2010

Secondary Low Volume Rural Road Safety: Segmentation, Crash Prediction, and Identification of High Crash Locations

Daniel Joseph Cook
Iowa State University

Secondary low volume rural road safety: Segmentation, crash prediction, and identification of high crash locations

by

Daniel Joseph Cook

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Civil Engineering (Transportation Engineering)

Program of Study Committee
Reginald Souleyrette, Major Professor
Konstantina Gkritza
Alicia Carriquiry

Iowa State University

Ames, Iowa

2010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to Dr. Souleyrette for giving me the guidance on this and other projects I have worked on. Additional thanks to Zach Hans for providing me with GIS support for this project. Also thanks to Massiel Orellana for providing me with statistical help when it was needed. I want to thank my program of study committee for assisting me.

Special thanks to Dr. Kannel for recognizing my interest in transportation while I was an undergraduate student, and introducing me to CTRE and helping me enroll in graduate school as a concurrent graduate student. And my greatest thanks go to my wife Ashley for supporting me through my college studies and work.

ABSTRACT

Traffic safety research is important to understand the interactions and relationships between crashes and the roadway. Methods have been established for segmenting roadways for safety analysis, creating safety performance functions, and identifying high crash locations. However, little work or reasoning is available to provide guidance for segmenting and modeling secondary low volume rural roads (LVRRs). This study investigated the effect of secondary LVRR segment length on segment analysis. Safety performance models were also examined and created for secondary LVRRs. Using previously proposed tests, four different high crash identification methods (crash frequency, crash rate, empirical Bayes and crash reduction potential) were compared for use on secondary LVRRs in Iowa. Analysis of the secondary LVRR system identifies a trend showing as segment length increases, so does the statistical reliability of the average annual crash frequency as compared to the variance in crash frequencies from year to year. Serious and total crash prediction models are recommended for use on four different classes of mainline secondary LVRRs: paved and unpaved 1-99 AADT, and paved and unpaved 100-400 AADT. Lastly, empirical Bayes is recommended as the best available method for identifying high crash locations on secondary LVRRs in Iowa. Care is advised when developing candidate high crash location lists for secondary LVRRs based on segmented systems where systemic treatment may be more appropriate.

CHAPTER 1. GENERAL INTRODUCTION

1.0 INTRODUCTION

Secondary low volume rural roads (LVRRs) in Iowa comprise a large portion of the state's roadway length: 79,771 miles out of the state's total of 115,371 miles. Table 1-1 shows different characteristics about a few selected road classes in Iowa. These statistics were derived from the 2008 Iowa DOT's Geographic Information Management Systems (GIMS) road database. Primary roads only make up 8.2 percent of all roadway mileage in Iowa while 78.0 percent are composed of rural secondary roadways. 69.1 percent of Iowa roadways are secondary low volume rural roads (LVRRs) and 57.2 percent are secondary LVRRs with fewer than 100 AADT. Figure 1-1 shows a visual representation of just how expansive the secondary LVRR system is in Iowa. Outside principal urbanized areas, secondary LVRRs appear nearly everywhere in the state.

Table 1-1. Characteristics of selected road classes in Iowa.

Road Class	Total Centerline Mileage	Percent of Iowa Total	2008 VMT	Percent of Total VMT
All Primary	9,432	8.2%	18,770,131,000	60.0%
All Rural Secondary	89,957	78.0%	5,438,613,000	17.4%
Two-lane Rural Secondary 1-400 AADT	79,771	69.1%	1,901,399,000	6.1%
Two-lane Rural Secondary 1-99 AADT	66,022	57.2%	861,008,000	2.8%
All Iowa	115,371	100.0%	31,301,615,000	100.0%

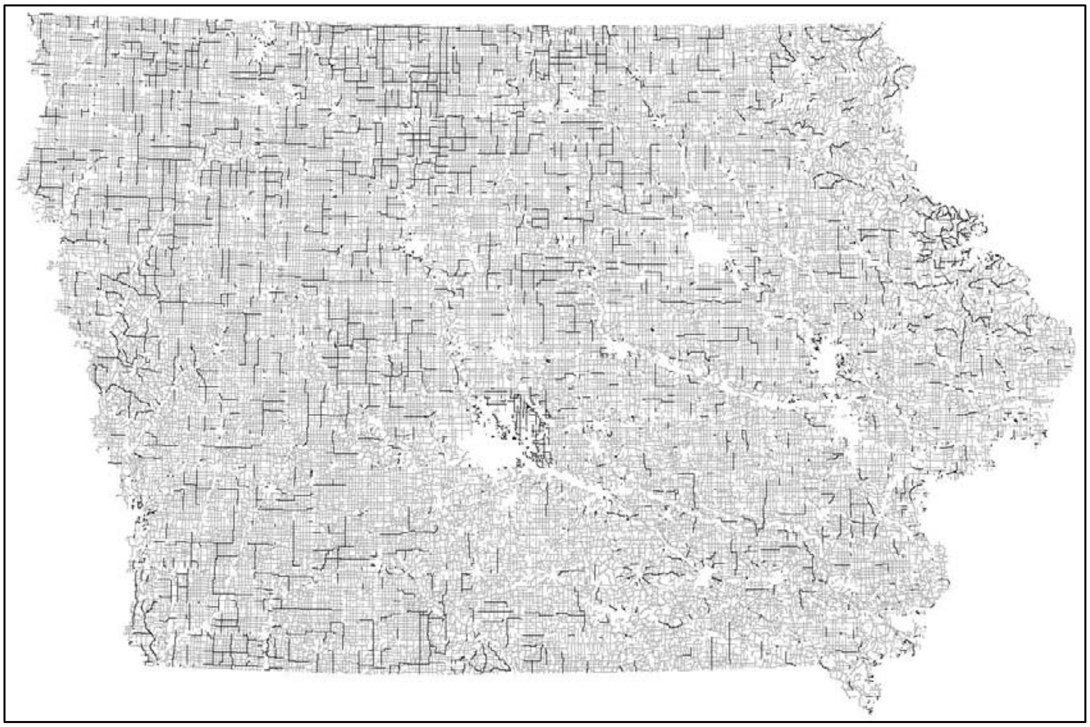

Figure 1-1. Secondary low volume rural roads in Iowa, with paved roads in darker color.

With the large expanse of secondary LVRRs, it is understandable that safety is a very important issue on these roadways. Unfortunately, low crash densities on these roadways make it difficult to identify high crash locations. Methods need to be established to properly identify high crash locations on secondary LVRRs in order to spend funds for safety improvements wisely. This report addresses key issues for identifying at-risk secondary LVRRs. The effect of segment length on safety analysis is first discussed. Second, safety performance functions are developed for estimating crash frequencies on secondary LVRRs. Last, four methods for identifying high crash locations are compared for use on secondary LVRRs: crash frequency, crash rate, empirical Bayes and crash reduction potential (Cheng and Washington, 2008).

2.0 THESIS ORGANIZATION

This thesis is divided into five chapters. The first chapter (this chapter) serves as an introduction of the thesis as well as a literature review for low volume rural roads. The second chapter is a paper titled "Effect of Segmentation Length on Safety Analysis," which will be presented at the Transportation Research Board 90th Annual Meeting in January 2011. The paper was written by Dan Cook, Reginald Souleyrette and Justin Jackson. The paper covers the effects of segment lengths on both two-lane rural primary roads and secondary low volume rural roads.

The third chapter covers the development of safety performance functions for predicting mainline crashes on secondary LVRRs. The fourth chapter examines four methods of identifying and selecting high crash locations on secondary LVRRs and compares the performance of each method. Lastly, the fifth chapter summarizes the conclusions of the previous three chapters and gives final recommendations for secondary LVRRs.

3.0 REVIEW OF LITERATURE

In *Guidelines for Geometric Design of Very Low-Volume Local Roads (ADT ≤400)* published by The American Association of State highway and Transportation Officials (AASHTO) defines very low-volume local roads as "a road that is functionally classified as a local road and has a design average daily traffic volume of 400 vehicles per day or less" (p. 1 AASHTO, 2001). While AASHTO uses the term "very low-volume local roads," this project will simply use "low volume roads." By definition, both terms are identical, since this project is looking at roads with an annual daily traffic (ADT) equal to or less than 400 vehicles per day and also fall into the local jurisdiction (non-state or federal facilities).

The purpose of low volume roads is much different than those of state and federal highways. AASHTO states the primary function of a low volume road "is to provide access to residences, farms, businesses, or other abutting property, rather than to serve through traffic" (p. 5 AASHTO, 2001). Essentially low volume roads are mainly collector and local roads, with a few exceptions in low populated areas. Another fact to consider is that local roads are maintained differently from state to state. Local roads can be under the control of Federal, state, or local agencies. In Iowa, local roads fall under local control.

It is important to keep in mind that most users of a low volume road have used it before. "Geometric design features that might surprise an unfamiliar driver will be anticipated by the familiar driver" (p. 11 AASHTO, 2001). There are low volume roads that do not comply with the design guidelines in AASHTO's guidelines for low volume roads, but this does not mean that these roads should be reconstructed to meet these guidelines. In most cases system-wide safety is not affected by reconstructing low volume roads to these guidelines. "Although treatments that are safety effective on higher traffic volume facilities should also improve safety on low-volume roads, they may not be cost-effective" (p. 1 Hall, 2003). With the small amounts of safety funds for local agencies to use in combination with the vast amount of mileage that exists for low volume roads, it is simply cost prohibitive to consider reconstruction system-wide.

While the data shows that there are many crashes on low volume roads in Iowa, crashes are spread out over a large amount of roadway as compared to crashes on state and Federal roads. Statistically it is difficult to assess crashes on low volume roads. Using a before-and-after study of a segment of low volume road carries the regression to the mean problem. Also, multi-vehicle crashes are extremely rare events on low volume roads since very low traffic volumes exist and the probability of two vehicles meeting are lower than those on higher volume roadways. Most crashes are single vehicle crashes, in which the majority is lane departure related.

Safety performance functions (SPF) are used to estimate the average number of expected crashes on a segment of roadway, at an intersection, or other special road feature. The equation is a function of certain trait values (AADT, section length, lane width, etc.) and of several regression parameters. Originally it was thought that crashes have a Poisson distribution, but now the negative binomial distribution is assumed for the empirical Bayes method (Hauer, 2001). The primary reason for using the negative binomial distribution for crashes is that it does not restrict the mean to equal the variance of the population. The negative binomial distribution allows for overdispersion.

In *Accident Models for Two-Lane Rural Segments and Intersections* (Vogt, 1998), crash prediction models were developed for segments and intersections (both 3 and 4-leg) for Washington and Minnesota. Variables used in the final segment model were the intercept, state, lane width, shoulder width, roadside hazard rating, driveway density, degree of curvature, crest curve grade rate, and vertical grade. Variables used in the 3-leg intersection model were the intercept, log(ADT of the minor road), log(ADT of the major road), crest curve grade rate of the major road, degree of curvature of major road, posted speed on major road, roadside hazard rating for the major road at the intersection, and the presence of a channelized right turn. Variables used in the 4-leg intersection model were the intercept, log(ADT of the minor road), log(ADT of the major road), crest curve grade rate of the major road, the adjusted intersection angle from 90 degrees, and the number of driveways in the vicinity of the intersection. The road data available for the low volume roads project do not include horizontal or vertical alignment information, but these previous models may give an idea of parameters to include in the low volume roads model.

Safety Conscious Planning in Indiana: Predicting Safety Benefits in corridor Studies, Volume 1, Research Project (Tarko, 2007) also presents a SPF for rural two-lane segments. The variables included in the model include lane width, shoulder width, average grade for vertical curves in the segment, and average degree of curvature in the segment. Again, horizontal and vertical alignment data will not be available for this project. Equation 1-1 shows the general form used for the safety performance functions.

Equation 1-1:

$$A = \exp(k) L Q^{\beta} \exp^{\left(\sum \gamma_i \chi_i\right)}$$

$A =$ number of crashes in a year,

$L =$ length of the section in miles,

$Q =$ AADT of the section,

$\chi =$ explanatory variables,

$k, \beta, \gamma =$ constants.

In *Measuring the Goodness-of-Fit of Accident Prediction Models* Miaou recommends the use of R^2_α for a goodness of fit predictor on SPFs. The criterion uses the dispersion parameter to figure how well the variance is explained in the data (Miaou, 1996). Several goodness-of-fit measures were examined in this study, with the dispersion parameter-based R^2 being recommended for use.

The Akaike's information criterion (AIC) is a goodness of fit measure of a statistical model. The AIC is a relative value to be used when selecting the best model to use for a set of data. The AIC takes into account the amount of information lost or gained when different models are constructed. The model having the lowest AIC value is best model (Hu, 2007).

There are several methods that exist for identifying high crash locations on roadways, but there is not much justification for which method is better. Cheng and Washington propose five different tests to use to decide which method (crash frequency, crash rate, empirical Bayes or crash reduction potential) is better for selecting high crash locations (Cheng and Washington, 2008). Four of the five tests include a test statistic that can be used to compare the performance of each method. The site consistency test calculates the total number of crashes identified from the high crash locations. The method consistency test determines the number sites identified as high crash locations in two adjacent time periods for each method. The total rank differences test calculates the total difference in rankings of sites between two adjacent time periods for each method. Lastly, the Poisson mean differences test determines the total true Poisson mean difference of the false identifications for each method.

CHAPTER 2. EFFECT OF SEGMENTATION LENGTH ON SAFETY ANALYSIS

1.0 INTRODUCTION

With increasing traffic and urban sprawl, safety is an increasingly significant concern for two-lane rural roads. These roads are amongst the most at-risk for fatalities and major injury crashes based on rate. Overall in the United States, fatality rates have been falling for the last few decades, with 2007 seeing a rate of 1.36 fatalities per 100 million vehicle miles traveled (HMVMT) (NHTSA, 2010). However in 2007, the fatality rate for rural arterial roads was 2.23 fatalities per HMVMT, 2.79 per HMVMT for rural collector roads, and 3.18 per HMVMT for rural local roads (FHWA, 2007). Even though the crash rates are high for rural two-lane roads, crashes are spread over a large network of roadways and are relatively rare. This may make the statistically proper identification of high crash locations difficult or impossible. High crash location identification and reliability of crash estimates depend on the method of segmentation used to segment the roadway network.

2.0 REVIEW OF LITERATURE

Typically, analysis segments are defined in two fundamental ways with respect to composition and length. Usually, to provide for modeling fidelity and implementation of results, analysis segments are defined to be relatively homogenous with respect to road geometry, traffic characteristics, safety, and other roadway characteristics. This results in variable lengths unless segments are very short (e.g., 0.01 miles.). Defining segments by longer fixed lengths result in heterogeneous characteristics. A number of approaches have been implemented to define roadway segments for identifying high crash locations.

Use of several criteria for defining segments allows testing of specific attributes as predictors of safety performance. Studies suggest that risk conditions can vary rapidly over a fairly short highway length (Papageorgiou, 2002). However, as segment length decreases, the number of segments containing zero crashes increases. Longer segments are generally more appropriate when conditions are fairly constant over extended distances.

Two types of segmentation are possible: predetermined length and sliding scale. In each type one may use either fixed or variable length segmentation. Predetermined fixed length results in analysis segments of almost all the same length (naturally, roads or routes are not always even-multiples of a given fixed length). Predetermined variable length obviously results in many or all segments of different lengths. The Iowa DOT segments the Iowa roadway network, called GIMS, using variable length segmentation based on homogenous attributes of the roadway. GIMS segments range in length from very short segments (0.001 mi or 5 feet) to considerably long segments (>>1.0 mi).

Following predetermined variable length segmentation, short segments may be combined (aggregated). To do this, a user may prescribe a minimum segment length. If a segment's length is less than the predetermined length, the next adjoining segment is added to that segment until the new segment's length meets or exceeds the predetermined length. As cases in point, Washington uses 0.1 mile or less segments and New York uses 0.3 mile segments (Geyer, 2005).

Sliding scale segmentation uses a moving window that "slides" along the virtual roadway. Again there are two types of sliding-scale segmentation possible: fixed length and variable length. To implement, the segment inside the moving window is first analyzed. If the segment meets or exceeds the defined crash rate threshold, the segment is included in an output file. If the predefined threshold is not met, the moving window advances along the roadway at an incremental length and the resulting segment is analyzed. This step is repeated until the user's definition of a segment is achieved. As an example, the Utah DOT uses one-mile segments, although, the UDOT system has the ability to use sliding scale segmentation. The Florida DOT system can also perform sliding scale analysis (Geyer, 2005).

The California DOT (Caltrans) currently uses a fixed length sliding scale in the analysis of roadway segments with high numbers of crashes. In the Caltrans system, analysis of a particular roadway starts at mile 0.0. The first 0.2 mile segment of the roadway is then analyzed. If the subject segment exceeds a predetermined number of crashes, the segment is defined and added to an output table. If not, the 0.2 mile segment advances along the roadway by an increment of 0.02 mile and this portion of the roadway is analyzed. The segment keeps sliding along the roadway until a segment is found to be significantly at risk. When a segment exceeds a predefined number of crashes it is added to the output table. The next segment to be analyzed is started at the end of the segment that was added to the output table (Geyer, 2005). A problem identified by Caltrans is that segments containing the highest number of crashes possibly may not be identified, as segments are defined when a predetermined number of crashes is attained (high crash segments therefore may be broken into two pieces, neither of which may be amongst the highest in the system).

The Wisconsin DOT (WisDOT) also identified the problem that sequential segmentation (sliding scale) has a bias towards not identifying high crash concentrations at either side of a jurisdictional or other border. To reduce this potential, WisDOT developed a floating highway segment algorithm, PRÈCIS (Drakopoulos, 2005). The process starts by identifying the first 0.01 mile segment with a crash during the analysis period. The algorithm then analyzes 0.01 mile segments upstream and downstream of the crash in order to identify segments containing the highest number of crashes.

Kentucky uses a program that allows the user to select segment length and define the minimum number of crashes per segment. The program advances from the beginning of the road to the first crash. This length of road defined by the user is then analyzed. If the segment’s crash frequency meets or exceeds the user defined number of crashes, the segment is exported into an output table. The program then advances from the first crash identified to the next crash along the route. Allowing the program to start the next segment analysis from the next crash location will ensure that the segments with the highest number of crashes will be identified (Agent, 2003).

The Highway Safety Information System (HSIS) has developed a variable sliding scale analysis tool for identification of high crash roadway segments. The sliding scale in this case has a variable rather than fixed length. The tool allows a user to define both segment and incremental length. Using the HSIS tool, the first segment of the roadway is analyzed. An incremental length will keep being added until the user defined crash rate is exceeded. Next, an incremental length still will be added until the crash rate drops below the threshold, and only then will the tool output the segment. This allows for the whole continuous section of roadway with high crash rates be identified as one segment (FHWA, 2000).

The European Road Assessment Program (EuroRAP) suggests guidelines for segmenting roadways for safety analysis. Section boundaries are chosen such that a section will typically have at least 20 fatal or major injury crashes over a period of three years. In Great Britain, for example, this results in sections averaging around 12 miles in length. If this criterion is not met, sections are combined under the following criteria: the combined segments have the same road number; they are adjacent; they are part of the same network; or they have similar average daily traffic volumes (ADT) with differences up to 10,000 being acceptable. However, due to lower crash densities in less populated areas such as Sweden, route segments can average only five fatal and major injury crashes over a period of three years (usRAP, 2006).

Using crashes as a threshold for segmentation on LVRRs is challenging as the number of crashes is relatively low. The American Association of State Highway and Transportation Officials (AASHTO) defines low-volume local roads as those roads with ADTs of 400 or less, and functionally classified as a local road (AASHTO, 2001). A study of LVRRs in New Mexico by Hall, Rutman, and Brogan, segmented roads with an ADT between 150 and 400. To do this, a minimum segment length of 15 miles was used, in order to provide for a statistically meaningful sample size (Hall, 2003).

3.0 METHODOLOGY

Two different analyses were performed to test the effects of segmentation length: one for rural two-lane primary roads in Iowa and one for two-lane secondary LVRRs in Iowa.

3.1 Rural two-lane primary roads

A sensitivity analysis was performed using three different segment lengths (two-mile, one-mile and one-half mile). Segments were defined and ranked according to the Iowa DOT Office of Traffic and Safety prioritization procedure, which weights cost of crashes (severity) by 60 percent, frequency by 20 percent and rate by 20 percent. The rank of each segment was compared to the rank of the overlapping segments of different length.

The rural primary system in the northwest portion of Iowa was used for the sensitivity test. Interstate 35 and Interstate 80 comprised the east and south boundaries of the study area. This portion of the Iowa system was first segmented into two-mile segments using dynamic segmentation in ArcGIS 9.1. Within each two-mile segment, two concurrent one-mile and four half-mile segments were also created. 1,535 two-mile segments were created, which were then split into one-mile and half-mile segments. Not all the network was converted into two-mile segments due to network topology. Segments for example were terminated at corporate boundaries and other jurisdictional boundaries.

After segments were identified, crashes were assigned using a spatial join and 50 meter tolerance to allow for accuracy of the crash location database. This process was repeated for one-mile and half-mile segments. After crashes were assigned, segments were rated and ranked for each segment length category using the Iowa DOT scoring method.

