TECHNICAL REPORT STANDARD PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/LA.13/503		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
Developing Inexpensive Crash Countermeasures for Louisiana Local Roads	5. Report Date: March 2014 6. Performing Organization Code: **LTRC Project Number: 10-5SS** **SIO/State Project Number: 30000240**
7. Author(s) Dr. Helmut Schneider, Dr. Xiaoduan Sun, and Dr. Peter Kelle	8. Performing Organization Report No.
9. Performing Organization Name and Address Louisiana State University Baton Rouge, LA 70803	10. Work Unit No. 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Louisiana Department of Transportation and Development P.O. Box 94245 ,Baton Rouge, LA 70804-9245	13. Type of Report and Period Covered **Final Report 1/17/11 – 1/16/2013** 14. Sponsoring Agency Code

15. Supplementary Notes

Conducted in Cooperation with the U.S. Department of Transportation, Federal Highway Administration

16. Abstract

Although 40% of all crashes in Louisiana are on local roads, local road safety improvement programs have not received the attention needed to reduce crashes. Local road crash countermeasures are an important part of the overall efforts to reduce crashes and their severity in Louisiana. The efforts to develop a local road safety program are hampered by the lack of appropriate risk assessment that enables local agencies to reduce crashes using low cost countermeasures. This paper provides a methodology that can be used by local agencies to deploy countermeasures based on a risk assessment and optimization to meet a fixed budget. First, a statistical model is presented to assess the risk of local road segments taking into account AADT and geometric features of the road segment or intersection. Secondly, low cost countermeasures are recommended for individual road segments and costs for improvements are assessed. Thirdly, a score which allows the ranking of road projects is developed for each road segment. This score incorporates the risk associated with the observed number of crashes, the benefits of improvements, and the total cost of a project. Finally, guidelines for a local road safety improvement program are presented to allow local agencies to institute procedures for a systematic system-wide road improvement methodology. The deliverables include an Excel application that uses OLAP to obtain a ranking of candidates for road improvements. This application makes use of crash data, engineering features, and AADT to compute empirical Bayes (EB) estimates and tail probabilities for each road segment and intersection. Road segments and intersections with a tail probability below 5% are selected as candidates for countermeasures. These candidates are evaluated using Google Earth, countermeasures are suggested, and costs and benefits of the countermeasures are obtained using published information. The resulting road improvement projects are then ranked using multi criteria DEA including costs, benefits and crash risks.

17. Key Words	18. Distribution Statement
Crash countermeasures	**Unrestricted. This document is available through the National Technical Information Service, Springfield, VA 21161.**

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

Project Review Committee

Each research project will have an advisory committee appointed by the LTRC Director. The Project Review Committee is responsible for assisting the LTRC Administrator or Manager in the development of acceptable research problem statements, requests for proposals, review of research proposals, oversight of approved research projects, and implementation of findings.

LTRC appreciates the dedication of the following Project Review Committee Members in guiding this research study to fruition.

LTRC Manager

Kirk Zeringue
Special Studies Research Manager

Members

Marie Walsh
Jason Taylor
Dan Magri
Jim Champagne
John Broemmelsiek
Betsey Tramonte

Directorate Implementation Sponsor

Dennis Decker
DOTD Assistant Secretary, Office of Multimodal Planning

Developing Inexpensive Crash Countermeasures for Louisiana Local Roads

by

Helmut Schneider, Ph.D.
Xiaoduan Sun, Ph.D.
Peter Kelle, Pd.D.

BEC 2200A Nicholson Extension
Baton Rouge, LA 70803
Phone: (225) 578-2516
http://hsrg.lsu.edu

LTRC Project No. 10-5SS
SIO No. 30000240

conducted for

Louisiana Department of Transportation and Development
Louisiana Transportation Research Center

March 2014

ABSTRACT

Although 40% of all crashes in Louisiana are on local roads, local road safety improvement programs have not received the attention needed to reduce crashes. Local road crash countermeasures are an important part of the overall efforts to reduce crashes and their severity in Louisiana. The efforts to develop a local road safety program are hampered by the lack of appropriate risk assessment that enables local agencies to reduce crashes using low cost countermeasures. This paper provides a methodology that can be used by local agencies to deploy countermeasures based on a risk assessment and optimization to meet a fixed budget. First, a statistical model is presented to assess the risk of local road segments taking into account AADT and geometric features of the road segment or intersection. Secondly, low cost countermeasures are recommended for individual road segments and costs for improvements are assessed. Thirdly, a score which allows the ranking of road projects is developed for each road segment. This score incorporates the risk associated with the observed number of crashes, the benefits of improvements, and the total cost of a project. Finally, guidelines for a local road safety improvement program are presented to allow local agencies to institute procedures for a systematic system-wide road improvement methodology. The deliverables include an Excel application that uses OLAP to obtain a ranking of candidates for road improvements. This application makes use of crash data, engineering features, and AADT to compute empirical Bayes (EB) estimates and tail probabilities for each road segment and intersection. Road segments and intersections with a tail probability below 5% are selected as candidates for countermeasures. These candidates are evaluated using Google Earth, countermeasures are suggested, and costs and benefits of the countermeasures are obtained using published information. The resulting road improvement projects are then ranked using multi criteria DEA including costs, benefits and crash risks.

ACKNOWLEDGMENTS

The significant contributions by the graduate assistants Christian Raschke and Charles Cavalier who supported the research of each of the tasks, Emily Pfetzer for editing the manuscript, and project manager Maria Mathews who kept the team on schedule for each task are acknowledged.

The research team also appreciates the dedication of the Project Review Committee Members in guiding this research study to fruition.

IMPLEMENTATION STATEMENT

This project includes an Excel application that was applied to local roads of two parishes in Louisiana. The application requires either access to the LSU domain or access via a virtual private network (vpn) to get access to the databases used for this project. The application can be applied to other parishes as well as information becomes available. The full implementation for the local roads in the state of Louisiana requires similar information as available for state routes. These road features included lane and shoulder width, curves, driveway density, and intersection features such as turn lane and traffic controls. Since there are no predefined road segments for local roads, segments of approximately 500 feet are recommended. Also crashes need to be map spotted to the intersections or road segments. After this information has been collected the program can be easily implemented for other parishes.

TABLE OF CONTENTS

ABSTRACT iii
ACKNOWLEDGMENTS v
IMPLEMENTATION STATEMENT vii
TABLE OF CONTENTS ix
LIST OF TABLES xi
LIST OF FIGURES xiii
INTRODUCTION 1
OBJECTIVE 3
SCOPE 5
METHODOLOGY 7
- Defining Hazardous Locations 7
- Identifying Hazardous Locations - General Approaches 10
  - Identifying Black Spots without Models 10
  - Statistical Models in Hazardous Location Identification 11
- Identifying Hazardous Locations - Safety Performance Functions 12
  - Minimum Data Requirements for SPF Estimation 12
  - Missing Data 13
  - Calibration vs. Estimation 15
- Safety Performance Function Estimations - Parametric Models 16
  - Poisson-Gamma Mixture Models (Negative Binomial Models) 16
  - Derivation of the Negative Binomial Model 17
  - Negative Binomial Models in the Literature 19
  - Common Sources of Error 20
  - Multivariate Models 21
- Empirical Bayesian Analysis 22
  - Empirical Bayes Estimates 25
  - Empirical Bayesian Analysis and Local Roads 27
  - Empirical Bayes Data Needs 29
  - Statistical Evaluation of SPF - Classical Analysis 29
  - Inexpensive Countermeasures 30
  - Targeted Areas 31
  - Specific Countermeasure 31
  - Observed Benefits 31
  - Selection Guideline 31
  - Multi Criteria Decision Making 32

Summary and Conclusion from Literature Review 36
Data…… 39
Road Grid.. 39
Crashes.. 40
Matching Crashes with Road Segments 41
Intersections 41
Identifying Curves 42
Other Road Characteristics and Design Features... 44
Average Annual daily Travel (AADT) 46
Using Microsoft Excel Application for Site Ranking... 48
Underlying Data Architecture……………………………………………...48
Application Overview...49
DISCUSSION OF RESULTS...59
Selection of Factor Level Settings 59
Selection of Road Segments and Intersections 61
Selecting Countermeasures.. 61
Crash Modification Factor (CMF) 61
Cost Estimation... 62
Ranking of Projects.. 62
CONCLUSIONS...67
RECOMMENDATIONS...71
ACRONYMS, ABBREVIATIONS, AND SYMBOLS 73
REFERENCES 75
APPENDIX A.. 85
Calculating Distances and Matching Points with Line Segments 85
Calculating Distances between Coordinates.. 85
Matching Points with Line Segments 86
APPENDIX B 89
Tables for Countermeasures and Estimated Costs.. 89
APPENDIX C 97
Example of Data Problems 97
Crash Coded as Segment Crash but too close to Intersection.......................... 97
Roadway Segment with no Google Street View (4 Sites Shown Here) 97

LIST OF TABLES

Table 1 Factors and factor levels 60
Table 2 Number of combinations investigated 60
Table 3 Comparison of the benefit and risk improvement for the different ranking methods for 30 intersection sites 64
Table 4 Comparison of the benefit and risk improvement 65
Table 5 Engineering features and settings 69
Table 6 Section levels and intersections analyzed 89
Table 7 Matrix of roadway segment levels and crash information 90
Table 8 Countermeasures, potential CMF values, and estimated costs for intersections 91
Table 9 Countermeasures, potential CMF values, and estimated costs for roadway segments 94

