Dorado	70	82	05B	2.4	0.02	11,000	9	1.121E-04
12	Dardanelle	75	7	13	14.55	0.05	8,300	16	1.056E-04
13	Fayetteville	72	71	16B	2.59	0.03	22,000	25	1.038E-04
14	North Little Rock	60	67	10	0.84	0.36	3,341	44	1.002E-04
15	Paragould	28	49	2	17.06	0.03	15,000	16	9.741E-05
16	Fort Smith	65	271	1	0	0.1	2,300	8	9.529E-05
17	Hamburg	2	82	8	24.03	0.01	8,700	3	9.447E-05
18	Marked Tree	56	140	1	0	0.01	2,900	1	9.447E-05
19	Pocahontas	61	62	19	10.59	0.02	4,400	3	9.340E-05
20	North Little Rock	60	70	13	0.66	0.03	11,000	11	9.132E-05
21	Pine Bluff	35	63	13B	1.34	0.05	9,100	15	9.032E-05
22	Sherwood	60	67	10	3.09	0.27	3,241	28	8.766E-05
23	Hope	29	67	2	14.76	0.03	4,200	4	8.698E-05
24	Sheridan	27	35	2	13.44	0.07	2,300	5	8.508E-05
25	Marion	18	77	5	15.84	0.02	6,600	4	8.302E-05

*Length and ADT have units of miles, while Frequency has units of cashes. All other numbers are references

Table 20: 2004 Crash Rate Rankings

Rank	City	County	Route	Section	BegLogmile	Length*	ADT*	Frequency*	Crash Rate
1	Jacksonville	60	67	10	10.89	0.28	2,124	44	2.027E-04
2	Marion	18	64	17	19.26	0.02	7,500	11	2.009E-04
3	Bryant	62	30	22	122.69	0.46	2,952	95	1.917E-04
4	Fort Smith	65	271	1	0	0.1	2,000	12	1.644E-04
5	El Dorado	70	82	05B	2.4	0.02	10,500	11	1.435E-04
6	Fort Smith	65	64	1	0.07	0.08	8,900	35	1.347E-04
7	North Little Rock	60	30	23	140.99	0.18	3,563	30	1.282E-04
8	Blytheville/Osceola	47	61	3	4.22	0.01	4,300	2	1.274E-04
9	Lockesburg	66	371	1	0	0.02	2,200	2	1.245E-04
10	Pangburn/Searcy	73	16	13	0.89	0.02	2,300	2	1.191E-04
11	Fayetteville	72	71	16B	2.59	0.03	23,900	28	1.070E-04
12	North Little Rock	60	70	13	0.66	0.03	12,300	14	1.039E-04
13	Pine Bluff	35	63	13B	1.34	0.05	9,200	17	1.013E-04
14	Dardanelle	75	7	13	14.55	0.05	7,200	13	9.893E-05
15	Fort Smith	65	22	1	0.06	0.01	11,300	4	9.698E-05
16	Marked Tree	56	63	08B	0.9	0.02	2,900	2	9.447E-05
17	Blytheville	47	18	6	15.99	0.02	5,800	4	9.447E-05
18	Paragould	28	412	9	0.19	0.03	14,600	15	9.383E-05
19	Paragould	28	412	9	0.16	0.03	14,900	15	9.194E-05
20	Magnolia	14	82	03B	0.24	0.03	8,200	8	8.910E-05
21	North Little Rock	60	67	10	0.84	0.36	3,341	39	8.884E-05
22	Harrison	5	65	01B	1.73	0.05	12,400	19	8.396E-05
23	Osceola	47	140	2	14.49	0.06	3,300	6	8.302E-05
24	Jacksonville	60	294	1	1.44	0.02	8,300	5	8.252E-05
25	Pine Bluff	35	79	9	11.91	0.03	9,000	8	8.118E-05

*Length and ADT have units of miles, while Frequency has units of cashes. All other numbers are references

Table 21: Final Observation Locations

Location #	County	City	Route	Section	Logmile	Type of Intersection
1	Crawford	Alma	71	15	0.00	Interstate/Highway
2	Mississippi	Blytheville	18	6	15.99 - 16.01	Highway/Highway
3	Mississippi	Blytheville	55	12	67.33	Interstate/Frontage
4	Mississippi	Blytheville	61	3	13.58	Interstate/Highway
5	Saline	Bryant	30	22	122.69 - 123.15	Highway/Frontage
6	Johnson	Clarksville	103	0	1.60 - 1.70	Interstate/Highway
7	Yell	Dardanelle	7	13	14.55 - 14.60	Highway/Highway
8	Desha	Dumas	54	2	0.68 - 0.69	Highway/City Street
9	Washington	Fayetteville	71	16B	2.6	Highway/City Street
10	Crawford	Fort Smith	22	1	0.60 - 0.70	Highway/Highway
11	Crawford	Fort Smith	255	3	5.28 - 5.32	Interstate/Highway
12	Crawford	Fort Smith	271	1	0.00 - 0.10	Highway/Highway
13	Pulaski	Jacksonville	67	10	10.89 - 11.17	Highway/City Street
14	Pulaski	Little Rock	365	12	0.69	Interstate/Highway
15	Pulaski	Little Rock	430	21	7.68	Interstate/City Street
16	Crittenden	Marion	64	17	19.26	Frontage/Access Road
17	Crittenden	Marion	77	5	15.85 - 15.86	Interstate/Highway
18	Pulaski	North Little Rock	67	10	1.16	Highway/City Street
19	Pulaski	North Little Rock	70	13	0.66 - 0.69	Highway/City Street
20	Pulaski	North Little Rock	70	13	0.98 - 1.00	Interstate/Highway
21	Jefferson	Pine Bluff	63	13B	1.34 - 1.39	Highway/City Street
22	Jefferson	Pine Bluff	79	9B	0.00	Interstate/Highway
23	Crawford	Van Buren	40	11	7.38	Interstate/Highway
24	Crawford	Van Buren	59	5	25.14	Interstate/Highway
25	Crawford	Van Buren	59	6	0.94	Interstate/Highway
26	Jefferson	White Hall	270	11	6.84	Interstate/Highway

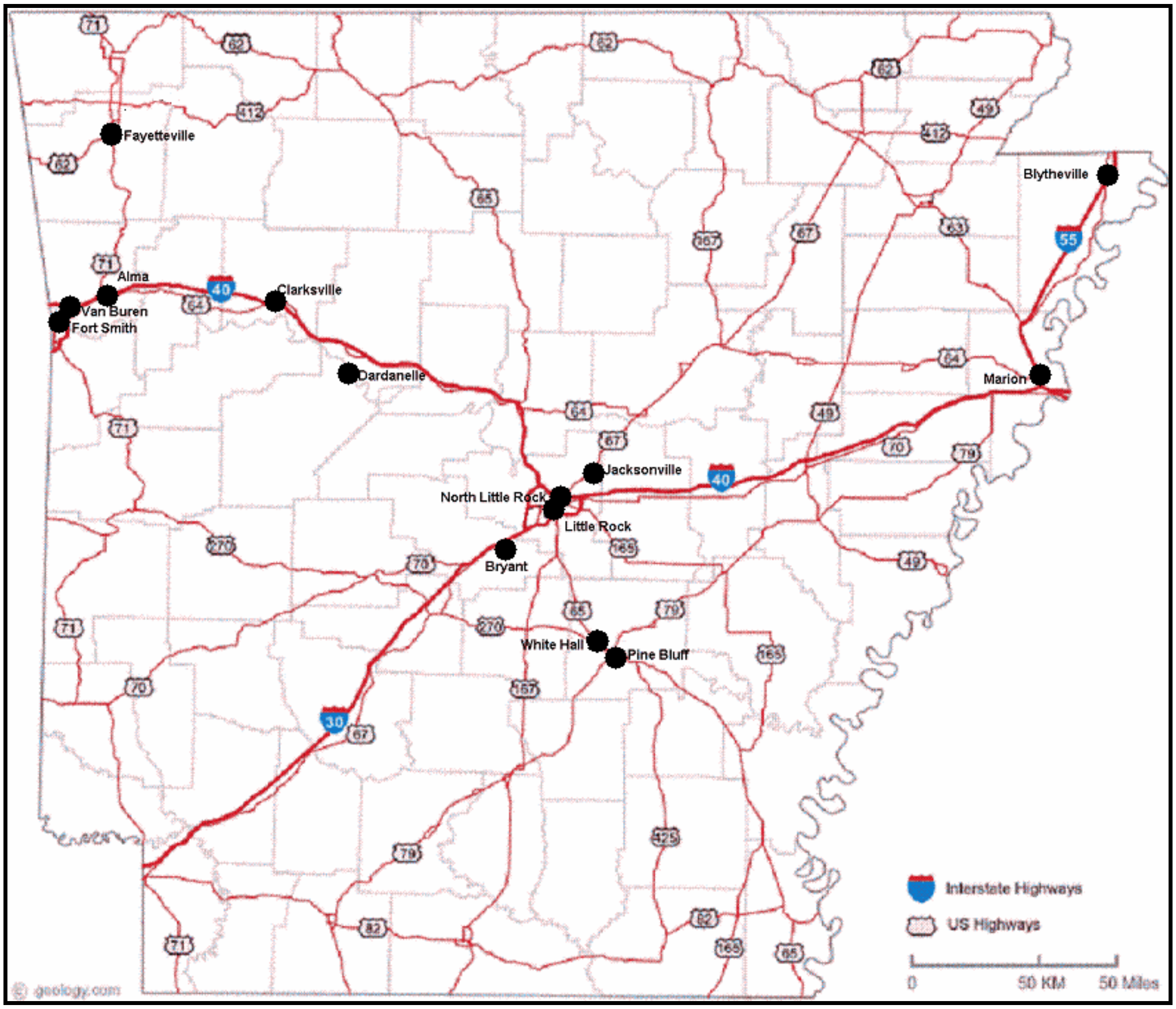


Figure 3: Observation Locations Overview

5.2.2 Location Analysis

The observational analysis, which covered a total of 26 road segments, led to many discoveries concerning road layouts and traffic behavior throughout Arkansas. Each road segment displayed some positive aspects, and some negative aspects. However, because these intersections were chosen based on their high crash rate ranking over the years between 2002 and 2004, it was found that there were several more aspects considered potentially dangerous rather than safe. This is especially true in the area of traffic signs, lane markings, driver visibility, and turning lanes. The crash and roadway databases do not have extensive information about these aspects along each road segment. Therefore these findings can potentially help give additional insights for why some areas are more dangerous than others. Table 22 is a display of many of the crucial findings discovered across these Arkansas intersections. It overviews a number of concerns along with a description and specific locations associated with them.

Table 22: Key Problems Among Locations

Major Problem Category		Description of Intersection Issues	Example Locations
C1	Poor Lane Markings	Markings are not visible, difficult to interpret, worn down, or non existent	6, 12, 14, 15, 17
C2	Poor Signage	Signs that are not visible, difficult to interpret, or non existent	6, 7, 10, 14, 17, 23, 25
C3	Conflicting Information	Signage, lane markings, signals, or infrastructure with conflicting driver information	7, 8, 14, 25
C4	Poor Turning Lanes	Turning Lanes are too short, too narrow, too crooked, not visible, or non existent	1, 6, 7, 9, 11, 12, 14, 20, 22, 24
C5	Medians	Medians along or within the road, serving as obstacles	1, 2, 3, 7, 15, 16, 17, 21, 25
C6	Poor Traffic Signal Layout	Signals are not located directly above road or on poles	6, 8, 12, 13, 14, 24
C7	Crosswalks/Other Obstacle	Crosswalks, railroad tracks, trolley tracks, or any other obstacle crossing the intersection	8, 12, 18, 19
C8	Poor Visibility	Field of view obstructed by objects, sharp turns in the road, or elevated roads	2, 3, 5, 6, 7, 9, 13, 15, 18, 21, 23
C9	Poor Merging	Roads merge too quickly or in dangerous conditions such as high speeds	3, 5, 10, 15, 16, 20, 21, 23, 26
C10	Timed Traffic Signals	Signals are not activated by sensors, affecting the responsiveness of traffic	10, 11, 19
C11	Traffic Signal Duration	Signals with longer than usual red light durations, causing several cars to run yellow lights	11, 13, 20
C12	No Traffic Signal	No traffic signal existing at intersection, only signs	4, 9
C13	Angled Intersections	Intersecting roads are not perpendicular, and form difficult angles of cross traffic	2, 9, 10, 12, 13, 17

Figures 4-7 show some examples of bad signage and lane markings (C1 and C2) found throughout these intersections. Instances like the ones shown in these figures are comparable to many intersections observed throughout this study. Figure 4 and 5 show two intersections that do not even have lane markings to direct the traffic. It was observed at these intersections that many drivers were not sure where to move their vehicles, because there were no obvious directions or separations on the road. In Figure 4, the main road consists of what looks like a three lane road, where one side of the road is wide enough to contain two lanes. Yet, this road quickly narrows to a one lane road, without any signage or markings to allow drivers to merge properly.

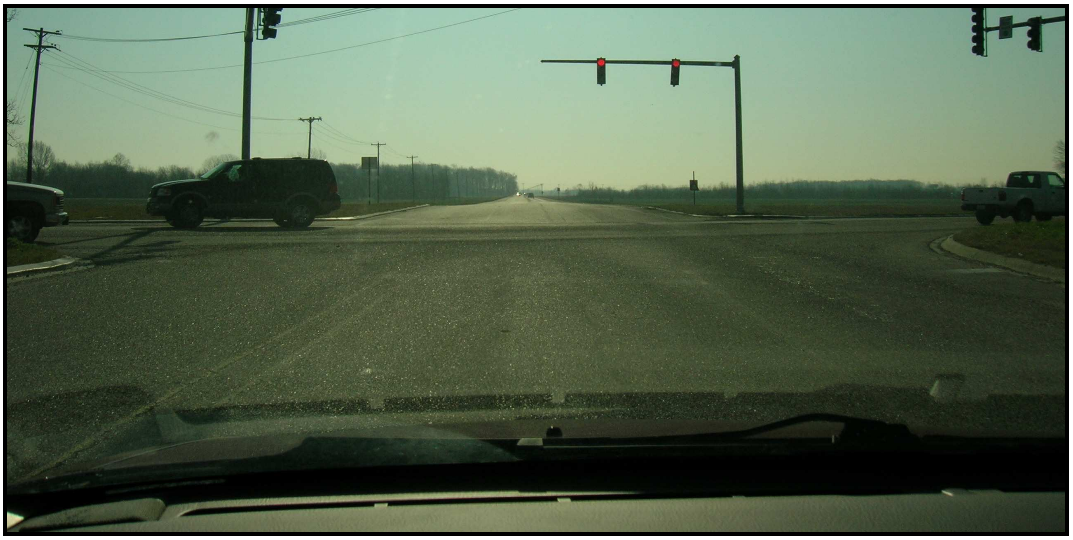


Figure 4: Worn Lane Markings (C1) – Location 17, Marion, Crittenden County

Figure 5: No Visible Lane Markings (C1) – Location 6, Clarksville, Johnson County

The situation in Figure 5 features an implied left turning lane along with two other lanes on either side of it. There are clues that this intersection used to be marked, but they have faded beyond recognition. Lanes that cannot be visibly seen should always be kept in good condition or drivers run the risk of misinterpreting where they should drive. This misinterpretation may easily lead to an accident as soon as other vehicles with the same misinterpretation enter these roads. Vehicles turning onto a road with no marking may cause them to cut the turn too short or too wide, which could cause a potential interaction with oncoming traffic on the other road.

Signs, like lane markings, are also a large source of driver information. However, there are many places that do not have signs or that do have signs that are not visible (C2). During this study, it was found that most locations contained a fair amount of signage. Yet, many of these signs were not located in the best locations. For example, Figure 6 shows an intersection where the turning directions for the traffic signal were displayed along the overpass. These signs were extremely dirty and nearly blended in with the surrounding infrastructure. They did not stand out like proper signage should, and therefore went unobserved by many drivers.

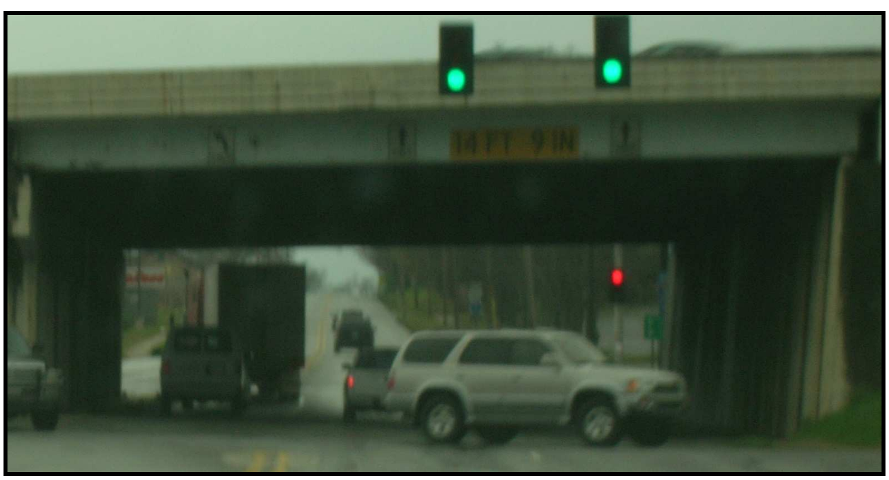


Figure 6: Poor Signage (C2) – Location 14, Little Rock, Pulaski County

Often times, intersections that contain good signage and lane markings will use both together to emphasize proper driver behavior. For example, a traffic signal may display signs that portray the turning conditions of the lanes ahead, while the lanes contain similar directional markings. At these good intersections, drivers have the opportunity to find out what lane they need to be in first with the road markings and then again with the signs. However, it is common that drivers do not pay attention to both. Some drivers may only pay attention to what is on signs, where as others may pay closer attention to the road itself and lane markings. Therefore, intersections that contain only signs or only lane markings contain significantly less information for drivers to be aware of. Figure 6 is an example of a road that contains only signage to direct traffic flows. On this road there are no lane markings for left or right turn lanes, even though there are left turn lanes. Drivers who fail to see the signs have no other way of knowing they are in the right or wrong lane until perhaps they get involved in an accident. Figure 7 is another example, but with the opposite conditions (C1 and C2). This road segment contains an arrow lane marking right before the traffic signal, but no signs above the traffic light. Here drivers may not see the arrow on the road and attempt to turn left; a potential crash situation. Both lane markings and signs are good pieces of information, but it increases safety if they both exist at an intersection and if they are both visible. This gives the driver the most awareness of the actions they will need to take.