3.2 Secondary two-lane low volume rural roads

Secondary rural roads with 0 to 400 ADT were selected from the Iowa GIMS database. Dynamic segmentation could not be used on secondary roads as they possess no linear referencing data. To accomplish aggregation, contiguous GIMS segments of common route number and county were combined. Next, these aggregated segments were then split into even-mile fixed length sections from two miles to 15 miles in length. After the roadway was split, remaining sections of uneven length were not included in the output. Non-intersection crashes were then assigned to these sections based on spatial proximity (again, 50 meters). Since the crash frequencies on these roads are very low, 8 years of data were used. The total number of crashes for each section was then compiled based on crash severity. Next, the standard deviation of crash frequency for each section was calculated using Equation 2-1 as proposed by Hauer (Hauer, 2001). The standard deviation was compared to the average annual number of crashes on each segment to determine the reliability of the crash estimate.

Equation 2-1:

$$\sigma\left(\text{standard deviation}\right) = \sqrt{\frac{\text{Annual Crash Frequency}}{Y}}$$

4.0 RESULTS OF ANALYSIS

A sensitivity analysis was performed on the concurrent segments of the rural primary road system. The effect of segment length on safety analysis was also tested for secondary LVRRs. Results differ between the two tests.

4.1 Rural primary road segmentation sensitivity analysis

While different state systems utilize various segment lengths for static segmentation, rationale could not be identified in the published literature. To demonstrate the effect of segment length on safety analysis, segments of Iowa primary highways were ranked using three different segment lengths: two miles, one mile and one-half mile.

4.1.1 Two-mile segments

Two-mile segments were chosen as a baseline. Segment ranks were then compared to average, high, and low ranks of corresponding one-mile and one-half mile segments.

Ranking List Shifts. The top 50, 100 and 200 high crash locations were first identified using two-mile segments with the Iowa ranking method. Then, crash scores were computed for one-mile and half-mile segments, and each segment for each length was assigned a rank. For each of the two-mile segments ranked in the top 200, there exist two corresponding one-mile segments and four corresponding half-mile sections. Figure 2-1 gives an example of how the average, high, and low ranks are defined for the corresponding one and half-mile segments. The nomenclature used in the following tables follow the definitions given in Figure 2-1. The concurrent or corresponding high rank one-mile segment for the two-mile segment shown in Figure 2-1 is ranked 12, while the corresponding low rank one-mile segment is ranked 19. The corresponding average rank one-mile segment is ranked 15.5, which is the average value of the two one-mile segments. The same rules apply to the four concurrent half-mile segments.

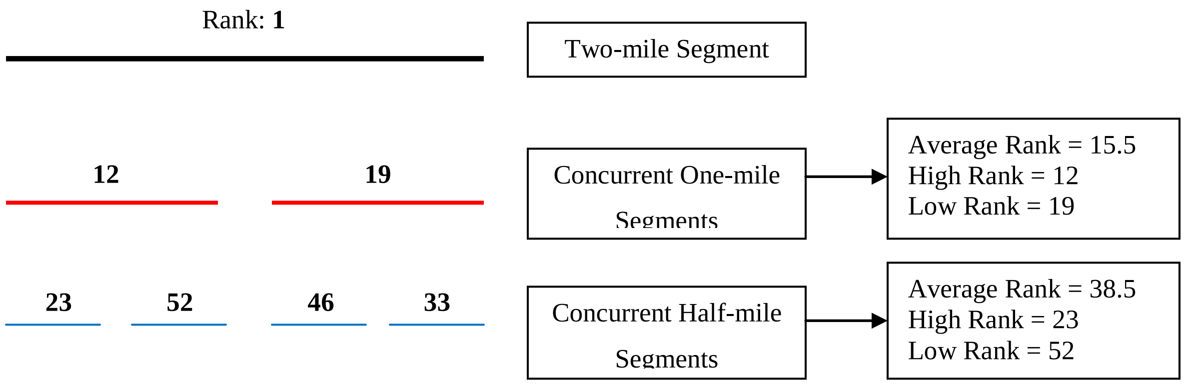

Figure 2-1. Example of how average, high and low rankings are defined.

If segment length did not affect ranking, it would be expected that the top 50 two-mile segments correspond to the top 100 one-mile sections and top 200 half-mile segments. For example, the top-ranked two-mile segment would be comprised of the top two ranked one-mile segments and the top four half-mile segments. In this case, the average one-mile concurrent segment rank of the top two-mile segment would be (1+2)/2 = 1.5 and the average half-mile concurrent segment rank of the top two-mile segment would be (1+2+3+4)/4 = 2.5. If all the lower ranked segments followed suite, the top 50 ranked two-mile segments would have rank scores of 1, 2, 3, …, 50. The "corresponding" one-mile pairs of the top 50 two-mile segments would have average scores of 1.5, 3.5, 5.5,…, 99.5, "highest" ranks of 1, 3, 5,…, 99, and "lowest" ranks of 2, 4, 6,…, 100. The "corresponding" half-mile quadruples of the top 50 two-mile segments would have average scores of 2.5, 6.5, 10.5,…, 198.5, "highest" ranks of 1, 5, 9,…, 197, and "lowest" ranks of 4, 8, 12,…, 200. In this case, all three segmentations of the top 50 two-mile segments could be said not to "shift" at all.

Clearly, upon inspection of Table 2-1, one can see that this is far from the case. In fact, only 4 of the top 50 high crash locations remained in the top 50 when "average rank" is used to compute one-mile ranks (as compared to two-mile ranks), and none of the highest "average rank" sites computed using half-mile segments match the top 50 as identified using two mile segmentation. The case is not so different if only the "highest ranked" segment of the pair or quadruple is used (66 percent of high crash locations are identified using one-mile and 60 percent using half-mile). Table 2-1 presents similar results for the top 100 and 200 high crash locations as identified using two-mile segmentation.

Table 2-1. Two-mile segment shifts in rank.

	Top 50 Locations		Top 100 Locations		Top 200 Locations	
Concurrent Segment	**Shift out**	**Percentage Shift**	**Shift out**	**Percentage Shift**	**Shift out**	**Percentage Shift**
Average Rank 1-mile	46	92%	92	92%	176	88%
High Rank 1-mile	17	34%	27	27%	44	22%
Low Rank 1-mile	47	94%	94	94%	188	94%
Average Rank ½ -mile	50	100%	100	100%	199	99.5%
High Rank ½ -mile	20	40%	41	41%	58	29%
Low Rank ½ -mile	50	100%	100	100%	200	100%

Absolute Value of Ranking. Next, the absolute value of ranking shift was calculated for corresponding one and half-mile segments, again using average, high and low rankings of corresponding pairs and quadruples for the top 50, 100, and 200 two-mile locations. Table 2-2 presents the results of this analysis classified into 0, 1-25, 26-100, 101-200, and (>200) change in the various ranking positions. For example, for the top 50 two-mile sections, no average one-mile segment rank changed by 0, 1 (2 percent) changed by between 1 and 25 ranks, 10 (20 percent) changed by between 26 and 100 ranks, etc.

Table 2-2. Two-mile segments absolute value change in ranks.

Concurrent Segment	Absolute Value Rank Shift	Top 50 Locations		Top 100 Locations		Top 200 Locations	
Average Rank 1-mile	0	0	(0%)	0	(0%)	0	(0%)
	1 - 25	1	(2%)	1	(1%)	1	(1%)
	26 - 100	10	(20%)	11	(11%)	13	(7%)
	101 - 200	15	(30%)	23	(23%)	35	(18%)
	> 200	24	(48%)	65	(65%)	151	(76%)
High Rank 1-mile	0	2	(4%)	2	(2%)	3	(2%)
	1 - 25	32	(64%)	50	(50%)	73	(37%)
	26 - 100	14	(28%)	34	(34%)	85	(43%)
	101 - 200	2	(4%)	13	(13%)	28	(14%)
	> 200	0	(0%)	1	(1%)	11	(6%)
Low Rank 1-mile	0	0	(0%)	0	(0%)	0	(0%)
	1 - 25	1	(2%)	1	(1%)	1	(1%)
	26 - 100	7	(14%)	8	(8%)	8	(4%)
	101 - 200	4	(8%)	5	(5%)	11	(6%)
	> 200	38	(76%)	86	(86%)	180	(90%)
Average Rank ½ -mile	0	0	(0%)	0	(0%)	0	(0%)
	1 - 25	0	(0%)	0	(0%)	0	(0%)
	26 - 100	0	(0%)	0	(0%)	0	(0%)
	101 - 200	1	(2%)	1	(1%)	1	(1%)
	> 200	49	(98%)	99	(99%)	199	(100%)
High Rank ½ -mile	0	0	(0%)	0	(0%)	0	(0%)
	1 - 25	28	(56%)	37	(37%)	51	(26%)
	26 - 100	18	(36%)	42	(42%)	90	(45%)
	101 - 200	2	(4%)	14	(14%)	32	(16%)
	> 200	2	(4%)	7	(7%)	27	(14%)
Low Rank ½ -mile	0	0	(0%)	0	(0%)	0	(0%)
	1 - 25	0	(0%)	0	(0%)	0	(0%)
	26 - 100	0	(0%)	0	(0%)	0	(0%)
	101 - 200	0	(0%)	0	(0%)	0	(0%)
	> 200	50	(100%)	100	(100%)	200	(100%)

Maximum Ranking. The maximum ranking (largest rank value) of corresponding one and half-mile segments are listed in Table 3 for the top 50, 100, and 200 two-mile locations. For example, of the top 50 two-mile segments, the maximum rank of the set of average one-mile ranks is 1072.5. However, only considering the set of corresponding one-mile high ranks, the maximum rank is only 163. Also, only considering the set of corresponding one-mile low ranks, the maximum rank is 1982. It is expected that the maximum rank of the “low rank segment” be the largest compared to the maximum rank of the “average” and “high rank” segments. The same trend appears in the results of the maximum rank of the corresponding half-mile segments. Unlike the results in Table 2-3 for the top 50 two-mile segments, ideally, the maximum rank of the set of average one-mile ranks would be 99.5, high rank one-mile ranks would be 99, and low rank one-mile ranks would be 100. Also, the maximum rank of the set of average half-mile ranks would be 198.5, high rank half-mile ranks would be 197, and low rank half-mile ranks would be 200.

Table 2-3. Two-mile segment maximum rank.

Top Locations	Average Rank 1-mile Segment	High Rank 1-mile Segment	Low Rank 1-mile Segment	Average Rank ½ - mile Segment	High Rank ½ -mile Segment	Low Rank ½ -mile Segment
50 Locations	1072.5	163	1982	2096	278	2793
100 Locations	1072.5	371	1982	2116	487	2793
200 Locations	1073.5	821	1982	2120	913	2793

4.1.2 One-mile segments

One-mile segments were chosen as a baseline. Similar to the previous section, segment ranks were compared to the rank of the concurrent two-mile segment and the average, high, and low ranks of the concurrent one-half mile segments.

Ranking List Shifts. The top 50, 100 and 200 high crash locations were next identified using one-mile segments with the Iowa ranking method. For each one-mile segment, there are two corresponding half-mile segments and only one corresponding two-mile segment. Since there is only one corresponding two-mile segment, there is only one ranking list to compare. However, since there are two corresponding half-mile segments, comparisons are made to the half-mile segments’ average, high and low rank. As previously mentioned, if segment length did not affect ranking, it would be expected that the top 100 one-mile segments correspond to the top 50 two-mile segments and the top 200 half-mile segments.

Table 2-4 shows otherwise. 38 of the top 50 high crash locations remained in the top 50 when comparing one-mile ranks to the concurrent two-mile segment. However, only 4 of the top 50 high crash locations remained in the top 50 when using “average rank” to compute half-mile ranks. Many more segments remained in the top 50 using "high rank" to compute half-mile ranks, but less segments remained in the top 50 using "low rank" to compute half-mile ranks. Similar results hold true for the top 100 and 200 high crash locations as identified using one-mile segmentation.

Table 2-4. One-mile segment shifts in rank.

	Top 50 Locations		Top 100 Locations		Top 200 Locations	
Concurrent Segment	**Shift out**	**Percentage Shift**	**Shift out**	**Percentage Shift**	**Shift out**	**Percentage Shift**
2-mile Rank	14	28%	20	20%	31	15.5%
Average Rank ½ -mile	46	92%	95	95%	188	94%
High Rank ½ -mile	17	34%	30	30%	42	21%
Low Rank ½ -mile	47	94%	96	96%	194	97%

Absolute Value of Ranking. The absolute value of ranking shift was calculated for corresponding two and half-mile segments, only using average, high and low rankings of corresponding quadruples for the top 50, 100, and 200 one-mile locations. Table 2-5 presents the results of this analysis in similar form as Table 2 where the results are grouped into 0, 1-25, 26-100, 101-200, and (>200) change in the various ranking positions.

Table 2-5. One-mile segments absolute value change in rank.

Concurrent Segment	Absolute Value Rank Shift	Top 50 Locations		Top 100 Locations		Top 200 Locations	
2-mile Rank	0	2	(4%)	2	(2%)	3	(2%)
	1 - 25	33	(66%)	53	(53%)	73	(37%)
	26 - 100	15	(30%)	45	(45%)	101	(51%)
	101 - 200	0	(0%)	0	(0%)	23	(12%)
	> 200	0	(0%)	0	(0%)	0	(0%)
Average Rank ½ -mile	0	0	(0%)	0	(0%)	0	(0%)
	1 - 25	1	(2%)	1	(1%)	1	(1%)
	26 - 100	4	(8%)	4	(4%)	4	(2%)
	101 - 200	9	(18%)	11	(11%)	17	(9%)
	> 200	36	(72%)	84	(84%)	178	(89%)
High Rank ½ -mile	0	1	(2%)	1	(1%)	1	(1%)
	1 - 25	35	(70%)	53	(53%)	74	(37%)
	26 - 100	11	(22%)	34	(34%)	95	(48%)
	101 - 200	2	(4%)	9	(9%)	17	(9%)
	> 200	1	(2%)	3	(3%)	13	(7%)
Low Rank ½ -mile	0	0	(0%)	0	(0%)	0	(0%)
	1 - 25	1	(2%)	1	(1%)	1	(1%)
	26 - 100	3	(6%)	3	(3%)	3	(2%)
	101 - 200	2	(4%)	2	(2%)	3	(2%)
	> 200	44	(88%)	94	(94%)	193	(97%)

Maximum Rankings. The maximum ranking (largest rank value) of corresponding two and half-mile segments are listed in Table 2-6 for the top 50, 100 and 200 one-mile locations. Again, since there is only one corresponding two-mile segment, the average, high and low rank cannot be computed.

Table 2-6. One-mile segment maximum rank.

Top Locations	2-mile Rank Segment	Average Rank ½-mile Segment	High Rank ½-mile Segment	Low Rank ½-mile Segment
50 Locations	120	1410	278	2793
100 Locations	175	1439	448	2793
200 Locations	296	1520	720	2793

4.1.3 One-half mile segments

Half-mile segments were chosen as a baseline. Now, segment ranks were compared to the rank of the concurrent two-mile and one-mile segments.

Ranking List Shifts. The top 50, 100 and 200 high crash locations were identified using half-mile segments with the Iowa ranking method. For each half-mile segment, there exists one corresponding one-mile and two-mile segments. Ideally, if segment length did not affect ranking, it would be expected that the top 200 half-mile segments correspond to the top 100 one-mile segments and top 50 two-mile segments. Results are contrary to this rationale. The portion of the concurrent two-mile segments shifting out of the top locations ranges from 18 to 28 percent. The concurrent one-mile segments had a high percentage of locations shifting out of the top ranked sites with a range from 20 to 34 percent.

Absolute Value of Ranking. The absolute value of ranking shift was calculated for corresponding rank of the two and one-mile segments for the top 50, 100 and 200 two-mile locations. Table 2-7 shows the results of this analysis classified into 0, 1-25, 26-100, 101-200, and (>200) change in the various ranking positions.

Table 2-7. One-half mile segments absolute value change in rank.

Concurrent Segment	Absolute Value Rank Shift	Top 50 Locations		Top 100 Locations		Top 200 Locations	
2-mile Rank	0	0	(0%)	0	(0%)	0	(0%)
	1 - 25	27	(54%)	38	(38%)	53	(27%)
	26 - 100	19	(38%)	49	(49%)	105	(53%)
	101 - 200	4	(8%)	13	(13%)	40	(20%)
	> 200	0	(0%)	0	(0%)	2	(1%)
1-mile Rank	0	1	(2%)	1	(1%)	1	(1%)
	1 - 25	37	(74%)	54	(54%)	77	(39%)
	26 - 100	12	(24%)	44	(44%)	108	(54%)
	101 - 200	0	(0%)	1	(1%)	13	(7%)
	> 200	0	(0%)	0	(0%)	1	(1%)

Maximum Rankings. The maximum ranking of corresponding two and one-mile segments were calculated for the top 50, 100 and 200 half-mile locations. The maximum rank of the set of concurrent two-mile segment ranks was 164 for the top 50 locations and 464 for the top 200 locations. The maximum rank of the set of concurrent one-mile segment ranks was 143 for the top 50 locations and 381 for the top 200 locations.

4.2 Secondary LVRR segmentation length analysis

As mentioned in the sensitivity analysis, there was no rationale found in the published literature for choosing a fixed length for segmentation. Two, one, and one-half mile lengths were analyzed for the two-lane rural primary road system in Northwest Iowa. Secondary LVRRs may require longer lengths in order to capture enough crashes to make the average annual crash frequency greater than the crash frequency standard deviation (precision). If the number of annual crashes is less than the variance of crash totals from year to year on that segment, then statistically those crashes are not over-represented. It should be noted that this measure is only used as a comparison between segments of different length and not to explicitly identify a segment as a high crash location. Paved and unpaved roads were analyzed separately.

4.2.1 Paved roads

All paved secondary two-lane roads in Iowa with an ADT of 0 to 400 were split into even-mile fixed length sections from two miles to 15 miles in length. Table 2-8 displays the number of sections for each fixed length subset that were found to have an average annual crash frequency larger than the standard deviation of the annual crash frequencies. First, only fatal and major injury crashes (K+A) were used. Next, all injury crashes were included. Lastly, all crashes were assigned to each section.

Table 2-8. Paved LVRR length analysis results.

Length	Total Sections	Mean > Standard Deviation			% Mean > Standard Deviation		
		K+A	All Injury	All Crashes	K+A	All Injury	All Crashes
2 mi	2370	23	320	1193	0.97%	13.50%	50.34%
3 mi	1326	24	294	878	1.81%	22.17%	66.21%
4 mi	804	22	233	614	2.74%	28.98%	76.37%
5 mi	525	17	195	430	3.24%	37.14%	81.90%
6 mi	357	13	153	308	3.64%	42.86%	86.27%
7 mi	232	12	113	203	5.17%	48.71%	87.50%
8 mi	152	8	83	136	5.26%	54.61%	89.47%
9 mi	109	6	60	100	5.50%	55.05%	91.74%
10 mi	71	1	41	65	1.41%	57.75%	91.55%
11 mi	55	1	36	51	1.82%	65.45%	92.73%
12 mi	38	2	27	36	5.26%	71.05%	94.74%
13 mi	30	2	21	29	6.67%	70.00%	96.67%
14 mi	25	2	18	23	8.00%	72.00%	92.00%
15 mi	21	1	15	20	4.76%	71.43%	95.24%

The highest percentage of paved sections that have an average annual crash frequency greater than the standard deviation using only fatal and major injury crashes is 8.00 percent using 14-mile sections. Using all injury crashes also results in the highest percentage occurring at 14-mile sections (72.00 percent). All crashes included gives way to the highest percentage of 96.67 percent at 13-mile sections. Overall, a trend appears with the increase in the percentage of statistically reliable sections being directly proportional to the increase in section length. Figure 2-2 shows this trend graphically.

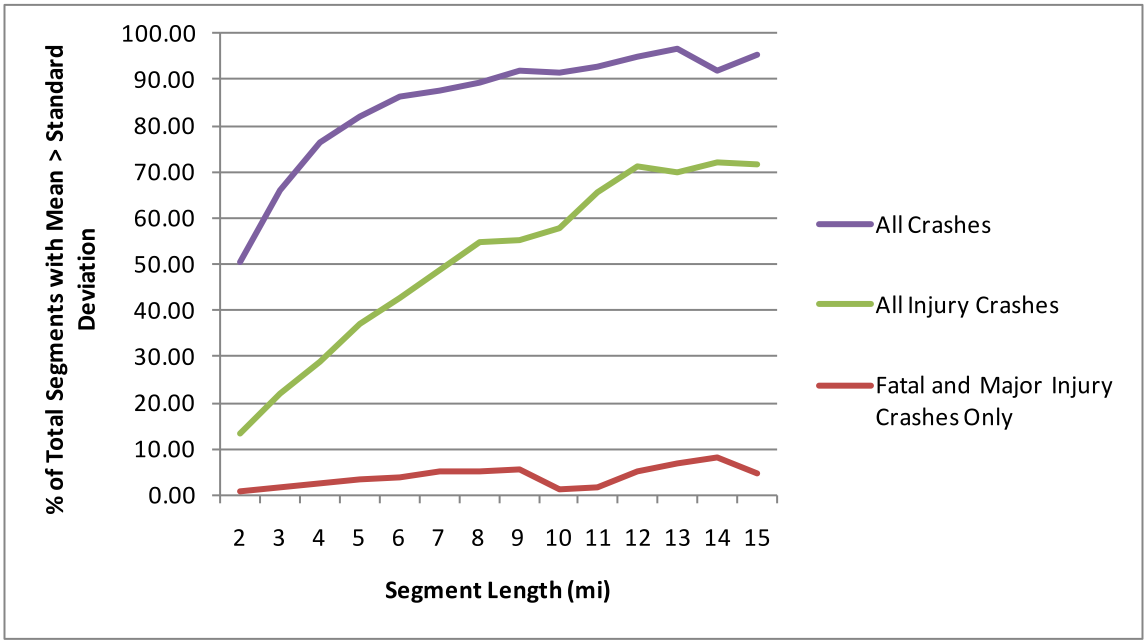

Figure 2-2. Percentage of paved segments by length with the average annual crash frequency greater than the standard deviation of the yearly crash frequencies.