LIST OF FIGURES

Figure 1 DEA example 35
Figure 2 Example of intersection identification........ 42
Figure 3 Example of curve identification 44
Figure 4 Example of road characteristics spotting program 46
Figure 5 Example of AADT measures........ 47
Figure 6 Microsoft Excel black spot analysis application ribbons 50
Figure 7 Microsoft Excel black spot analysis example 51
Figure 8 Pivot table fields........ 52
Figure 9 Optional fields 53
Figure 10 Selected road segment/intersection 54
Figure 11 Google Earth picture for selected intersection 55
Figure 12 Detailed information for selected crash........ 56
Figure 13 Intersection view 56
Figure 14 Options for selecting types of crashes 57
Figure 15 Comparison of benefit and risk improvement 65
Figure 16 Comparison of benefit and risk improvement for different ranking methods for road segments........ 66
Figure 17 Google view of Mystic Blvd. (Picture Id- 12)........ 97
Figure 18 Google view of Arms St. (Picture Id- 37) 97
Figure 19 Google view of Sanders Road (Picture Id- 9) 97
Figure 20 Google view of N J Theriot Road (Picture Id- 36)........ 97
Figure 21 Google view of Clendenning Road (Picture Id- 41)........ 98
Figure 22 Google view of White Rose Dr. (Picture Id- 38)........ 98

INTRODUCTION

Local roads make up 73% of all road miles in Louisiana and experience 40% of the 160,000 yearly crashes on Louisiana roads. Over the past five years, 951 fatal crashes, over 92,000 injury crashes, and over 242,000 property damage-only crashes occurred on local roads. The total cost for these crashes added up to over $10 billion, over the five years. Louisiana has one of the highest insurance premiums of all 50 states, and is also one of the ten worst states in regard to safety performance. To improve these matters, Louisiana has launched an ambitious Strategic Highway Safety Plan aiming at zero deaths, with the interim goal of reducing traffic fatalities and serious injuries by 50% by 2030. The accomplishment of such tall objectives calls for effective crash countermeasures in all aspects, including reducing the number of crashes taking place on local roads. Thus, local road crash countermeasures are an important part of the overall efforts to reduce crashes in Louisiana and their impact on lives, health, and the economy.

Currently, no local road improvement program takes into account the risk at road segments or intersections based on average annual daily travel (AADT) and geometric design features. In order to develop a safety improvement program for local roads, several obstacles must be overcome. Among them are: lack of available information on clearly marked road segments through mile posts, the average daily travel for clearly identifiable road segments, clearly marked crash location (i.e., GPS coordinates), and a road inventory database that allows easy linkage to the crash database. Local agencies responsible for maintaining these roads often lack the resources to analyze crash frequencies or to identify locations in need of safety improvements, and many lack the funds necessary to make major improvements to the roads. To enable local agencies to identify high-risk crash locations, estimate costs for safety improvements, and recommend low-cost solutions to implement safety measures within the constraints of the information available, a process is needed that incorporates all necessary steps to implement low cost improvements successfully.

Other states have investigated potential solutions that will allow local agencies to perform crash data analysis and implement engineering solutions. For example, in Illinois, Jo et al., developed software that integrates maps and crash data in order to illustrate patterns and facilitate analysis of crashes by local agencies *[1]*. There is also an ongoing project at North Dakota State University with the goal to improve local road safety by Vachal and Johnson *[2]*. In a more general context, Sohn has applied quality functions to local roads in order to analyze crash data in Korea *[3]*. While local police stations are responsible for traffic accident management, crash data are housed and analyzed centrally in Korea. Sohn models different locations in the jurisdiction of a given police station in terms of the “quality of service” that customers (i.e., traffic participants) receive at those locations *[3]*.

Transportation agencies must consider both safety and security when allocating funds to potential road improvement projects *[4]*. Due to the fact that resources are often scarce, it is important to locate projects where intervention would be most beneficial. Comparative evaluation of locations requires a systematic approach, beginning with a standardized definition for classifying locations by hazard concerns, followed by a theoretically sound approach to statistically identity these locations. There are multiple methods of doing this, each with differing data and resource requirements. This report provides a detailed discussion of each available method, as well as each method's applicability to certain types of situations.

OBJECTIVE

The objective of this research project is to develop inexpensive crash countermeasures for local roads in the state of Louisiana. The current state-of-the-art approach to roadway safety and intersection safety prescribes an evaluation of the relevant roadway segments or intersection based on a Safety Performance Function (SPF). The purpose of an SPF is to gauge the safety of a given section of roadway compared to other roads with the same traffic volume (AADT) so that problem locations and problem areas may be identified. However, not all jurisdictions use an SPF when analyzing their crashes. Additionally, data constraints sometimes make it impossible to develop a proper SPF. This report will present the current state of the literature on safety performance functions and address the following questions: What are the advantages of using a Safety Performance Function over other types of methods to identify the top percentage (p%) of problem locations or problem areas that should be investigated further? The p% may be between 5% or 10% depending on the resource availability. What characteristics make for a “good" Safety Performance Function? How do other approaches to developing or calibrating a Safety Performance Function differ from each other and what are the benefits and data/resource requirements associated with each approach? These questions will be discussed specifically in the context of local roads and the challenges an analyst faces when examining local roads in Louisiana.

SCOPE

The scope of the project includes countermeasures for local roads in two parishes, Lafourche and Terrebonne. The project includes road segments and intersections. Intersections of local roads with state routes are not included because they are maintained by LADOTD.

METHODOLOGY

The literature review is laid out as follows: Section 1 presents a discussion of different ways to define what constitutes a hazardous location in the context of traffic safety. Section 2 provides some general background on identifying hazardous locations, while Section 3 specifically addresses safety performance functions. Section 4 provides the details of parametric statistical models used to fit crash data in order to estimate safety performance, and Section 5 pays special attention to the Empirical Bayes approach of identifying hazardous locations and provides estimates for the Empirical Bayes. Section 6 identifies inexpensive countermeasures that will be proposed for selected hazardous locations. Section 7 provides an overview of Multi Criteria Decision Making models that will be used for ranking the locations according to several criteria including risk, cost, and benefits. Section 8 provides conclusions from the literature search and sets the stage for the methodology applied in the spreadsheet application provided as a result of this study. Section 9 describes the data collection including crash data, road information, and exposure measured by average annual daily travel (AADT). Section 10 provides a low cost solution for ranking sites.

1. Defining Hazardous Locations

The ultimate goal of this project is to develop inexpensive crash countermeasures for local road locations selected for their safety concerns, relative to other similar locations. A location's safety level classification considers basic factors like the expected number of crashes, measures of exposure, and road design features. However, the previous literature on identifying and ranking hazardous locations has not always been in agreement regarding how to define or even what terminology to use, when explaining hazardous road locations. Terminology used includes "black spots," "hot spots," and "sites with promise," among other variations on the same *[5],[6],[7]*.These terms are fairly vague and tend to be used quite loosely throughout the literature, without any clarification or distinction among them as to what should designate a particular road section hazardous *[8]*. Given the lack of clear difference, the researchers use the above terms interchangeably throughout this report. As addressed in more detail in Section 5, it is not possible to identify black spots by statistically analyzing crash data alone; researchers believe that site visits and engineering evaluations are necessary to conclude that a pre-screened location is truly a black spot. The term "top p% of crash locations that warrant further investigation" is preferred.

The question of what exactly constitutes a black spot, a hot spot, or a site with promise is partly a philosophic one. For example, researchers can observe a relatively large number of crashes on interstate highways, compared to smaller city streets. However, the number of vehicles that pass over a certain road segment on an interstate highway is large, and therefore the risk of each individual being involved in a crash is low. On a small city street, there might be fewer crashes observed; however, due to the lower number of average daily traffic, the risk of each individual traffic participant being involved in a crash is higher than on interstates. In this example, which road deserves more attention and should have priority in terms of safety improvements? On one hand, one could argue that reducing the total number of accidents is a reasonable goal, as it will provide the largest benefit in terms of the total number of lives saved or injuries prevented. On the other hand, this approach might neglect some of the riskiest road sections in the traffic network.

Regardless of how one defines a "safe" or "unsafe" location, in practice, this distinction is never a truly binary one. Safety is a continuum where "safe" and "unsafe" sites are located alongside one another *[5]*. The analyst is therefore required to exercise careful judgment in order to make a distinction between the two on a practical level. It is difficult to find a single measure that reflects risk appropriately, even on a continuum. There is not one single criterion that is able to incorporate the wide array of individuals' hazard perceptions, based on the available data of crash reports. Some may consider a road to be hazardous if there are too many crashes, or too many injury crashes, or too many fatal crashes, or the crash rate is too high, or the percentage of fatal crashes to all crashes is too high; some safety professionals consider the case of too many run-off road crashes, or too many side impact crashes, and so on. Clearly, a benchmark to determine what constitutes "too many" is needed. One approach of obtaining benchmarks is to develop statistical models for crash counts. Elvik used the previous work of Hauer, Kononov, et al. and Persaud and Lyon to categorize the different definitions of black spots into three groups:

- Numerical definitions,
- Statistical definitions
- Model-based definitions *[9] [10] [11]*.