Figure 7: Lane Markings Only (C1, 2) – Location 7, Dardanelle, Yell County

The next two figures display an interesting issue discovered during the intersection observations. The issue here is quite the opposite of the most usual case of not enough signage and lane markings. Instead, this particular road segment contained an over abundance of markings. Figures 8 and 9 feature the same road segment, which is at the intersection of a state highway and an intestate. Therefore, there were several exit ramps coming to and from the interstate; many of which were one-way roads. No Entrance signs are common at these types of intersections, yet this intersection contained 6. The two roads intersecting the main highway were the on and off ramps for the interstate; one is allowable to enter, the other is not.

However, with the current layout of signs and lane markings, it appears that neither road is approachable. The information on this road segment was too complex, leading to driver confusion (C3). It was observed that drivers took a lot longer to make turns at this intersection, perhaps due to more decision making on the driver's part. In fact, the main problem area at this intersection had conflicting information between its signs and lane markings (Figure 8).

Figure 8: Conflicting Information (C3) – Location 25, Van Buren, Crawford County

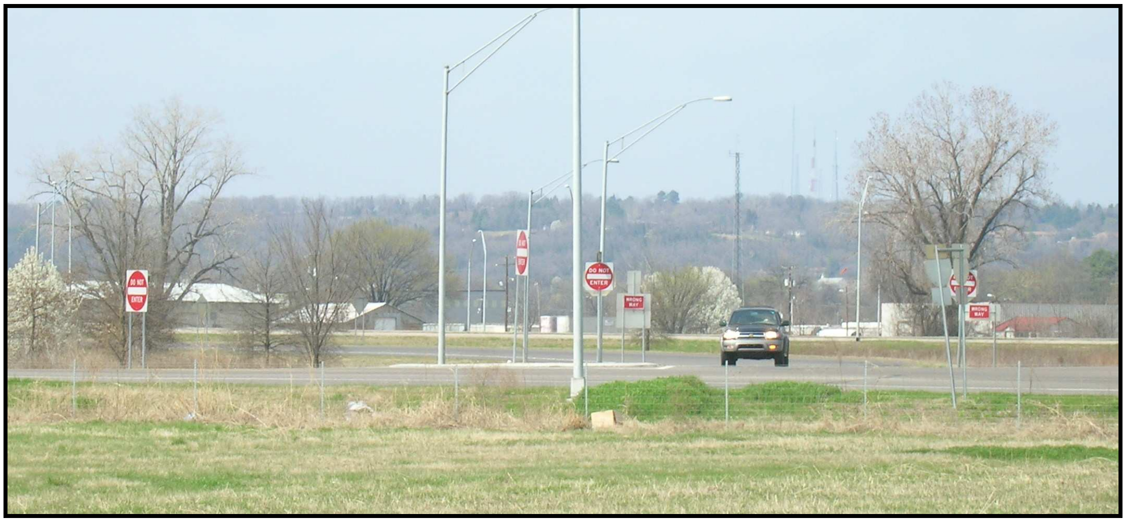


Figure 9: Unnecessary Signage (C1, 3) – Location 25, Van Buren, Crawford County

Infrastructure was another key aspect of this observational study. Specific issues that were noticed at these intersections included the placement of lanes, medians, and traffic signals. Many of these observations are intersection specific, such as the turning lane shown in Figure 10 (C4). Almost every intersection has a different type of layout for their turning lanes. The majority of these turning lanes appear on road segments that are intersecting an interstate. Traffic traveling down the main roads needing to get onto the interstates must then get into one of these turning lanes. Figure 10 is an example of a particularly bad turning lane, because of its size and crookedness. This turning lane is barely large enough to contain one regular sized vehicle at best. This may potentially cause traffic to back up in the main traffic lanes simply because there is not enough space for vehicles to pull out of the main line of traffic to turn.

Figure 10: Small Turning Lane (C4) – Location 11, Fort Smith, Crawford County

Figure 11: No Turning Lane (C4) – Location 20, North Little Rock, Pulaski County

Many of the better intersections contained longer turning lanes. Some of these stretched completely under the overpass so that a long line of cars could build up in the turning lane without interfering with the main traffic flow. Figure 11 shows an example of a busy intersection that does not contain a turning lane at all (C4). Traffic is still directed by a traffic light, and the light allows for cars in the left lane to turn left. Yet, all the cars travelling along this highway must wait for these turning cars to turn before they proceed further. This in turn causes severe congestion and traffic build up along this road.

Figure 12: Median Before On Ramp (C5) – Location 1, Alma, Crawford County

Figure 13: Median in Intersection (C5) – Location 21, Pine Bluff, Jefferson County

Other than the placement of lanes, another infrastructural issue discovered during observations was the placement of medians along the roads (C5). These medians were originally designed to separate lanes and to better direct traffic. For the most, the medians do this job well. Nevertheless, these medians are also obstacles placed on the road. Some medians found around Arkansas do not even seem to have a practical purpose, as shown in Figures 12 and 13. They may be attempting to direct traffic, but they seem to get in the way of traffic more than they should. This is evident by the several tire marks and cracks along the medians showing that vehicles drive over or hit the obstacle frequently. During the observation, this was even verified by a number of cars continued to run over these medians.

Figure 14: Larger Median (C5) – Location 15, Little Rock, Pulaski County

Because medians are an obstacle in the road, there is a potential for them to be involved with many accidents. For smaller medians like those in Figures 12 and 13, drivers may end up losing control of their vehicles if they run over them. Larger medians, like the one shown in Figure 14, may present a greater danger for damage if they are run into. Also, as an unintended side effect of these large medians, which often stretch for great distances, is that they typically do not allow vehicles to turn around very easily. This type of barrier may increase the situations where drivers ignore street signs and lane markings when there is a break in the median, thus increasing illegal turning situations. Several illegal turns were witnessed during this study around roads containing medians. These medians simply block the drivers from driving in the ways they are comfortable.

Because all of the observed sites were intersection related, a large proportion of the issues discovered were traffic signal related (C6). These traffic signals are the main source of directing right-of-way situations. However, in order for these signals to direct traffic well, again they need to be clearly visible. In general, a four-way signaled intersection will contain four distinct traffic signals above each road. This is the design that most drivers come to expect. Figure 15 shows a deviation from this design that might cause some confusion to drivers. The figure shows a four-way intersection, which contains four traffic lights, but with a slightly different orientation. One of these traffic lights was placed over the corner of two roads, rather than above the road. The light is no longer directly in front of the driver, where their vision is hopefully more concentrated. Also, because the light is at an angle, there is an increased glare which makes the visibility worse.

Older intersections and street lights also appeared to cause some issues. Along with its odd orientation, the signals shown in Figure 15 appeared to be much older than most traffic signals. One observation made at this intersection was that these traffic signals were not very sturdy. The wind caused the poles, on which the signals were located, to move and bend quite frequently. This constant movement of the signal made focusing on the lights much more difficult. More modern traffic signals are reinforced with materials that prevent this movement. Another example of this issue is shown in Figure 16. This is an intersection located directly over a railroad track (C7). Due to its location, the traffic signals were suspended by a cord, rather than a pole. This situation caused even more movement with the lights. Often times during the observations, these signals would turn sideways out of view from the drivers directly ahead of the lights. When the wind blew strong enough, drivers could not even focus on the color or state of the traffic signal.

Figure 15: Traffic Signal on Corner (C6) – Location 12, Fort Smith, Crawford County

Figure 16: Traffic Signals on Cord (C6, 7) – Location 8, Dumas, Desha County

Figure 17 shows a condition where the traffic signal infrastructure did not match up with the road infrastructure (C3 and C6). This intersection was the source of a lot of poor driver behavior, simply because drivers at this intersection did not know what to do.

Figure 17: Poor Signal Layout (C3, 6) – Location 14, Little Rock, Pulaski County

The figure shows a road which is actually a two lane road. There is not a turning lane for traffic traveling in the direction pictured. Yet, above this road there is a traffic signal with four lights, implying at least three lanes, possibly four lanes. The light to the far left is actually a light designed for a turning lane, for which there is none. Cars traveling in the left lane have two traffic lights that they can potentially follow.

However, these lights often do not work well together. The light to the far left may be red to imply that the vehicles cannot turn left, whereas the light next to it may be green to show that cars can still travel straight. The problem is that there exists both a red and green light for one lane, which confuses drivers. The vehicles wanting to travel straight, and who have the right-of-way, may feel impulse to stop due to the red light. The situation is overall not a good one for communicating information to the drivers and keeping them aware.

Visibility was often limited at areas such as overpasses, access roads, and on/off ramps (C8). The overpass in Figure 17 is comparable to many in the state of Arkansas, which contain barriers directly under the bridge for support. Often times, these columns and barriers obstruct the view of the drivers along the main road, or even drivers getting on the highways from the off ramps. They restrict the visibility of the road and specifically the traffic flowing in the opposite direction. Drivers who pull out onto the road may not notice cars coming from under the overpass at high speeds, further increasing the chance of a collision. Another similar case of this issue was shown on access roads and on/off ramps. Vehicles are constantly trying to merge onto high speed highways or interstates where there is little or no room to do so (C9). Small merging lanes cause problems for vehicles that cannot gain enough speed or that do not have the capacity of entering a stream of traffic. Also, blind merging lanes or lanes that are located around curved roadways, large buildings, or other objects cause the problem of visibility for these merging vehicles.

Other minor issues observed during the study included the affect of traffic lights on driver behavior (C10 and C11). Today, many traffic signals change from green to red based on a sensor that moderates the traffic volume at each road. Because the sensors are based on the arrival of the vehicles, they are much more responsive to traffic patterns. Older traffic signals do not use this system, and use timed traffic signals instead (C10). These traffic signals have a specific duration for each red and green light at each end of the intersection, which continue to cycle throughout the day. However, this system does not take into account the volume of traffic at the intersection. A timed intersection may have no traffic at all, but the lights will still cycle through. This causes a problem, however, when traffic volumes are heavy on one road, and not on the intersecting road. The timed light will cause the main flow of traffic to start and stop when there may be no need. A signal working on a sensor system will not stop the main traffic until one or more cars arrive along the intersecting street.

Traffic light duration is another issue that affects driver behavior along traffic signals (C11). Location 11 (Fort Smith, Crawford County), Location 13 (Jacksonville, Pulaski County), and Location 20 (North Little Rock, Pulaski County) are all examples of lights which exhibit significantly long red lights along their intersections. Some of these lights range up to a minute or a minute and a half. The reason this was observed as an issue was due to the number of drivers that ran yellow and red lights at these intersections. It was observed that long red lights usually led to short green lights. Drivers who are aware of this and who are in a hurry may feel more inclined to run the light than at lights with more moderate light durations. At the Fort Smith intersection (Location 11), five different vehicles were observed running through a red light.

The final and most important portion of the observational roadway analysis was focused on driver behavior. Studying driver behavior, like many other studies in the past, was shown to be quite difficult. The observations that were made resulted in some interesting trends in driver behavior with regards to many situations. For the most part, these trends dealt with the situations that were previously mentioned. Driver behavior was found to be strongly tied to the conditions of the road and intersection, including signage, lane markings, traffic signals, and infrastructure. Some examples include the drivers' actions at the traffic signals given the amount of information that was presented to them. Figure 17 is a good example of what drivers tended to do in situations where there were no signs, no lane markings, and a poor infrastructure. The white car in this figure needed to turn left, yet they only realized that there was no turning lane after they had pulled into the middle of the intersection. This situation did not cause an accident, but shows how there could potentially be an accident. Improper turns, signals, lane changes, and stops are all common behaviors observed at this type of intersection. These issues were not observed nearly as much at the better intersections that provided drivers with a lot of information.

Figure 18: Cell Phone Distraction

Driver distraction was the most difficult aspect of these observations to measure. There exist far too many obstacles to make an accurate estimate of this driver behavior. However, some trends were found during this study. Cell phone use, as expected, was a frequent issue at every intersection (see Figure 18).

There was a small trend of cell phone usage depending on the size of the city observed. The Little Rock and Fayetteville areas of the state had several more instances of cell phone usage than smaller cities such as Van Buren, or Pine Bluff. In fact, there was more of a trend with the number of passengers throughout these cities. Smaller cities tended to have more drivers with at least one passenger, whereas the larger cities tended to have more single persons driving around. There appeared to be a direct correlation with the number of passengers and whether the driver was talking on a cell phone or not. One obstacle in the way of studying these driver distractions was the fact that several cars now have tinted windows. Drivers cannot be seen through these windows, let alone the distractions going on inside the vehicle. Observations were also difficult to make because of the weather during the week long study. The weather was cold and wet throughout most of the week, which caused many drivers to have their windows rolled up, further preventing accurate depictions of the distracting behavior.

Despite these obstacles, several instances of driver distraction were observed. These include eating, drinking, reading, texting, watching movies, smoking, searching around the vehicle, talking to passengers, talking with other drivers or pedestrians, applying makeup, and driving with a pet or animal in the front seat. These results are comparable to studies done in the past. It is still important, though, to emphasize these as problems that are still happening and are still dangerous.

CHAPTER VI – GENERAL DISCUSSION AND CONCLUSION

6.1 General Discussion

By combining the results of both the statistical and observational analyses, several insights towards the road traffic safety of Arkansas can be made. On the one hand, these two methods work together and help emphasize the aspects of road traffic safety that they have in common. Both studies have shown that road infrastructure and the road's surroundings are significant factors to the safety of the driver. Statistically, factors such as road condition, road width, horizontal curvature, and vertical curvature were all shown to be significant in determining crash frequencies. In a corresponding manner it was discovered during observations that key aspects of dangerous intersection locations were poor road conditions, narrow roads, and visibility obscured by horizontal or vertical curvature.

On the other hand, each method gives its own unique perspective of traffic safety. The limitations of the statistical study may actually be the strength of the observational study, or vice versa. But they can be used together to compensate their limitations by filling in some of the gaps found in their stand-alone results. Statistical analyses are great for determining predictability and trends between the numerous factors involved in road traffic accidents. This predictability is difficult to simply observe in any roadway setting, and therefore is a limitation of the observation analysis that is compensated in the statistical methods. Not only can the mathematical methods develop predictive models and trends, but it can also detect changes within these trends. The Arkansas data analysis showed that the County in which crashes occurred followed a significant trend for 2003, whereas it the trend failed to be significant in 2002 and 2004. Quantitative aspects like these are difficult to physically observe. These quantitative capabilities are the strong point of statistical analyses. Physical observations, however, have the increased ability of finding potential crash hazards that are not represented within the data. Information regarding sign visibility, driving patterns, and detailed driver behavior are all aspects that can be observed through a proactive method of roadway examination. The data is limited and does not contain these highly detailed factors. Human factors are always questionably recorded in historical data, because officers at the scene of an accident may not know if the driver was talking on their cell phone or falling asleep at the wheel. Many human factors are based on the actual observation of drivers in their vehicles.

Overall, the combined effects of the statistical and observational analyses show vastly superior results as compared to any of the method’s stand-alone results. Together, these methods have resulted in an improved understanding of road safety with Arkansas.

6.1.1 Recommendations

Together, the statistical results and the observational results show that the physical design of the road is essential to road traffic safety. Based on these results, it should be recommended that extra attention and care be made to the design and construction of roads throughout the state. Specifically, road and lane widths need to be made large enough for drivers to feel comfortable and also to reduce potential contact between vehicles. Vertical and horizontal curves obstruct the driver’s field of vision, and so it should also be recommended that future roads avoid as much curvature as possible. Roads should be kept up regularly to avoid poor conditions such as potholes, cracks, or even worn lane markings. Because crash frequency increases with additional traffic volume, extra lanes should be considered for roads with particularly high ADT values. Of course these infrastructural designs come at a price, but because of the number of potential lives saved as a result, the benefits should automatically outweigh the costs.

On the other hand, human factors are the cheapest to effectively change. However, it is not up to a design, but rather the person to make the changes. It was found that the number of vehicle passengers corresponds to the crash severity. The additional passengers may serve as potential distractions to the driver, which may even be comparable to cell phone conversations. It would therefore be a recommendation of this study to avoid these potential distractions when at all possible for the driver. Other simple recommendations would be for drivers to always wear their seat belts and never drink alcohol and drive.

To improve the effect that road infrastructure and road surroundings have on human factors, it should be recommended that signage, signals, and lane markings be made as clear and visible as possible. Traffic signals should be based on traffic signals in order to be more responsive to traffic flow patterns and avoid potential issues of vehicles violating the signals.

The final recommendation for this research is further described as a limitation of the current analysis. The recommendation is for the improvement in data quality collected throughout Arkansas. The reasoning for this need is described more fully in the following section containing the limitations of the study. Basically, more complete data is required to develop more complex and more meaningful models. The more improved the data becomes, the better fit the statistical models, and the better the knowledge of road traffic safety in Arkansas will become.