4.2.2 Unpaved Roads

The same process was used for splitting unpaved roads as was for paved roads. Table 2-9 displays the same data as Table 2-8, but now only for unpaved roads instead of paved roads.

Table 2-9. Unpaved LVRR length analysis results.

Length	Total Sections	Mean > Standard Deviation			% Mean > Standard Deviation		
		K+A	All Injury	All Crashes	K+A	All Injury	All Crashes
2 mi	19654	44	927	3332	0.22%	4.72%	16.95%
3 mi	10363	46	844	2671	0.44%	8.14%	25.77%
4 mi	6297	40	704	2058	0.64%	11.18%	32.68%
5 mi	4149	29	613	1669	0.70%	14.77%	40.23%
6 mi	2771	29	494	1280	1.05%	17.83%	46.19%
7 mi	1905	22	383	973	1.15%	20.10%	51.08%
8 mi	1345	21	281	725	1.56%	20.89%	53.90%
9 mi	977	20	233	556	2.05%	23.85%	56.91%
10 mi	726	17	197	441	2.34%	27.13%	60.74%
11 mi	546	14	163	353	2.56%	29.85%	64.65%
12 mi	370	11	113	248	2.97%	30.54%	67.03%
13 mi	267	8	89	184	3.00%	33.33%	68.91%
14 mi	201	6	75	138	2.99%	37.31%	68.66%
15 mi	143	4	49	100	2.80%	34.27%	69.93%

The highest percentage of unpaved sections that have an average annual crash frequency greater than the standard deviation of the yearly crash frequencies using fatal and major injury crashes occurs at 13-mile sections with 3.00 percent. 14-mile sections have the highest percentage using all injury crashes with 37.31 percent. Lastly, using all crashes gives way to the highest percentage occurring at 15-mile sections with 69.93 percent. The trend discussed for paved roads seems to hold true also with unpaved roads where an increase in the percentage of statistically reliable sections is directly proportional to the increase in section length. Figure 2-3 shows the relationship between segment length and the percentage of statistically reliable segments.

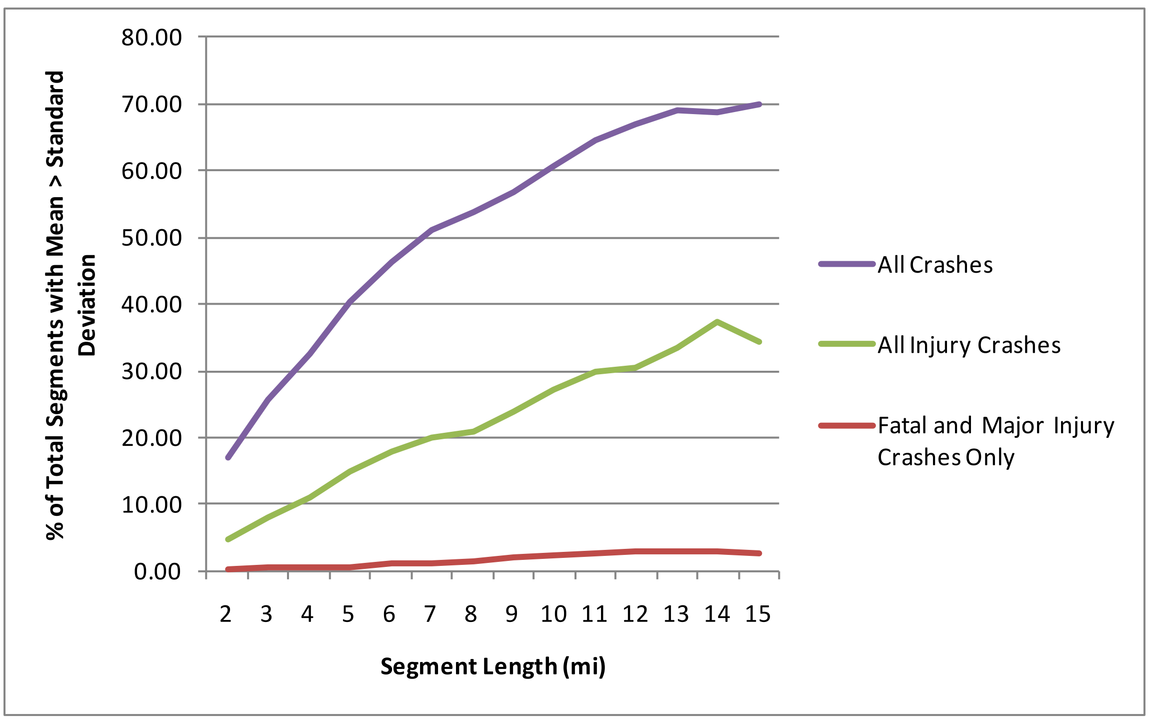

Figure 2-3. Percentage of unpaved segments by length with the average annual crash frequency greater than the standard deviation of the yearly crash frequencies.

5.0 CONCLUSIONS AND RECOMMENDATIONS

Identifying high crash locations is an important step in improving the safety of the highway network. This study has investigated the effects of segmentation length on identification of these locations.

5.1 Rural primary road segmentation

The effect of segmentation was tested using three different static, predefined lengths: two miles, one mile, and one-half mile. Locations were ranked by the Iowa DOT scoring method using each of the three lengths. Using two-mile segments as a baseline, significant shifts in rank (average and lowest ranked segment) were observed as compared to the use of one-mile and half-mile segmentation. Limited shifting was observed using the highest rank of the corresponding segments. When using one-mile segments as a baseline, a similar effect was observed with the lowest rank of the concurrent half-mile segments experiencing the largest shift in rank. The smallest effects were observed where one-half mile segmentation was used as the baseline.

It is recommended that shorter segments (half-mile in this study) be used in safety analysis of rural two-lane primary roads. However, segments that are too short may lead to difficulties in developing statistically robust models for crash location and analysis. This problem is addressed with the secondary LVRR segmentation analysis. Further studies of the effect of variable segment lengths and fixed and variable length sliding scale are recommended.

5.2 Secondary LVRR segmentation

Secondary LVRRs were split into even-mile fixed length sections from two miles to 15 miles to test the effect of segmentation length on whether or not the average annual crash frequency would be larger than the standard deviation of the yearly crash frequencies. It was found that as the segment length increased, so did the number of segments with a statistically reliable crash estimate. Also, expanding the number of crashes from only using fatal and major injury crashes to using all injury and property damage only crashes increases the number of segments with statistically reliable crash estimates. This makes sense, because adding more crashes and/or making the segments longer which also adds more crashes increases the reliability of the crash estimate.

The results show it is better to use all crash severities rather than just fatal and major injury crashes because using all crash severities produce a more robust dataset, but practitioners would argue against using all crash severities. The crashes that need to be reduced the most are fatal and major injury crashes. Safety engineers identifying high crash locations would argue that if one section with five total crashes that are all property damage only crashes should not show up as a higher crash location than another section with four total crashes in which two are fatal and/or major injury crashes. Thus, the focus of this study was to see how long secondary LVRR segments needed to be to have enough fatal and major injury crashes to make the segments' crash estimates statistically reliable.

The study shows that even with segment lengths approaching 15 miles in length, the percentage of segments in which fatal and major injury crash predictions are reliable is not much larger than 5 percent. It is recommended that further studies are needed to see how variable segment lengths and sliding scale segmentation can be implemented on secondary LVRRs.

CHAPTER 3. CRASH MODELS

1.0 INTRODUCTION

Crash frequencies and crash rates are common and straightforward methods to identify the safety performance of roadways. Another method for estimating the safety of roadways is the empirical Bayes (EB) method. This method takes into account both crash history and the safety performance of similar roadway segments. The latter is produced through the use of safety performance functions or also referred to as crash prediction models. In order for the EB method to be used on low volume rural roads (LVRRs), crash prediction models need to be developed. This study was conducted to create safety performance functions for LVRRs based on Iowa data. First a review of literature was prepared in order to explore the development of safety performance functions. Secondly, descriptive statistics were used to explore the LVRR crash data. Models for both serious crashes (fatal and major injury crashes) and all crashes were considered for three different segmentation techniques. Lastly, the models from the three segmentations were compared to test the reliability of the safety performance function in the EB process.

2.0 DESCRIPTIVE STATISTICS

The LVRR system is an extensive system of roadways in the state of Iowa, totaling nearly 80,000 miles. It is important to use descriptive statistics to get an idea of what is happening on these roadways in terms of safety. The follow sections explore several attributes of the LVRR system and its crash history. Both paved and unpaved roads are viewed as combined and separate entities. Note that all crash data is for 2001-2008.

2.1 All secondary low volume rural roads

This section looks at statistics from all secondary LVRRs. Figure 3-1 shows the total length in miles of secondary LVRRs by their average annual daily traffic (AADT). The AADT of each road was rounded up to the nearest 10 for graphical purposes. Secondary LVRRs with AADT below 100 vehicles per day (veh/day) represent 85 percent of the whole system. The remaining 15 percent of the total length consists of secondary LVRRs with an

AADT of 100 to 400 veh/day.

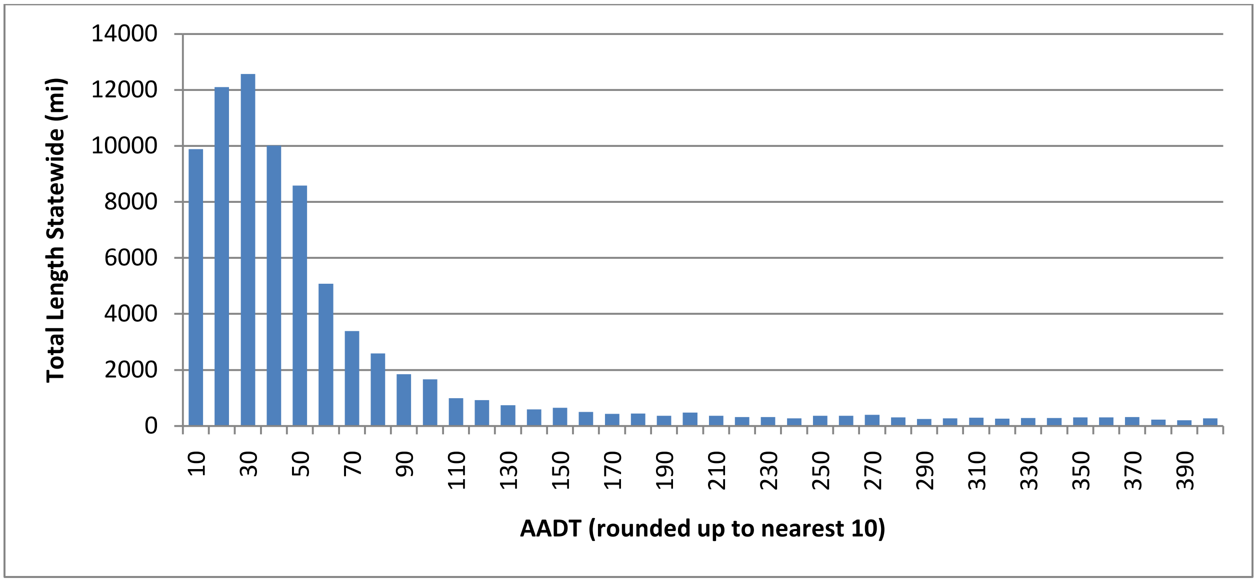

Figure 3-1. Total length of secondary LVRRs per AADT.

Figure 3-2 displays the total number of mainline crashes by their AADT for all severities. Also shown is the crash rate in units of crashes per million vehicle miles traveled (Crashes per MVMT) by AADT for all secondary LVRRs in Iowa. Nearly 60 percent of all crashes occur on roadways with less than 100 AADT. Besides the high crash rate for roads with 10 AADT and below, the crash rate is between 2 and 3 crashes per MVMT for roads with an AADT below 180. The crash rate drops below 2 crashes per MVMT for roadways with over 180 AADT.

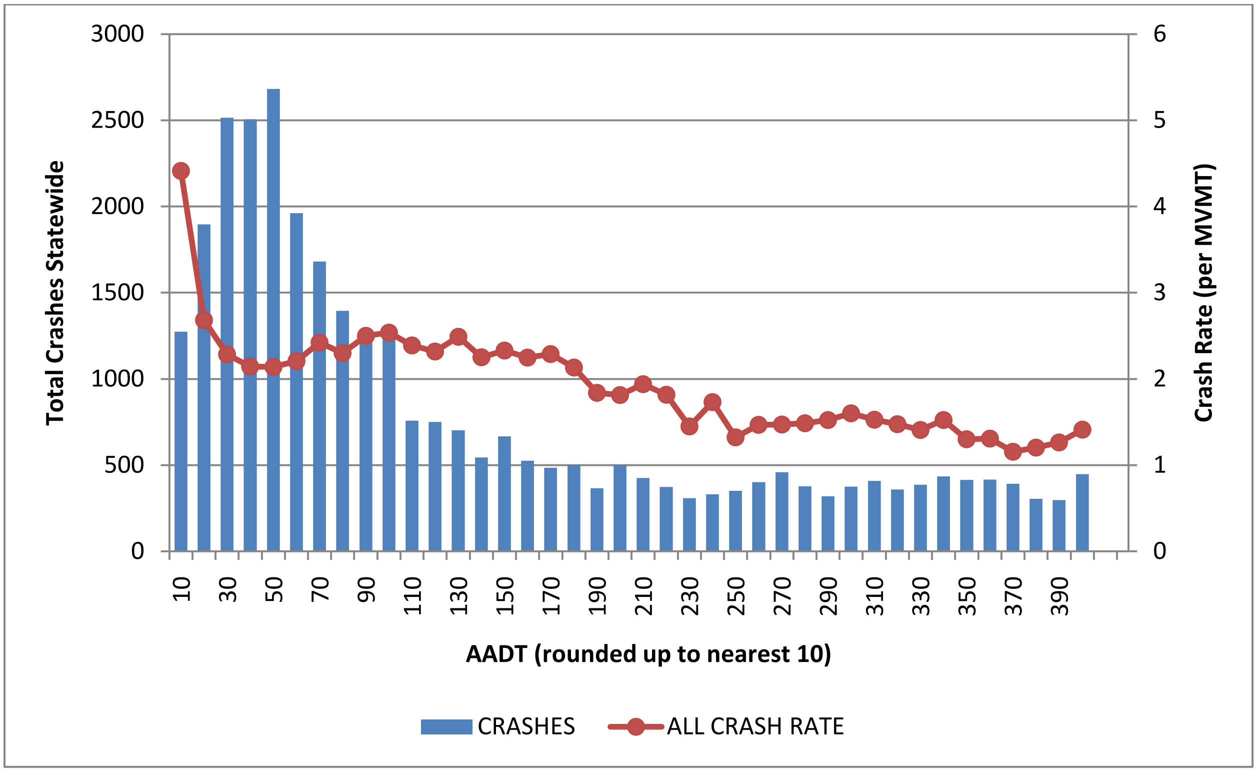

Figure 3-2. Total mainline crashes (all severities) of secondary LVRRs and crash rate (all severities) per AADT.

Figure 3-3 plots the number of mainline serious crashes statewide for all secondary LVRRs and serious crash rates per AADT. Notice the shapes of the distributions of both Figure 3-2 and Figure 3-3 are similar as well as the slight decrease in crash rates as AADT increases. Nearly 60 percent of all fatal and major injury crashes on secondary LVRRs occur on roadways with less than 100 AADT. This proportion is the same when considering all crash severities.

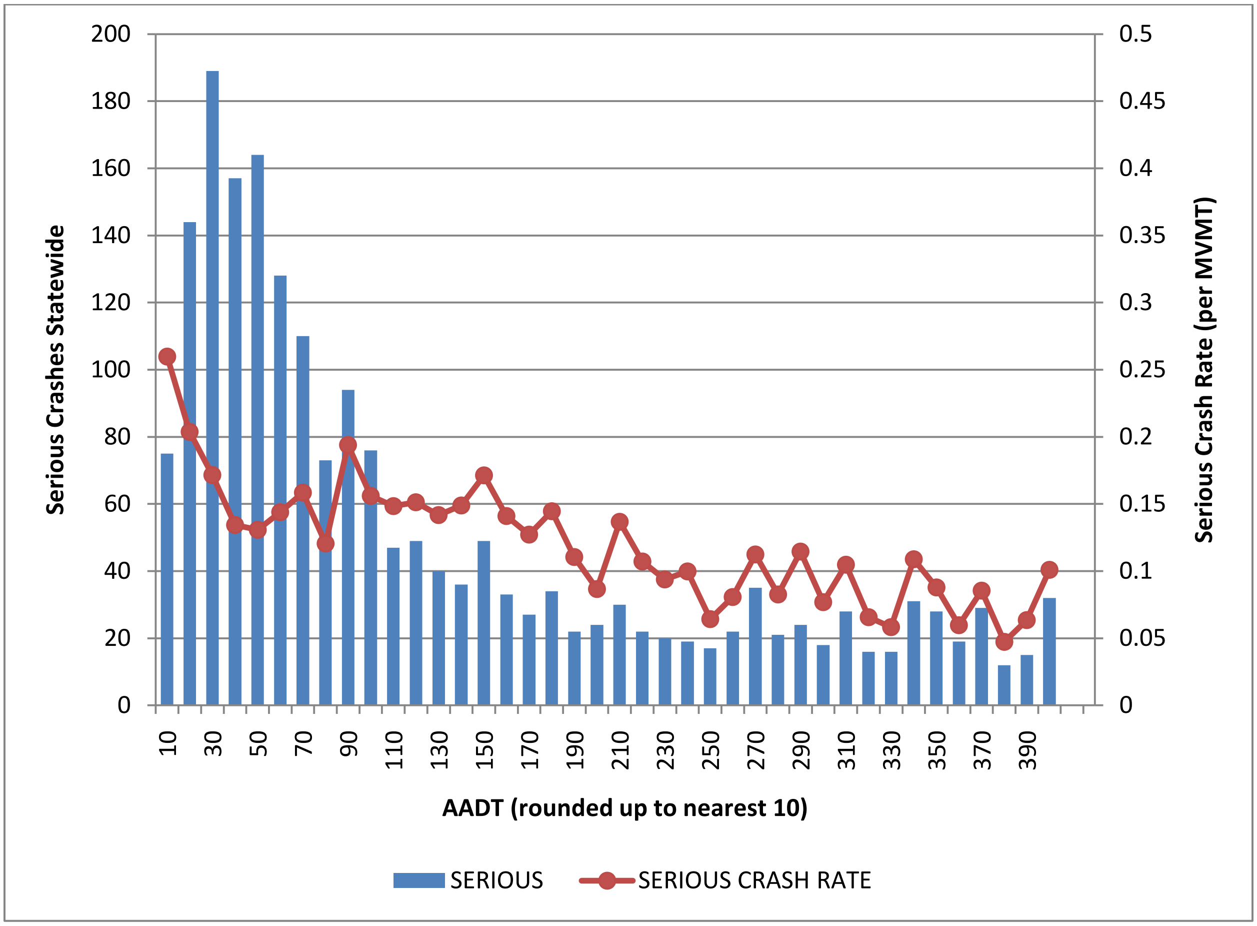

Figure 3-3. Mainline serious crashes (fatal and major injuries) of secondary LVRRs and serious crash rate per AADT.

2.2 Paved secondary low volume roads

The secondary LVRR system consists of both paved and unpaved roadways. Figure 3-4 graphs the total length of paved LVRRs in Iowa by AADT. Again, the AADT of each road was rounded up to the nearest 10 for graphical purposes. About 87 percent of the total length of paved secondary LVRRs belongs to roadways with over 100 AADT, while 85 percent of the total length for all secondary LVRRs belongs to roadways with fewer than 100 AADT. This difference is nearly the exact opposite.