A numerical definition is straightforward. If the number of crashes at a particular location exceeds a certain threshold, then that location is considered a black spot. This definition also encompasses accident rate definitions, i.e., defining as a black spot all locations where the number of crashes per certain number of vehicles exceeds a predefined threshold level. Some European countries employ this type of definition in the analysis of their crash data and some academic studies use this definition as well. For example, Morency and Cloutier, who studied injury crashes involving pedestrians in the city of Montreal, Canada considered a location where there were at least eight pedestrian victims within a 5-year period *[9] [12]*. The current practice of identifying black spots in Flanders, Belgium, is similar, but takes into account the severity of the recorded crashes: a weighting scheme based on the severity of the crashes observed at a given location is implemented. Severe injuries receive three times more weight than crashes involving no injuries, and fatal crashes receive five times more weight *[13]*. Clearly, the selection of weights associated with the levels of severity influence which locations are identified as black spots *[14]*.

The data requirements for this method are comparatively low. The number of crashes at different locations is needed, along with any other characteristic of the crash or of the road that will be used in possible weighting schemes or other adjustments. The way that the threshold levels are defined often depends on resources. When there are more resources available, then the engineering team will be able to visit more sites and evaluate potential improvements; therefore, the required number of crashes or the crash rate to be observed at any location in order for that location to be considered for further evaluation will be smaller.

Statistical definitions make reference to a “normal” number of accidents or a “normal” rate of accidents for a type of location. A statistical definition of a black spot will take into consideration the number of accidents that will normally be observed for a particular road section, given its characteristics. Statistically significant departures from this norm identify black spots. This approach will often rely on a distributional assumption; for example, that the number of crashes is Poisson distributed. Comparing observed crash counts to a certain critical percentile of the assumed distribution is fairly straightforward. Note that emphasis must be placed on the term “statistically significant.” Simply observing a higher number of crashes than an expected value or above the mean of an assumed distribution is not sufficient to identify a hazardous location. By the very definition of an expected value, there must be some locations where higher crash counts will be observed.

Statistical models are often formulated in order to calculate the expected number of accidents for a location, conditional on the characteristics of the roadway and the traffic observed at that location. A safety performance function falls under this approach. Models have the strictest data requirements for identifying black spots, since there must be ample quality data available to formulate a statistical model. Both data regarding the crashes and their locations, as well as covariate data, are needed to estimate or calibrate a model. Using statistical models allows the analyst to consider many local risk factors in evaluating locations.

Elvik proposed the following generic definition of a black spot: "*A road accident black spot is any location that has a higher expected number of accidents than other similar locations, as results of local risk factors" [9]*. There are several important things to note here. First, the researchers were primarily concerned with the expected number of crashes, not the observed number. This was so that random variations in the data would not result in a location being wrongfully declared a black spot. The researchers were interested in picking up on systematic variations between crash locations and therefore the expected number of crashes was of primary importance. Secondly, the expected number of crashes at different locations should be compared to similar locations in order to identify black spots. The implications of these provisions are not trivial. It allowed for locations that exhibited fewer total crashes than other sites to be deemed less safe. The final point, again, ascertains that the increased number of crashes was systematic and not idiosyncratic. In the end, it was only worthwhile to identify black spots so that appropriate engineering measures may be implemented to improve the safety of the location. Obviously, it is only possible to evaluate and implement engineering measures when a particular potential for improvement can be identified. Most of the time, the statistical identification of black spots cannot draw conclusions regarding the particular local risk factor that contributed to the elevated number of crashes. Rather, it is only able to identify that a certain location should be evaluated further, possibly by means of site visits.

2. Identifying Hazardous Locations - General Approaches

Crash data can come in various forms and formats. Any analysis of data strongly depends on how the data were collected, and conversely, the type of analysis one is expected to perform with them depends on how the data are collected. The following will provide a brief discussion of how some jurisdictions following the numerical definition or the statistical definition of black spots identify black spots, and what type of data they require to do so. Then, it will be discussed in more detail the data requirements to estimate a safety performance function.

Identifying Black Spots without Models

Hungary uses a sliding window definition of a black spot. If there is a section of road that does not exceed 100 meters in length, and along that section there are more than four accidents observed in a three year period, then that section is identified as a black spot. One can think of this approach as a window of a certain width (100 meters in Hungary's case) that slides along each road segment of the road network to be examined. If a certain number of crashes within a certain time period (four crashes within five years, in Hungary's case) are visible in the window, then the location of the window is identified as a hazardous location. This is a simple numerical definition and other characteristics of the roadway and potential traffic patterns are not considered. The four “allowable” accidents are an absolute measure and no attempt is made to calculate an expected number of crashes of a given section given the road's characteristics.

Applying a sliding window technique does not necessarily imply the analyst is bound to use a simple numerical definition of black spots. Other countries, such as Austria, Denmark, Norway, and Portugal, use some variation of a sliding window technique to identify black-spots, but at the same time consider a “normal level of safety” for certain roadway elements before designating a location as hazardous *[9]*. Moreover, sliding windows may be only the first stage to conducting a more sophisticated analysis. For example, Hamidi, Fontaine, and Demetksy combined a sliding window approach with a statistical model-based identification method in the discussion of their network screening procedures *[15]*. Possibly, with only the simplest of cases in mind, the discussion found in Elvik, Sorensen, and Sorensen and Elvik nonetheless discourages analysts from using a sliding window approach to identifying black spots *[9], [16], [5]*. The researchers could not, however, find any argument against using sliding windows in combination with different approaches to black spot identification.

Statistical Models in Hazardous Location Identification

In some sense, the Highway Safety Manual (HSM) allows an evaluation of crash locations without having to specify a model *[17]*. Data regarding specific locations were simply converted into a calibration factor by looking up values in a table, and then an expected number of crashes is obtained for those characteristics. While this relieves the analyst of the burden of specifying his or her own model, the HSM still does implicitly use a model-based approach. The only difference is that in the HSM, the models have already been estimated by previous literature and the results are presented as “plug-in” estimators to the analysts. Of course, there is no guarantee that the models estimated to form the basis of the HSM are compatible with the data at hand, or if one should expect to arrive at very different results in a particular locality that is analyzed. One important thing to consider in the context of Louisiana roads, specifically, is that it is possible that no data from Louisiana were included in development of the Safety Performance Function (SPF) in the HSM. Also, the SPF in the

HSM was developed for base conditions, i.e., standard width of the road and other engineering factors. The expected number of crashes at a specific location is then obtained by multiplying the SPF function with the crash modification factors. The assumption for this modeling is that the crash modification factors are valid for Louisiana and no interaction between factors exists. Also, when an SPF function is calibrated for a state, one must keep in mind that if some factors are not available, then the calibration constant may be severely biased. For instance, if the driveway density is not available for Louisiana roads then the calibration constant may be inflated, depending on how many roads with high driveway density are included in the sample.

It has been noted that it is of primary importance to control for the specific characteristics of a location when deciding whether that location should be considered extraordinarily hazardous. The issue of philosophy has been highlighted before in this report. Regardless of philosophical viewpoints, it is instructive to calculate the expected number of crashes at given locations. An SPF is a mathematical function that models the expected number of crashes at a particular location using average annual daily travel (AADT). Yet, it is important to be careful about how to interpret the results of an SPF estimation or calibration exercise.

It is not enough to calculate the expected number of crashes or the expected crash rate at a given location, and then observe whether the actual number or rate of crashes is higher. Although some authors do take this approach, it is conceptually problematic and likely to result in the identification of more sites requiring further engineering evaluation than is feasible for any given agency (e.g., *[18]*). In fact, by definition of an expectation, it is necessary for some observed values to fall above the expectation as well as for some to fall below.

3. Identifying Hazardous Locations - Safety Performance Functions

Minimum Data Requirements for SPF Estimation

Sorensen lists several data prerequisites that must be available to the analyst in order to perform a state-of-the-art analysis of black spots:

- Locality
- Accident type
- Severity
- Time

- Road elements and the surrounding environment
- Circumstances and vehicles involved
- Records must have an acceptable level of reporting *[16]*.

Locality is, naturally, of prime importance. The analyst must be able to clearly identify precisely where the crash occurred in order to assign the crash to a specific road segment or intersection. The way this is defined strongly depends on the crash report data at hand. For example, mileposts along state routes and interstate highways, street names and distances to intersections, or GPS information may all be used to precisely pinpoint the accident. Some agencies, for example the German police, use large wall-maps and colored pins on which crashes are continuously recorded as they occur *[9]*.

Accident type refers to whether the crash was a head-on collision, a rear-end crash, sideswipe, etc. This information is necessary for successful black spot management, yet is only of minor importance when estimating a safety performance function. Severity is not actually used directly in the identification process *[16]*. However, many newer methods that attempt to account for the correlation structure of the different severities observed in crashes make extensive use of this, e.g., *[19], [20]*. A detailed discussion of these approaches follows in Section 4.

All of the above listed items are certainly necessary to successfully manage black spots. The researchers propose that the following prerequisite not listed by Sorensen deserves some attention in this context: A measure of exposure must be available *[16]*. In general, this means that Annual Average Daily Traffic must be available. Though this is often not explicitly outlined in practice, the variables that will be included in the specification of a safety performance function should have a solid theoretical foundation *[21]*. As identified by previous literature, exposure is the most important factor to include *[22], [11], [9], [23]*. Elvik asserts that inadequate, incomplete, and erroneous data on exposure are major weaknesses of current accident prediction models *[9]*. The fact that exposure is of utmost importance is not surprising; while there are some proposed remedies to estimating safety performance functions when traffic volume is measured with error in general, the most sophisticated techniques cannot be applied when the data are simply not available *[23]*.