6.1.2 Limitations of Study

The statistcal analyses of this research were to some extent very limited. The study was performed based on the data retrieved from the crash and roadway databases for Arkansas during the periods between 2002 and 2004. Both databases were extremely vast and comprehensive, but very often incomplete or insufficient. For the Poisson and Negative Binomial models, data was used based on crash frequencies along US highways, state highways, and interstates only. This was due to the fact that crash locations were only given sufficient detail for these roads. Specifically, these were roads that contained unique values for the route, section, and logmile categories. City and county roads, which make up the vast majority of the traffic accidents throughout Arkansas, only contain a route value. Often times this is a single road, street, or avenue and does not have a standard format. One street may potentially contain four or five different variations on its name, and therefore aggregating crash frequencies along these roads are made nearly impossible. Many entries had to be left out of the analysis due to this limitation.

Another important limitation was the fact that some aspects of the crashes have yet to be recorded at all. This includes the driver who takes the fault of the collision, which could potentially help in determining the true factors involved with causing an accident. Currently, all individuals present at a car accident are recorded within the crash database. Essentially, the best results can be found using those drivers who were to blame for the collision, rather than biasing the data by including all of the innocent bystanders.

6.1.3 Future Areas of Study

One major area that should be explored in addition to the current study is the possibility of more complex statistical models. The statistical fits of the three developed models were relatively fair, however some alterations to the models may present better or more accurate results than those found in this study. Models that may present a better fit to the Arkansas crash data are the Hierarchical Logistic or Poisson models which take into account the natural nesting of passengers within vehicles within crashes. The current analysis did not use this nesting feature, which is an important application to road safety models. Also, Zero-Inflated Poisson models may also show better fits to crash frequency data with excess zeros in the response variable. Due to the limitations of the current study and the software packages available, neither of these complex models was developed for this data.

Also, due to time constraints the observational analyses only took place at intersections that were deemed dangerous based on their crash rating. A future area of study could be the examination and observation of more varieties of road intersections. This could include some good intersections, some poor, and some in between. Also, future studies should explore the option of observing different road sections. The current study focused completely on intersections, whereas future areas could focus on all road junction types. Essentially, the more locations observed, the better the resulting insights on road traffic safety become.

6.2 Conclusion

The risk of road traffic accidents, which has been increasing tremendously throughout the past decade, is a major issue that calls for improved road traffic safety measures. Arkansas, which ranks third highest in traffic fatalities nationally, is one key area that calls for an evaluation of traffic safety. As of yet there has not been an extensive study to evaluate the traffic safety trends and factors for the state. This is an important area to focus on, mainly because road traffic accidents are not consistent throughout the country, and each state has its own trends and issues (US Census Bureau, 2004). Overall trends for the United States may not be representative of the individual state. Historically, the two main methods of evaluating road traffic safety are through the use of statistical analysis of historical data and the experimental, observational based analysis of road systems. Both methods are used in determining the potential root causes of road traffic accidents, which can in turn be prevented through proper information, planning and road design.

Within each accident there are hundreds of potential factors that could have an effect on the drivers and the vehicles involved. These factors can include temporal, environmental, geographical, infrastructural, vehicular, or human elements that were present at some point during the accident. To improve road traffic safety, it is important to understand these factors, and somehow determine which of these factors have the largest effect on the accidents that occur on the roads. Statistically this is done through the use of statistical models which take historical data and use it to predict crash outcomes. Several potential models can be used to evaluate crash outcomes, and the proper choice of model is of the most importance. For the Arkansas crash and roadway data, collected between 2002 and 2004, three models were used to evaluate these potential crash factors. The Poisson and Negative Binomial regression models were used to evaluate crash frequencies along road segments of US highways, State highways, and interstates. Also, a Logistic regression model was used to predict crash and injury severities among all roads in Arkansas. The results from these tests showed that several factors are significant contributors to crash frequencies and injuries in Arkansas. Road width, number of lanes, pavement condition, horizontal curvature, and vertical curvature were all shown to be significant infrastructural factors effecting road traffic accidents, whereas the type of surface was insignificant. Geographically, the county and urban level of a location also showed no statistical significance in the models. Weather and light conditions were shown to be highly significant. Significant human factors include the use of seat belts, consumption of alcohol, the driver’s license state, age, and the number of passengers involved. Driver gender was shown to be insignificant in predicting crashes.

The second method used in evaluating road traffic safety is through the use of observational studies, which survey the entire road, its surroundings, and the actual driver behaviors. The quality of the findings from these studies is much greater than those found using mathematical models, but it is also more costly. There are several practical applications with using this model, as it gives the observer firsthand knowledge of how the road and its users operate. Along these lines, this study demonstrated a brief evaluation of roads throughout Arkansas. In particular, the study focused on intersections along road segments that have ranked highly in crash occurrences. Several aspects of the road and driver behavior were analyzed at these intersections, including the infrastructure, signage, signals, and driver reactions to road and its surroundings. In general, it was found that many dangerous locations were due to poor signage, worn lane markings, roadway obstacles, and unclear right-of-way cues. In summary, road traffic safety in the state of Arkansas was examined and evaluated using the current methods of statistical and observational analyses. These results give important insights and highlight particular areas of driver behavior and roadway characteristics that effect road traffic accidents throughout the state. With the knowledge of these results and their limitations, steps can now be taken to further study these key areas and begin the growing need for road traffic safety in Arkansas.

References

(1979-2006). Highway Statistics Publications. *US Department of Transportation – Federal Highway Administration*. http://www.fhwa.dot.gov/policy/ohpi/hss/hsspubs.htm.

(2000-2006). *Traffic Safety Facts.* National Highway Traffic Safety Administration, U.S. Department of Transportation. Washington, D.C.

(2008). Cellular Telecommunications and Internet Association: Industry Information. http://www.ctia.org/media/industry_info/index.cfm/AID/10319

Abdel-Aty, M., Radwan, A. E. (2000). Modeling Traffic Accident Occurrence and Involvement. *Accident Analysis & Prevention. 32(5)*, 633-642.

Abdel-Aty, M., Keller, J., Brady, P.A. (2005). Analysis of the Types of Crashes at Signalized Intersections Using Complete Crash Data and Tree-based Regression. In *Proceedings of the TRB 84th Annual Meeting, Washington, DC*.

Ahmad, H.Z.B., Rahman, M.Y.B.A. (2003). Traffic Calming Approaches to Road Safety. In *Proceedings - Conference of the Australian Road Research Board. 27*, 1623-1638.

Al-Ghamdi, A. (2002). Using Logistic Regression to Estimate the Influence of Accident Factors on Accident Severity. *Accident Analysis & Prevention. 34 (6)*, 729-741.

Allsop, R. (1997). Road Safety Audit and Safety Impact Assessment. *European Transport Safety Council, Brussles.*

Anastasopoulos, P.C., Tarko, A.P., Mannering, F.L. (2007). Tobit Analysis of Vehicle Accident Rates on Interstate Highways. *Accident Analysis & Prevention. 40(2),* 768-775.

Baulk, S.D., Biggs, S.N., Reid, K.J., van den Heuvel, C.J., Dawson, D. (2007). Chasing the Silver Bullet: Measuring Driver Fatigue Using Simple and Complex Tasks. *Accident Analysis & Prevention. 40(1),* 396-402.

Berhanu, G. (2004). Models Relating Traffic Safety with Road Environment and Traffic Flows on Arterial Roads in Addis Ababa. *Accident Analysis & Prevention*. *33(1),* 99-109.

Carson, J., Mannering, F. (2000). The Effect of Ice Warning Signs on Ice-Accident Frequencies and Severities. *Accident Analysis & Prevention*. *36(5),* 697-704.

Chang, B.-H., Graham, J.D. (1993). A New Method for Making Interstate Comparisons of Highway Fatality Rates. *Accident Analysis & Prevention. 25(1)*, 85-90.

Chang, L., Wang, H. (2006). Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques. *Accident Analysis & Prevention. 38(5),* 1019-1027.

de Leur, P., Sayed, T. (2003). A Framework to Proactively Consider Road Safety within the Road Planning Process. *Can. J. Civ. Eng. 30*, 711-9.

Escalera, A., Moreno, L.E., Salichs, M.A., Armingol, J.M. (1997). Road Traffic Sign Detection and Classification. *IEEE Transactions on Industrial Electronics. 44(6).* 848-859.

Evans, L. (2003). A New Traffic Safety Vision for the United States. *American Journal of Public Health. 93(9).* 1384-1386.

Evans, L. (2004). *Traffic Safety*. Science Serving Society: United States of America.

Flahaut, B. (2003). Impact of Infrastructure and Local Environment on Road Unsafety Logistic Modeling with Spatial Autocorrelation. *Accident Analysis & Prevention*. *36(6),* 1055-1066.

Gårder, P. (2005). Segment Characteristic and Severity of Head-On Crashes on Two-Lane Rural Highways in Maine. *Accident Analysis & Prevention*. *38(4),* 652-661.

Goodman, M., Tijerina, L. Bents, F.D., Weirwille, W.W. (1999). Using Cellular Telephones in Vehicles: Safe or Unsafe? *Transportation Human Factors. 1(1),* 3-42.

Greibe, P. (2003). Accident Prediction Models for Urban Roads. *Accident Analysis & Prevention, 35(2),* 273-285.

Hakim, D. (2003, November 27). Once World Leader in Traffic Safety, U.S. Drops to No. 9. *New York Times.*

Horne, D.A. (1999). Road Safety Audits: The Federal Highway Administration Perspective. *Transportation Frontiers for the Next Millennium*

Horrey, W.J., Wickens, C.D. (2006). Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques. *Human Factors 48 (1),* 196–205.

Hutchings, C., Knight, S., Reading, J.C. (2003). The Use of Generalized Estimating Equations in the Analysis of Motor Vehicle Crash Data. *Accident Analysis & Prevention*. *35(1),* 3-8.

Janson, B.N., Karimkhani, E. (2001). Cost Effectiveness of Infrastructure Improvement versus Behavioral Programs to Reduce Traffic Accidents. Draft Literature Review. Transportation Research Center: Denver, Colorado.

Jones, A.P., Jørgensen, S.H. (2003). The Use of Multilevel Models for the Prediction of Road Accident Outcomes. *Accident Analysis & Prevention. 35(1),* 59-69.

Karlaftis, M.G., Golias, I. (2002). Effects of Road Geometry and Traffic Volumes on Rural Roadway Accident Rates. *Accident Analysis & Prevention. 34 (3),* 357-365.

Kim, D.-G.,Washington, S., Oh, J. (2006). Modeling crash outcomes: New insights into the effects of covariates on crashes at rural intersections. *Journal of Transportation Engineering. 132(4),* 282-292.

Kim, D.-G., Lee, Y., Washington, S., Choi, K. (2007). Modeling Crash Outcome Probabilities at Rural Intersections: Application of Hierarchical Binomial Logistic Models. *Accident Analysis & Prevention. 39(1),* 125-34.

Klauer, S.G., Neale, V.L., Dingus, T.A., Ramsey, D.J., Sudweeks, J.D. (2006). Driver Inattention: A Contributing Factor to Crashes and Near-Crashes. Proceedings from *Human Factors and Ergonomics Society Annual Meeting.* 1922-1926.

Lamble, D., Kauranen, T., Laakso, M., Summala, H. (1999). Cognitive Load and Detection Thresholds in Car Following Situations: Safety implications for using mobile (Cellular) Telephones While Driving. *Accident Analysis and Prevention. 31(6),* 617–623.

Lee, J., Mannering, F. (2002). Impact of Roadside Features on the Frequency and Severity of Run-Off-Roadway Accidents: An Empirical Analysis. *Accident Analysis & Prevention*. *34(2),* 149-161.

Lewis-Beck, M.S. (1980). *Applied Regression: An Introduction.* University of Iowa: Sage Publications Inc.

Lord, D., Persaud, B.N. (2000). Accident Prediction Models With and Without Trend: Applications of the Generalized Estimating Equations Procedure. *Transportation Research Record. 1717.* 102-108.

Lord, D., Washington, S.P., Ivan, J.N. (2005). Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis & Prevention. 37(1),* 35-46.

Mashaw, J. L., Harfst, D.L. (1991). The Struggle for Auto Safety. *Harvard Journal of Law & Technology. 4, 307*-312.

McCormick, J. (2003). The Human/Machine Interface. *Automotive Industries. 183(1).* 39-41.

Milton, J.C., Shankar, V.N., Mannering, F.L. (2008). Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. *Accident Analysis & Prevention. 40(1),* 260-266.

Nantulya, V.M., Reich, M.R. (2002). The Neglected Epidemic: Road Traffic Injuries in Developing Countries. *British Medical Journal. 324*, 1139-1141.

Neyens, D.M., Boyle, L.N. (2008). The Influence of Driver Distraction on the Severity of Injuries Sustained by Teenage Drivers and Their Passengers. *Accident Analysis & Prevention. 40(1),* 254-259.

Noland, R.B. (2003). Traffic Fatalities and Injuries: The Effect of Changes in Infrastructure and Other Trends. *Accident Analysis & Prevention*. *35(4),* 599-611.

Noy, Y.I. (1997). Human Factors in Modern Traffic Systems. *Ergonomics. 40(10),* 1016-1024.

Peden, M.M. (2001). *Proceedings of WHO Meeting to Develop a 5-year Strategy for Road Traffic Injury Prevention*. Geneva: World Health Organization.

Peden, M.M. (2004). *World Report on Road Traffic Injury Prevention: Summary*. Geneva: World Health Organization.

Petridou, E., Moustaki, M. (2000). Human Factors in the Causation of Road Traffic Crashes. *European Journal of Epidemiology. 16,* 819–826.

Pickering, C.A. (2004). Driver Distraction Trends and Issues. *IEE Computing & Control Engineering.* 28-30.

Poch, M., Mannering, F. (1996). Negative Binomial Analysis of Intersection-Accident Frequencies. *Journal of Transportation Engineering. 122(2),* 105-113.

Rasmussen, J., Nixon, P., Warner, F. (1990). Human Error and the Problem of Causality in Analysis of Accidents. *Philosophical Transactions of the Royal Society of London. 327(1241),* 449-462.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association Between Cellular-Telephone Calls and Motor Vehicle Collisions. *New England Journal of Medicine. 336,* 453–458.

Ricketts, T.C., Johnson-Webb, K.D., Taylor, P. (1998). Definitions of Rural: A Handbook for Health Policy Makers and Researchers. *American Journal of Public Health. 95(7).* 1149-1155.

Rumar, K. (1988). Collective Risk But Individual Safety. *Ergonomics. 31(4),* 507-518.

Sabey, B., Taylor, H. (1980). The Known Risks We Run: The Highway. Transport and Road Research Laboratory (TRRL). Supplementary Report. Crowthorn, UK.

Saccomanno, F.F., Nassar, S.A., Shortreed, J.H. (1996). Reliability of Statistical Road Accident Injury Severity Models. *Transportation Research Record. 1542,* 14-23.

Salvucci, D.D., Markley, D., Zuber, M., Brumby, B.P. (2007). iPod Distraction: Effects of Portable Music-Player Use on Driver Performance. Proceedings from *Human Factors in Computing Systems: Task and Attention.* 243-250.

Salvucci, D.D. (2006). Modeling Driver Behavior in a Cognitive Architecture. *Human Factors. 48(2),* 362-380.

Scurfield, R. (2002). Poor Quality Data Are Major Obstacle to Improving Road Safety, Says World Bank. *British Medical Journal. 324,* 1116.

Shankar, V., Chayanan, S., Sittikariya, S., Shyu, M.-B., Juvva, N.K., Milton, J. C. (2004). Marginal Impacts of Design, Traffic, Weather, and Related Interactions on Roadside Crashes. *Transportation Research Record. 1897,* 156-163.

Sheridan, T. (2004). Driver Distraction Form a Control Theory Perspective. *Human Factors. 46(4),* 587–599.

Strayer, D.L., Drews, F.A., Albert, R.W., Johnston, W.A. (2001). Cell Pone Induced Perceptual Impairments During Simulated Driving. Department of Psychology: University of Utah.

Sundeen, M. (2007). Cell Phones and Highway Safety. *National Conference of State Legislatures – 2006 State Legislative Update.*

Tijerina, L. (2000). Issues in the Evaluation of Driver Distraction Associated with In-Vehicle Information and Telecommunications Systems. *Transportation Research Inc.*

Trbovich, P., Harbluk, J.L. (2003). Cell Phone Communication and Driver Visual Behavior: The Impact of Cognitive Distraction. Proceedings from *Human Factors and Ergonomics Society Annual Meeting.* 1980-1983

US Census Bureau. (2004). State Rankings - Statistical Abstract of the United States. http://www.census.gov/compendia/statab/ranks/rank39.htm

US Census Bureau. (2008). International Data Base: World Population Information. http://www.census.gov/ipc/www/idb/worldpopinfo.html

Van Beeck, E.F., Mackenbach, J.P., Looman, C.W.N., Krunst, A.E. (1991). Determinants of Traffic Accident Mortality in the Netherlands: A Geographic Analysis. *International Journal of Epidemiology. 20(3).* 698-706.

Wang X., Abdel-Aty, M., Brady, P.A. (2006). Crash Estimation at Signalized Intersections: Significant Factors and Temporal Effect. In *Transportation Research Record: Journal of the Transportation Research Board, No 1953,* Transportation Research Board of the National Academies, Washington, D.C. 10-20.

Weirwille, W., Tijerina, L., Kiger, S., Rockwell, T., Lauber, E., Bittner Jr., A. (1996). Heavy Vehicle Driver Workload Assessment. Heavy Vehicle Driver Workload Assessment – Executive Summary. (Report DOT HS 808 466). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Wogalter, M.S., Mayhorn, C.B. (2005). Perceptions of Driver Distraction by Cellular Phone Users and Nonusers. *Human Factors. 47(2).* 455-467.