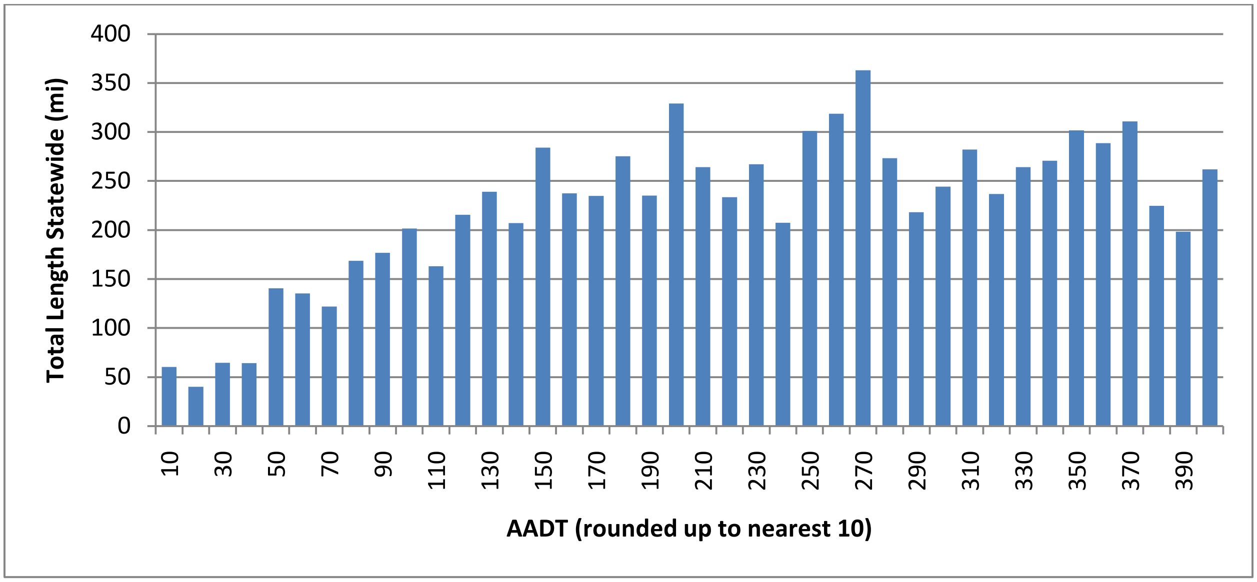

Figure 3-4. Total length of paved secondary LVRRs by AADT.

The number of all mainline crashes and crash rates for all severities by AADT is shown in Figure 3-5. 92 percent of all crashes on paved secondary LVRRs occur on roads with greater than 100 AADT, compared to only 40 percent on all secondary LVRRs with greater than 100 AADT. The crash rate slightly decreases as AADT increases besides the very high crash rates for roads with less than 30 AADT.

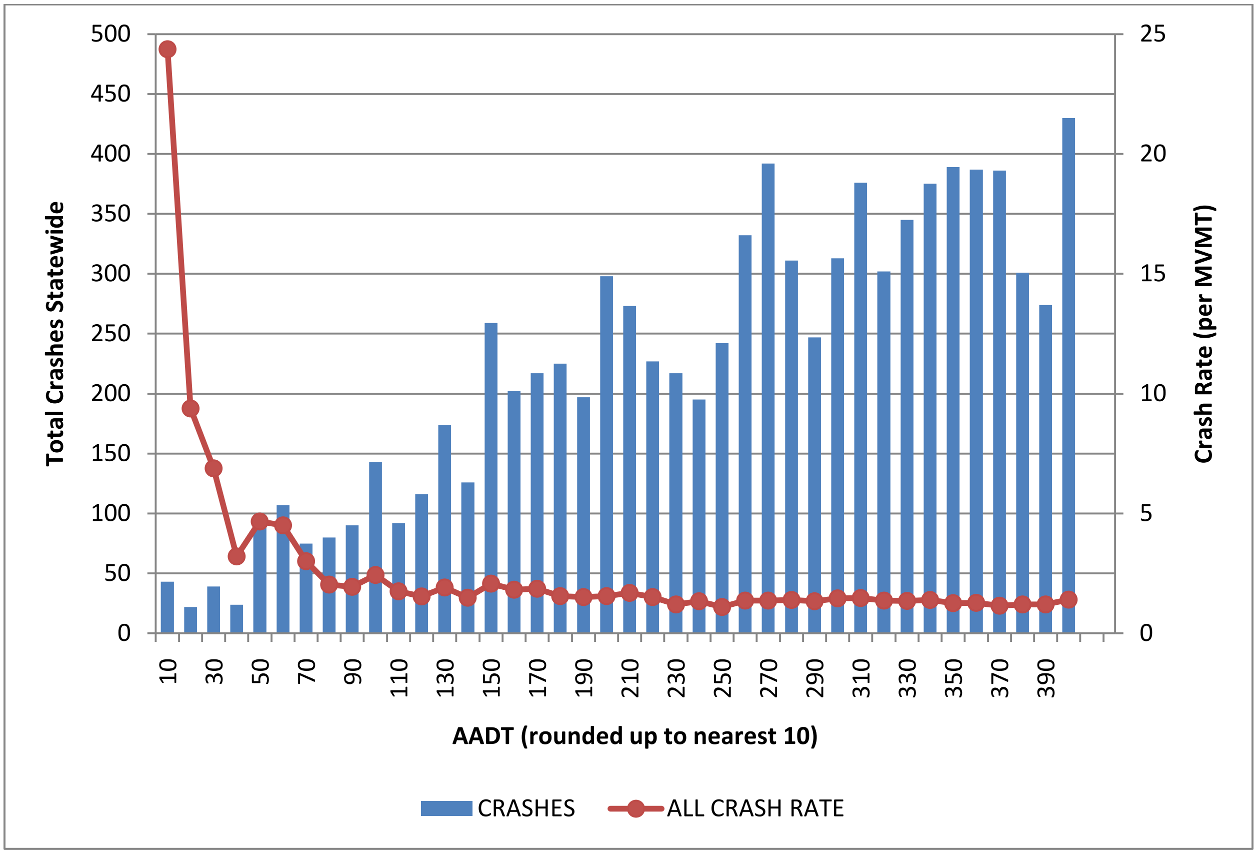

Figure 3-5. Total mainline crashes (all severities) of paved secondary LVRRs and crash rate (all severities) per AADT.

Figure 3-6 displays the number of mainline serious crashes and serious crash rates on paved secondary LVRRs by AADT. The distribution of serious crashes is comparable to the distribution of total crashes shown in Figure 3-5. 91 percent of all serious crashes occur on paved secondary LVRRs with greater than 100 AADT. Below 100 AADT, the serious crash rate is highly variable. It should be noted that because of the very low exposure of these very low volume roads, it may only take one or two serious crashes to create a very large crash rate.

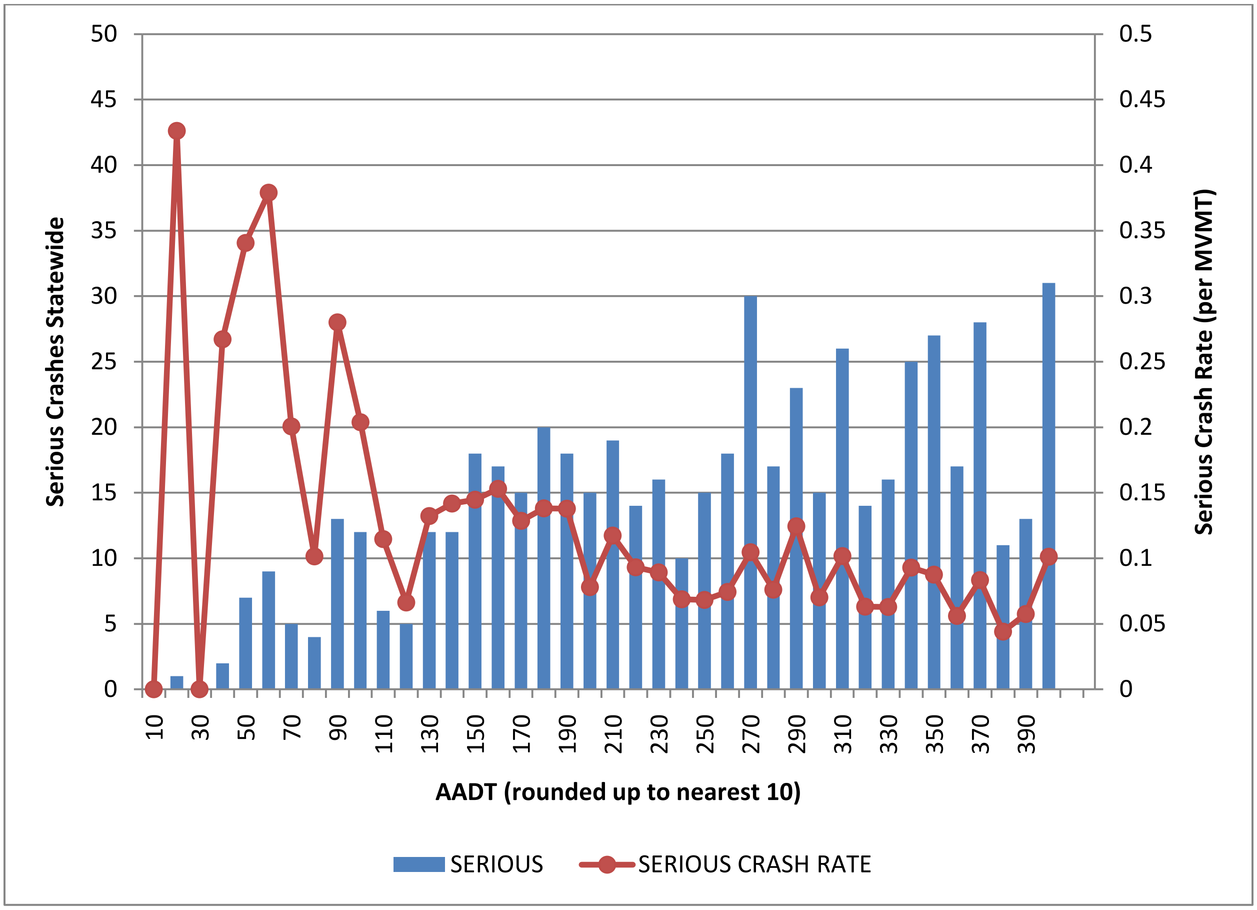

Figure 3-6. Serious mainline crashes (fatal and major injuries) of paved secondary LVRRs and serious crash rate per AADT.

Table 3-1 shows the total number of mainline crashes on paved secondary LVRRs with 1-99 AADT by crash severity and several roadway characteristics. The shoulder and land widths are in feet. The shoulder type is split into four categories: earth, gravel, paved and no shoulder. Terrain is separated into three classifications: flat, rolling and hilly. The terrain refers to the land surrounding the roadway. It is important to keep in mind the classes of each attribute are not equally represented, so it is difficult to see any trends.

Table 3-1. Number of mainline crashes by crash severity and shoulder width, shoulder type, land width and terrain of the roadway for paved secondary LVRRs with 1-99 AADT.

Paved 1-99																				
	Shoulder Width (ft)						Shoulder Type				Lane Width (ft)						Terrain			
Crash Severity	0'	1'-2'	3'-4'	5'-6'	7'-8'	>8'	None	Earth	Gravel	Paved	<9'	9'	10'	11'	12'	>12'	Flat	Rolling	Hilly	Total
Fatal	0	5	1	3	0	0	0	8	0	1	0	1	0	3	5	0	4	5	0	**9**
Major Inj	0	17	7	5	2	1	0	26	6	0	0	0	3	15	11	3	14	18	0	**32**
Minor Inj	1	38	17	6	6	3	1	59	10	1	0	1	5	31	22	12	36	31	4	**71**
Poss/Unk	5	26	25	4	10	2	5	54	12	1	1	2	3	28	25	13	26	44	2	**72**
PDO	18	191	81	28	62	12	18	297	75	2	7	24	23	142	151	45	125	251	16	**392**
Total	**24**	**277**	**131**	**46**	**80**	**18**	**24**	**444**	**103**	**5**	**8**	**28**	**34**	**219**	**214**	**73**	**205**	**349**	**22**	**576**

Table 3-2 displays the total number of mainline crashes by crash severity and several roadway characteristics for paved secondary LVRRs with 100-400 AADT. As explained before, it is difficult to detect trends in the data because the classes of each attribute are not equally represented.

Table 3-2. Number of mainline crashes by crash severity and shoulder width, shoulder type, land width and terrain of the roadway for paved secondary LVRRs with 100-400 AADT.

Paved 100-400																				
	Shoulder Width (ft)						Shoulder Type				Lane Width (ft)						Terrain			
Crash Severity	0'	1'-2'	3'-4'	5'-6'	7'-8'	>8'	None	Earth	Gravel	Paved	<9'	9'	10'	11'	12'	>12'	Flat	Rolling	Hilly	Total
Fatal	0	20	50	38	11	0	0	95	24	0	0	3	6	80	28	2	35	72	12	**119**
Major Inj	7	135	154	87	29	4	7	318	88	3	0	8	21	291	68	28	139	243	34	**416**
Minor Inj	4	338	345	235	79	25	4	801	215	6	1	23	61	686	192	63	300	636	90	**1026**
Poss/Unk	10	343	384	219	92	18	10	834	213	9	3	33	70	716	189	55	383	581	102	**1066**
PDO	76	1508	1970	1442	569	171	76	4225	1409	26	6	121	263	4038	989	319	1733	3542	461	**5736**
Total	**97**	**2344**	**2903**	**2021**	**780**	**218**	**97**	**6273**	**1949**	**44**	**10**	**188**	**421**	**5811**	**1466**	**467**	**2590**	**5074**	**699**	**8363**

2.3 Unpaved secondary low volume roads

The total length of unpaved secondary LVRRs by AADT is shown in Figure 3-7. About 94 percent of the total length of unpaved secondary LVRRs in Iowa has AADT less than 100. Because 89 percent of the secondary LVRR system is unpaved, the distribution of Figure 3-7 heavily influences the distribution of Figure 3-1.

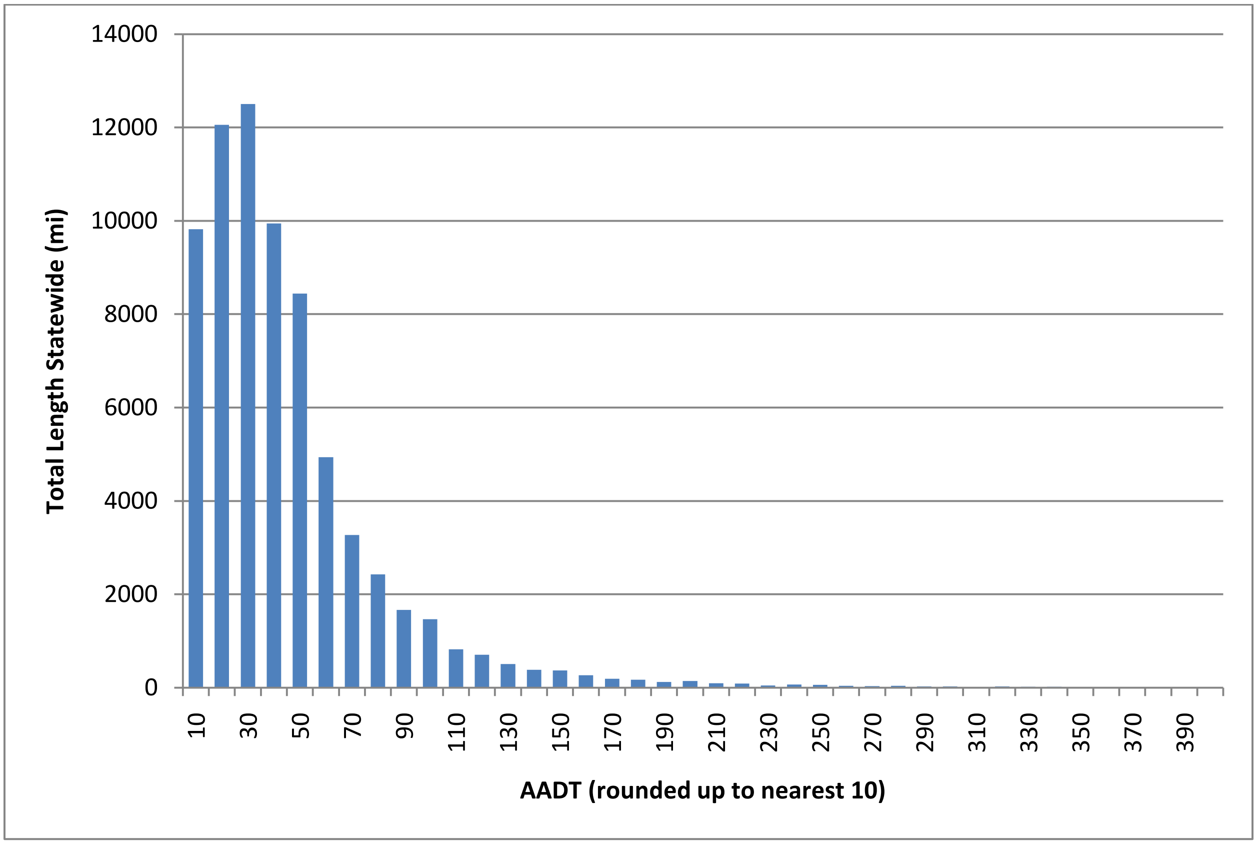

Figure 3-7. Total length of unpaved secondary LVRRs by AADT in Iowa.

Figure 3-8 plots total mainline crashes and crash rates of unpaved secondary LVRRs by AADT. 77 percent of all crashes on the unpaved secondary LVRR system occur on roads with less than 100 AADT. The crash rate appears to fluctuate between 2 and 3 crashes per MVMT for most of the AADT groups.

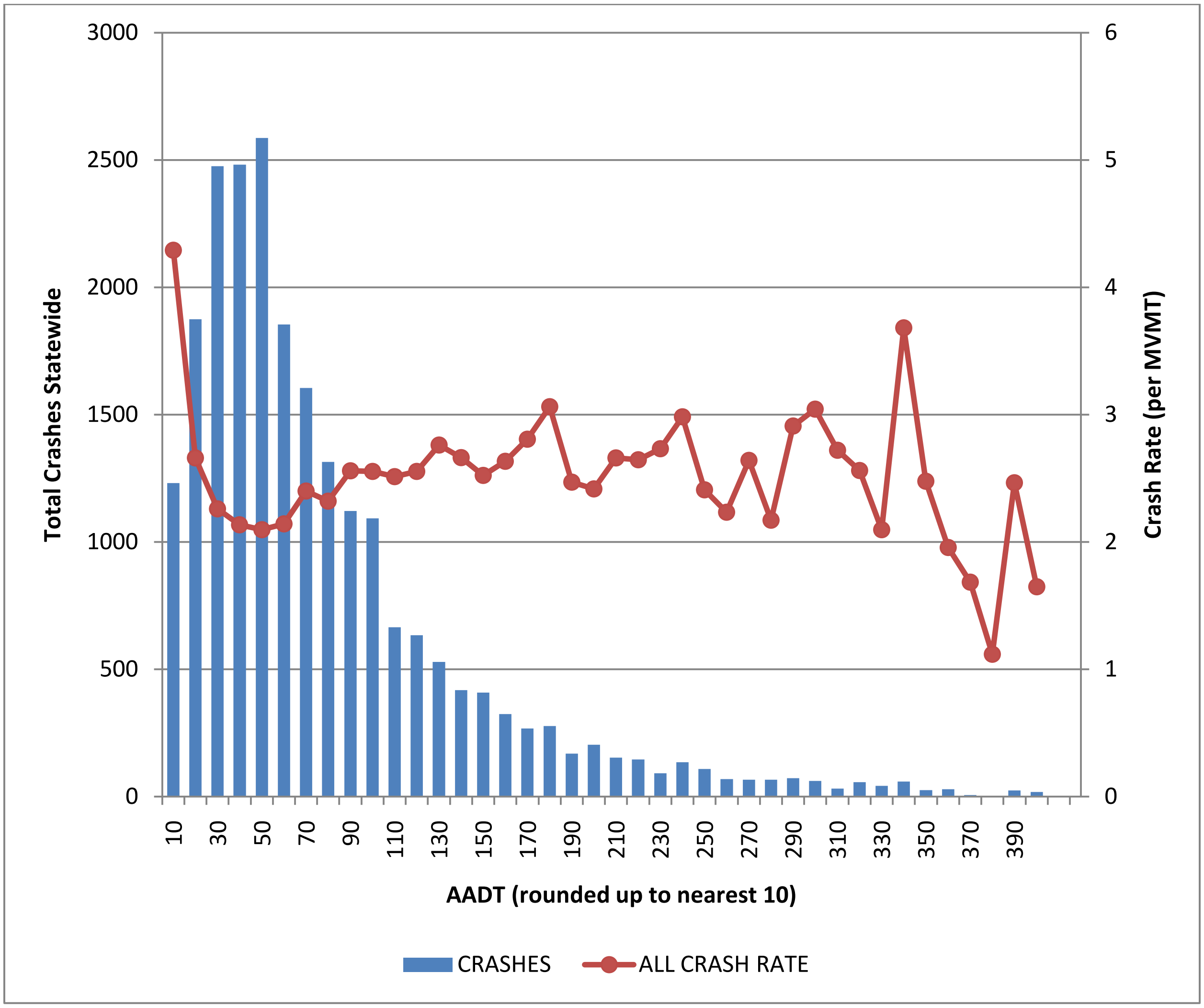

Figure 3-8. Total mainline crashes (all severities) of unpaved secondary LVRRs and crash rate (all severities) per AADT.

Figure 3-9 shows the total number of serious mainline crashes and serious crash rates for unpaved secondary LVRRs by AADT. Notice the distribution of crashes in both Figure 3-8 and Figure 3-9 are similar. 80 percent of serious crashes on unpaved secondary LVRRs occur on roadways with less than 100 AADT. The serious crash rate is quite variable when the AADT exceeds 150.