Missing Data

One of the major problems to overcome in any statistical analysis of crash data is missing data. There are two different scenarios: Either, the data on crashes are incomplete, or data on covariates are not fully available. These two situations have different technical implications, as well as different approaches to remedy them.

There are few papers that explicitly discuss data shortcomings and the way in which those shortcomings were overcome. Abdel-Aty and Pande reported that when they used three lanes of traffic, which they planned to analyze separately, there were rarely valid data available for all three lanes simultaneously *[24]*. The data collected here were the number of vehicles, as well as their speeds for certain time intervals. The authors chose to use an average of the valid data for a location and assign this value to all lanes in that location. They justify this decision by noting that most of the crashes they analyzed were rear-end crashes and therefore lateral variation between the lanes was not of primary importance. Assigning averages is a very crude measure, but it is nonetheless a valid option when the nature of the missing data and the problem at hand allows it. For instance, most models use average annual daily travel for a road or highway; even so, the traffic varies substantially from month to month, from day to day, and even hour to hour. Thus, two roads with the same average annual daily travel are considered to have the same exposure, even though one road may have rush hour traffic while the other may not. More complicated methods, such as the imputation techniques developed by Zhang and Liu for dynamic traffic control systems, are likely too complex to implement for comparatively simple applications such as black spot identification *[25]*.

Missing crashes are another data issue to contend with. The police must be called to the scene of a crash for a record of the crash to exist. While it is reasonable to assume that very severe accidents (especially fatal crashes) will result in the involvement of police, for crashes that are less severe, the probability of having a crash report is decreased. Individuals are reluctant to get insurance companies and the police involved when the damage and injury are minor, or when the involved parties are not insured and, therefore, prefer to settle matters between themselves. Louisiana requires all motorists to be covered by valid liability insurance, yet non-compliance with this requirement in Louisiana is notorious.

This is a well-known problem, the extent of which has been widely studied (e.g., *[26], [27], [28], [29], [30], [31])*. These studies have focused mainly on differential reporting by crash severity or by the type of traffic participant involved (pedestrians, cyclists, etc.). It is likely, however, that there is some differential reporting in terms of local vs. state, or federal roads, as well, possibly as a result of a significant correlation of typical state/federal road characteristics and severity-potential. Speeds on local roads are often slower, compared to interstate routes or state roads. Thus, any crash that might occur is likely to result in less damage, and is, therefore, less likely to be reported.

While the literature cited above is concerned with identifying the discrepancy between crash reports and actual crashes, they all use hospital records linked to police records to identify this discrepancy. Those crashes that did not result in a hospital visit by at least one of the involved parties are never picked up by this methodology. In addition, there is no existing literature, of which the researchers are aware, that proposes a solution to the problem of missing accident records when attempting to identify hazardous locations or estimate safety performance functions. This is not an issue specific to local roads.

Lack of clear road segmentation is another issue that is entirely different in terms of issues that arise due to it. Generally, it is assumed that there is some sort of natural segmentation present in road data; i.e., that roads can be broken up into different segments where geometric characteristics do not vary within a segments. This may not always be present, and especially in the context of local roads in Louisiana, there is no unifying system, such as mile posts, that would allow appropriate segments to be created by the analyst. Indeed, most road segments are identified by name, rather than by numeric identifiers: e.g., “Main St between 4th and 5th." Since the length of the road segment cannot directly be inferred from such a descriptor, the analyst must calculate or measure it. If the segment is straight, then the length can be calculated from the GPS coordinates of the end points. The GPS coordinates are easily obtainable from the Internet (e.g., Google Earth) and the distance between the coordinates can then be calculated using standard distance formulas.

Calibration vs. Estimation

The data requirements to estimate a Safety Performance Function are generally higher than the data requirements to calibrate a model. This is primarily an issue at the intensive margin. One needs data on many locations to reliably estimate a model with many different characteristics between locations. There is, of course, no golden number of observations that needs to be achieved, but the number of locations for which a local jurisdiction can produce data may be not enough in most cases. In order to be able to effectively analyze crashes, data on as many engineering features of the crash location as possible, as well as all circumstances of the crash itself, are needed. The issue here is regarding a mere quantity of crashes needed to make SPF estimation reliable.

Since calibration relies on the coefficients of previous estimations, it is a viable alternative when jurisdictions are small and unable to estimate their own SPF. The HSM follows this approach, and its advantages and potential problems apply to local roads as well. A hybrid method of the two is also possible by first estimating a baseline model for a base condition using only AADT to model crashes, then using the crash modification factors (CMF) provided by the HSM to adjust for further local engineering features.

The more factors researchers were able to include, the more accurate the predicted crash count. So the researchers wished to include all available crash modification factors; however, it should be made aware that crash modification factors are estimates with a limited precision. Some crash modification factors have higher precision than others. Unfortunately, little is known about the precision of multiple factors when they are combined. If the SPF is estimated along with the crash modification factors, then the estimate of the variance will include the variation due to these factors. When a SPF from the HSM is calibrated and CMF's from the HSM are used then there is no simple way of computing a variance of the expected number of crashes. This is important for both the Empirical Bayes analysis as well as for classical hypothesis testing.

4. Safety Performance Function Estimations - Parametric Models

Poisson-Gamma Mixture Models (Negative Binomial Models)

Negative Binomial Models are very common in analysis of crash data. In fact, they are probably the most used specification in the recent literature. While the Negative Binomial distribution has its own interpretation, as the number of successful outcomes in a series of Bernoulli trials until a predetermined, non-random number of consecutive failures occurs, in the context of Negative Binomial regressions models, it is often derived as a Poisson-Gamma mixture model. Its properties, then, are very similar to that of the Poisson regression model, but it overcomes the restriction of equi-dispersion that is imposed in Poisson models. This allows for greater flexibility and applicability to different datasets, while retaining the positive characteristics that make Poisson-based models a preferred specification in integer response data problems.

The Poisson distribution has been used to model count data since the late 19th century. It has the important restriction, however, that the mean of the distribution is equal to the variance. This restriction on the variance is also called "equi-dispersion," and in order to overcome it, the researchers could model the mean of the distribution by using a gamma distribution with one additional parameter to be estimated, as the gamma distribution actually has two parameters, which are restricted to equal the reciprocal of each other. These result in a Negative Binomial model and due to its derivation in this context, the model is sometimes called a Poisson-Gamma mixture model.

Derivation of the Negative Binomial Model

The following is a brief discussion of the derivation of the Negative Binomial Model. A more detailed discussion can be found in Hilbe *[32]*. Let

$$P[\mathrm{Y} = \mathrm{y_i} \mid \mathrm{x_i}] = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\Gamma(1+y_i)} \qquad (1)$$

denote the probability of $y_i = 0, 1, 2, ...$, crashes at a location i, given specified vector of covariate x_i describing exposure and engineering features. The researchers specify the mean of the distribution by the equation

$$\lambda_i = \exp\,(x_i'\beta) \qquad (2)$$

where, x_i is a vector of covariates. In this standard Poisson model, the researchers obtain the conditional mean and conditional variance for the number of crashes at location i as

$$E[y_i|x_i] = var(y_i|x_i) = \lambda_i \qquad (3)$$

Among others, Greene and Cameron and Trivedi proceeded to derive the Negative Binomial model by introducing latent heterogeneity into the specification of the conditional mean *[33], [34]*:

$$E[y_i|x_i, \epsilon_i] = \exp(\mathrm{x_i}\beta + \epsilon_i) = \exp(x_i\beta)\exp(\epsilon_i) \qquad (4)$$

Assuming that $\exp(\epsilon)$ is Gamma distributed with mean 1 and variance θ^{-1}, the researchers can integrate $\exp(\epsilon)$ out of the joint distribution and obtain

$$P[Y = y_i|x_i] = \frac{\Gamma(\theta+y_i)}{\Gamma(1+y_i)\Gamma(\theta)}\left(\frac{\theta}{\theta+\lambda_i}\right)^{\theta}\left(\frac{\lambda_i}{\theta+\lambda_i}\right)^{y_i} \qquad (4)$$

This results in the model no longer being equi-dispersed, but rather over-dispersed with

$$E[y_i|x_i] = \lambda_i \tag{5}$$

$$var(y_i|x_i) = \lambda_i + \frac{{\lambda_i}^2}{\theta} \tag{6}$$

Since θ determines how much the variance will exceed the mean, θ is also called the "over-dispersion parameter." In fact, as $\theta \to 0$, the more the data will seem to be over-dispersed, and as $\theta \to \infty$ the model will again approach the Poisson specification and be equi-dispersed. Some authors prefer to specify the over-dispersion parameter as the inverse of θ, say $\alpha = 1/\theta$ *[32]*. While it is generally accepted that crash data are over-dispersed, it is possible to perform a statistical test for over-dispersion. Cameron first developed a method that involves only simple least-squares regressions to test whether the over-dispersion parameter (α, in this specification) is statistically significantly different from 0 *[35]*.