Yamamoto, T., Shankar, V.N. (2004). Bivariate Ordered-Response Probit Model of Driver's and Passenger's Injury Severities in Collisions with Fixed Objects. *Accident Analysis & Prevention*. *36(5),* 869-876.

Yau, K., Lo, H.P., Fung, S. (2006). Multiple-Vehicle Traffic Accidents in Hong Kong. *Accident Analysis & Prevention. 38(6),* 1157-1161.

Zhang, W., Huang, Y.-H., Roetting, M., Wang, Y., Wei, H. (2006). Driver's Views and Behaviors About Safety in China—What Do They NOT Know About Driving? *Accident Analysis & Prevention. 38(1),* 22-27.

APPENDIX A: SPSS Coding and Output

(1) Poisson Regression 2002 Data

```
DATASET ACTIVATE DataSet1.
* Generalized Linear Models.
GENLIN Frequency WITH DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID REC
ONS TERAN RDWID PAVCO
  /MODEL DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID RECONS TERAN R
DWID PAVCO INTERCEPT=YES
 DISTRIBUTION=POISSON LINK=LOG
  /CRITERIA SCALE=1 COVB=MODEL PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-
012 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FUL
   L
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION CORB.
DataSet1] C:\Documents and Settings\Jacob Mercer\Desktop\SPSS\Poisson2002I.sav
```

(2) Poisson Regression: 2003 Data

```
DATASET ACTIVATE DataSet2.
* Generalized Linear Models.
GENLIN Frequency WITH DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID REC
ONS TERAN RDWID PAVCO
  /MODEL DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID RECONS TERAN R
DWID PAVCO INTERCEPT=YES
 DISTRIBUTION=POISSON LINK=LOG
  /CRITERIA SCALE=1 COVB=MODEL PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-
012 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FUL
   L
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION CORB.
DataSet1] C:\Documents and Settings\Jacob Mercer\Desktop\SPSS\Poisson2003I.sav
```

(3) Poisson Regression: 2004 Data

```
DATASET ACTIVATE DataSet3.
* Generalized Linear Models.
GENLIN Frequency WITH DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID REC
ONS TERAN RDWID PAVCO
  /MODEL DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID RECONS TERAN R
DWID PAVCO INTERCEPT=YES
 DISTRIBUTION=POISSON LINK=LOG
  /CRITERIA SCALE=1 COVB=MODEL PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-
012 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FUL
   L
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION CORB.
DataSet1] C:\Documents and Settings\Jacob Mercer\Desktop\SPSS\Poisson2004I.sav
```

(3) Negative Binomial Regression: 2002

```
DATASET ACTIVATE DataSet1.
* Generalized Linear Models.
GENLIN Frequency WITH DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID REC
ONS TERAN RDWID PAVCO
  /MODEL DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID RECONS TERAN R
DWID PAVCO INTERCEPT=YES
 DISTRIBUTION=NEGBIN(1) LINK=LOG
```

```
  /CRITERIA SCALE=1 COVB=MODEL PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-
012 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FUL
   L
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION CORB.
DataSet1] C:\Documents and Settings\Jacob Mercer\Desktop\SPSS\ NegBin2002I.sav
```

(5) Negative Binomial Regression: 2003

```
DATASET ACTIVATE DataSet2.
* Generalized Linear Models.
GENLIN Frequency WITH DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID REC
ONS TERAN RDWID PAVCO
  /MODEL DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID RECONS TERAN R
DWID PAVCO INTERCEPT=YES
 DISTRIBUTION=NEGBIN(1) LINK=LOG
  /CRITERIA SCALE=1 COVB=MODEL PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-
012 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FUL
   L
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION CORB.
DataSet1] C:\Documents and Settings\Jacob Mercer\Desktop\SPSS\NegBin2003I.sav
```

(6) Negative Binomial Regression: 2004

```
DATASET ACTIVATE DataSet3.
* Generalized Linear Models.
GENLIN Frequency WITH DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID REC
ONS TERAN RDWID PAVCO
  /MODEL DSTNO CONTY Length POPGR URBAN FNCLA ADT NOLAN SURTY LNWID RECONS TERAN R
DWID PAVCO INTERCEPT=YES
 DISTRIBUTION=NEGBIN(1) LINK=LOG
  /CRITERIA SCALE=1 COVB=MODEL PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-
012 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FUL
   L
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION CORB.
DataSet1] C:\Documents and Settings\Jacob Mercer\Desktop\SPSS\NegBin2003I.sav
```

(7) Logistic Regression: All Variables

```
LOGISTIC REGRESSION VARIABLES CRASHSEVERITY_BIN
  /METHOD=ENTER YEAR ATMOSPHERICCONDITIONS LIGHTCONDITIONS RURALURBAN ROADSURF
ACECONDITION ROADSYSTEM ROADWAYALIGNMENT ROADWAYPROFIL
   E WEEKDAY NUMBERINVOLVED ALCOHOLINVOLVED RESTRAINTCODE SEX AGE INJURYSEVERIT
Y_ORD LICENSESTATE COUNTYNUMBER
  /PRINT=CORR
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
[DataSet1] C:\Documents and Settings\Jacob Mercer\Desktop\SPSS\Logistic Regression.sav
```