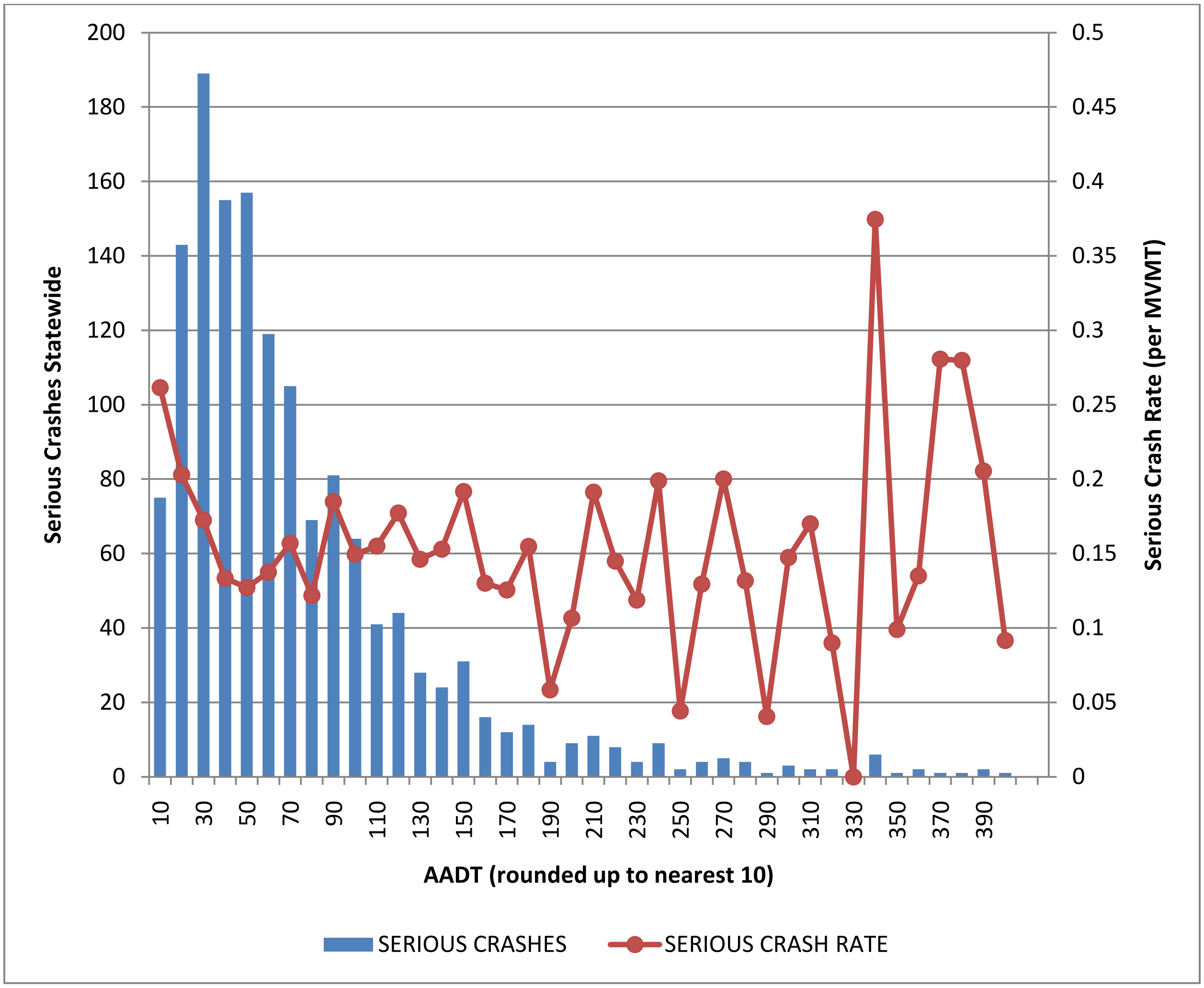

Figure 3-9. Serious mainline crashes (fatal and major injuries) of paved secondary LVRRs and serious crash rate per AADT.

The number of mainline crashes on unpaved secondary LVRRs with 1-99 AADT is shown in Table 3-3 by crash severity and two roadway attributes. Table 3-4 shows the same statistics as Table 3-3, except unpaved secondary LVRRs with 100-400 AADT data is given. Similar to Table 3-1 and Table 3-2, it is difficult to detect any trends in these tables because the classes of each attribute are not equally represented.

Table 3-3. Number of mainline crashes by crash severity and land width and terrain of the roadway for unpaved secondary LVRRs with 1-99 AADT.

Unpaved 1-99										
	Lane Width (ft)						Terrain			
Crash Severity	<9'	9'	10'	11'	12'	>12'	Flat	Rolling	Hilly	Total
Fatal	5	10	27	70	62	45	64	136	19	**219**
Major Inj	11	23	93	248	286	213	224	598	52	**874**
Minor Inj	31	71	251	804	843	675	823	1685	167	**2675**
Poss/Unk	36	90	312	847	850	661	812	1781	203	**2796**
PDO	150	313	1083	3034	3146	2255	2672	6490	819	**9981**
Total	**233**	**507**	**1766**	**5003**	**5187**	**3849**	**4595**	**10690**	**1260**	**16545**

Table 3-4. Number of mainline crashes by crash severity and land width and terrain of the roadway for unpaved secondary LVRRs with 100-400 AADT.

Unpaved 100-400										
	Lane Width (ft)						Terrain			
Crash Severity	<9'	9'	10'	11'	12'	>12'	Flat	Rolling	Hilly	Total
Fatal	0	2	0	13	21	25	18	37	6	**61**
Major Inj	0	2	16	65	92	120	80	195	20	**295**
Minor Inj	4	8	50	245	296	382	261	645	79	**985**
Poss/Unk	3	12	73	314	330	440	321	754	97	**1172**
PDO	18	46	258	1024	1080	1315	950	2354	437	**3741**
Total	**25**	**70**	**397**	**1661**	**1819**	**2282**	**1630**	**3985**	**639**	**6254**

2.4 Relationship between serious crashes and all crashes

Serious crash frequency is important to consider when addressing the safety of a roadway. Using serious crashes to identify locations for crash mitigation measures is common practice, however identifying high serious crash locations on secondary LVRRs is very difficult. Because of the large extent of the secondary LVRR system and low traffic volumes, serious crash densities are very low. Using all crashes to identify high crash locations is more practical on secondary LVRRs. If one could use total crashes to predict the number of serious crashes, then it may be possible to use all crashes instead of serious crashes to identify high crash locations and still represent the locations that are high serious crash locations.

Figure 3-10 shows the number of total mainline crashes and serious mainline crashes by AADT for secondary LVRRs in Iowa. Also plotted on the graph is the percentage of total crashes that are serious. The percent serious line is fairly constant at around 5 to 7 percent. In Figure 3-11 the same fairly constant percent serious line is seen for paved secondary LVRRs between 100-400 AADT, however in this case, it fluctuates between 5 and 9 percent. There are not many crashes occurring on paved secondary LVRRs with fewer than 100 AADT, so the percent serious crash line has a large variation because it is more sensitive to the existence of a serious crash. Only considering unpaved secondary LVRRs in Figure 3-12 gives way to a fairly constant percent serious line varying between 5 and 8 percent for roadways under 180 AADT. As for paved roads under 100 AADT, there are very few crashes occurring under 200 AADT on unpaved secondary LVRRs, so the percent serious crash line has a larger variation as it is more sensitive to the occurrence of a serious crash.

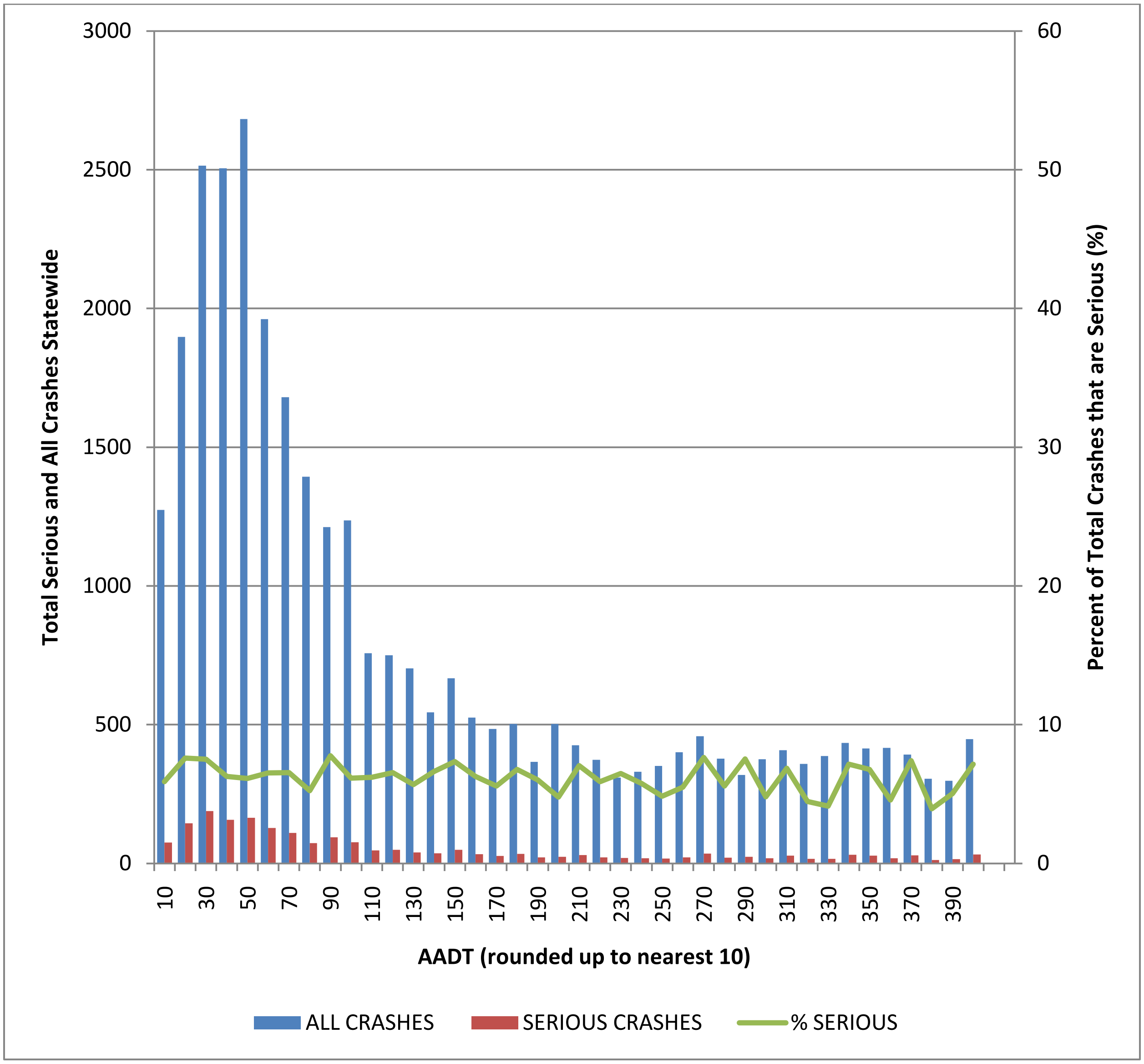

Figure 3-10. Total serious and all mainline crashes statewide and percent of total crashes that are serious per AADT for all secondary LVRRs in Iowa.

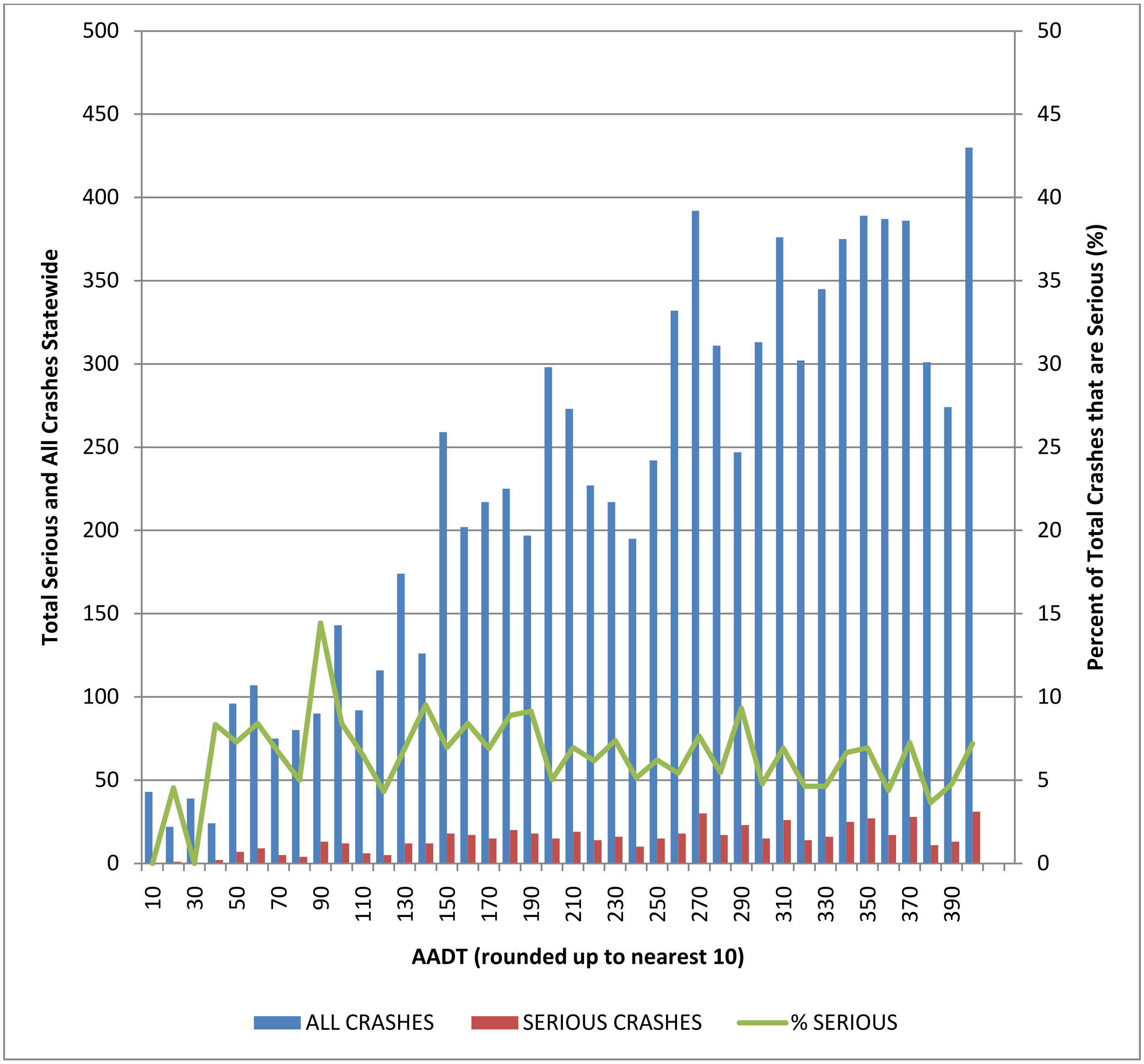

Figure 3-11. Total serious and all mainline crashes statewide and percent of total crashes that are serious per AADT for paved secondary LVRRs in Iowa.

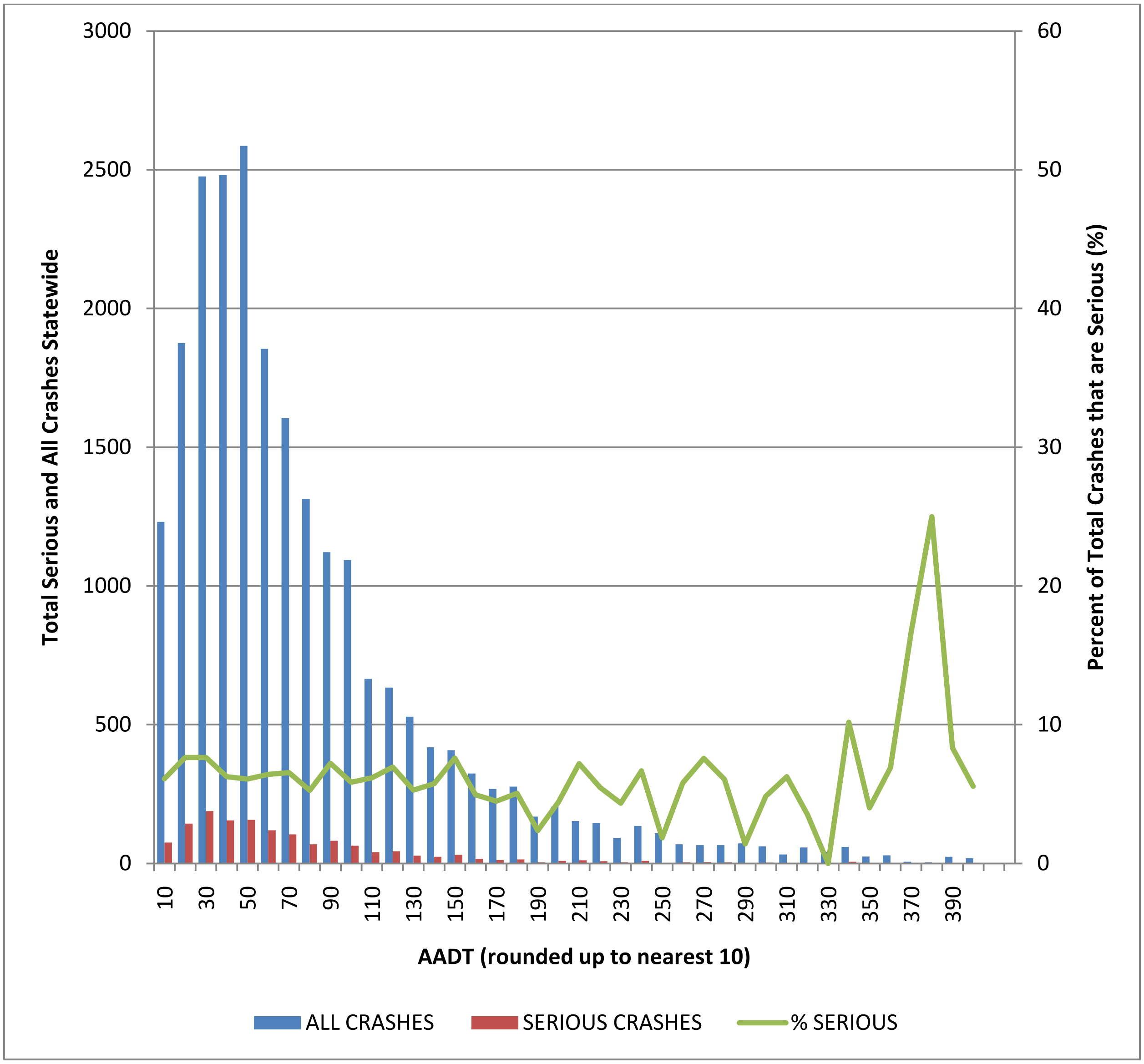

Figure 3-12. Total serious and all mainline crashes statewide and percent of total crashes that are serious per AADT for unpaved secondary LVRRs in Iowa.

The data shown in these figures indicate there may be a direct relationship between serious crashes and all crashes. Of course more work needs to be done to prove if this relationship exists. The reason for showing there may be a connection between total crash and serious crash occurrence is because all crashes were used to examine different high crash identification methods instead of serious crashes in Chapter 4. This will be further explained in that chapter. It seems for secondary LVRR mainlines, the percentage of total crashes that are serious is around 5 to 10 percent. However, this correlation may be different for other types of roadways such as urban intersections, rural freeways, urban arterials, etc.

2.5 General thoughts of descriptive statistics

The descriptive statistics shown in the previous sections have shown there is a definite difference between crash occurrence on paved and unpaved secondary LVRRs. There is a direct relationship between crash frequency and traffic volume. Crash rates appear to be higher on unpaved roadways than paved. It seems that 100 AADT is the point in which roads begin to be paved, as there are very few roads with fewer than 100 AADT. There are also very few roads with over 100 AADT. It may be better to separate paved and unpaved roadways as well as further separate these two into roads with 1-99 AADT and 100-400 AADT for modeling.

3.0 METHODOLOGY

3.1 Data collection/preparation

Roadway and crash data were collected for all secondary LVRRs in Iowa. Crash data was obtained from the Iowa SAVER crash database for 2001-2008. Roadway data was gathered from the Iowa GIMS roadway database for 2008. Crashes were assigned to the GIMS roadway network using a spatial join in ArcGIS. These crashes are only mainline crashes. No intersection crashes were included in the database.

3.2 Segmentation

Before crash models can be produced, the road network needs to be segmented. The previous chapter explored the effect of segment length on roadway safety analysis and discusses the different types of segmentation in practice. For that study, a segmentation was developed for secondary LVRRs that grouped the segments from the GIMS roadway database into larger continuous segments based on continuity of the roadway. Along with this segmentation, another segmentation was developed for crash modeling that is a discontinuous segmentation that focuses on grouping GIMS database segments based on homogenous roadway characteristics. Finally, the GIMS segmentation will also be used for developing crash models as a comparison to the continuous and discontinuous segmentations.