In this simplest form of a Negative Binomial model, the over-dispersion parameter will not vary from site to site, but is fixed for all locations i = 1,2,3,4,…. Hardin and Hilbe, as well as Greene suggest that the over-dispersion parameter itself could be modeled, for example as $\alpha_i = \exp(z_i\delta)$ *[36]*, *[37]*. Here the covariates z can contain, though not necessarily, the same variables as are contained in the vector of covariates x to model the mean. Lord and Park called this the Generalized Negative Binomial (GNB) model and investigated whether a better goodness-of-fit to the data may be obtained by using this GNB model compared to the traditional Negative Binomial specification *[38]*. Their conclusion was that the GNB does indeed provide better statistical properties and they suggested that, specifically in models where only traffic flow data are available, this specification outperformed the traditional Negative Binomial specification. El-Basyouny and Sayed used a very similar setup that they call the "Modified Negative Binomial" model *[39]*. They found little difference in the performance of the traditional Negative Binomial model and the modified version. However, they do suggest that as the data become more over-dispersed, the benefits of using the modified Negative Binomial model are expected to become clearer.

Some data are characterized by "excess zeros." This means that there were many locations that had no crashes during the time of observations. The distributional assumptions of either the Poisson model or the Negative Binomial model and its variants support "zero-observations;" in fact, both distributions require that some locations are observed that did not record any crashes. There is no clear threshold of what constitutes "too many" cases of zero crashes, but there are some techniques that the analyst is able to consider when a problem of this nature observed. These models are called Zero Inflated models and they exist for both the Poisson specification, as well as the Negative Binomial specification *[37]*.

While Zero Inflated models are useful and valid in many contexts, there are several strong arguments against using those models in crash data analysis. Lord, Washington, and Ivan argued that carefully selecting the time/space scales for analysis, obtaining better or more data as explanatory variables, as well as introducing latent heterogeneity terms in standard, is preferred *[40]*. The authors argued that while providing a better statistical fit to the data, there are severe logical fallacies in applying Zero Inflated models to crash data, as one implicit assumption will be that certain locations are inherently safe. Statistical goodness of fit is not the prime criterion when specifying safety performance functions, as potential over-fitting can lead to dangerous and misleading conclusions. Lord, Washington, and Ivan provided some intuitive illustrations of the implicit assumptions one must entertain when estimating Zero Inflated models *[41]*. The title of their work, "Many zeros does not mean zero inflation," succinctly sums up the issue *[42]*.

In the particular case of local roads in Louisiana, the problem is likely on the opposite end of the spectrum. Excess zeros or simply many zeros only arise when a true inventory of road segments is available, and there are many road segments without crashes. If such road inventory does not exist, then the road segments are likely to include only those sites where some crashes were observed. Those roads segments or intersections that had no crashes during the time frame of the analysis will not be included in the data. If road segment and traffic characteristics are not readily available and must be compiled, then they will be compiled for only those sections that the analyst knows about: those where at least one crash occurred. In essence, this is a zero-truncated model, which technically violates the distributional assumptions of the Negative Binomial (as well as the Poisson) regression model. In order to alleviate this problem, some researchers employ a zero-truncated Poisson regression model *[3]*. A zero-truncated Negative Binomial model exists as well and is readily available in standard statistical software; however, the researchers could not find any research study in traffic safety or in related fields that makes use of the truncated negative binomial specification.

Negative Binomial Models in the Literature

While in crash analysis the Negative Binomial specification is most often used in the context of an Empirical Bayesian Analysis (to be discussed in a following section), there are some papers that make use of the direct Negative Binomial specification or one of its variants. For example, some used Poisson regressions and Negative Binomial regressions to investigate whether the coefficients in the model differ for various crash types *[43]*. It is reasonable to presume that the impact of an increase in traffic density is different in head-on collisions compared to rear-end collisions. The study found that there were large differences in the size of the coefficients of annual average daily traffic, the presence of turning lanes, the number of driveways, and other geometric road design features.

Hadi et al. investigated cross section designs using Negative Binomial regression *[44]*. They also tested for over-dispersion and concluded that indeed the dispersion parameter is statistically different from indicating a Poisson model using their data. Others used a Negative Binomial model to arrive at their concept of "level of service of safety" *[18]*.

Zhang, Ye, and Lord proposed a bootstrap method to improve the estimation of the over-dispersion parameter *[45]*. They conducted Monte Carlo simulations to show that estimation of the over-dispersion parameter can be improved by using a bootstrap resampling method combined with maximum likelihood estimation whenever the sample size is small and the sample means are low. Applying the same method to un-signalized intersections in Toronto, Canada, validated the simulation results. While not the immediate focus of the study, Chang *[46]* used a Negative Binomial specification as a benchmark to evaluate the performance of an Artificial Neural Network as an alternative method for analyzing accident frequencies.

Common Sources of Error

There are several issues and potential sources of problems when trying to fit a safety performance function to data. Elvik proposed three common forms of error:

- Omitted-variable bias
- Co-linearity between explanatory variables
- Incorrect specification of the functional relationship between outcomes and covariates used in the model *[9]*.

Omitted variables in developed models create a bias in predicted crashes for all locations. This bias also occurs in the calibration coefficients when variables are omitted. This bias has been analyzed and a general consensus on which factors to include seems to exist *[9]*. Still, in practice the question of what to include in the model is not a theoretical one, but is rather dictated by the availability of data *[5]*. At a minimum, well defined road segments, a measure of exposure and some engineering features must be available in order to conduct any kind of analysis. For instance the HSM specifies which factors are required for the calibration of the SPFs provided in the HSM and which factors are desirable. Most research that the researchers surveyed has been concerned with finding ways to analyze or correct problems that occur due to incorrect specification of the functional relationship between outcomes and covariates used in the model.

Multivariate Models

While the Poisson-Gamma model is certainly the most popular of the Poisson mixture models, there have been several attempts made to refine its approach or to tailor the analysis to a very specific problem at hand. Specifically, it is unclear how to treat the issue of different accident severities in the negative binomial framework, so some analysts have modeled the accident severity categories separately instead of pooling them together. However, this neglects the important fact that the count or the rate of different types of accidents may be correlated. Therefore, several attempts have been made to model the crash count or the crash rate jointly with crash severity. The model structure allows for unspecified correlation between the severity and counts/crash rate that will be estimated as well and will overall lead to more reliable point estimates *[19]*. The favorite specification of such multivariate models in the literature is the Multivariate Poisson model, or the multivariate Poisson-lognormal mixture model.

Ma and Kockelman used a Gibbs sampler to estimate a multivariate Poisson model for highway crashes in Washington *[47]*. They compared parameter estimates and goodness of fit of the model with a set of independent Poisson regressions and concluded that their multivariate Poisson model has promise. The authors also proposed that a multivariate Negative Binomial Model could be applied to this problem, as was previously developed in the context of budget constrained activity demand analysis by Kockelman in order to allow for a more flexible treatment of over-dispersion *[48]*.

A more frequently used multivariate model is the multivariate Poisson-Lognormal mixture regression model. Examples of this include Park and Lord who modeled crash frequency and severity in a simultaneous framework, as well as Ma, Kockelman, and Damien, and Aguero-Valverde and Jovanis, who modeled crash counts and severity simultaneously *[19], [49], [20]* . El-Basyouny and Sayed provided a detailed analysis of the performance of the multivariate Poisson-lognormal mixture model *[50]*. Their results indicated that the estimates of additional Poisson dispersion parameters were considerably smaller in the multivariate Poisson-lognormal mixture model compared to univariate models. They concluded that using the multivariate specification increases precision, and that this improvement in precision is mainly due to the fact that correlations between latent variables in the reduced form of low-severity and high-severity crashes are accounted for.

The issue with their models is the tradeoff between data quality and the level of dis-aggregation that is useful in an analysis. When the data are very noisy, then there is little use in breaking the individual observations up into increasingly smaller units, as the signal to noise ratio deteriorates. This may be counteracted by attempting to perform a type of site aggregation *[15]*. In the context of local roads, however, there is likely not a sufficient number of sites available to the analyst to perform such an aggregation. Moreover, as identification of hazardous locations was the goal of the estimations, aggregating sites distinctly misses the point of what the researchers tried to accomplish.

While, as noted above, there have been a number of academic studies regarding the joint modeling of crash counts and severities, the use of these models is limited in practice. Typically, analysts perform separate estimations or calibrations by severity level, or they completely disregard crash severities for much of the analysis *[1], [17]*.

5. Empirical Bayesian Analysis

Bayesian and Frequentist methods have always been on opposite ends of the statistical philosophical spectrum *[51], [52], and [53]*. In transportation safety research, arguments for or against Bayesian or Frequentist approaches have not been rooted in philosophy. Interestingly, there have been many arguments for the Bayesian approach that base their reasoning on comparing predictive power and goodness of fit of a particular model, after conducting Monte Carlo Simulations. This approach is problematic, as will be discussed later in this report. First, the researchers provide a general description of Bayesian data analysis, as well as some details of the commonly used methods in the context of roadway and intersection safety.

Bayesian data analysis in general consists of the following three steps (for more detail, see *[54])*.

1. Setting up a full probability model. The analyst specifies a joint probability distribution for all observable and latent quantities in a problem. To this effect, the analyst will make use of the current state of knowledge regarding the problem and should also be conscious of the details of the data collection process.
2. Conditioning on observed data. This means calculating the conditional probability distribution of the latent quantities under study, conditional on the observed data.
3. Examining the model fit and interpreting the obtained posterior distribution. Also the sensitivity of the results to the modeling assumptions made during Step 1 should be examined.