(8) Logistic Regression: Correlation Removed

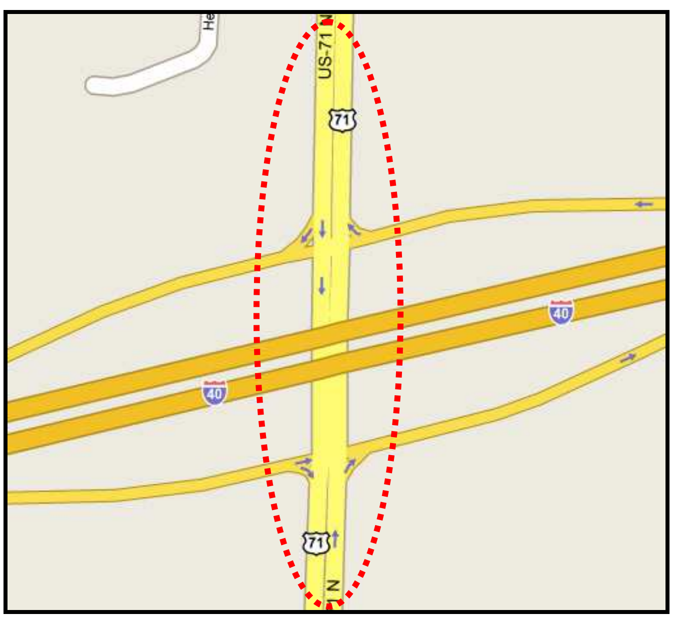
```
LOGISTIC REGRESSION VARIABLES CRASHSEVERITY_BIN
  /METHOD=ENTER YEAR LIGHTCONDITIONS RURALURBAN ROADSURFACECONDITION ROADSYSTE
M ROADWAYALIGNMENT ROADWAYPROFILE WEEKDAY NUMBERINVOLV
   ED ALCOHOLINVOLVED RESTRAINTCODE SEX AGE LICENSESTATE COUNTYNUMBER
  /PRINT=CORR
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

APPENDIX B: Observation Locations

**All Locations Provided by Google Maps

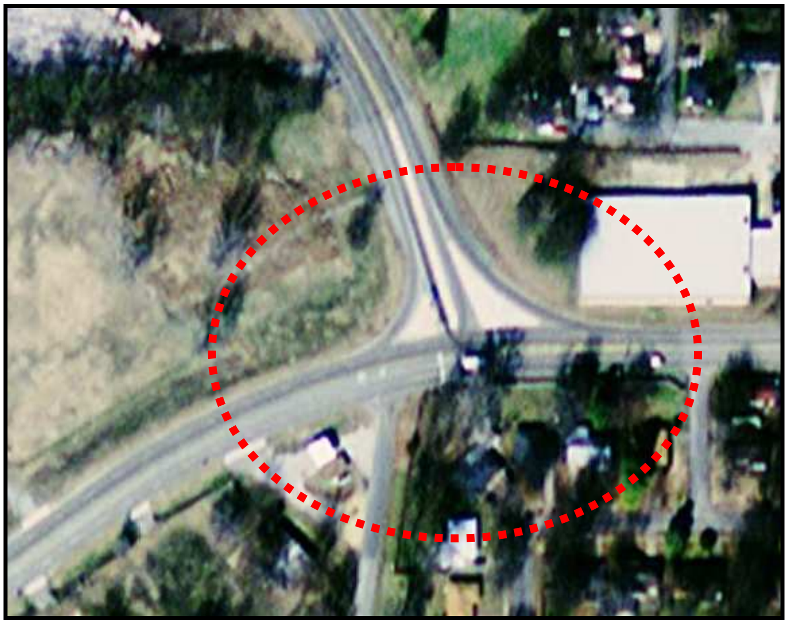
Location 1: Alma, Crawford County

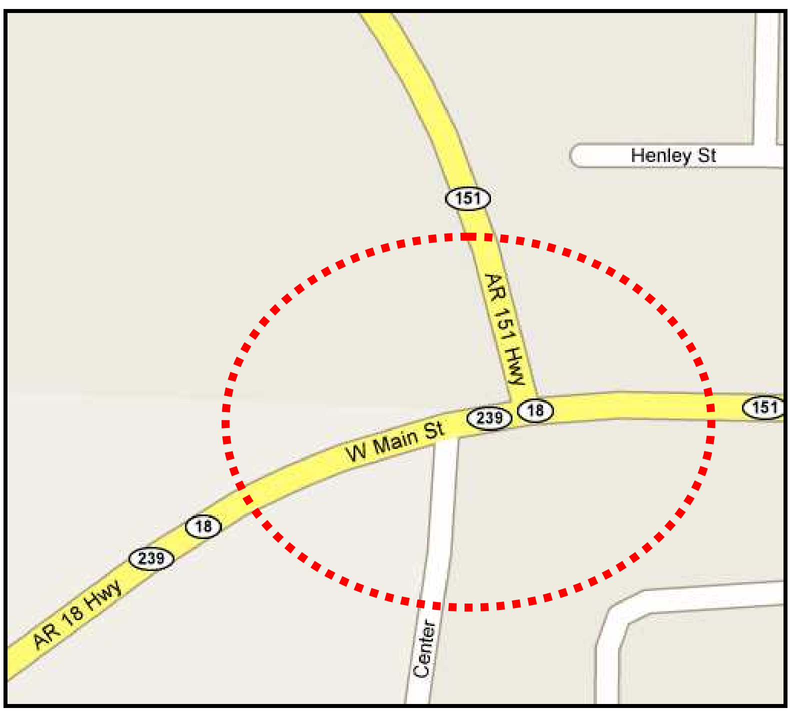
- **Intersection of US Highway 71 and Interstate 40**
- **Observations made along US Highway 71**
- **Details: Good visibility along roadway; Turning lane for intersecting street not long enough for more than one vehicle; Median within the roadway designed for the turning lane acts as an obstacle**



Location 2: Blytheville, Mississippi County

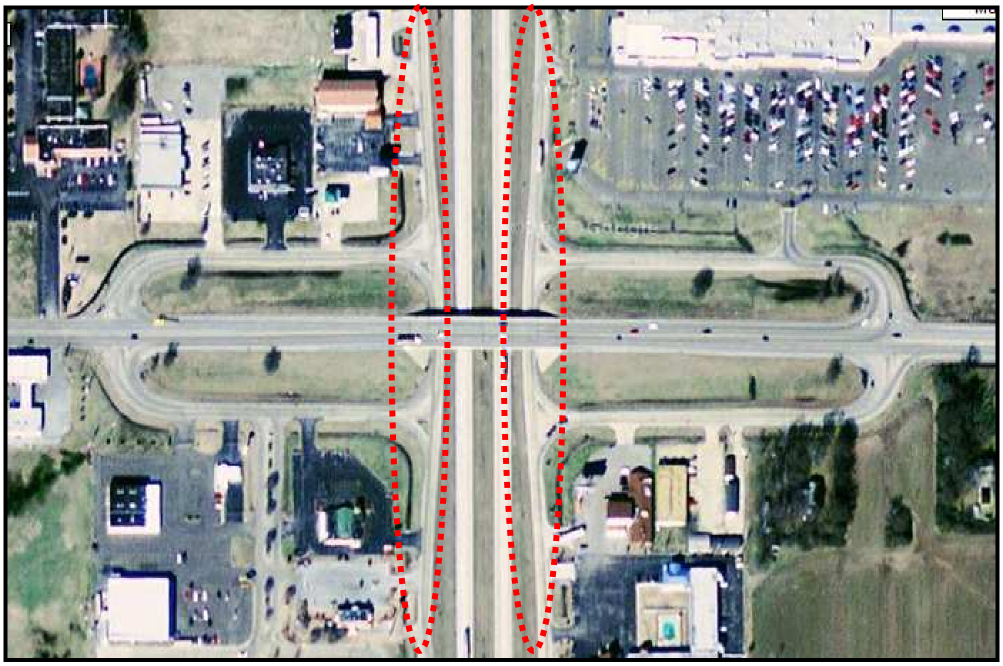
- **Intersection of State Highway 18 and State Highway 151**
- **Observations made along State Highway 18**
- **Details: Good signage; Poor visibility due to horizontal curvature of road; Several medians surround the traffic signal; Intersecting roads are not perpendicular; Good merging conditions with separate lanes; Near to airport**

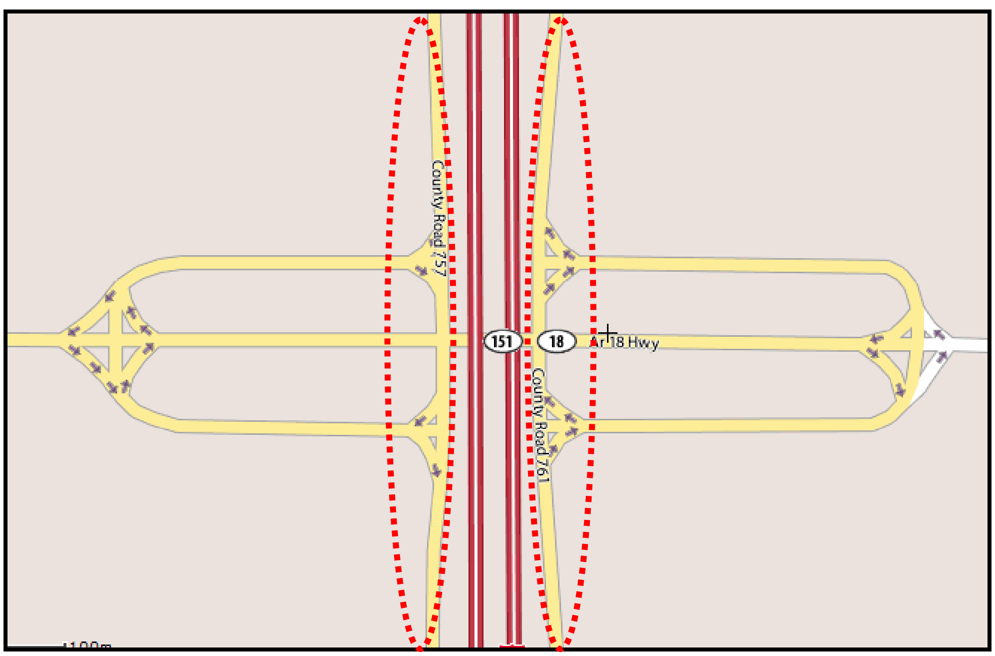




Location 3: Blytheville, Mississippi County

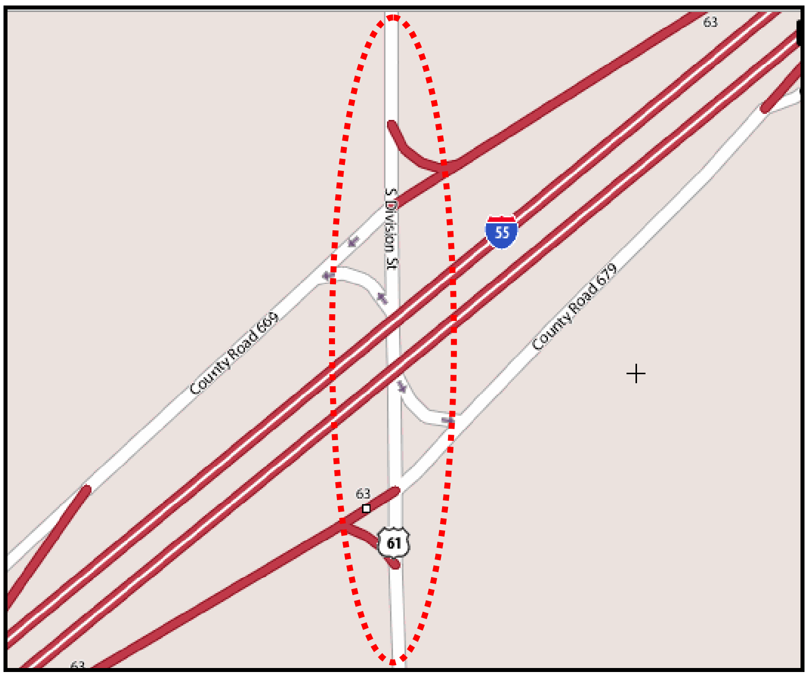
- **Intersection of State Highway 18 and Interstate 55**
- **Observations made on access roads alongside Interstate 55**
- **Details: Dangerous merging along access roads and interstate; Poor visibility surrounding the overpass; Several medians along access road and interstate on/off ramps**





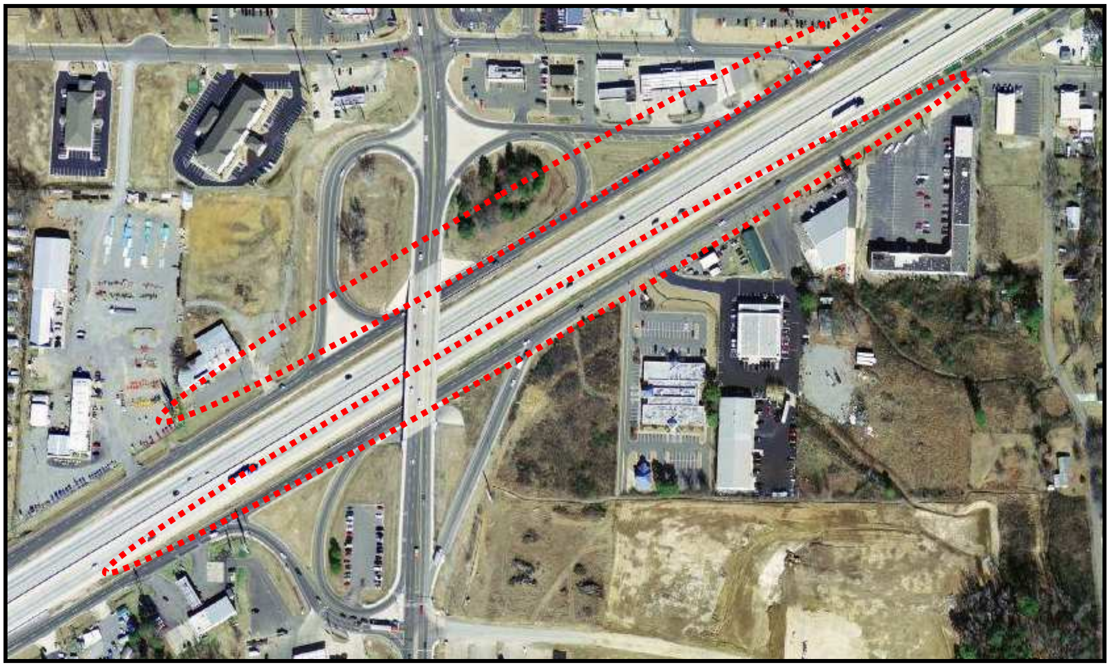
Location 4: Blytheville, Mississippi County

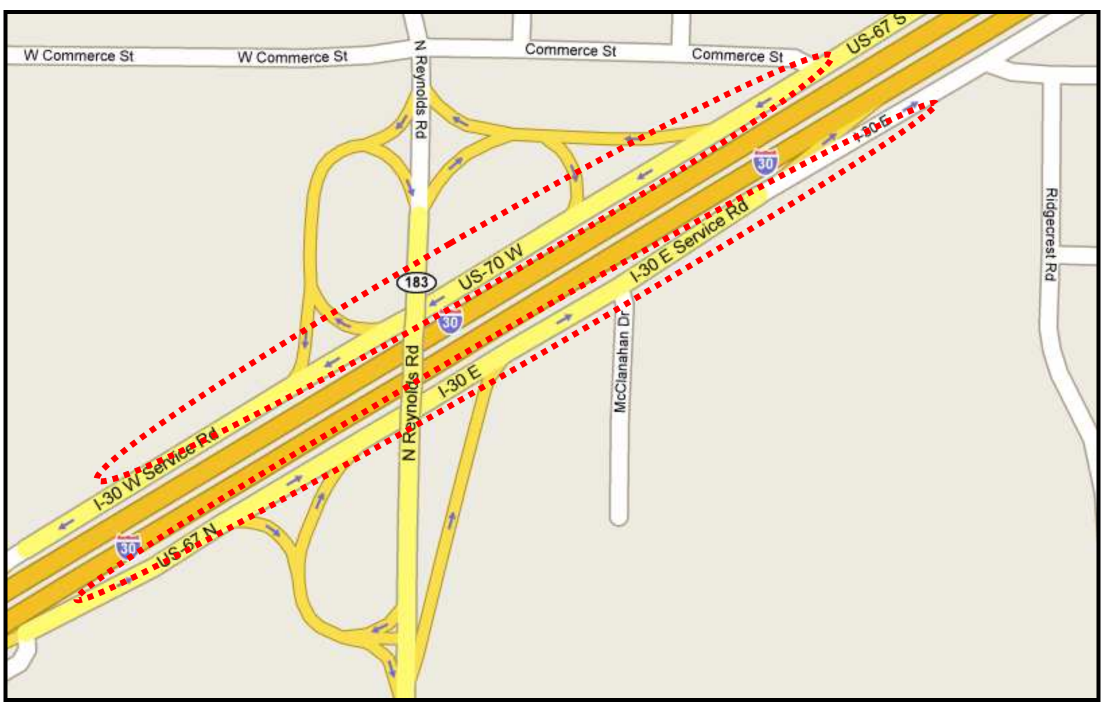
- **Intersection of US Highway 61 and Interstate 55**
- **Observations made along US Highway 61**
- **Details: Good signage; Good lane markings; Good visibility; No traffic signal located at on/off ramps**



Location 5: Bryant, Saline County

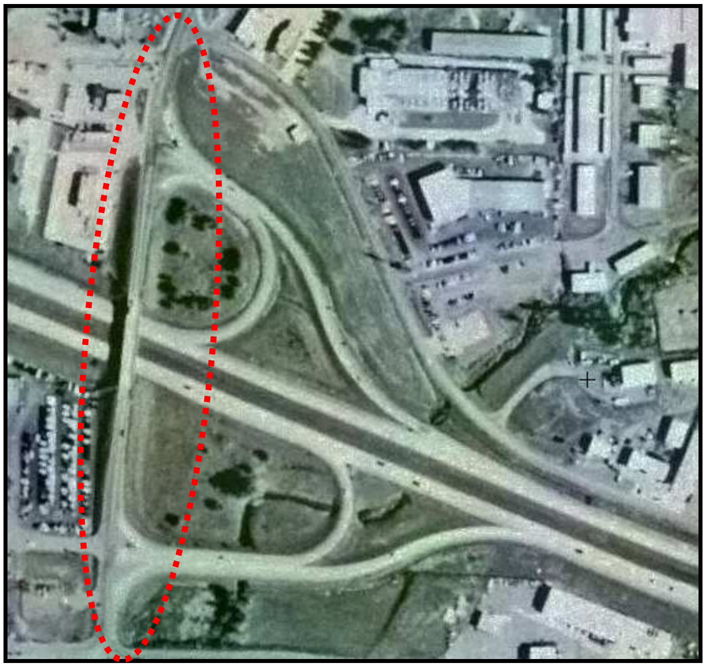
- **Intersection of State Highway 183 and Interstate 30**
- **Observations made on access roads alongside Interstate 30**
- **Details: Dangerous merging along access roads and interstate; Several intersecting streets along access road; High amounts of cell phone use observed; Failure to yield also observed**

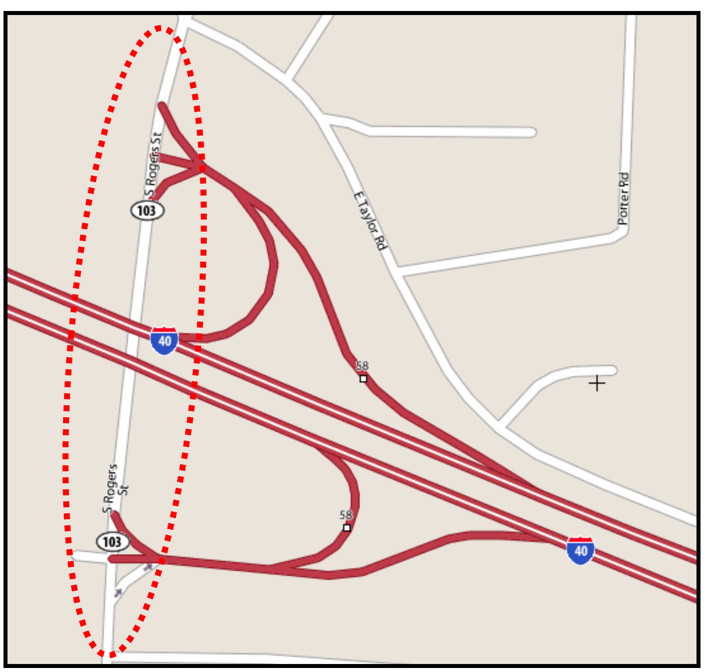




Location 6: Clarksville, Johnson County

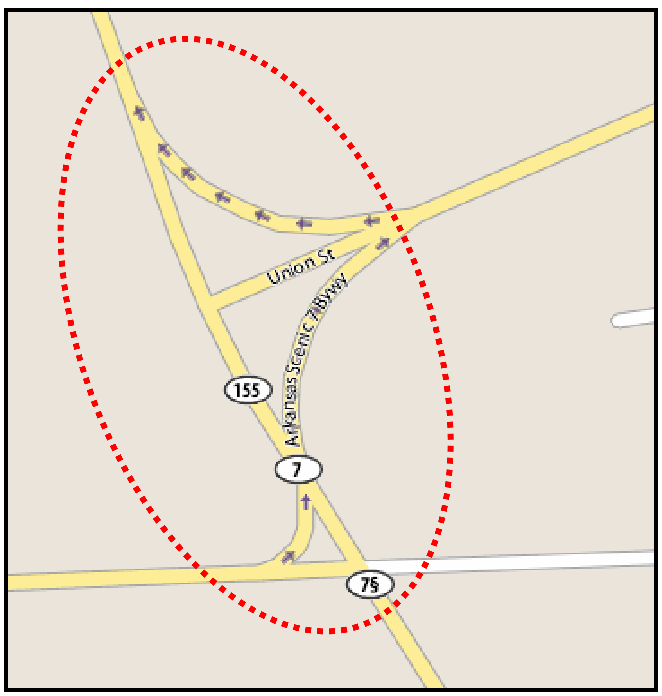
- **Intersection of State Highway 103 and Interstate 40**