3.2.1 Continuous segments

Continuous segments were combined from the GIMS roadway database segments under the conditions that the GIMS segments have the same road name or route number, are within the same county, have the same road surface (paved or unpaved) and are continuous. The purpose of segmenting the roadways in this manner is to have the longest segments possible to capture the most crashes possible. The analysis of both paved and unpaved secondary LVRRs in the previous chapter reveal that as segment length increases, so does the percentage of segments with a statistically reliable crash frequency. Also, continuous segments are practical for safety mitigation measures to be installed in one continuous segment instead of a discontinuous segment in which the segment exists in several pieces not necessarily in the same proximity. The continuous segmentation was split into four groups for creating safety performance functions: paved 1-99 AADT, paved 100-400 AADT, unpaved 1-99 AADT, and unpaved 100-400 AADT.

3.2.2 Discontinuous segments

As explained in the previous section, continuous segments are more practical for crash mitigation measures to be installed; however, discontinuous segments have more homogenous roadway characteristics. Unpaved GIMS segments were combined based on having the same following characteristics: terrain, surface width and AADT. Paved GIMS segments were combined based on having the same following characteristics: terrain, surfaced width, shoulder type, shoulder width and AADT.

Paved 1-99 AADT roads. Several rules were established for combining GIMS segments into discontinuous segments. These rules are listed in order of importance. First, the terrain had to be the same (flat, rolling, or hilly). Next the shoulder type had to be constant (earth, gravel, paved or no shoulder). Third, the shoulder width needed to be the same, which was based on six shoulder width bins: 0 feet, 1 foot, 2-3 feet, 4-5 feet, 6-7 feet and 8 or more feet. Fourth, the surface width of the roadway (distance between shoulders) had to be the same, which was also based on 7 surface width bins: 16 or under feet, 17-19 feet, 20-21 feet, 22-23 feet, 24-25 feet, 26-27 feet, and 28 or more feet. Lastly, the AADT needed to be the same, which was based on 5 AADT bins: 1-20, 21-40, 41-60, 61-80, and 81-99 AADT.

The GIMS segments were first sorted in order of the categories described in the previous paragraph. The GIMS segments were lastly sorted by road name or route number, whichever one applied to the segment. The first discontinuous segment was created by starting at the top of the sorted database, and GIMS segments were combined for this discontinuous segment until either there were no more GIMS segments that contained the same roadway features or the length of the discontinuous segment exceeded 10 miles. This 10 mile limitation was to prevent discontinuous segments from becoming very large and to keep the focus on creating segments short enough to be identified as high crash locations. The only exception to the length rule is if there were more GIMS segments with the same road name or route number as the GIMS segment that caused the discontinuous segment to exceed 10 miles. If this was the case, then the GIMS segments with the same road name or route number would be added and the next discontinuous segment would be started.

Paved 100-400 AADT roads. The procedure that was used for aggregated the paved secondary 1-99 AADT LVRRs applies for combining paved secondary LVRRs with 100-400 AADT with a few exceptions. Because there are much more paved roads with 100-400 AADT than 1-99 AADT (Figure 3-4), some of the attribute bins were changed to represent the characteristics of the paved 100-400 AADT LVRRs. First, the shoulder width bins were changed to the following: 0 feet, 1 foot, 2 feet, 3 feet, 4 feet, 5 feet, 6 feet, 7 feet, 8 feet, and 9 or greater feet. Second, surface width bins were changed to the following: 1-17 feet, 18-19 feet, 20-21 feet, 22-23 feet, 24-25 feet, 26-27 feet, and 28 or greater feet. Lastly, the AADT bins were changed to the following: 100-199, 200-299, and 300-400 AADT.

Unpaved 1-99 AADT roads. Once again, the same procedure was used to aggregate unpaved secondary LVRRs with 1-99 AADT into discontinuous segments. Shoulder width and type were not used as a constraint. The surface width bins were changed to represent the characteristics of the unpaved 1-99 AADT LVRRs. Surface width bins were changed to the following: 1-18 feet, 19-20 feet, 21-22 feet, 23-24 feet, 25-26 feet, and 27 or greater feet.

Unpaved 100-400 AADT roads. The same attribute bins were used for unpaved secondary LVRRs with 100-400 AADT that were used for the unpaved 1-99 AADT roads.

3.2.3 GIMS segments

The GIMS database segmentation was also used for creating safety performance functions as a comparison to the longer continuous segmentation and the discontinuous segmentation. The GIMS segmentation was also split into four groups for modeling: Unpaved 1-99 AADT, unpaved 100-400 AADT, paved 1-99 AADT, and paved 100-400 AADT.

3.3 Negative binomial regression

The crash data for each of the 12 segmentations discussed in the previous section were fitted to a generalized linear model using negative binomial regression in SAS. Based on the literature, using the negative binomial generalized linear model is the preferred method for creating safety performance functions. See Equation 3-1 for the general form of the safety performance function.

Equation 3-1:

$$\mu = LENG\_MI * AADT^{\alpha} * e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k}$$

μ = expected number of crashes

$LENG\_MI$ = length of the segment in miles

$AADT$ = AADT of the segment

X_k = model covariates (roadway attributes)

α, β_k = model coefficients

The LENG_MI is an offset variable and is considered directly proportional to the expected number of crashes. Equation 3-1 was derived from Equation 3-2. For modeling purposes, the natural log of the AADT and length for each segment was used in the regression process in order to use the model form in Equation 3-1.

Equation 3-2:

$$\mu = e^{\ln(LENG\_MI) + \alpha \ln(AADT) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k}$$

3.4 Variables

The following variables were included in the regression process:

LENG_MI The segment length was considered an offset variable in the model.

logAADT The natural log of the segment AADT.

SHDWID The width in feet of the roadway shoulder (only paved models).

LANEWID The width in feet of the roadway lane (half the surface width)

TERRAIN 1 Equals one if terrain is flat. Zero if not.

TERRAIN 2 Equals one if terrain is rolling. Zero if not.

SHDTYPE 0 Equals one if there is no shoulder. Zero if otherwise (only paved models).

SHDTYPE 1 Equals one if the shoulder type is earth. Zero if not (only paved models).

SHDTYPE 2 Equals one if the shoulder type is gravel. Zero if not (only paved models).

LIMMPH Speed limit of the roadway in miles per hour.

PASSREST Length of passing restriction per mile of the segment in miles per mile.

There are other variables that were considered for use in the safety performance functions, but were not selected because the data could not be easily obtained for this study.

Both vertical and horizontal alignments were desired to be included in the crash models; however, Iowa does not maintain this information. Using the terrain variable helped to represent the general vertical profile of the roadway. Driveway density was also considered to be included in the crash model. This data is not included in the Iowa GIMS roadway database.

4.0 ANALYSIS

24 different safety performance functions were attempted to be created. These 24 were derived from 12 different segmentations. A serious crash model and a total crash model were formed for each segmentation. Akaike's information criterion (AIC) was used as a goodness of fit measure to select the best model. Variables were selected as long as their respective p-value indicated the coefficient of the variable was statistically significant (not zero). Four models could not be fit: GIMS paved 1-99 serious crash model; continuous paved 1-99 serious crash model; discontinuous paved 1-99 total crash model; and GIMS paved 1-99 total crash model. Each crash model is described in the following sections.

4.1 Crash Models

The general form for the crash models is shown in Equation 3-1. Each safety performance function gives an 8 year crash estimate for mainline crashes only. Dispersion factors were found to be statistically significant in all models.

4.1.1 Continuous unpaved 1-99 AADT

Table 3-5 shows the negative binomial regression output for predicting serious crashes using continuous unpaved 1-99 AADT segmentation. Included in the output is the coefficient estimate of the parameters as well as their p-values that are used to test the variables' significance in the model. The dispersion parameter and the AIC are also shown. The only covariate found significant was logAADT. Equation 3-3 shows the safety performance function for serious crashes on unpaved secondary LVRRs with fewer than 100 AADT using continuous unpaved 1-99 AADT segmentation.

Table 3-5. Serious crash model using continuous unpaved 1-99 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-7.0802	<.0001	0.5147	8509.1795
logAADT	0.8434	<.0001		

Equation 3-3:

$$\mu = LENG\_MI * AADT^{0.8434} * e^{-7.0802}$$

Table 3-6 shows the negative binomial regression output for predicting total crashes using continuous unpaved 1-99 AADT segmentation. More variables were found significant in the total crash model than in the serious crash model. Equation 3-4 shows the safety performance function for predicting total crashes on unpaved secondary LVRRs with fewer than 100 AADT using the continuous unpaved 1-99 AADT segmentation.

Table 3-6. Total crash model using continuous unpaved 1-99 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-2.8503	<.0001	0.3951	48677.6561
logAADT	0.8778	<.0001		
LANEWID	-0.0227	0.0093		
TERRAIN 1	-0.4212	<.0001		
TERRAIN 2	-0.3422	<.0001		
LIMMPH	-0.0199	<.0001		
PASSREST	0.1133	0.0143		

Equation 3-4:

$$\mu = LENG\_MI * AADT^{0.8778} * e^{-2.8503-0.0227 LANEWID-0.4212 TERRAIN1-0.3422 TERRAIN2-0.0199 LIMMPH+0.1133 PASSREST}$$

4.1.2 Continuous paved 1-99 AADT

No appropriate model was derived for predicting serious crashes on paved secondary LVRRs with fewer than 100 AADT using the continuous paved 1-99 AADT segmentation. Table 3-7 displays the negative binomial regression output for predicting total crashes using continuous paved 1-99 AADT segmentation. Equation 3-5 shows the safety performance function for predicting total crashes on paved secondary LVRRs with 1-99 AADT using the continuous paved 1-99 AADT segmentation.

Table 3-7. Total crash model using continuous paved 1-99 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-2.9013	0.0009	1.2512	1422.0357
logAADT	0.2424	0.0663		
LANEWID	0.1674	0.0016		
SHDWID	0.0622	0.0428		
LIMMPH	-0.0128	0.0442		

Equation 3-5:

$$\mu = LENG\_MI * AADT^{0.2424} * e^{-2.9013+0.1674 LANEWID+0.0622 SHDWID-0.0128 LIMMPH}$$

4.1.3 Continuous unpaved 100-400 AADT

Table 3-8 shows the negative binomial regression output for predicting serious crashes using continuous unpaved 100-400 AADT segmentation. No covariates were found significant besides logAADT. Equation 3-6 presents the safety performance function for predicting serious crashes on unpaved secondary LVRRs with 100-400 AADT using the continuous unpaved 100-400 AADT segmentation.

Table 3-8. Serious crash model using continuous unpaved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-7.2111	<.0001	0.1846	1492.0976
logAADT	0.9008	<.0001		

Equation 3-6:

$$\mu = LENG\_MI * AADT^{0.9008} * e^{-7.2111}$$

The negative binomial regression output for predicting total crashes using continuous unpaved 100-400 AADT segmentation is shown in Table 3-9. Along with logAADT, TERRAIN was found significant in the model. Because TERRAIN is a categorical variable, two dummy variables were created called TERRAIN 1 and TERRAIN 2. Equation 3-7 gives the safety performance function for predicting total crashes on unpaved secondary LVRRs with 100-400 AADT using the continuous unpaved 100-400 AADT segmentation.

Table 3-9. Total crash model using continuous unpaved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-4.2806	<.0001	0.4191	7829.1834
logAADT	0.9594	<.0001		
TERRAIN 1	-0.3107	0.0011		
TERRAIN 2	-0.2490	0.0067		

Equation 3-7:

$$\mu = LENG\_MI * AADT^{0.9594} * e^{-4.2806-0.3107 TERRAIN1-0.249 TERRAIN2}$$

4.1.4 Continuous paved 100-400 AADT

Table 3-10 displays the negative binomial regression output for predicting serious crashes using continuous paved 100-400 AADT segmentation. LogAADT, SHDWID and TERRAIN are found significant in the model. See Equation 3-8 for the safety performance function for predicting serious crashes on paved secondary LVRRs with 100-400 AADT using the continuous paved 100-400 AADT segmentation.

Table 3-10. Serious crash model using continuous paved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-5.4663	<.0001	0.2104	2442.7611
logAADT	0.6612	<.0001		
SHDWID	-0.1216	<.0001		
TERRAIN 1	-0.5610	0.0039		
TERRAIN 2	-0.3768	0.0417		

Equation 3-8:

$$\mu = LENG\_MI * AADT^{0.6612} * e^{-5.4663-0.1216SHDWID-0.5610TERRAIN1-0.3768TERRAIN2}$$

Table 3-11 shows the negative binomial regression output for predicting total crashes using continuous paved 100-400 AADT segmentation. In addition to the serious crash model, SHDTYPE 0, LIMMPH and PASSREST were found significant in the total crash model. SHDTYPE 0 is a dummy variable from the categorical variable SHDTYPE. SHDTYPE 0 means no shoulder exists. The other two dummy SHDTYPE variables (SHDTYPE 1 and SHDTYPE 2) were not found significant. The safety performance function for predicting total crashes on paved secondary LVRRs with 100-400 AADT using the continuous paved 100-400 AADT segmentation is shown in Equation 3-9.

Table 3-11. Total crash model using continuous paved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-3.2573	<.0001	0.3377	10783.6639
logAADT	0.7596	<.0001		
SHDWID	-0.0562	<.0001		
TERRAIN 1	-0.4363	<.0001		
TERRAIN 2	-0.3056	<.0001		
SHDTYPE 0	0.3834	0.0139		
LIMMPH	-0.0102	<.0001		
PASSREST	0.5711	<.0001		

Equation 3-9:

$$\mu = LENG\_MI * AADT^{0.7596} * e^{-3.2573-0.0562SHDWID-0.4363TERRAIN1-0.3056TERRAIN2+0.3834SHDTYPE0-0.0102LIMMPH+0.5711PASSREST}$$

4.1.5 Discontinuous unpaved 1-99 AADT

The negative binomial regression output for predicting serious crashes using discontinuous unpaved 1-99 AADT segmentation is given in Table 3-12. Only logAADT was found significant in the continuous unpaved 1-99 AADT, while for this discontinuous unpaved 1-99 AADT model, TERRAIN and PASSREST were also found significant.

Equation 3-10 shows the safety performance function for predicting serious crashes on unpaved secondary LVRRs with 1-99 AADT using discontinuous unpaved 1-99 AADT segmentation.

Table 3-12. Serious crash model using discontinuous unpaved 1-99 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-5.8410	<.0001	0.4389	6236.4660
logAADT	0.7075	<.0001		
TERRAIN 1	-0.6626	0.0001		
TERRAIN 2	-0.2780	0.0356		
PASSREST	-0.5324	0.0160		

Equation 3-10:

$$\mu = LENG\_MI * AADT^{0.7075} * e^{-5.8410-0.6626TERRAIN1-0.2780TERRAIN2-0.5324PASSREST}$$

Table 3-13 shows the negative binomial regression output for predicting total crashes using discontinuous unpaved 1-99 AADT segmentation. Every variable found significant in the continuous unpaved 1-99 AADT model are significant in the discontinuous unpaved 1-99 AADT model except for PASSREST. Equation 3-11 displays the safety performance function for predicting total crashes on unpaved secondary LVRRs with 1-99 AADT using discontinuous unpaved 1-99 AADT segmentation.

Table 3-13. Total crash model using discontinuous unpaved 1-99 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-1.9332	<.0001	0.2019	25344.3208
logAADT	0.7944	<.0001		
LANEWID	-0.0292	0.0002		
TERRAIN 1	-0.4794	<.0001		
TERRAIN 2	-0.3729	<.0001		
LIMMPH	-0.0271	<.0001		

Equation 3-11:

$$\mu = LENG\_MI * AADT^{0.7944} * e^{-1.9332-0.0292LANEWID-0.4794TERRAIN1-0.3729TERRAIN2-0.0271LIMMPH}$$

4.1.6 Discontinuous paved 1-99 AADT

The negative binomial regression output for predicting serious crashes using discontinuous paved 1-99 AADT segmentation is shown in Table 3-14. SHDWID and TERRAIN are found significant in the model along with logAADT. Equation 3-12 gives the safety performance function for predicting serious crashes on paved secondary LVRRs with 1-99 AADT using discontinuous paved 1-99 AADT segmentation.

Table 3-14. Serious crash model using discontinuous paved 1-99 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-7.4268	0.0004	0.0276	194.9705
logAADT	1.0547	0.0369		

Equation 3-12:

$$\mu = LENG\_MI * AADT^{1.0547} * e^{-7.4268}$$

No appropriate model was fit using discontinuous paved 1-99 AADT segmentation to predict total crashes on paved secondary LVRRs with 1-99 AADT.

4.1.7 Discontinuous unpaved 100-400 AADT

Table 3-15 displays the negative binomial regression output for predicting serious crashes using discontinuous unpaved 100-400 AADT segmentation. Only logAADT is found significant in the model. Note the dispersion parameter is very small. This will cause the model to have practically no weight in the empirical Bayes estimate, which will be discussed later. See Equation 3-13 for the safety performance function for predicting serious crashes on unpaved secondary LVRRs with 100-400 AADT using discontinuous unpaved 100-400 AADT segmentation.

Table 3-15. Serious crash model using discontinuous unpaved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept logAADT	-6.9487 0.8503	<.0001 <.0001	Near Zero	1171.4930

Equation 3-13:

$$\mu = LENG\_MI * AADT^{0.8503} * e^{-6.9487}$$

The negative binomial regression output for predicting total crashes using discontinuous unpaved 100-400 AADT segmentation is shown in Table 3-16. LogAADT and PASSREST are found significant. Only logAADT was found significant in the continuous unpaved 100-400 AADT model. Equation 3-14 presents the safety performance function for predicting total crashes on unpaved secondary LVRRs with 100-400 AADT using discontinuous unpaved 100-400 AADT segmentation.

Table 3-16. Total crash model using discontinuous unpaved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept logAADT PASSREST	-4.7915 0.9546 0.3105	<.0001 <.0001 0.0011	0.0631	3169.9668

Equation 3-14:

$$\mu = LENG\_MI * AADT^{0.9546} * e^{-4.7915+0.3105 PASSREST}$$

4.1.8 Discontinuous paved 100-400 AADT

The negative binomial regression output for predicting serious crashes using discontinuous paved 100-400 AADT segmentation is given in Table 3-17. The same variables found significant in the continuous paved 100-400 AADT model are also significant in the discontinuous paved 100-400 AADT model. Equation 3-15 shows the safety performance function for predicting serious crashes on paved secondary LVRRs with 100-400 AADT using discontinuous paved 100-400 AADT segmentation.

Table 3-17. Serious crash model using discontinuous paved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-4.0545	<.0001	0.1754	1851.5021
logAADT	0.4063	0.0025		
SHDWID	-0.1022	<.0001		
TERRAIN 1	-0.6181	0.0005		
TERRAIN 2	-0.4333	0.0108		

Equation 3-15:

$$\mu = LENG\_MI * AADT^{0.4063} * e^{-4.0545-0.1022SHDWID-0.6181TERRAIN1-0.4333TERRAIN2}$$

Table 3-18 displays the negative binomial regression output for predicting total crashes using discontinuous paved 100-400 AADT segmentation. The continuous paved 100-400 AADT model only found SHDTYPE 0 significant while the discontinuous paved 100-400 AADT also found SHDTYPE 1 and SHDTYPE 2 significant. Equation 3-16 shows the safety performance function for predicting total crashes on paved secondary LVRRs with 100-400 AADT using discontinuous paved 100-400 AADT segmentation.

Table 3-18. Total crash model using discontinuous paved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-2.8136	<.0001	0.1336	5743.9232
logAADT	0.6672	<.0001		
SHDWID	-0.0453	<.0001		
TERRAIN 1	-0.6426	<.0001		
TERRAIN 2	-0.4227	<.0001		
SHDTYPE 0	0.7182	0.0009		
SHDTYPE 1	0.4054	0.0161		
SHDTYPE 2	0.5439	0.0015		
LIMMPH	-0.0125	0.0001		
PASSREST	0.1884	<.0001		

Equation 3-16:

$$\mu = LENG\_MI * AADT^{0.6672} * e^{-2.8136 - 0.0453 SHDWID - 0.6426 TERRAIN1 - 0.4227 TERRAIN2 + 0.7182 SHDTYPE0 + 0.4054 SHDTYPE1 + 0.5439 SHDTYPE2 - 0.0125 LIMMPH + 0.1884 PASSREST}$$

4.1.9 GIMS unpaved 1-99 AADT

Table 3-19 gives the negative binomial regression output for predicting serious crashes using GIMS unpaved 1-99 AADT segmentation. Only logAADT was found to be significant in the model. The safety performance function for predicting serious crashes on unpaved secondary LVRRs with 1-99 AADT using GIMS unpaved 1-99 AADT segmentation is shown in Equation 3-17.

Table 3-19. Serious crash model using GIMS unpaved 1-99 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-6.4754	<.0001	0.0282	11790.8282
logAADT	0.6816	<.0001		

Equation 3-17:

$$\mu = LENG\_MI * AADT^{0.6816} * e^{-6.4754}$$

The negative binomial regression output for predicting total crashes using GIMS unpaved 1-99 AADT segmentation is presented in Table 3-20. The same variables were found significant in both the discontinuous and GIMS unpaved 1-99 AADT models; however, PASSREST was additionally found significant in the continuous unpaved 1-99 AADT model. Equation 3-18 gives the safety performance function for predicting total crashes on unpaved secondary LVRRs with 1-99 AADT using GIMS unpaved 1-99 AADT segmentation.