The above is fairly generic, so this report will discuss in some more detail the specific models that have been applied to crash data analysis. First to be described and discussed will be the Empirical Bayes Approach, since several authors have identified it as the preferred method of estimating safety performance functions *[9], [16], [5], [55]*. Then, will be discussed some full Bayesian approaches. The latter are often highly complex models that require a significant amount of mathematics in their derivation; some of them are also associated with significant computing and programming effort. This report will, therefore, concentrate more detailed discussion of those models that are more commonly used in safety engineering practice and those approaches that will provide a reasonable balance between model complexity and model performance.

Empirical Bayesian Analysis has been the recommended method for conducting before-after observational studies *[55]*. This is due to the fact that it minimizes issues regarding regression to the mean. “Regression to the mean” is a term used to indicate that a particularly high number of observed crashes at a certain location may be the result of random statistical variation, and is not attributable to any location-specific issues with the roadway or intersection design. If the number of observed crashes is high due to random statistical variation, then the researchers would expect the number of crashes in the following period of observation to be lower. The converse applies, as well. In the context of before-after studies, particular sites may have been selected for treatment exactly because they exhibited high crash counts or a high crash rate. In the after period, the researchers would therefore expect the number of crashes to decrease. This decrease may falsely be attributed to the treatment of the site, and may not take into account random statistical variation. The Empirical Bayes method attempts to correct for precisely this effect.

The researchers will briefly summarize the derivation of the Empirical Bayes approach *[56]*. A tutorial that introduces the practitioner to Empirical Bayesian analysis of crash data, as well as further examples can be found in *[10]*. Casella provided a short and highly intuitive introduction to general Empirical Bayes data analysis *[52]*.

The Bayes Theorem states that the conditional probability of the expected value λ given observed values y is proportional to the conditional probability of y given λ times the prior distribution of λ.

$$p(\lambda|y) \propto f(y|\lambda) \times g(\lambda) \tag{7}$$

In the end, the goal is to obtain a good estimate of the mean of the data generating process of the data; i.e., the researchers are interested in the expected value $\hat{\lambda}(y)$. Assuming a Mean Squared Loss Function, the Bayesian Estimator of the quantity of interest is the expected value of λ conditional on y; i.e., $\hat{\lambda}(y) = \int \lambda \cdot p(\lambda|y)d\lambda = E[\lambda|y]$. This is very intuitive. The best estimate for the expected values of y is that value that has the highest probability of being the mean of the data generating process for y, conditional on the values of y that was observed. Some assumptions regarding the probability densities are necessary to derive this equation. For this example, assume that the likelihood and the prior density are Gaussian, i.e., then

$$y|\lambda \sim N(\lambda, \sigma^2) \quad and \quad \lambda \sim N(\mu, \xi^2) \tag{8}$$

where, $N(\cdot,\cdot)$ indicates the Gaussian probability density function. The σ^2 is within sample variance and ξ^2 is the between sample variance. With the proper substitutions for the mean and the variance, plug this function in place of $f(\cdot)$ and $g(\cdot)$ from the Bayes equation (7). It is now easy to show that $p(\lambda|y)$ is actually still a Gaussian probability density, though with different means and variance.

$$\lambda|y \sim N\left(\frac{\sigma^2\mu + \xi^2\bar{y}}{\sigma^2 + \xi^2}, \frac{\sigma^2\xi^2}{\sigma^2+\xi^2}\right) \tag{9}$$

And therefore,

$$\bar{y}^{Bayesian} = \widehat{\lambda}\,(y) = E[p(\lambda|y)] = \frac{\sigma^2\mu + \xi^2\bar{y}}{\sigma^2+\xi^2} \tag{10}$$

Or, after rearranging, the more familiar form:

$$\bar{y}^{Bayesian} = \frac{\sigma^2}{\sigma^2+\xi^2}\mu + \frac{\xi^2}{\sigma^2+\xi^2}\bar{y} = \omega\mu + (1-\omega)\bar{y} \tag{11}$$

The Empirical Bayes estimate is, therefore, a weighted average of the overall mean μ and the sample mean $\bar{y}$, with the weights the fraction of the within-sample variance and between-sample variance. The larger the between sample variance, the more weight is given to the individual sample mean. While this simple example was provided using the Gaussian distribution, a very similar result will be obtained when using different distributions and the characteristic appearance as a weighted average is preserved. Most common in traffic crash analysis is a Poisson distribution with a Gamma-distributed prior. This is closely related to the Negative Binomial regression model outlined in the previous section and is, therefore, favored by traffic safety analysts. The Gamma distribution is a conjugate prior to the Poisson distribution, and using L_i as the section length the within-sample variance is

$$\sigma^2 = \hat{\lambda} L_i \tag{12}$$

and the between-sample variance is:

$$\xi^2 = \frac{(\hat{\lambda} L_i)^2}{\theta L_i} \tag{13}$$

Therefore, the equation obtained through substitution is:

$$\bar{y}^{Bayesian} = \frac{1}{1+\frac{\hat{\lambda}}{\theta}} \hat{\lambda} + \left(1 - \frac{1}{1+\frac{\hat{\lambda}}{\theta}}\right) \bar{y} \tag{14}$$

where, θ is the over-dispersion parameter estimated for the negative binomial regression. Note, again, that some authors prefer to use $\alpha = \theta^{-1}$ in their specifications. For this purpose it is essential that the over-dispersion parameter is estimated per unit length for road segments (for example, over-dispersion per mile), and it is essential that the length units between crashes and the over-dispersion parameter are compatible *[55]*. Intersections do not technically have a length and therefore the length is taken to be unity. Moreover, in most studies the over-dispersion parameter is taken to be fixed. As discussed earlier and again later in this section, this need not be the case. The estimate $\hat{\lambda}$ is the regression estimate for the specific AADT and road characteristics. If an SPF is not available, the within and between sample variance can be estimated directly from the sample.

Empirical Bayes Estimates

Hauer suggested using the method of sample moments for estimating population parameters of the distribution. The moment estimators for the within and between sample variances are

$$\hat{\sigma}^2 = \bar{X} \tag{15}$$

and

$$\hat{\xi}^2 = s^2 - \bar{X} \tag{16}$$

respectively *[57]*. That is, since the variance is equal to the mean in the Poisson case, the within sample variance (and the overall mean) is best estimated by the overall sample mean and the between sample variance is best estimated by the sample variance minus the sample mean. Since these estimates come directly from observed data, they are increasingly accurate as sample sizes increase. Alternatively, it is possible to use the maximum likelihood estimates for $\hat{\lambda}$ and k obtained from the model using the relationship in the analysis; compare both methods for estimating the mean and variance and also compare the Empirical Bayes estimates therein produced.

In addition to providing Empirical Bayes estimates, a tail probability using the estimates can also be calculated. This tail probability provides a way to rank highway segments with different AADT and engineering features. This tail probability is calculated using the (posterior) distribution of the Empirical Bayes with parameters

$$\beta_1 = \frac{\mathrm{EB}}{\mathrm{VAR(EB)}} \tag{17}$$

and

$$\alpha_1 = \beta_1 \cdot \mathrm{EB} \tag{18}$$

Then the tail probability is calculated as

$$\int_0^{\hat{\mu}} \frac{\beta_1 \lambda^{\alpha_1 - 1} e^{-\beta_1 \lambda}}{\Gamma(\alpha_1)} d\lambda \tag{19}$$

Using a particular cut-off for this tail–probability value, the analyst can narrow down his or her list of sections that may need attention, as well as rank them, regardless of their characteristics. Again, α and β in the above equations can be determined using the method of moments (MME) or maximum likelihood estimates (MLE).

There are several assumptions that must be satisfied in order for the Empirical Bayes approach to be valid. The most important one is probably that there must not be a significant time trend in the data. This is relevant, for example, when admitting that the demographic characteristics of a location's typical driver might change. Suppose, for example, that over time there are more mature drivers in an area and teenagers move out, little by little. Then, there is an inherent trend in the expected value of crashes that occur in that area. If one considers five years of crash data in order to arrive at the estimate of the expected value, then the weighting scheme will result in a systematic overstatement of the expected value by the Empirical Bayes estimate.

In addition to before-after studies, the Empirical Bayes specification is also used to identify black spots. The identification of black spots also has the potential to be influenced by regression to the mean; therefore, it is not only useful to take into consideration the condition of the geometric road features and traffic volume, but also the inherent randomness of crash counts. In essence, the expectation of the crashes is modeled as a negative binomial model and combined as a weighted average with the average crash count of that location in order to assess the level of safety at that particular location.

Empirical Bayesian Analysis and Local Roads

Some researchers conclude that the Empirical Bayes Method is preferred over all other methods *[6], [9]*. Indeed, using the Empirical Bayes approach is the current standard practice for traffic safety professionals. With the introduction of the Highway Safety Manual, the Empirical Bayes was attempted to be made more accessible to a larger number of traffic engineers and states' Departments of Transportation.

In the specific context of local road safety performance functions, Jo et al., in their data analysis software for local agencies in Illinois, first developed a safety performance function using Negative Binomial regression, and then used the Empirical Bayes procedure to arrive at the estimated number of crashes at each segment or intersection *[1]*. They estimated a Negative Binomial model for all crashes of a particular severity type (they included only fatal, major injury, and moderate injury crashes) for a particular peer group of roads. A peer group was a 12-category system of functional classes. The coefficients of the covariates, as well as the estimated dispersion parameter per mile of roadway length were stored in a database. The local agencies were then able to input their own data for a particular roadway into a provided computer program and an Empirical Bayes adjustment was performed in order to assess the safety performance of the local road in comparison to other similar roads.