- **Observations made along State Highway 103**
- **Details: Poor lane markings along on/off ramps; Poor turning lanes along entire road; Narrow roads and turning lanes; Short yellow light durations for turning vehicles; Poor traffic signal infrastructure**





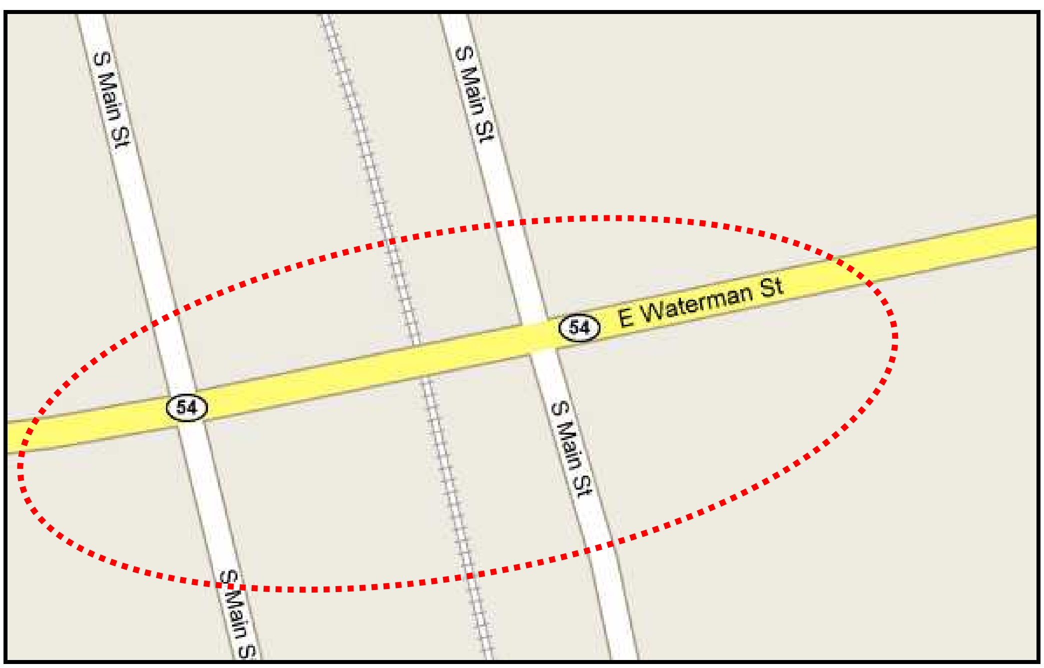
Location 7: Dardanelle, Yell County

- **Intersection of State Highway 22 and State Highway 7**
- **Observations made along State Highway 22**
- **Details: Large medians surrounding as well as along the entire roadway; Poor signage and turning lane markings; Poor visibility due to the width of the intersection and ramps**



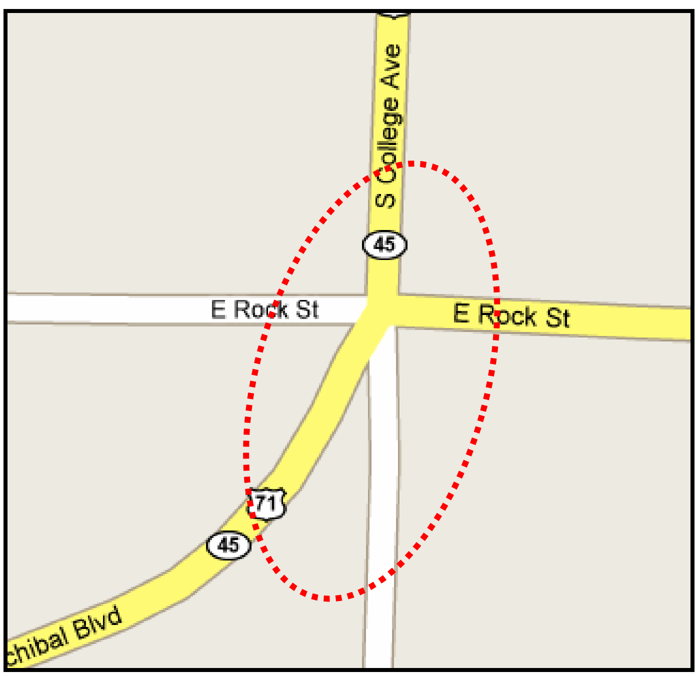
Location 8: Dumas, Desha County

- **Intersection of State Highway 54 and Main Street**
- **Observations made along State Highway 54**
- **Details: Intersecting roads separated by an active railroad; Crosswalks along each road; Downtown area; Traffic signals suspended from cords, which sway in windy conditions and are not visible**



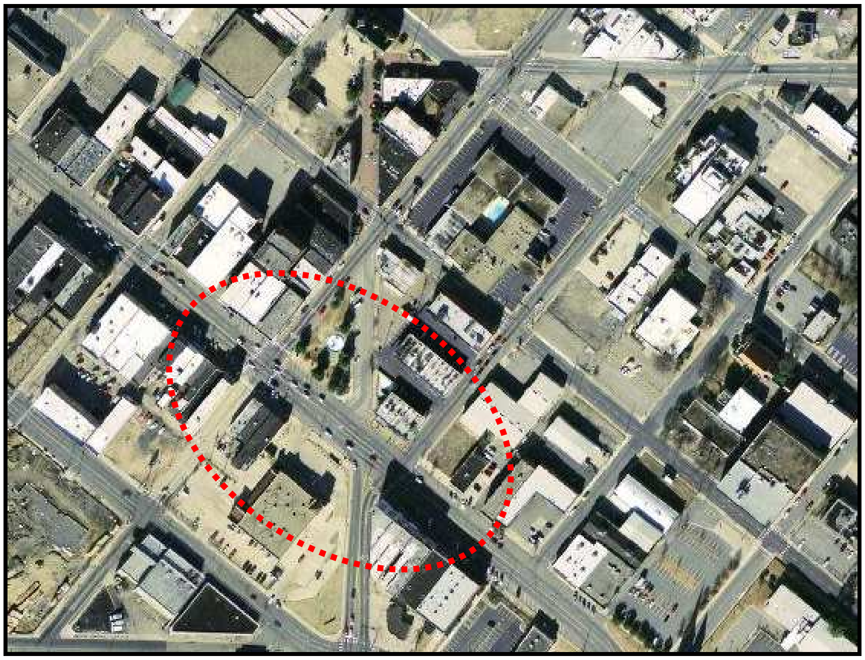
Location 9: Fayetteville, Washington County

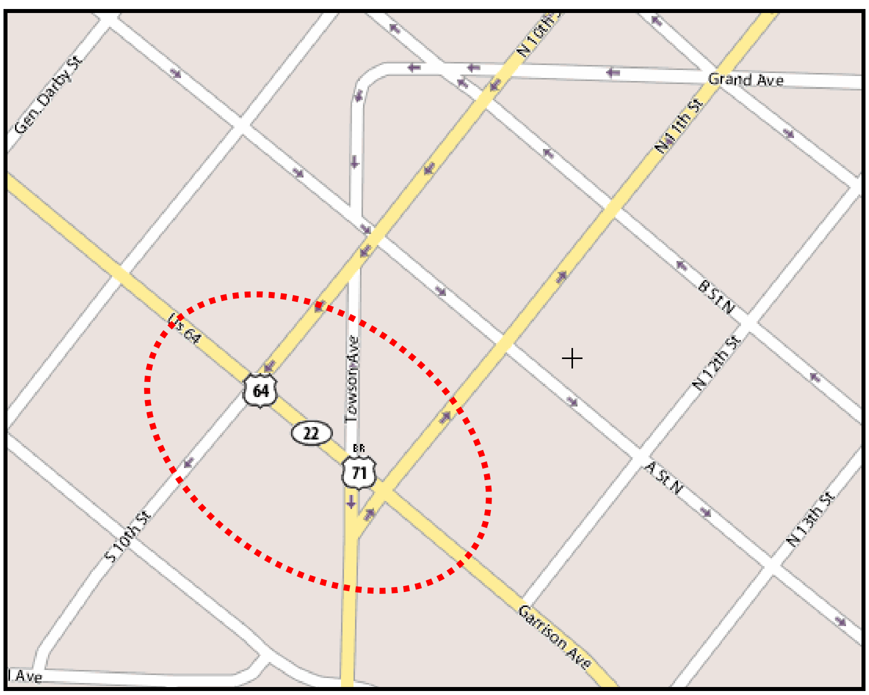
- **Intersection of US Highway 71B and Rock Street**
- **Observations made along US Highway 71B**
- **Details: Poor visibility; Vertical and horizontal curvature at the top of a hill; No turning lane along the entire road; No traffic signal; Failure to yield and improper turning prevalent along this road**



Location 10: Fort Smith, Crawford County

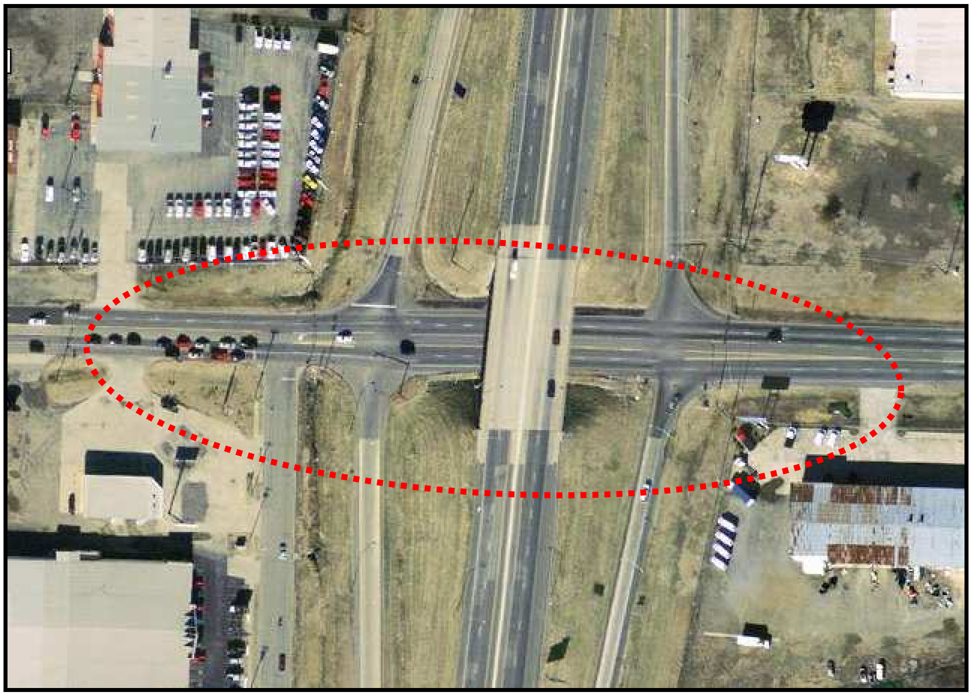
- **Intersection of US Highway 71 and US Highway 64**
- **Observations made along US Highway 64**
- **Details: Traffic signals are on timers and are synchronized with each other; One-way traffic along most of the roads; Poor signage; Poor merging along south part of road; Several crosswalk areas located across the road**

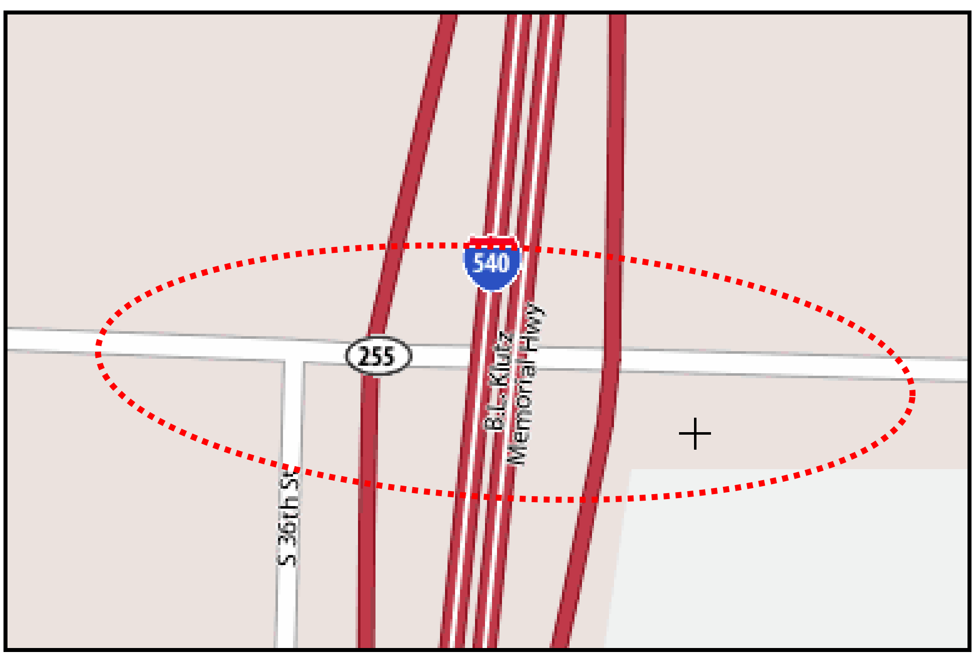




Location 11: Fort Smith, Crawford County

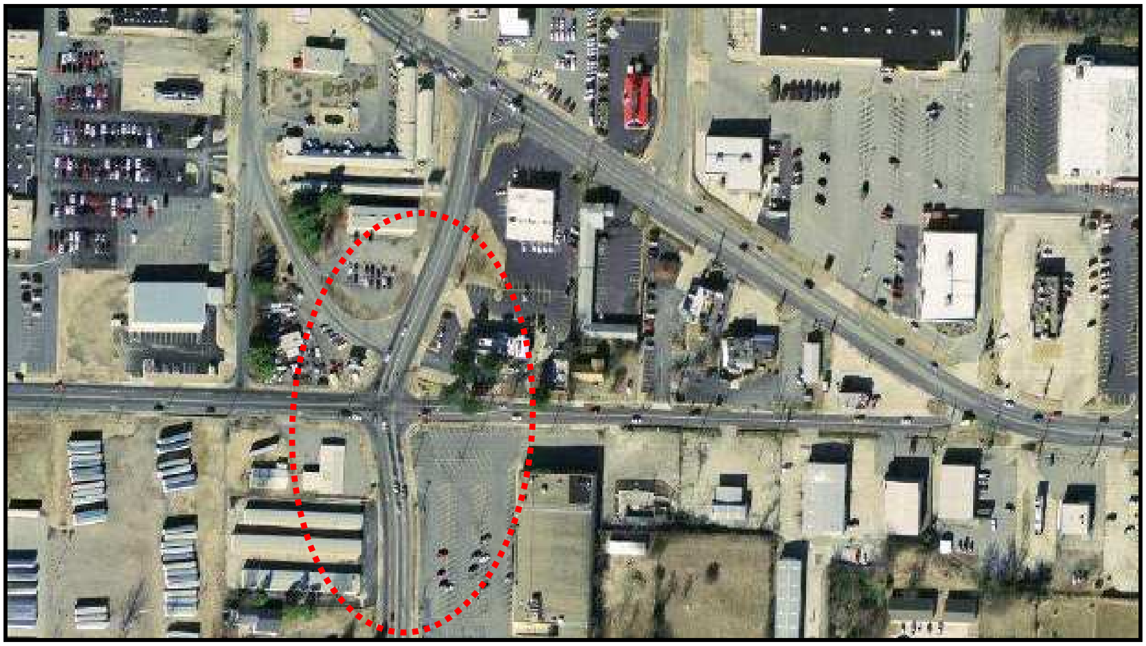
- **Intersection of State Highway 255 and Interstate 540**
- **Observations made along State Highway 255**
- **Details: Poor traffic signal durations with long red lights and short green lights; Parallel turning lanes under overpass; Short, crooked turn lane at on ramp; Improper turning and running yellow lights observed at intersection**

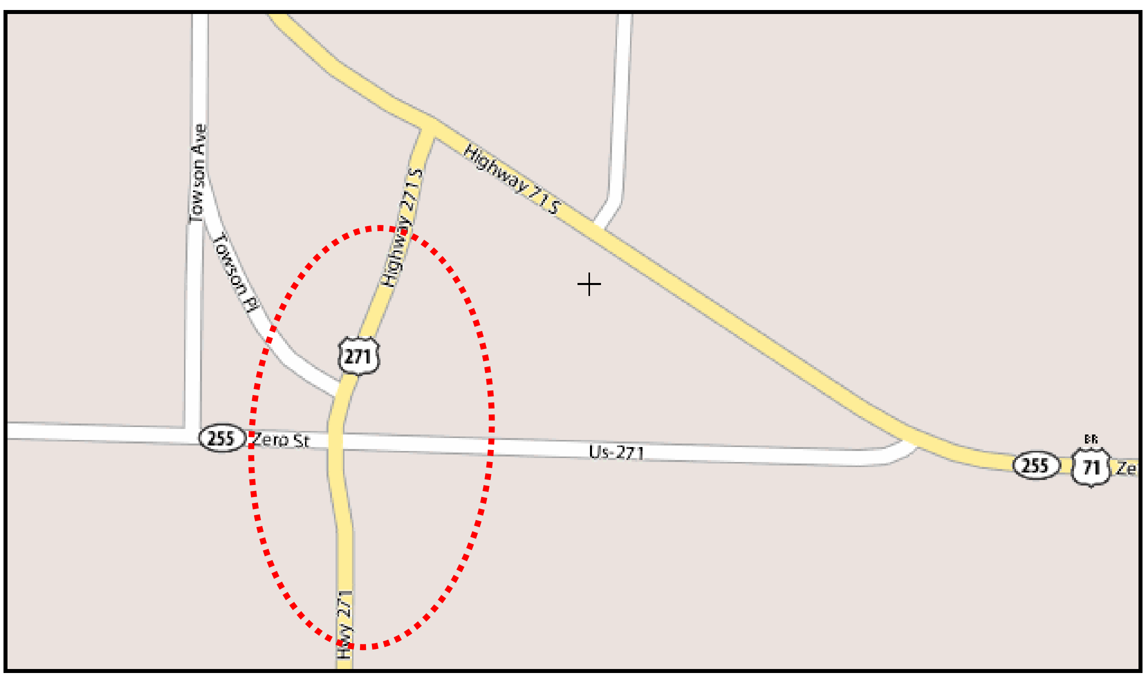




Location 12: Fort Smith, Crawford County

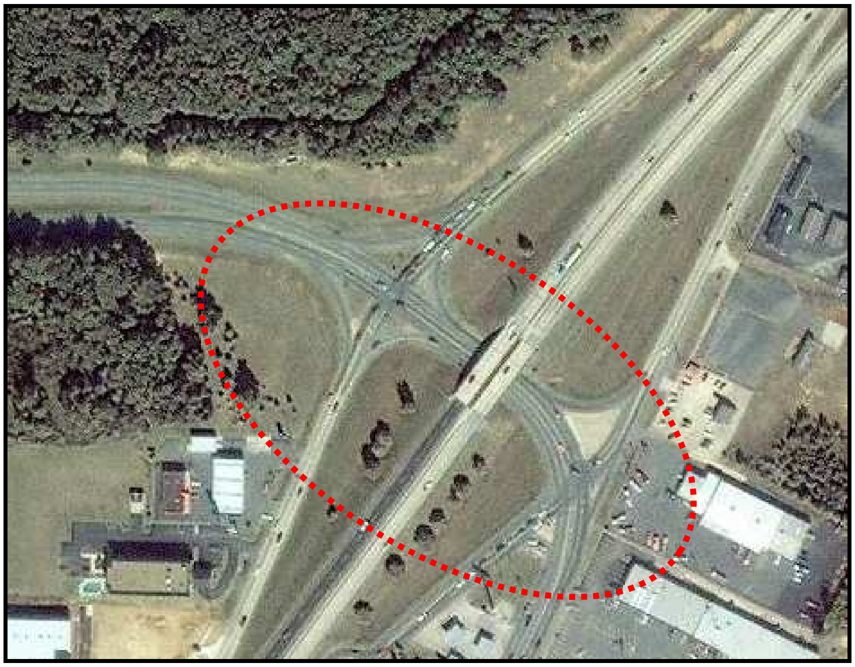
- **Intersection of State Highway 255 and US Highway 271**
- **Observations made along US Highway 271**
- **Details: Angled intersection; Traffic signals are crooked and sway in the wind; Lane markings only; No crosswalks along road, despite high pedestrian traffic**