Table 3-20. Total crash model using GIMS unpaved 1-99 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-2.2226	<.0001	0.8936	91725.1011
logAADT	0.7319	<.0001		
LANEWID	-0.0142	0.0379		
TERRAIN 1	-0.4976	<.0001		
TERRAIN 2	-0.4002	<.0001		
LIMMPH	-0.0208	<.0001		

Equation 3-18:

$$\mu = LENG\_MI * AADT^{0.7319} * e^{-2.2226-0.0142LANEWID-0.4976TERRAIN1-0.4002TERRAIN2-0.0208LIMMPH}$$

4.1.10 GIMS paved 1-99 AADT

No appropriate models were derived for predicting both serious and total crashes on paved secondary LVRRs with 1-99 AADT using GIMS paved 1-99 AADT segmentation.

4.1.11 GIMS unpaved 100-400 AADT

The negative binomial regression output for predicting serious crashes using GIMS unpaved 100-400 AADT segmentation is shown in Table 3-21. Only logAADT was found significant in the model as was the case for both the discontinuous and continuous unpaved 100-400 AADT models. Equation 3-19 displays the safety performance function for predicting serious crashes on unpaved secondary LVRRs with 100-400 AADT using GIMS unpaved 100-400 AADT segmentation.

Table 3-21. Serious crash model using GIMS unpaved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-6.8717	<.0001	0.0327	3084.0297
logAADT	0.8254	<.0001		

Equation 3-19:

$$\mu = LENG\_MI * AADT^{0.8254} * e^{-6.8717}$$

The negative binomial regression output for predicting total crashes using GIMS unpaved 100-400 AADT segmentation is given in Table 3-22. LogAADT, LANEWID and TERRAIN were all found significant in the GIMS unpaved 100-400 AADT model, while only logAADT and TERRAIN were significant in the continuous unpaved 100-400 AADT model and only logAADT was significant in the discontinuous unpaved 100-400 AADT model. The safety performance function for predicting total crashes on unpaved secondary LVRRs with 100-400 AADT using GIMS unpaved 100-400 AADT segmentation is presented in Equation 3-20.

Table 3-22. Total crash model using GIMS unpaved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-3.9163	<.0001	0.6430	21214.4148
logAADT	1.0031	<.0001		
LANEWID	-0.0287	0.0090		
TERRAIN 1	-0.2559	<.0001		
TERRAIN 2	-0.2087	0.0001		

Equation 3-20:

$$\mu = LENG\_MI * AADT^{1.0031} * e^{-3.9163-0.0287LANEWID-0.2559TERRAIN1-0.2087TERRAIN2}$$

4.1.12 GIMS paved 100-400 AADT

Table 3-23 shows the negative binomial regression output for predicting serious crashes using GIMS paved 100-400 AADT segmentation. The same variables were found significant in both continuous, discontinuous and GIMS paved 100-400 AADT models: logAADT, SHDWID and TERRAIN. Equation 3-21 gives the safety performance function for predicting serious crashes on paved secondary LVRRs with 100-400 AADT using GIMS paved 100-400 AADT segmentation.

Table 3-23. Serious crash model using GIMS paved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-4.6103	<.0001	0.5634	4578.1870
logAADT	0.5105	<.0001		
SHDWID	-0.1063	<.0001		
TERRAIN 1	-0.6205	0.0003		
TERRAIN 2	-0.4335	0.0073		

Equation 3-21:

$$\mu = LENG\_MI * AADT^{0.5105} * e^{-4.6103-0.1063SHDWID-0.6205TERRAIN1-0.4335TERRAIN2}$$

The negative binomial regression output for predicting total crashes using GIMS paved 100-400 AADT segmentation is shown in Table 3-24. In addition to all the covariates found significant in the discontinuous paved 100-400 AADT model, LIMMPH was also found significant in the GIMS paved 100-400 AADT model. Equation 3-22 shows the safety performance function for predicting total crashes on paved secondary LVRRs with 100-400 AADT using GIMS paved 100-400 AADT segmentation.

Table 3-24. Total crash model using GIMS paved 100-400 AADT segmentation.

Parameter	Estimate	p-value	Dispersion	AIC
Intercept	-3.2840	<.0001	0.5123	28777.6409
logAADT	0.7002	<.0001		
SHDWID	-0.0492	<.0001		
TERRAIN 1	-0.5438	<.0001		
TERRAIN 2	-0.3629	<.0001		
SHDTYPE 0	0.8134	<.0001		
SHDTYPE 1	0.4506	0.0081		
SHDTYPE 2	0.5837	0.0007		
LIMMPH	-0.0104	<.0001		
PASSREST	0.4026	<.0001		

Equation 3-22:

$$\mu = LENG\_MI * AADT^{0.7002} *$$

$$e^{-3.2840-0.0492SHDWID-0.5438TERRAIN1-0.3629TERRAIN2+0.8134SHDTYPE0+0.4506SHDTYPE1+0.5837SHDTYPE2-0.0104LIMMPH+0.4026PASSREST}$$

4.2 Model Comparisons

The empirical Bayes method of estimating the number of crashes at a location is based on both a safety performance function and crash history at that location. A weight is calculated to determine how much the safety performance function crash estimate contributes to the empirical Bayes estimate. The weight ranges in value from 0 to 1. Equation 3-23 shows this calculation. Since the models derived in the previous sections produce an 8 year crash estimate and 8 years of crash data were used as the crash history component of the empirical Bayes process, the equation turns into Equation 3-24 (Hauer, 2001).

Equation 3-23:

$$Weight = \frac{1}{1+\frac{(\mu * Y)}{\varphi}}$$

μ = safety performance model estimate
Y = number of years of crash data used
φ = dispersion parameter

Equation 3-24:

$$Weight = \frac{1}{1+\frac{\mu}{\varphi}}$$

The LVRR system was split into four categories for modeling: paved 1-99 AADT, paved 100-400 AADT, unpaved 1-99 AADT and unpaved 100-400 AADT. For each of these four groups three different models exist for the three different segmentations used to derive the models. Both serious and total crash prediction models were created for each segmentation. Simply stated, three safety performance functions exist for the same road type. In order to tell which of the models is best to use, the average weight was calculated and compared in Table 3-25. For example, for the GIMS unpaved 100-400 AADT segmentation, the GIMS unpaved 100-400 AADT model was used to calculate a weight for each segment. Then the average weight was calculated. This was done for both serious

(KA) and total (ALL) models. Next, the same procedure was done except using the discontinuous and continuous unpaved 100-400 AADT models. A model that produces a higher average weight will be represented more in the empirical Bayes crash estimate.

Table 3-25. Calculated average weights for comparing models usefulness in the empirical Bayes process.

			Comparison					
			Discontinuous Model		GIMS Model		Continuous Model	
	Average Weight		Average Weight		Average Weight		Average Weight	
Segmentation	KA Model	ALL Model	KA Model	ALL Model	KA Model	ALL Model	KA Model	ALL Model
Discontinuous Paved 1-99	0.6209	-			-	-	-	0.4934
GIMS Paved 1-99	-	-	0.7891	-			-	0.9002
Continuous Paved 1-99	-	0.8216	0.6774	-	-	-		
Discontinuous Paved 100-400	0.4773	0.1323			0.669	0.2517	0.5132	0.2092
GIMS Paved 100-400	0.9455	0.5712	0.8524	0.3162			0.8753	0.4867
Continuous Paved 100-400	0.6953	0.2991	0.6594	0.1766	0.8275	0.3666		
Discontinuous Unpaved 1-99	0.736	0.106			0.188	0.3087	0.7724	0.1854
GIMS Unpaved 1-99	0.7594	0.86	0.9764	0.6208			0.9807	0.7517
Continuous Unpaved 1-99	0.9337	0.558	0.9215	0.4341	0.1063	0.6881		
Discontinuous Unpaved 100-400	0.0019	0.0066			0.0541	0.0408	0.2282	0.0354
GIMS Unpaved 100-400	0.5692	0.5058	0.0487	0.2156			0.8587	0.4733
Continuous Unpaved 100-400	0.692	0.2893	0.0333	0.0927	0.367	0.3131		

For paved secondary LVRRs with 1-99 AADT, only one model was developed for both serious and total crashes, so there were no other models to compare. It appears the GIMS paved 100-400 AADT model produced the highest average weights for predicting both serious and total crashes on paved secondary LVRRs with 100-400 AADT. The greatest average weight that was produced for predicting serious crashes on both unpaved secondary LVRRs with 1-99 AADT and 100-400 AADT were the continuous unpaved 1-99 AADT and continuous unpaved 100-400 AADT models, respectively. The GIMS unpaved 1-99 AADT and GIMS unpaved 100-400 AADT models produced the largest average weights for predicting total crashes on both unpaved secondary LVRRs with 1-99 AADT and 100-400 AADT, respectively.

5.0 RECOMMENDATIONS/CONCLUSIONS

Modeling crashes on low volume rural roads is a challenge. Little research was found on the topic of modeling crashes on secondary LVRRs. Developing safety performance functions is an essential step in the empirical Bayes method of estimating crashes. A literature review was conducted to review the literature available on producing safety performance functions. Descriptive statistics were developed to visually display the characteristics of secondary LVRRs in Iowa. Also, there appears to be a constant correlation between serious crash and total crash frequency. This shows crash models that predict total crashes can possibly be used on secondary LVRRs instead of crash models that predict serious crashes, and still produce a statistically reasonable estimate of serious crashes.

Three different segmentations were used for creating safety performance functions: continuous segmentation, discontinuous segmentation, and GIMS segmentation. Based on the descriptive statistics, it was decided to split secondary LVRRs into four different categories: unpaved 1-99 AADT, unpaved 100-400 AADT, paved 1-99 AADT, and paved 100-400 AADT. Also, serious and total crash prediction models were developed for each category. Since there were three different segmentations, 24 safety performance functions were attempted to be produced. Only 20 models were able to be created. In order to compare models that were created from the three segmentations for the same category of roadway, the average weight from the empirical Bayes procedure was used as a statistic to show how much the model was represented in the final empirical Bayes crash estimate. Table 3-26 shows the recommended crash models to be used for predicting serious and total crashes on the four different categories of secondary LVRRs.

It is also recommended that future work be conducted to examine the specification of speed limit as a categorical variable (a continuous variable in the present study). The effect of land use should also be examined, although speed limit is likely correlated with land use to some degree.

Table 3-26. Recommended safety performance functions for each secondary LVRR road type category.

Secondary LVRR Category	Serious Crash Prediction	Total Crash Prediction
Paved 1-99 AADT	Discontinuous paved 1-99 AADT serious crash model	Continuous paved 1-99 AADT total crash model
Paved 100-400 AADT	GIMS paved 100-400 AADT serious crash model	GIMS paved 100-400 AADT total crash model
Unpaved 1-99 AADT	Continuous unpaved 1-99 AADT serious crash model	GIMS unpaved 1-99 AADT total crash model
Unpaved 100-400 AADT	Continuous unpaved 100-400 AADT serious crash model	GIMS unpaved 100-400 AADT total crash model

CHAPTER 4. HIGH CRASH LOCATION METHOD PERFORMANCE

1.0 INTRODUCTION

Several methods exist for identifying high crash locations on roadways. Each method produces a set of ranked segments in order of safety, but the order of the segments most likely will be different depending on which method is chosen. Crash frequency (CF) and crash rate (CR) are two different statistics used to identify high crash locations that are easily obtained with crash history and traffic data. Empirical Bayes (EB) and crash reduction potential (CRP) are two other methods that incorporate safety performance functions with crash history. The latter two are more difficult to use because of the necessity of having a safety performance function.

Safety analysts need to decide which method to use for identifying high crash locations, but which method should be chosen and why? The secondary LVRRs system in Iowa is a very large network of paved and unpaved roadways totaling nearly 80,000 miles in length. Determining the best technique for identifying high crash locations is crucial for deciding where crash mitigation measures should be placed. This chapter explores a methodology for testing the performance of each of the four high crash location identification methods (CF, CR, EB and CRP) on secondary LVRRs in Iowa.

2.0 METHODOLOGY

2.1 Data collection/preparation

12 different segmentations were used to develop safety performance functions in the previous chapter. Discontinuous, continuous and GIMS segmentations were each split into four different secondary LVRR categories: paved 1-99 AADT, paved 100-400 AADT, unpaved 1-99 AADT and unpaved 100-400 AADT. The performances of four high crash identification methods (CF, CR, EB and CRP) were tested on each of the 12 different segmentations. Crash data from 2001-2008 was assigned onto each segmentation. Only mainline crashes were included. The total number of crashes was summed for each of the two periods: 2001-2004 and 2005-2008.

Total crashes were used instead of just serious crashes to identify high crash locations because of a couple reasons. Even though there are many serious crashes that occur on the secondary LVRR system, they are spread over a very large network. Using the segmentations for this project, it was seen that if a segment experienced a serious crash in the study period, then the segment was automatically placed in the top 5 percent of high crash locations. If total crashes are used instead, there are far less segments that experience zero crashes, so there is more confidence in the top 5 percent of high crash locations. In chapter 3, it was shown that there appears to be a correlation between serious and total crash occurrence, so by using total crashes in this analysis, the high crash locations that result will closely represent the high serious crash locations as well. The following four sections detail the necessary data needed for each high crash location identification method.

2.1.1 Crash Frequency

The only data needed to rank segments by crash frequency are the number of total crashes in each of the two analysis periods (2001-2004, 2005-2008) for each segmentation.

2.1.2 Crash Rate

Traffic data and crash data are necessary for ranking segments by crash rate. The crash rate was calculated for each segment for the two analysis periods, 2001-2004 and 2005-2008. Equation 4-1 gives the calculation for crash rate in crashes per million vehicle miles travelled (crashes per MVMT). Note that four years of crash data were used.

Equation 4-1:

$$CR = \frac{\#Crashes * 1{,}000{,}000}{AADT * Y * 365}$$

Y = number of years of crash data

2.1.3 EB

Besides crash data, ranking segments using the empirical Bayes method requires a safety performance function. Chapter 3 developed safety performance functions for predicting both serious and total crashes on 12 different secondary LVRR segmentations. The total crash models were used for ranking secondary LVRRs using EB estimates for this study. Equation 4-2 through Equation 4-11 give the models that were produced from chapter 3. The dispersion parameter is also included with each model, which will be used in the EB process. No models were produced for either the discontinuous or GIMS paved 1-99 AADT segmentation.

Equation 4-2: Continuous paved 1-99 AADT model.

$$\mu = LENG\_MI * AADT^{0.2424} * e^{-2.9013+0.1674 LANEWID+0.0622 SHDWID-0.0128 LIMMPH}$$

$$\varphi = 1.2512$$

Equation 4-3: Discontinuous paved 100-400 AADT model.

$$\mu = LENG\_MI * AADT^{0.6672} * e^{\begin{array}{l}-2.8136-0.0453 SHDWID-0.6426 TERRAIN1-0.4227 TERRAIN2+0.7182 SHDTYPE0+0.4054 SHDTYPE1+0.5439 SHDTYPE2\\-0.0125 LIMMPH+0.1884 PASSREST\end{array}}$$

$$\varphi = 0.1336$$

Equation 4-4: GIMS paved 100-400 AADT model.

$$\mu = LENG\_MI * AADT^{0.7002} * e^{\begin{array}{l}-3.2840-0.0492 SHDWID-0.5438 TERRAIN1-0.3629 TERRAIN2+0.8134 SHDTYPE0+0.4506 SHDTYPE1+0.5837 SHDTYPE2\\-0.0104 LIMMPH+0.4026 PASSREST\end{array}}$$

$$\varphi = 0.5123$$

Equation 4-5: Continuous paved 100-400 AADT model.

$$\mu = LENG\_MI * AADT^{0.7596} * e^{-3.2573-0.0562 SHDWID-0.4363 TERRAIN1-0.3056 TERRAIN2+0.3834 SHDTYPE0-0.0102 LIMMPH+0.5711 PASSREST}$$

$$\varphi = 0.3377$$

Equation 4-6: Discontinuous unpaved 1-99 AADT model.

$$\mu = LENG\_MI * AADT^{0.7944} * e^{-1.9332-0.0292LANEWID-0.4794TERRAIN1-0.3729TERRAIN2-0.0271LIMMPH}$$
$$\varphi = 0.2019$$

Equation 4-7: GIMS unpaved 1-99 AADT model.

$$\mu = LENG\_MI * AADT^{0.7319} * e^{-2.2226-0.0142LANEWID-0.4976TERRAIN1-0.4002TERRAIN2-0.0208LIMMPH}$$
$$\varphi = 0.8936$$

Equation 4-8: Continuous unpaved 1-99 AADT model.

$$\mu = LENG\_MI * AADT^{0.8778} * e^{-2.8503-0.0227LANEWID-0.4212TERRAIN1-0.3422TERRAIN2-0.0199LIMMPH+0.1133PASSREST}$$
$$\varphi = 0.3951$$

Equation 4-9: Discontinuous unpaved 100-400 AADT model.

$$\mu = LENG\_MI * AADT^{0.9546} * e^{-4.7915+0.3105PASSREST}$$
$$\varphi = 0.0631$$

Equation 4-10: GIMS unpaved 100-400 AADT model.

$$\mu = LENG\_MI * AADT^{1.0031} * e^{-3.9163-0.0287LANEWID-0.2559TERRAIN1-0.2087TERRAIN2}$$
$$\varphi = 0.6430$$

Equation 4-11: Continuous unpaved 100-400 AADT model.

$$\mu = LENG\_MI * AADT^{0.9594} * e^{-4.2806-0.3107TERRAIN1-0.249TERRAIN2}$$
$$\varphi = 0.4191$$

The first step in producing an EB estimate is to calculate the expected crash estimate from the appropriate safety performance function. Next, a weight needs to be calculated that will decide how much the safety performance function will factor into the final EB estimate

(Equation 3-23 and Equation 3-24). Lastly, the EB estimate is calculated using Equation 4-12 (Hauer, 2001).

Equation 4-12:

$$Estimate = \mu * w + C * (1 - w)$$

μ = crash model estimate

w = weight

C = total crashes on that segment in the analysis period

2.1.4 Crash Reduction Potential

Crash reduction potential (CRP) is the difference between the EB estimate and the safety performance function estimate as shown in Equation 4-13. The same safety performance functions shown in Equation 4-2 to Equation 4-11 are also used for calculating the CRP. The sections of each of the segmentations were ranked from the least reduction potential to the greatest reduction potential. The reduction potential is used to highlight segments which are the most over-represented compared to other similar roadways.

Equation 4-13:

$$CRP = EB - \mu$$

$$CRP = \mu * w + C * (1 - w) - \mu$$

2.2 Performance Tests

Four different procedures developed and proposed in a study by Cheng and Washington were used to test the differences between high crash locations selected for two different time periods. The four tests are the site consistency test, method consistency test, total rank differences test and the Poisson mean differences test (Cheng and Washington, 2008).

2.2.1 Site consistency test

The site consistency test measures the ability of one of the high crash location identification methods (CF, CR, EB or ARP) to consistently identify a roadway segment as high risk over two time periods. A site identified as high risk during the first time period should also be identified as high risk during the second time period. The test statistic, shown in Equation 4-14, gives the sum of the crashes for the top α percentage of sites.

Equation 4-14:

$$T1_j = \sum_{k=n-n\alpha}^{n} C_{k,i} \text{ or } \sum_{k=n-n\alpha}^{n} C_{k,i+1}$$

n = total number of sites being compared,
C = crash count for site ranked site k,
α = threshold of identified high - risk sites (5% or 10%)
j = hot - spot identification method being compared
i = observation period (i = 2001 - 2004, i + 1 = 2005 - 2008)

2.2.2 Method consistency test

The method consistency test determines how many sites are identified as high crash locations in both time periods. The test statistic is the total number of sites for the top α percentage of sites that are considered high-risk in both observation periods (Equation 4-15).

Equation 4-15:

$$T2_j = \{k_{n-n\alpha}, k_{n-n\alpha+1}, ..., k_n\}_{j,i} \cap \{k_{n-n\alpha}, k_{n-n\alpha+1}, ..., k_n\}_{j,i+1}$$

2.2.3 Total rank differences test

The total rank differences test is similar to the method consistency test in that it also tests the method's ability to identify a site as high-risk for both time periods. The test statistic is the sum of the differences in rank value between the two observation periods (Equation 4-16).

Equation 4-16:

$$T3_j = \sum_{k=n-n\alpha}^{n} \left(\left| \Re\left(k_{j,i}\right) - \Re\left(k_{j,i+1}\right) \right| \right)$$

2.2.4 Poisson mean differences test

False identifications (FIs) are a problem that arises when a site is deemed safe when it is really not (false negative, FN) or when a site is deemed unsafe when it is really safe (false positive, FP). The Poisson mean differences test identifies these FIs and the magnitude of the FIs and computes a test statistic that reflects the amount and magnitude of the FIs. First the true Poisson mean (TPM) needed to be calculated for each segment. This is the 4-year crash mean for each segment. The critical TPM is the number of crashes at which a site becomes high-risk (either top 5 percent or 10 percent in this study). For each FI, the absolute difference of the TPM and the critical TPM was calculated (Equation 4-17).

Equation 4-17:

$$T4_j = \sum_{k=n-n\alpha}^{n} \left(\left| TPM_{k,j,i} - TPM_{critical,k,j,i} \right| \right) + \sum_{k=n-n\alpha}^{n} \left(\left| TPM_{k,j,i+1} - TPM_{critical,k,j,i+1} \right| \right)$$

3.0 ANALYSIS

3.1 Paved 1-99 AADT

Only one safety performance function was able to be created for paved secondary LVRRs with 1-99 AADT, which was using continuous segmentation.