In Virginia, Hamidi, Fontaine, and Demetksy, provided a more macroscopic perspective by aggregating together sites into sections of intermediate length for analysis *[15]*. The authors understood *site aggregation* to mean that road segments that are geographically adjacent and have identical characteristics in terms of traffic density and geometric design features are combined. This means that noise in the calibration/estimation process of those aggregated sites will be reduced, compared to using a larger number of smaller segments. They compared different SPF specifications for those aggregated sites with respect to applicability and performance specifically in Virginia. For site prioritization, the authors used a sliding window approach that calibrates the SPF at each sliding window location, and then applied an Empirical Bayes correction. The excess crash frequency for each site was then calculated by site, as well as by site-mile, and the top 5% of a ranked list of excess crash frequencies identified. This is a combination of many different methods and approaches to black spot identification. Still, the Empirical Bayes component in Hamidi, Fontaine, and Demetksy's analysis was considered to be very important *[15]*.

In terms of data constraints, Hamidi, Fontaine, and Demetksy provided some degree of discussion; however, since the roads considered in this study were part of the primary system in Virginia, it appears that a good road inventory system was present *[15]*. Still, the main discussion centered around the practice of matching crashes from a crash database to existing road inventories. Primary road numbers were available and were used to match crashes in a certain area. Due to the aggregation of sites, no sub-segment details (such as mile-posts) were considered in the analysis. Data requirements at the aggregated sites were also very broad; as they only used the length of the particular segments and AADT in their analysis *[15]*. Since the site data were aggregated, all the characteristics of the crashes aggregated together must be constant within the level of aggregation.

Other researchers explore Hilbe’s Generalize Negative Binomial model in the context of traffic data *[32]*, *[39], [58]*. The generalized model allows the variance function to be made conditional on parameters as well by means of modeling the over-dispersion parameter. This certainly has implications for Empirical Bayesian analysis, as the weights for the Empirical Bayes formula make use of the variance of the estimate in their theoretical derivation and explicitly involve the over-dispersion parameter. Lord and Park investigated exactly those implications and found that statistical properties of the estimates can be much improved using the Generalized Negative Binomial formulation *[38]*. While Lord and Park used a sample of roads from California in their analysis, the researchers were unable to find an application that used a generalized approach to crash data analysis in practice for the purpose of identifying black spots *[38]*. Despite its merits, the discussion of generalized negative binomial models seems to be restricted to academic papers and has not yet made its way into safety engineering practice.

Empirical Bayes Data Needs

There are no data needs *per se* that are imposed by the Empirical Bayes method itself. The constraint is the availability of good quality data that will allow the analyst to estimate (or calibrate) the expected number of crashes of similar locations that is to be entered into the weighted average formula of the Empirical Bayes estimator. The same is true for the variance of the estimate that is used in the weights themselves.

If the estimated number of crashes at similar locations is not a very good estimate, its variance will be large, and the Empirical Bayes procedure will account for this. However, in order to even run the Negative Binomial regression model that is recommended for crash data, or to calibrate a SPF, the researchers must, at least, have the section length, as well as the (Annual) Average Daily Traffic for road segments, and for intersection data, the researchers must have (1) major road AADT and (2), minor road AADT. Moreover, it will be necessary to be able to divide the road segments into different function classes. Finally, it is necessary to be able to link all accidents that occurred on a road segment back to that road segment.

All of the discussion regarding the data needs for Negative Binomial regression models applies in this context as well. While the Empirical Bayes procedure does alleviate some problems that may arise regarding regression to the mean, it does not solve any problems that arise due to data quality issues.

Statistical Evaluation of SPF - Classical Analysis

There has been some literature suggesting that Bayesian methods are a preferred approach to identify hazardous locations; however, the comparisons that are entertained by those papers are not always the appropriate ones. Elvik, for example, compares Empirical Bayesian methods to sliding window approaches and other numerical definitions of black spots *[9]*. Others base their conclusions on Monte Carlo Experiments where, by assumption, they have chosen a correct functional specification *[6]*. Although the Bayesian Approaches to evaluating a fitted safety performance function are popular, some arguments for using a Bayesian approach are not solid from a statistical viewpoint. The more valid comparison is to compare the Bayesian approach with classical hypothesis tests of whether an observed value is statistically significantly different from the mean of the distribution.

The classical approach also starts with estimating a Safety Performance Function based on the Negative Binomial Model. But instead of using the expected value from the SPF as part of a weighting scheme, a statistical test is performed to determine if the observed value is significantly different from the expected value from the SPF. There are many different suitable statistical tests that may be used to test whether the difference of the observed and the expected value is significant. Since the distributional assumptions and low means do not satisfy the criteria necessary for a standard t-test, it may be useful to consider non-parametric tests in this case. Note that critical value will vary with section length and all the covariates included in the model. Using a classical approach is subject to the same assumptions as used for the Empirical Bayesian analysis.

6. Inexpensive Countermeasures

The serious traffic crash problems on local roads (mainly rural) have been recognized in recent years. The Federal Highway Administration (FHWA) has been leading the way to establish programs and funds to improve local roadway safety as part of their Highway Safety Improvement Program (HSIP). In their Local Road Resources CD, FHWA provides all practical information and guidance to agencies responsible for roadway safety *[59]*. In addition to FHWA, Minnesota has also been at the forefront of promoting roadway safety of rural highways. The Center for Excellence in Rural Safety (CERS) at the University of Minnesota facilitates research, training, and outreach activities related to rural transportation safety *[60]*.

The information regarding low cost crash countermeasures from various sources including FHWA and CERS can be summarized into four different topics:

1. targeted areas;
2. specific countermeasures;
3. observed benefits;
4. selection guideline *[61]*.

Targeted Areas

All 4E areas have been documented as important in reducing the number of crashes and crash impact. Because low traffic local roads are often designed to a lesser standard than state and US highways with more traffic volume, engineering countermeasures are mostly used to make local roadways more forgiving. Enforcement and education are also considered important in changing unsafe travel behavior. Also, higher fatality rates at local rural roads call for a quick emergency response to crashes to increase the chance of survival in crashes.

Specific Countermeasure

The majority of the concrete crash countermeasures are from roadway engineering, which has been documented by different reports targeting crashes at different locations. For instance, to reduce the number of run-off-roadway crashes at a horizontal curve (one of the most common type of crashes on rural roads), the following low cost countermeasures have been widely used on local roads including those in Louisiana: provide advance warning signage, add chevrons along the curve, add embedded pavement markings and enhanced curve delineation, and add roadside reflectors to delineate curves. Crash countermeasures in targeting bad driving behavior on rural local roads include: developing safety policy directed at driver distraction, seat belt and alcohol enforcement campaigns in rural areas, focusing on increasing perception of risk, and using technologies such as speed monitors fitted to vehicles driven by teens in rural areas that are too isolated to police effectively.

Observed Benefits

The effectiveness of low cost crash countermeasures have been evaluated in both qualitative and quantitative terms by several studies, most of them led by FHWA *[61], [62], [63], [64], [65], [66], [67], [68], [69],[70], [71], [72].* The quantitative evaluation demonstrates the reduction in crashes through the observational before-and-after studies. Due to the difficulties in obtaining the cost information, very few studies mention the cost and benefit ratio information even though C/B analysis has been very much promoted in roadway safety.

Selection Guideline

When selecting a countermeasure, the following issues need to be considered:

<u>Technically feasible</u> – Is the countermeasure feasible for the particular location? Does it comply with existing guidelines and/or standards?

Advantageous Cost/Benefit – Does the benefit of the countermeasure outweigh the costs? Are there more cost-effective strategies to consider?

Affordable and Practical – Considering the identified problem, is the countermeasure practical? Can it be funded?

Acceptable – Will the public accept the countermeasure politically and within the community? Will there be educational needs for the public?

Legal – Is the countermeasure legal to use? For example, speed limits are regularly revised without proper authorization, and STOP signs are used without meeting the appropriate MUTCD warrants *[73]*.

7. Multi Criteria Decision Making

A highway improvement project consists of the countermeasures applied to a specific site, i.e., a road segment or intersection. Usually at each site, different types of countermeasures described above are possible with different costs of implementation and benefits in terms of crash cost reduction. While there are crash-reduction factors that have been developed by researchers for most of the countermeasures considered in this project, cost estimates for improvements are time consuming and may require costly engineering analysis. Therefore, in practice, only a small number of sites are considered for improvement. Given a limited budget, the question arises how to rank the projects according to costs, benefits, and risks. This leads to a multi-criteria decision problem. Solutions to this multi-criteria decision-making problem will be reviewed in this section.

A number of studies concentrate on the cost-benefit analysis of certain projects using various outcome measures. One study applied an incremental cost–benefit analysis approach toward highway projects *[74]*. Another analyzed the reduction in the expected number of accidents due to highway improvements *[75]*. Others considered the effect of improvements on the severity of accidents; evaluated the safety impacts of highway projects using various measures; and estimated the effectiveness of projects in reducing crashes *[76], [77], [78]*.

The above cited articles deal with the evaluation of single projects. Transportation departments have fixed budgets, and so they are only able to fund a limited number of road improvement projects each year. To find an optimum selection of projects, Melachrinoudis and Kozanidis applied a mixed integer knapsack solution to the selection of projects by maximizing the total reduction in the expected number of accidents under a fixed budget constraint *[79]*. Since most decisions on project selection involve immediate costs but benefits occur in the future, Brown applied dynamic programming to obtain a set of projects which provide an optimum solution which take into consideration present costs but benefits over several years into the future *[80]*.