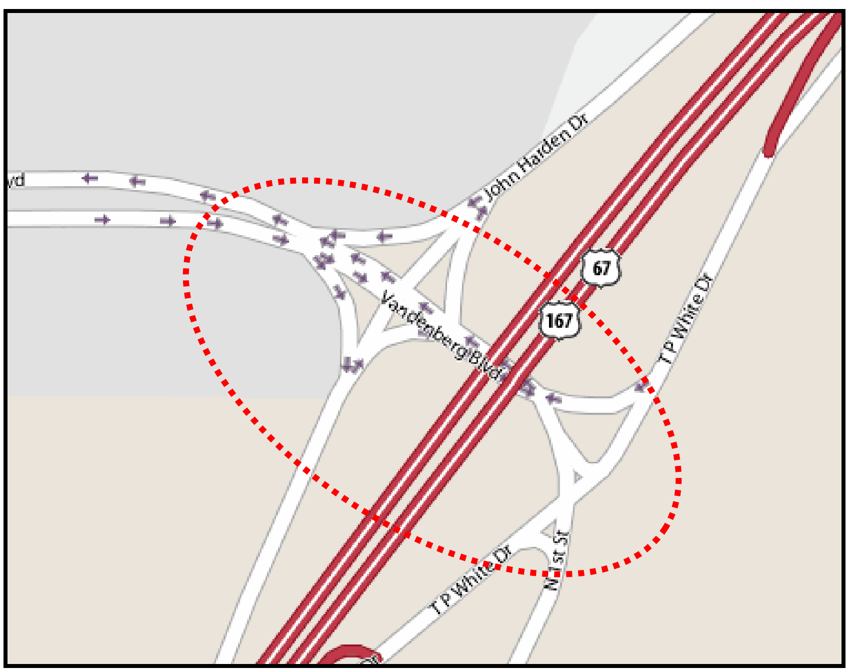




Location 13: Jacksonville, Pulaski County

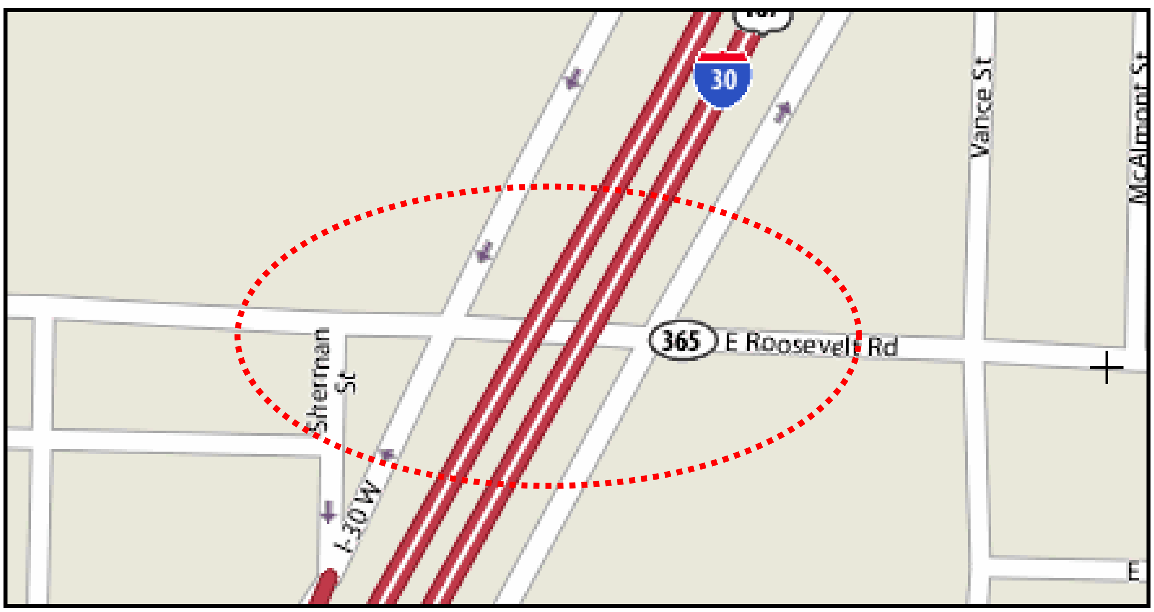
- **Intersection of US Highway 67 and 1st Street**
- **Observations made along 1st Street**
- **Details: Angled intersection due to large horizontal curvature; Poor visibility around overpass; Poor traffic signal infrastructure; High traffic volumes at peak periods; No turning lanes onto highway**





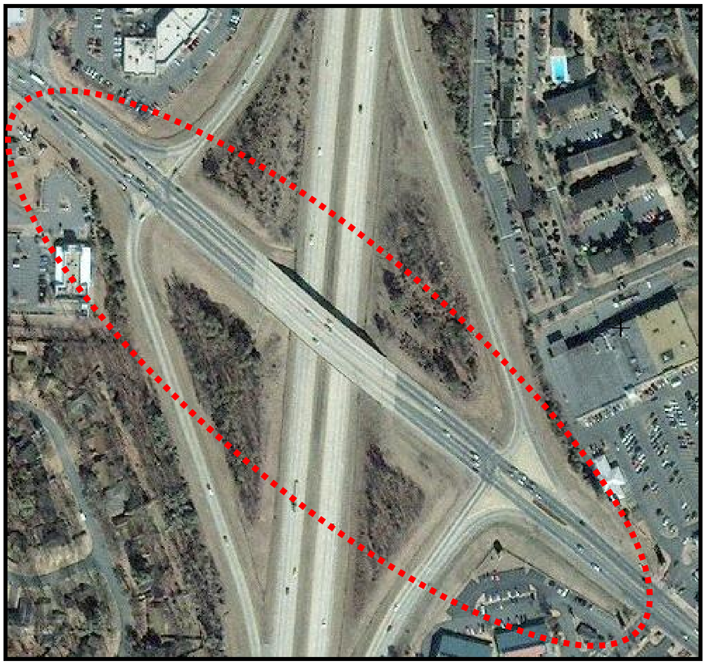
Location 14: Little Rock, Pulaski County

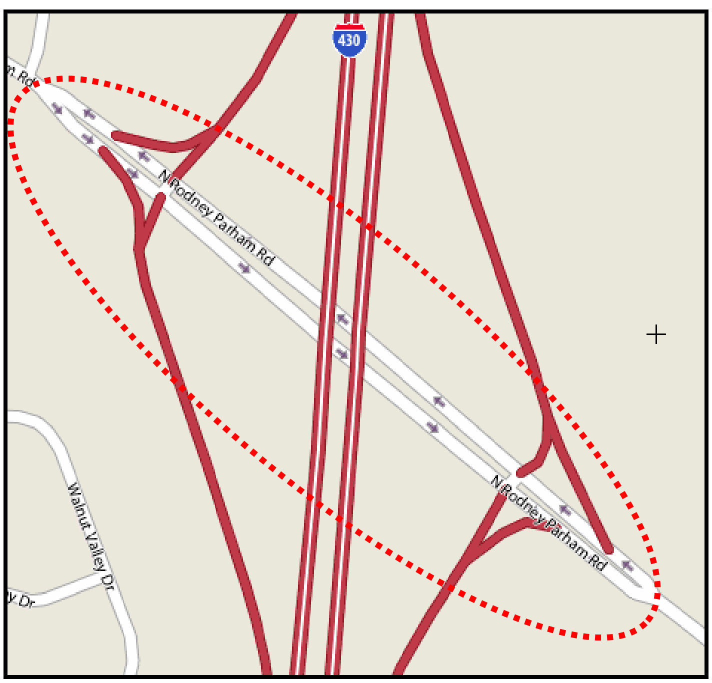
- **Intersection of State Highway 365 and Interstate 30**
- **Observations made along State Highway 365**
- **Details: Conflicting information with traffic signal and infrastructure; Poor signage and traffic signals**



Location 15: Little Rock, Pulaski County

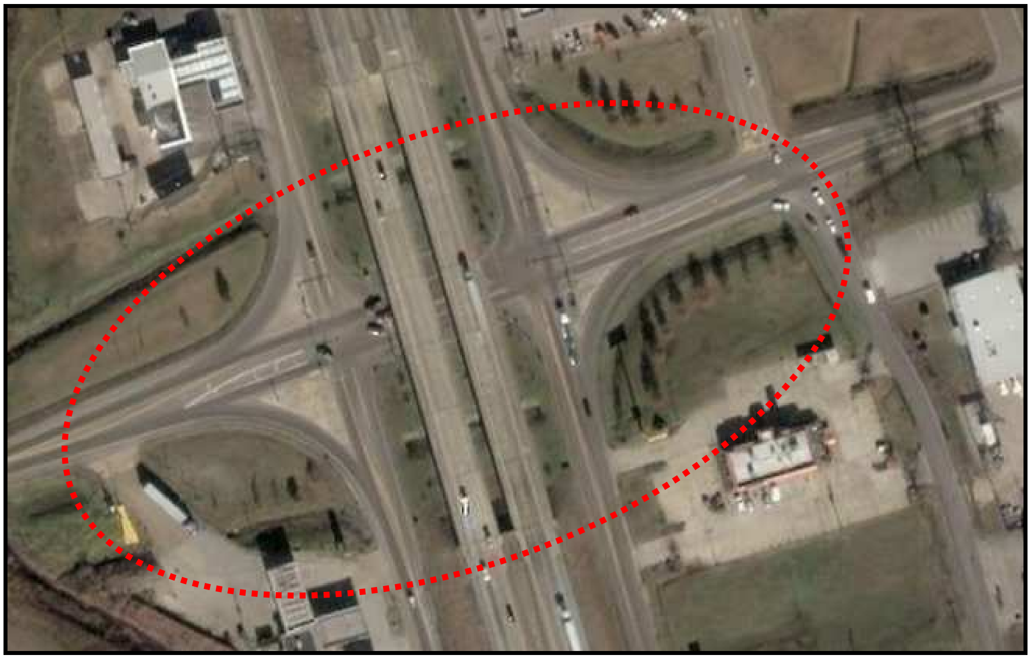
- **Intersection of Rodney Parham Road and Interstate 430**
- **Observations made along Rodney Parham Road**
- **Details: Poor visibility due to vertical curvature along highway and hills around on/off ramps; Several medians along the roadway acting as obstacles; Poor merging due to short lanes right along the off ramp**

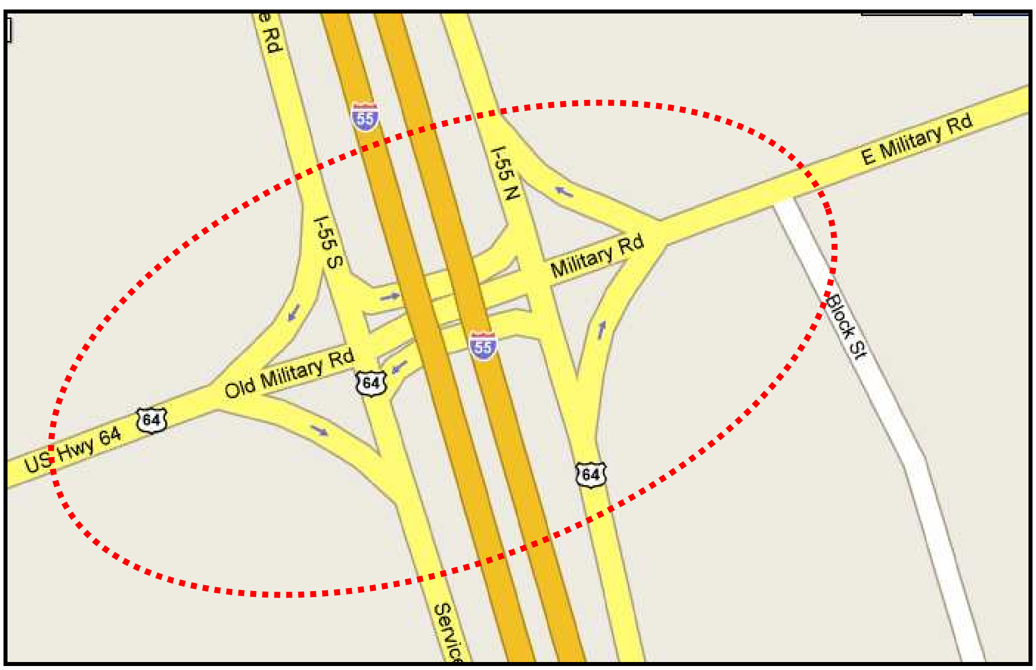




Location 16: Marion, Crittenden County

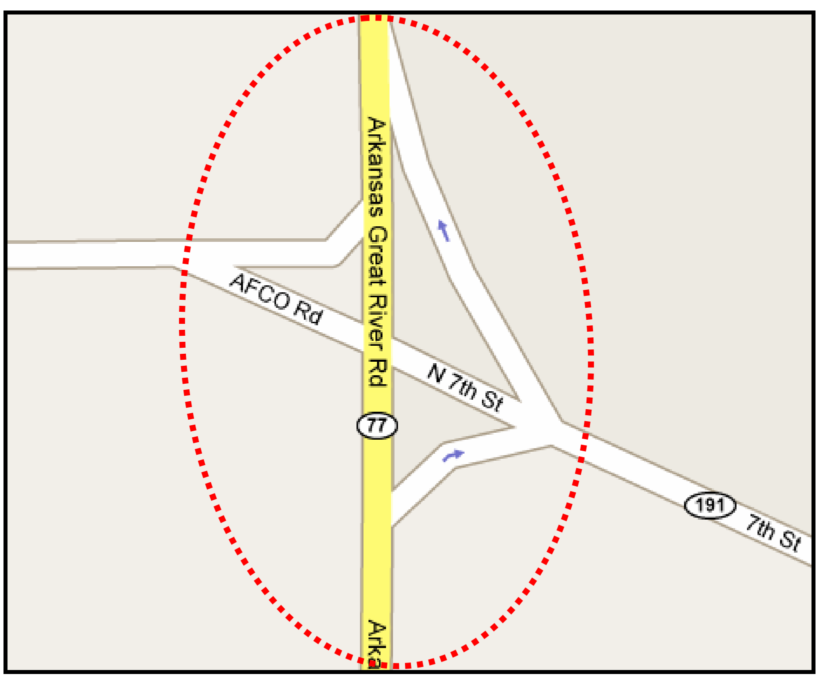
- **Intersection of US Highway 64 and Interstate 55**
- **Observations made on US Highway 64 and access roads alongside I-55**
- **Details: Several medians surrounding and along the highway; Poor traffic signal infrastructure; Poor merging when roads narrow on either side of road segment; Access roads along either side of interstate**





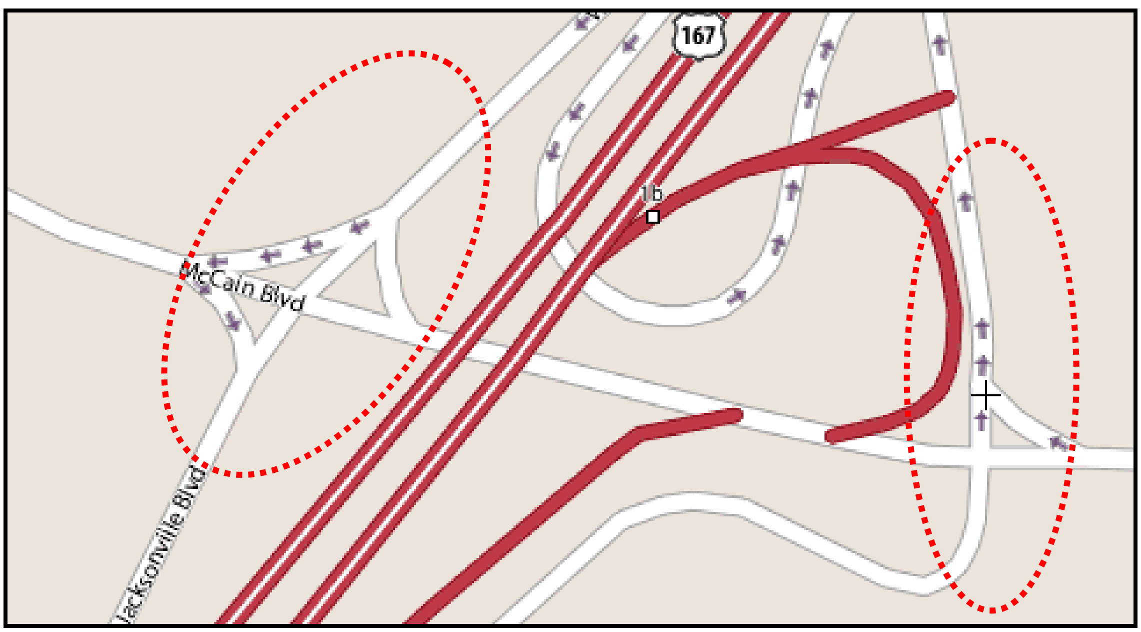
Location 17: Marion, Crittenden County

- **Intersection of State Highway 77 and State Highway 191**
- **Observations made on State Highway 77**
- **Details: Angled Intersection; Medians surrounding roadways; Lane markings worn beyond visibility; Two lane road appears to be only one lane; No signage located at these roads**



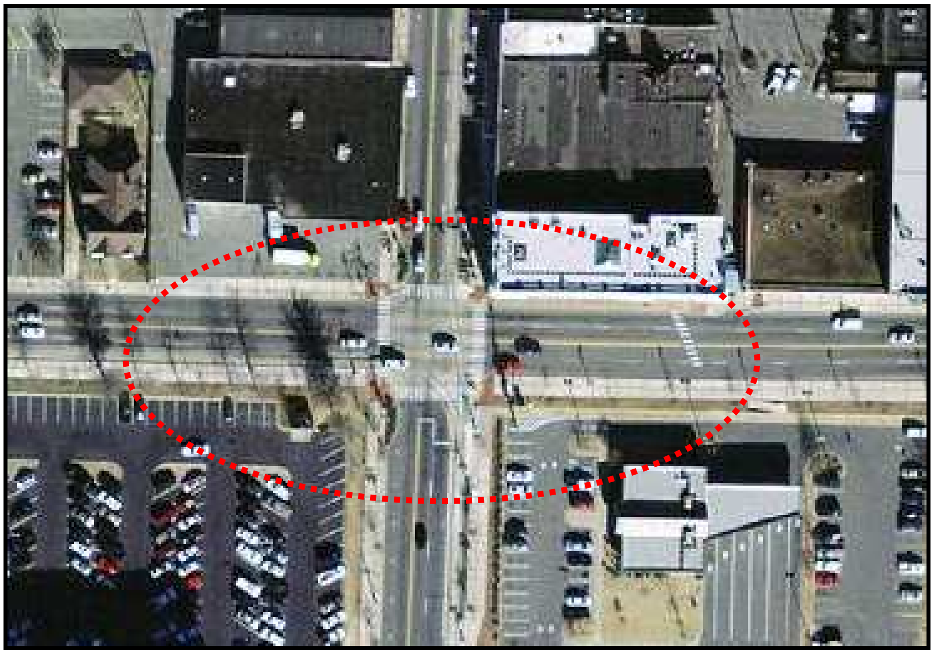
Location 18: North Little Rock, Pulaski County

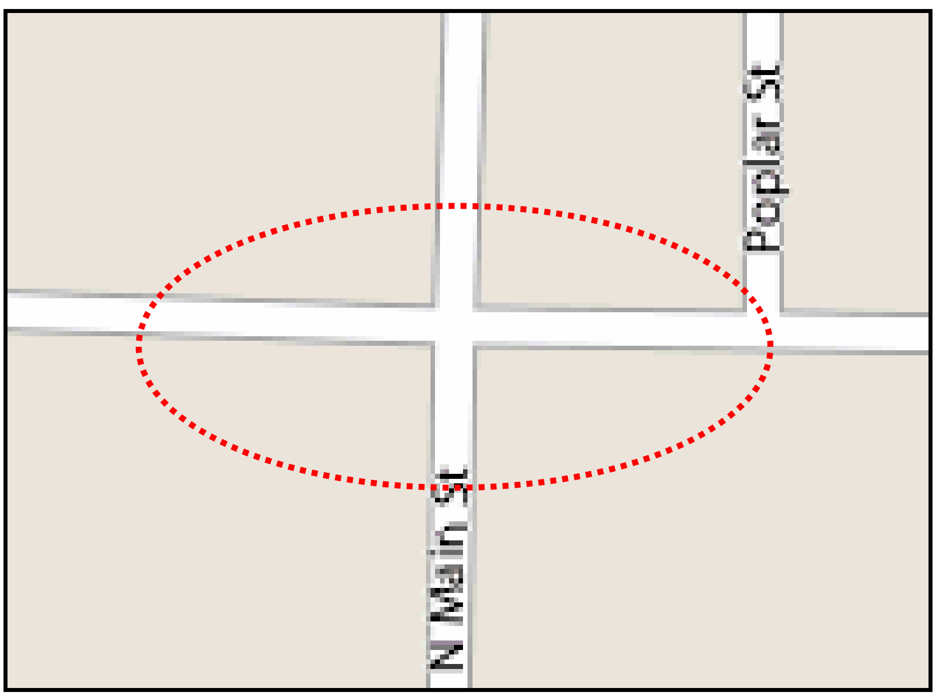
- **Intersection of US Highway 67 and McCain Blvd.**
- **Observations made on McCain Blvd. and access roads alongside US 67**
- **Key Problems: Good signage; Poor visibility due to vertical curvature; Numerous crosswalks along roadway; Failure to yield common during observation**



Location 19: North Little Rock, Pulaski County

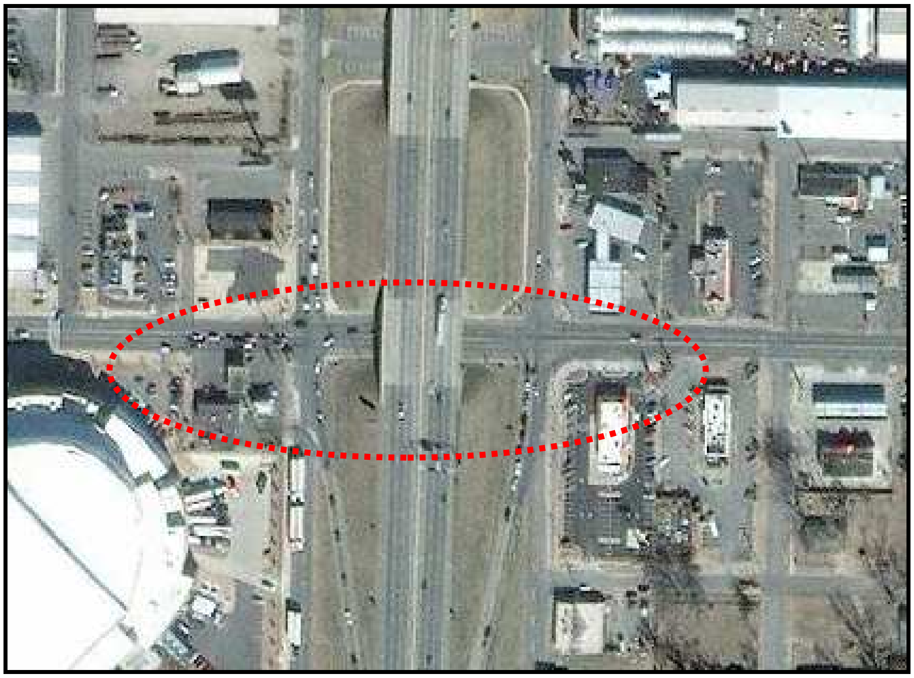
- **Intersection of US Highway 70 and Main Street**
- **Observations made on US Highway 70**