3.1.1 Continuous segmentation

Continuous paved 1-99 AADT segmentation was the only segmentation that produced a usable safety performance function. Discontinuous and GIMS paved 1-99 AADT segmentations were not considered for the four performance tests. Table 4-1 shows the results of the site consistency test. The highlighted cells represent the highest number of crashes in each time period. Crash frequency produces the best results for both time periods for both top 5 percent and 10 percent sites. Crash reduction potential follows crash frequency with the most crashes.

Table 4-1. Site consistency test results for continuous paved 1-99 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	217	195	217	195
CR	98	88	181	172
EB	107	98	136	127
CRP	129	119	186	173

Results from the method consistency test are presented in Table 4-2. Once again, crash frequency shows the best results of the four methods with 40 sites being considered high-risk in both observation periods. Second is EB followed by CRP and lastly CR.

Table 4-2. Method consistency test results for continuous paved 1-99 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	40	40
CR	6	18
EB	28	31
CRP	10	21

Table 4-3 gives the results of the total rank differences test. This test shows the total differences between ranks of a site from one time period to another time period. The best method is EB followed by CR, CF, and then CRP.

Table 4-3. Total rank differences test results for continuous paved 1-99 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	10106	10106
CR	4792	6867
EB	1065	2726
CRP	39229	70022

The final high crash location identification method performance test for continuous paved 1-99 AADT segmentation is the Poisson mean differences test. The results are shown in Table 4-4. EB produces the best results for the top 5 percent sites, yet produces the worst results for the top 10 percent sites. CF gives the best results for the top 10 percent sites, however shows the worst results for the top 5 percent sites.

Table 4-4. Poisson mean differences test results for continuous paved 1-99 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	342.625	156.625
CR	327.25	192.875
EB	174.125	265.625
CRP	217.125	194.625

3.2 Paved 100-400 AADT

The four performance tests were conducted for the discontinuous, GIMS and continuous segmentations of paved secondary LVRRs with 100-400 AADT.

3.2.1 Discontinuous segmentation

The results of the site consistency test are shown in Table 4-5. CF performs best on the site consistency test. Table 4-6 displays the results of the method consistency test results showing that CF is again the best performing method followed closely by EB.

Table 4-5. Site consistency test results for discontinuous paved 100-400 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	1161	1021	1870	1575
CR	139	165	549	552
EB	952	851	1618	1438
CRP	892	779	1513	1296

Table 4-6. Method consistency test results for discontinuous paved 100-400 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	36	80
CR	14	44
EB	31	69
CRP	19	44

Table 4-7 presents the total rank differences test results showing EB as the best performing method followed by CF. The results of the Poisson mean differences test are given in Table 4-8 where EB is again the best performing method followed by CF.

Table 4-7. Total rank differences test results for discontinuous paved 100-400 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	10107	33450
CR	35762	56615
EB	6663	17981
CRP	27341	58741

Table 4-8. Poisson mean differences test results for discontinuous paved 100-400 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	1728.625	2126.625
CR	5136	6868.375
EB	1397.5	1871.125
CRP	2495.625	3452.25

3.2.2 GIMS segmentation

Table 4-9 shows the results of the site consistency test. CF is the best performing method for both time periods and for both top 5 percent and 10 percent sites followed by CRP. CF is also the best performing method in the method consistency test followed by EB shown in Table 4-10.

Table 4-9. Site consistency test results for GIMS paved 100-400 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	4449	3914	4449	3914
CR	1192	1173	2424	2331
EB	1686	1556	2549	2395
CRP	1820	1640	2619	2438

Table 4-10. Method consistency test results for GIMS paved 100-400 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	1065	1065
CR	125	309
EB	224	605
CRP	142	443

Results of the total rank differences test are given in Table 4-11. CR is the best performing method followed by EB. The Poisson mean differences test shows that CRP and EB are the best performing methods (Table 4-12). CRP has the lowest TPM differences sum followed closely by EB for the top 5 percent sites. EB has the lowest TPM differences sum followed closely by CRP for the top 10 percent sites.

Table 4-11. Total rank differences test results for GIMS paved 100-400 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	5856470	5856470
CR	1708543	2881074
EB	1838425	3481516
CRP	6251229	12855059

Table 4-12. Poisson mean differences test results for GIMS paved 100-400 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	7953.75	7953.75
CR	6844.25	6845.25
EB	4435.5	5008.125
CRP	4183.375	5426.5

3.2.3 Continuous segmentation

The results of the site consistency test of the continuous paved 100-400 AADT segmentation are shown in Table 4-13. CF is the best performing method, capturing the most crashes for both time periods and both top 5 percent and 10 percent sites, while EB follows closely behind CF. CR shows the worst results by far. Table 4-14 gives the method consistency test results. CF once again shows the best results with EB showing good results as well. CR again shows the worst results.

Table 4-13. Site consistency test results for continuous paved 100-400 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	1530	1385	2407	2455
CR	304	296	809	757
EB	1446	1244	2215	1917
CRP	1367	1143	2051	1691

Table 4-14. Method consistency test results for continuous paved 100-400 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	91	244
CR	31	89
EB	83	196
CRP	57	131

Table 4-15 displays the results of the total rank differences test and Table 4-16 presents the results of the Poisson mean differences test. EB is the best performing method for both tests and for both top 5 percent and 10 percent sites. CF is the next best performing method for both tests. CRP shows the worst results in the total rank differences test and CR gives the worst results in the Poisson mean differences test.

Table 4-15. Total rank differences test results for continuous paved 100-400 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	44562	112323
CR	170891	281115
EB	40155	92333
CRP	202065	402552

Table 4-16. Poisson mean differences test results for continuous paved 100-400 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	1726.125	2546.625
CR	6129.875	7619.875
EB	1538	1797
CRP	2639.375	3677.625

3.3 Unpaved 1-99 AADT

The four performance tests were conducted for the discontinuous, GIMS and continuous unpaved 1-99 AADT segmentations.

3.3.1 Discontinuous segmentation

The results of the site consistency test for discontinuous unpaved 1-99 AADT segmentation are shown in Table 4-17. CF is the best performing method. EB and CRP share similar results with each other. CR gives the worst results.

Table 4-18 displays the results of the method consistency test. CF again is the best performing method with EB showing the second best results and CR and CRP showing the worst results.

Table 4-17. Site consistency test results for discontinuous unpaved 1-99 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	2947	2052	4972	3600
CR	997	987	1980	2020
EB	1874	1708	3140	2793
CRP	1843	1700	3059	2780

Table 4-18. Method consistency test results for discontinuous unpaved 1-99 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	149	427
CR	79	170
EB	97	255
CRP	66	170

Table 4-19 presents the results of the total rank differences test. Table 4-20 gives the results of the Poisson mean differences test. EB is the best performing method in both tests.

Table 4-19. Total rank differences test results for discontinuous unpaved 1-99 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	699259	1588162
CR	649944	1283768
EB	398706	895964
CRP	777877	1529199

Table 4-20. Poisson mean differences test results for discontinuous unpaved 1-99 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	4493.875	7349.375
CR	8555.75	11283.875
EB	3635	4817.125
CRP	4457.125	6137.125

3.3.2 GIMS segmentation

The results of the site consistency test for GIMS unpaved 1-99 AADT segmentation are shown in Table 4-21 while the results of the method consistency test are given in Table

4-22. CF shows the best results in both tests for the top 5 percent sites. However, when the top 10 percent sites are considered, CF, CR and CRP tie for the best performing methods in both tests. The reason the values are the same for the three different methods seems to be a coincidence.

Table 4-21. Site consistency test results for GIMS unpaved 1-99 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	9044	7501	9044	7501
CR	6112	5944	9044	7501
EB	5200	4599	6265	5335
CRP	6125	5918	9044	7501

Table 4-22. Method consistency test results for GIMS unpaved 1-99 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	1201	1201
CR	587	1201
EB	795	943
CRP	927	1201

Table 4-23 presents the results of the total rank differences test. CR is the best performing method according the results of this test. The results of the Poisson mean differences test in Table 4-24 are similar to the results of the method consistency test in Table 4-22 with CF showing the best results for the top 5 percent sites and CF, CR and CRP having the best performance when considering the top 10 percent sites.

Table 4-23. Total rank differences test results for GIMS unpaved 1-99 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	42318544	42318544
CR	20058892	22533691
EB	54899165	83762444
CRP	374111458	473124388

Table 4-24. Poisson mean differences test results for GIMS unpaved 1-99 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	10815.25	10815.25
CR	14507.25	10815.25
EB	16380.375	14899.5
CRP	14588.875	10815.25

3.3.3 Continuous segmentation

Table 4-25 presents the results of the site consistency test for continuous unpaved 1-99 AADT segmentation. CF is the best performing method in the site consistency test with EB and CRP sharing similar results with each other. The results of the method consistency test are given in Table 4-26. EB is the best performing method with CF trailing by only 2 sites while considering the top 5 percent sites. However, when considering top 10 percent sites, CF shows the best results with EB not closely following.

Table 4-25. Site consistency test results for continuous unpaved 1-99 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	5139	3957	9440	7851
CR	1994	1956	4403	4188
EB	4053	3573	5807	5103
CRP	4026	3643	5792	5063

Table 4-26. Method consistency test results for continuous unpaved 1-99 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	579	2559
CR	184	601
EB	581	1557
CRP	389	919

The results of the total rank differences test are shown in Table 4-27. EB is the best performing method by far. Table 4-28 provides the results of the Poisson mean differences test. As for the total rank differences test, EB shows the best results for the Poisson mean differences test.

Table 4-27. Total rank differences test results for continuous unpaved 1-99 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	4571885	15686986
CR	5601454	9021554
EB	3126904	6823550
CRP	17670311	38581894

Table 4-28. Poisson mean differences test results for continuous unpaved 1-99 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	6117.75	12058
CR	14913.375	14044.75
EB	5547.75	7279.875
CRP	6457.375	9565.5

3.4 Unpaved 100-400 AADT

The four performance tests were conducted for the discontinuous, GIMS and continuous unpaved 100-400 AADT segmentations.

3.4.1 Discontinuous segmentation

The results of the site consistency test are shown in Table 4-29 for discontinuous unpaved 100-400 AADT segmentation. CF is the best performing method with EB giving the next best results for the site consistency test. Table 4-30 displays the results of the method consistency test. CF once again shows the best results followed by EB. CR produces the worst results for each of the site and method consistency tests.

Table 4-29. Site consistency test results for discontinuous unpaved 100-400 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	506	461	892	894
CR	346	323	623	569
EB	441	401	760	694
CRP	411	374	721	635

Table 4-30. Method consistency test results for discontinuous unpaved 100-400 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	14	29
CR	5	10
EB	12	24
CRP	6	13

Table 4-31 presents the results of the total rank differences test. EB is the best performing method by far. EB also shows the best results for the Poisson mean differences test given in Table 4-32. CR and CRP provide the worst results for both the total rank and

Poisson mean differences tests.

Table 4-31. Total rank differences test results for discontinuous unpaved 100-400 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	2638	6665
CR	5246	10776
EB	1501	4624
CRP	4212	10876

Table 4-32. Poisson mean differences test results for discontinuous unpaved 100-400 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	795.35	1485.625
CR	1802.2	2812.375
EB	706.3	1154.25
CRP	1294.525	2246.75

3.4.2 GIMS segmentation

Table 4-33 shows the results of the site consistency test for GIMS unpaved 100-400 AADT segmentation. The results of the method consistency test are given in Table 4-34. CF is the best performing method for both consistency tests. CR offers the worst results for both consistency tests.

Table 4-33. Site consistency test results for GIMS unpaved 100-400 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	1355	3071	3183	3071
CR	926	922	1755	1747
EB	1214	1188	1848	1813
CRP	1319	1277	1884	1843

Table 4-34. Method consistency test results for GIMS unpaved 100-400 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	267	859
CR	102	240
EB	185	466
CRP	103	287

The results of the total rank differences test are given in Table 4-35. CF is the best performing method when only considering the top 5 percent sites. However, CR shows the best results for the top 10 percent sites. Table 4-36 presents the results of the Poisson mean differences test. CRP shows the best results when only the top 5 percent sites are considered, but when the top 10 percent sites are considered, EB is the best performing method.

Table 4-35. Total rank differences test results for GIMS unpaved 100-400 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	766566	3042705
CR	901047	1563749
EB	937964	1852431
CRP	2957692	5714431

Table 4-36. Poisson mean differences test results for GIMS unpaved 100-400 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	4278.75	5589.625
CR	4784.25	4849.5
EB	3310.625	3519.25
CRP	3075.75	4097.625

3.4.3 Continuous segmentation

The results of the site consistency test for continuous unpaved 100-400 AADT segmentation are shown in Table 4-37. CF is the best performing method while CR is the worst performing method. Table 4-38 gives the results of the method consistency test. CF once again provides the best results with CR showing the worst results.

Table 4-37. Site consistency test results for continuous unpaved 100-400 AADT segmentation.

SITE CONSISTENCY TEST				
	Top 5% Sites		Top 10% Sites	
Method	2001-2004	2005-2008	2001-2004	2005-2008
CF	965	977	1382	1379
CR	286	299	631	638
EB	725	693	1136	1119
CRP	702	684	1099	1089

Table 4-38. Method consistency test results for continuous unpaved 100-400 AADT segmentation.

METHOD CONSISTENCY TEST		
Method	Top 5% Sites	Top 10% Sites
CF	81	147
CR	33	70
EB	63	120
CRP	36	89

Table 4-39 presents the results of the total rank differences test. EB is the best performing method with CRP showing the worst results. EB once again provides the best results for the Poisson mean differences test in Table 4-40. CR shows the worst results for the Poisson mean differences test.

Table 4-39. Total rank differences test results for continuous unpaved 100-400 AADT segmentation.

TOTAL RANK DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	39486	91589
CR	73262	132735
EB	21095	56891
CRP	78944	177228

Table 4-40. Poisson mean differences test results for continuous unpaved 100-400 AADT segmentation.

POISSON MEAN DIFFERENCES TEST		
Method	Top 5% Sites	Top 10% Sites
CF	1428.75	1829.25
CR	3083.5	4002
EB	956.125	1340.75
CRP	1443.875	1925

4.0 RECOMMENDATIONS/CONCLUSIONS

The results of the four performance test given in Table 4-1 through Table 4-40 show interesting results for the 10 different segmentations considered. For all 10 segmentations, crash frequency (CF) was returned as the best performing method according the site consistency test. This means that when CF is used to identify the top 5 percent and top 10 percent sites for both 2001-2004 and 2005-2008, the high crash locations that result contain the most crashes than when using any other method. Also, CF consistently identified the most sites that were labeled as high crash locations during both observation periods for all 10 segmentations according to the method consistency test. The only exception to this is the continuous unpaved 1-99 AADT segmentation when CF fell two sites shy of EB while considering only top 5 percent sites. This difference should be considered negligible.

EB showed the least differences in rankings between sites from one time period to the next in 7 of the 10 segmentations according to the total rank differences test. The other three segmentations that did not show EB as the preferred method were all GIMS segmentations. EB also proved to be the best method in the majority of the segmentations for producing the least total Poisson mean differences.

It is fair to say that CF and EB showed the best results. Both CF and EB produce crash frequencies. EB has the advantage over CF by having the ability to factor in the safety performance of similar sites, which in turn negates the regression to the mean problem. It is recommended that either CF or EB be used as methods for identifying high crash locations on secondary LVRRs, however it is preferred to use EB because of its enhanced abilities over CF.

CHAPTER 5. GENERAL CONCLUSIONS

1.0 GENERAL DISCUSSION

Addressing safety on secondary LVRRs in Iowa has been and will be a major issue. It is difficult to identify high crash locations on secondary LVRRs because of the low crash densities that exist on these roadways. One argument against looking for black spots on secondary LVRRs is to look for systemic methods to implement safety treatments. An example would be to only look at horizontal curves and treat those curves whose characteristics have been shown to result in more run off the road crashes. This thesis focused on identifying high crash locations on secondary LVRRs, however safety treatments should not be limited to just high crash locations. Systemic methods for installing safety treatments should be considered as well.

Selection of locations to implement crash countermeasures depends not only on expected benefits but also on costs. This thesis has focused on selecting secondary LVRR high crash locations using crash data and models. However, selecting many short segments for treatment will likely lead to high mobilization costs. In order to make best use of available safety funds, care should be taken in selecting sites and economic as well as engineering analysis should be conducted on any candidate site. This thesis assists the safety analyst in selecting high crash locations on secondary LVRRs, but a balance needs to be determined between costs and exposure of safety treatments.

2.0 RECOMMENDATIONS/CONCLUSIONS

When examining fixed segment lengths on secondary LVRRs, the length affects the statistical reliability of the expected crash frequency of the segment. Longer segments allow for the opportunity for more crashes to be examined per segment, thus increasing the statistical confidence that the estimated annual crash frequency of the segment will be greater than the annual crash variance of that segment. It is recommended that longer segments be chosen for safety analysis of secondary LVRRs; however, no absolute length is recommended from this research.

Chapter three presents research on the development of safety performance functions for secondary LVRRs. Based on the descriptive statistics presented in this chapter, it is recommended that the secondary LVRR system in Iowa be split into four parts for crash models: paved 1-99 AADT, paved 100-400 AADT, unpaved 1-99 AADT and unpaved 100-400 AADT. Serious crash models and total crash models were developed originating from three different segmentation styles: discontinuous, continuous and GIMS segmentations. The final crash models were compared using their average weights computed during the empirical Bayes procedure. Serious and total crash models were recommended for each secondary LVRR road class in Table 3-26.

High crash location identification methods were compared using data from secondary LVRRs in Iowa. Results of the performance tests were fairly consistent across the different LVRR road classes and segmentations. Crash frequency and empirical Bayes methods gave the best results. It is recommended to use empirical Bayes for high crash location identification on secondary LVRRs in Iowa over crash frequency, because of the ability for empirical Bayes to incorporate safety performances of similar sites and eliminate the regression to the mean problem.

3.0 FUTURE RESEARCH

3.1 Additional variables

Additional variables should be considered for the development of safety performance functions for secondary LVRRs in future research. These may include vertical and horizontal alignment and driveway density. And, as mentioned before, speed limit could be specified as a categorical variable.

3.2 Intersection crash model

This thesis only focused on mainlines of secondary LVRRs and not intersections of secondary LVRRs. It is recommended future research be conducted for the development of intersection safety performance functions. An intersection database will need to be created for secondary LVRRs in Iowa before a model can be developed.

4.0 REFERENCES

Guidelines for Geometric Design of Very Low-Volume Local Roads (ADT ≥ 400). American Association of State Highway and Transportation Officials, 2001.

Agent, K., L. O'Connell, E. Green, D. Kreis, J. Pigman, and N. Tollner, *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, Lexington, KY, 2003.

Cheng, Wen and Simon Washington. *New Criteria for Evaluating Methods of Identifying Hot Spots*. Transportation Research Record 2083, pp. 76-85. 2008.

Drakopoulos, Ertan. *Addressing Non-Intersection Crashes Statewide: Identification and Ranking of Highway Segments in Need of Safety Upgrades Using a Floating Highway Segment Tool Integrated with GIS*. Proceeding of 2005 Mid-Continent Transportation Research Symposium, Ames, IA, 2005.

GIS Safety Analysis Tools Version 3.0 User Guide. GIS/Trans, LTD, Federal Highway Administration, 2000.

Federal Highway Administration. *Fatality Rates by Functional Class (2007)*. http://safety.fhwa.dot.gov/speedmgt/data_facts/docs/fataltbl2007.pdf. Accessed July 14, 2010.

Geyer, J., E. Lankina, D. Rangland, C. Chan, and A. Sharafsaleh. Literature Survey of Methods for Identifying High Concentration Collisions (HCCL). *Traffic Safety Center and California PATH*. University of California, Berkeley, 2005.

Hall, Jerome W., Elizabeth W. Rutman, James D. Brogan. *Highway Safety Challenges on Low-Volume Rural Roads*. Proceeding of 2003 Institute of Transportation Engineers Annual

Meeting.

Hauer, Ezra, Douglas W. Harwood, Forrest M. Council, Michael S. Griffith. *Estimating Safety by the Empirical Bayes Method: A Tutorial*. 2001.

Hu, Shuhua. *Akaike Information Criterion*. North Carolina State University. 2007.

Miaou, Shaw-Pin. *Measuring the Goodness-of-Fit of Accident Prediction Models*. Federal Highway Administration. 1996.

National Highway Traffic Safety Administration. *FARS: Fatality Analysis Reporting System Encyclopedia*. http://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed July 14, 2010.

Papageorgiou, Markos, *Fellow*, IEEE, and Apostolos Kostsialos. Freeway Ramp Metering: an Overview. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 3, No. 4, 2002.

Tarko, Andrew P., Muhammad Asif Iqbal, Michal Inerowicz, Hong Liang, and Gopal Panicker. *Safety Concious Planning in Indiana: Predicting Safety Benefits in Corridor Studies, Volume 1*. 2007.

usRAP: United States Road Assessment Program. *Feasibility Assessment and Pilot Program*. 2006.

Vogt, Andrew and Joe Bared. *Accident Models for Two-Lane Rural Segments and Intersections*. Transportation Research Record 1635, pp. 18-29. 1998.