- **Details: A Trolley line crosses this intersection, which runs periodically through traffic; Intersections are set with a timer; Crosswalks are located at each corner of the intersection; Pedestrian traffic signals also on a timer**

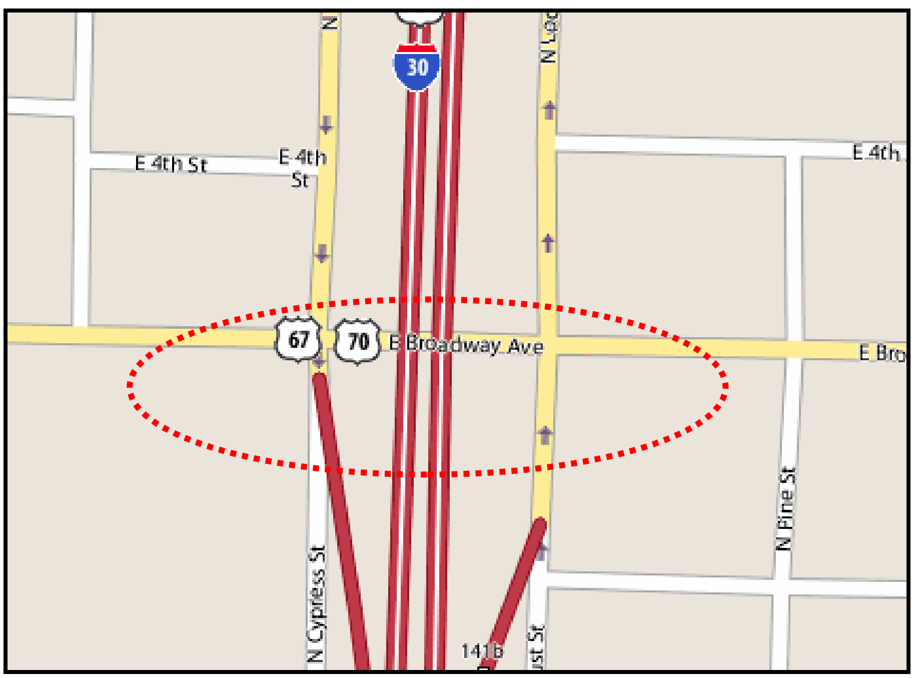




Location 20: North Little Rock, Pulaski County

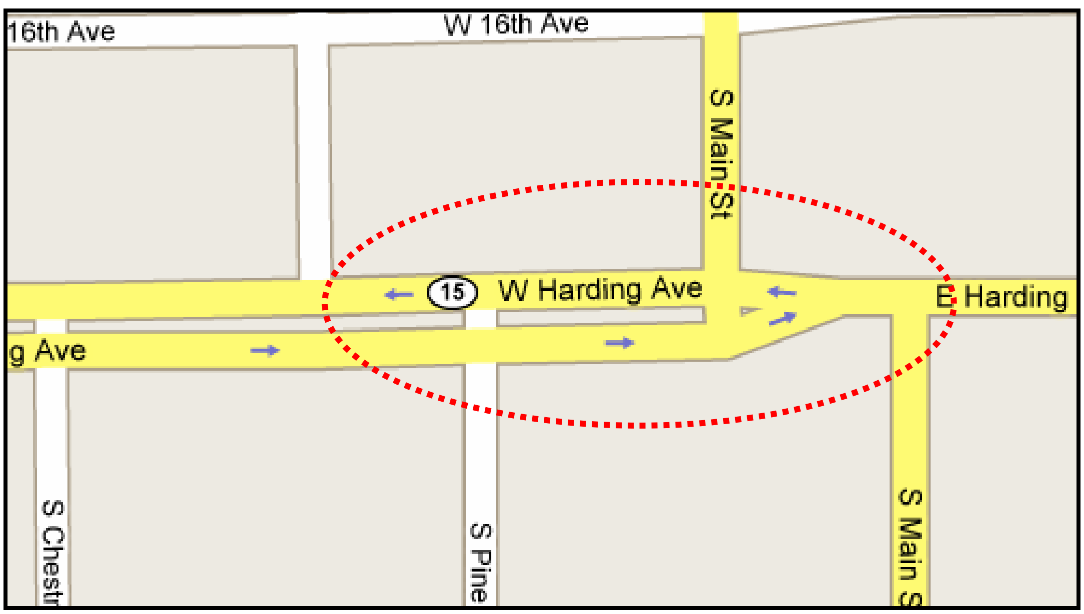
- **Intersection of US Highway 70 and Interstate 30**
- **Observations made on US Highway 70**
- **Details: No turning lanes towards on ramps; Difficult merging along access roads coming onto the highway; Long red light durations; Several observed drivers running yellow lights**





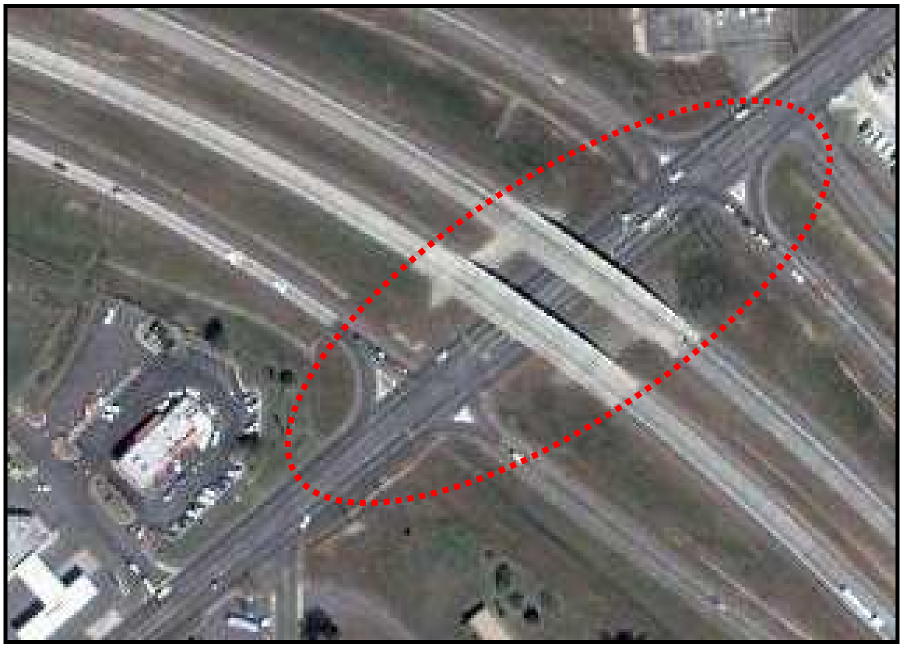
Location 21: Pine Bluff, Jefferson County

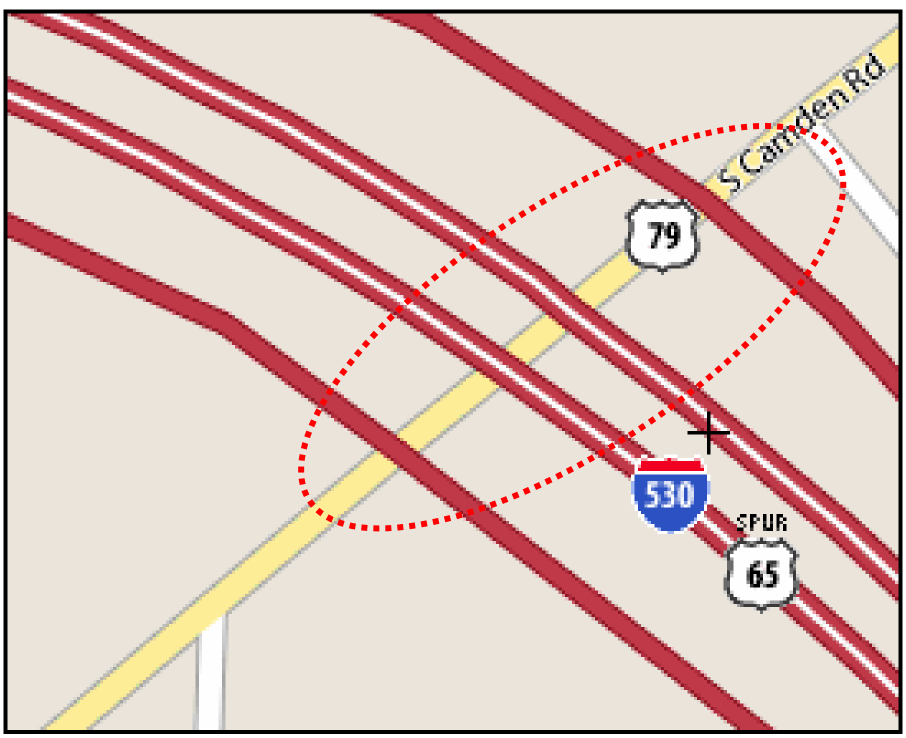
- **Intersection of State Highway 15 and East Harding Avenue**
- **Observations made on East Harding Avenue**
- **Details: Good lane markings; Two one-way roads parallel to each other; Poor merging of parallel roads; Medians along the roadway that act as obstacles; Poor visibility due to surrounding wooded areas**



Location 22: Pine Bluff, Jefferson County

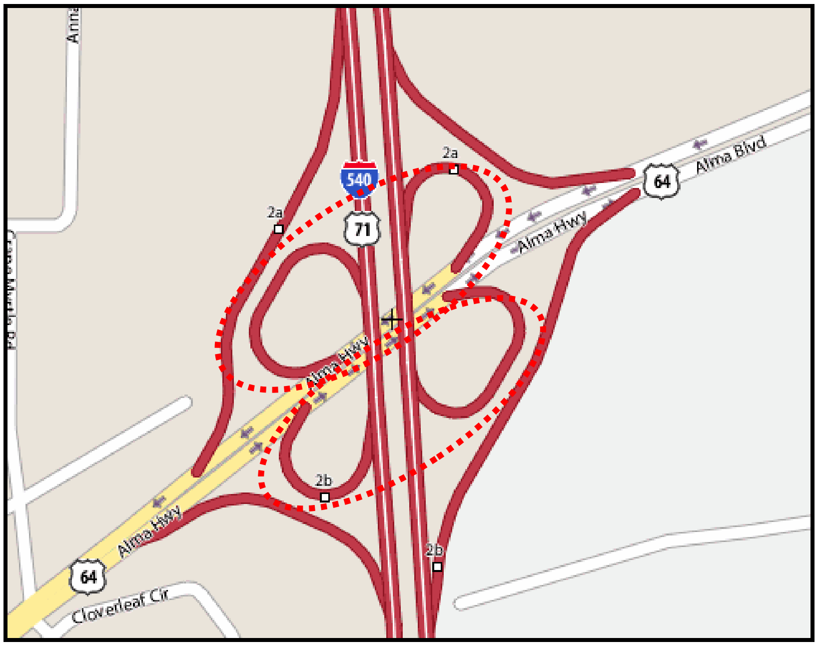
- **Intersection of US Highway 79 and Interstate 530**
- **Observations made on US Highway 79**
- **Details: Good signage and lane markings; Great visibility; Parallel turning lanes under overpass; Failure to yield prominent during observation**





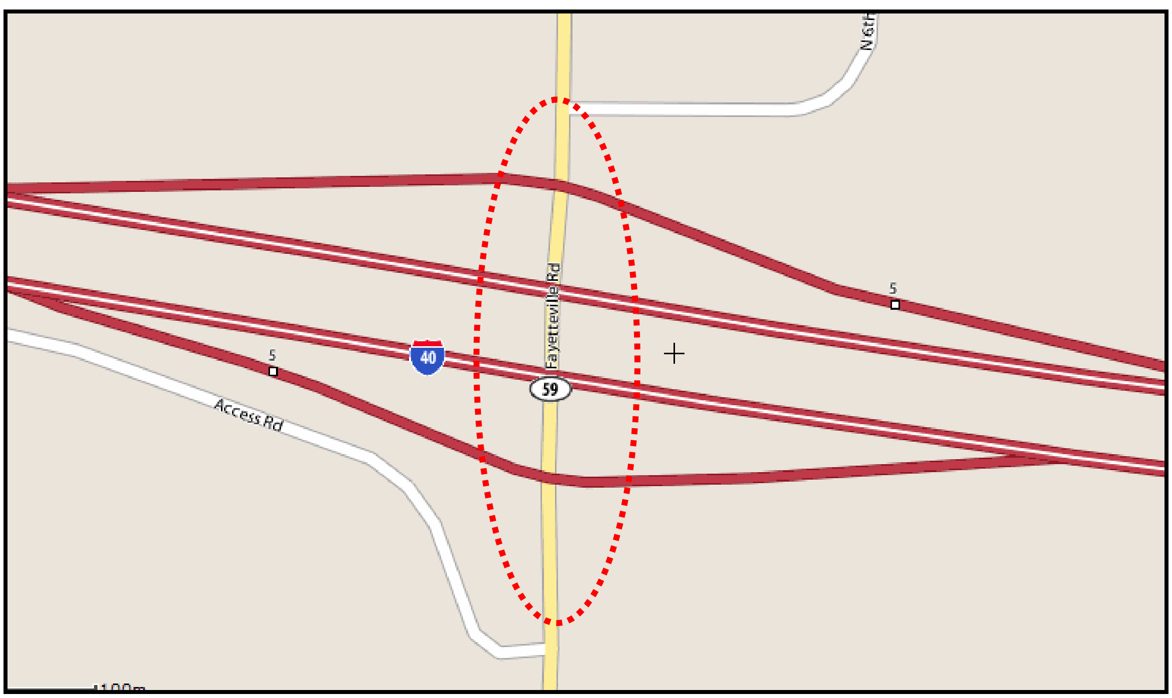
Location 23: Van Buren, Crawford County

- **Intersection of US Highway 64 and Interstate 540**
- **Observations made along ramps on both I-540 and US Highway 64**
- **Details: Four roundabout exit ramps for merging along each road; Difficult merging along these ramps; Poor visibility and roadside information also along these ramps**



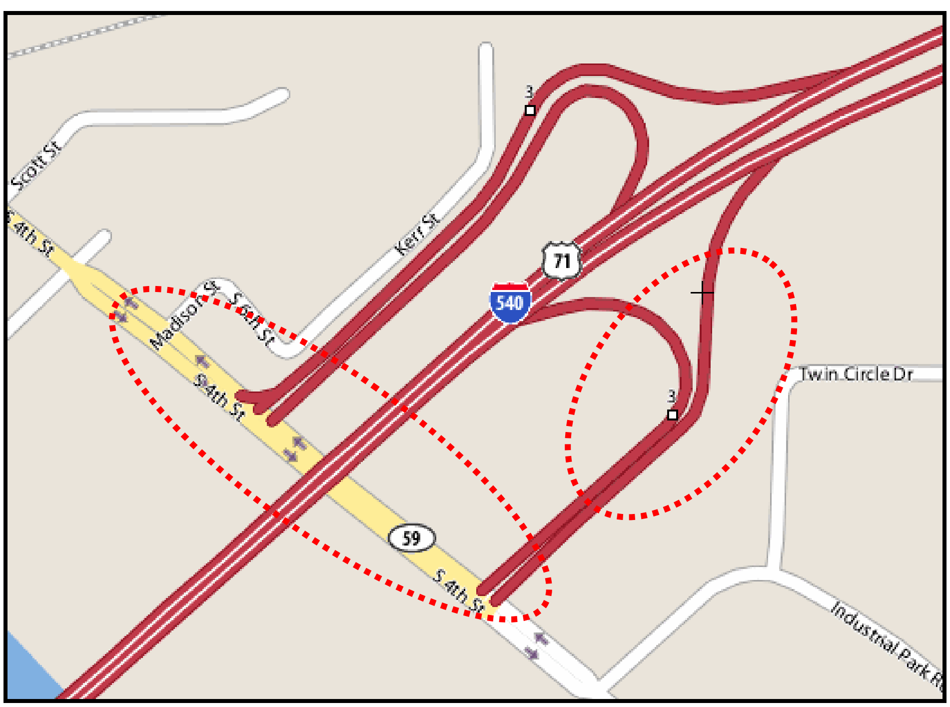
Location 24: Van Buren, Crawford County

- **Intersection of State Highway 59 and Interstate 40**
- **Observations made on State Highway 59**
- **Details: Poor traffic signal infrastructure at on/off ramps; Turning lane underneath the overpass too short for traffic traveling in both directions**



Location 25: Van Buren, Crawford County

- **Intersection of State Highway 59 and Interstate 540**
- **Observations made on State Highway 59 and exit ramps to I-540**
- **Details: Numerous signs along on/off ramp; Conflicting information between lane markings and signage; Several medians surrounding roadway**



Location 26: White Hall, Jefferson County

- **Intersection of US Highway 270 and Interstate 530**
- **Observations made on US Highway 270**
- **Details: Good signage and lane markings; Good visibility; Poor merging following on/off ramps; Failure to yield common during observations**