While most previous research articles examining the selection of public projects use cost-benefit analysis, there are a few notable exceptions. Norese and Viale used a multi-criteria sorting procedure to support public decisions *[81]*. Yedla and Shrestha suggested a selection method for environmentally sustainable transport systems *[82]* and Hinloopen et al. applied cardinal and ordinal judgment criteria to the planning of public transport systems [*83]*. Recently, Odeck used the Data Envelopment Analysis (DEA) approach for measuring target performance for traffic safety and Tudela et al. compared a cost-benefit analysis with multi-criteria decision methods for transportation projects *[84], [85]*.

Recently, Ghorbani and Rabbani published a new multi-objective algorithm for the project selection problem *[86]*. Two objective functions have been considered to maximize total expected benefit of selected projects and minimize the variation of allotted resources. Kozanidis solved a knapsack problem with two objectives: profit and equity. The second objective minimizes the maximum difference between the resource amounts allocated to any two sets of activities *[87]*. Zongzhi et al. developed a heuristic approach for system-wide highway project selection to achieve maximal total benefits *[88]*. Teng et al. published an empirical study of the budget allocations in northern Taiwan *[89]*.

Although the literature of multi-criteria ranking and selection methods is very rich, the majority of those methods are not applicable to this situation because it requires user input in most cases. Instead, the researchers propose the use of Data Envelopment Analysis (DEA) which is based on preference weights without user input. DEA was first developed by Farrell and consolidated by Charnes et al. as a non-parametric procedure that compares decision units using performance indicators *[90], [91]*. The DEA method has been applied for ranking in several different areas (see, e.g., *[92]*, which contains a list of over 1500 references). Recent extensions and applications include a matrix-type network DEA algorithm and its application for the performance measurement of a transportation network *[93]*, *[94]*.

In DEA, the preference weights are calculated by linear programming, an optimization method which can easily be applied to evaluate and rank thousands of projects. DEA is an extreme point method; i.e., it compares each project to all other projects with weights calculated to be the most favorable for the particular project being evaluated. This is the major advantage of this goal because it ensures that none of those projects, which may be advantageous in any one of the important criteria (or in any weighted combination of the criteria), are ranked low. Only those projects that are *not* preferable for any weighted combination of the criteria are at the bottom of the ranking.

In DEA, the efficiency of a project is the weighted output over weighted input. The objective of the DEA is to identify projects that produce the largest values of outputs by consuming the least amount of inputs. For instance, the input would be the cost of the project, and the output would be the gain in safety which can be measured by the crash cost reduction. Let *s* be the number of different output criteria, *m* be the number of input criteria used, and *N* be the number of different projects. Consider a specific project k = 1, 2, … , N, where R_{ik} represents the measure of the *i*th output criterion (i = 1, 2, …, s) and V_{jk} represents the measure of the *j*th input criterion (j = 1, 2, …, m) for project k. The efficiency of project *k* is measured as the weighted sum of outputs over the weighted sum of inputs (as in productivity measures)

$$E_{kk} = \frac{\sum_{i=1}^{s} v_{ik} R_{ik}}{\sum_{j=1}^{m} u_{jk} V_{jk}} \tag{13}$$

By using DEA, the researchers attempted to find optimal weights, u_{jk}, and v_{ik}, for each project that maximizes the project efficiency, E_{kk}, by comparing each project, *k* , with all other projects subject to the restriction such that the weights are nonnegative and all $E_{kn} \leq 1$ for n=1, 2, …, N. For each project, *k*, the above optimization problem can be described as the equivalent linear programming problem

$$\begin{aligned} & E_k^* = \max\ E_{kk} \\ & \text{s.t.}\quad \mathrm{E}_{kn} = \frac{\sum_{i=1}^{s} v_{ik} R_{in}}{\sum_{j=1}^{m} u_{jk} V_{jn}} \leq 1 \quad (\mathrm{n=1,...,N}) \\ & \quad v_{ik} \geq 0 \quad (i = 1,...,s) \quad u_{jk} \geq 0 \quad (j = 1,...,m) \end{aligned} \tag{14}$$

with E_{kk} defined in (13). The efficiency of a project, the E^*_k value, is the *DEA efficiency measure* for project *k*. There are two basic cases:

$E^*_k = 1$ project k is efficient (Pareto-optimal project), and

$E^*_k < 1$ project k is not efficient; it is *dominated* by other project(s).

To illustrate this concept, consider two output criteria, such as the odds for a crash count and the percentage of injury crashes at each road segment. For two criteria, the idea of DEA can be shown in a graph. In Figure 1, the points A, P, and B represent the road segments. The line y2-A-B-x2 is the efficient envelope. Any road segment that has a weighted average of the two criterions inside the envelope is not as efficient as the points on the envelope. The road segment P has an efficiency computed as the distance OP divided by the distance OD.

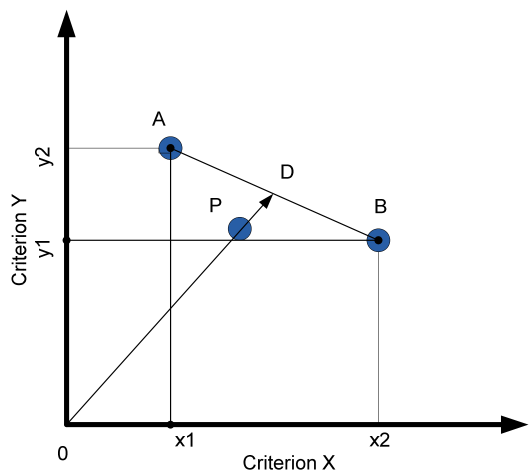


Figure 1
DEA example

Note that higher efficiency of a project in this context means better selection for safety improvement. Based on the DEA method of (14), for each project, the researchers calculate the $\boldsymbol{HE_k} = E^*_k$ value. This is considered as the *Efficiency Measure of project k.* The literature mentions the potential disadvantages in using DEA ranking, such as too many efficient projects, the sensitivity of the selection of the projects included in the list of projects, and the data estimation error.

The method of combining the DEA ranking with Multi-Criteria Decision Making has been used for different applications. For instance, Golany combined interactive, multiple-objective linear programming with DEA; Stewart compared the concepts of efficiency and Pareto Optimality in DEA and MCDM *[95], [96]*. Furthermore, Belton and Stewart stated that MCDM is generally applied to ex-ante problem areas where data is not readily available, such as in the case of future technologies *[97]*. DEA, on the other hand, provides an ex post analysis of the past as a basis to learn. Since project selection case is based primarily on existing statistical data related to such areas as accident risk, cost, and benefits the DEA ranking gives valuable information for selecting the future improvement projects.

The advantage of this methodology is that it can be applied to project selections in which (1) tens of projects are to be selected out of hundreds of potential projects pre-selected from thousands of potential projects and (2) the selection is based on multi-criteria objectives which require no user input information regarding the weights or *pairwise* comparison of the projects.

8. Summary and Conclusion from Literature Review

In this study, the researchers reviewed the existing literature on the identification of black spots (among other considerations) and identified a procedure that will allow analysts to identify black spots on local rural roads. The researchers presented a detailed review of the available methods and each of the requirements for all of the methods and concluded that estimating the number of crashes in an Empirical Bayesian framework based on the Negative Binomial Statistical model is the preferred approach to determine expected crash counts for highway sections and intersections. However, the researchers will not use the terminology "black spots" or "abnormal locations," as they do not wish to imply that statistics can determine which locations have safety hazards. Statistics can only rank the crash locations according to some established criteria and select the top p% for further investigation for safety hazards. The point of using the SPF and crash modification factors in the modeling is to discriminate between crash locations with different exposure levels and engineering features, and to take this into account when computing expected crash counts for locations. The regression model provides a better predicted crash count than the average crash count for this location because it uses information from all locations. Provided that the model is appropriate for the data analyzed, a regression line is usually a better predictor for the true mean crash count at a location than the average crash count at the location, assuming that the regression modeling is done simultaneously for exposure and crash modification factors. In the HSM’s two-stage approach the SPF is “calibrated” and multiplied with the CMF’s, it is impossible to guarantee that the estimate is superior to the actual average, as it is not clear that these are valid assumptions for a given location in the first place. For instance, interactions between factors are ignored and the factors are assumed independent. Another issue is the standard error for the CMFs which can be considerable. For point estimation, the standard errors are often ignored. However, when the objective is to identify the top p%, then the standard errors of the CMFs cannot be ignored. It is well recognized by researchers that obtaining the combined standard error for multiple CMFs poses an unsolved problem.

The best estimate is obtained from a model that simultaneously estimates all factors and provides a total variance estimate as well. However, for local roads, much of the information needed for building a comprehensive regression model is not available. Thus, until more information is available the researchers recommend a simpler approach to modeling the crash counts. This approach is based on the fact that the expected crash counts are neither very sensitive to small change in exposure, nor to small changes to engineering features. Thus the researchers will discretize all factors including AADT, road width, and shoulder width, and then group road segments and intersections with the same features into classes of similar locations. For each of these classes, the researchers rank the expected crash count by average and EB to obtain the top p% of locations. Provided there are enough locations in each class, the researchers will also be able to compute a tail probability and odds for exceeding a certain crash count, given certain road features and exposure levels. While the expected crash counts should not be compared across classes, the odds can be used to rank all expected crash counts. This is because the odds are independent of the magnitude of the average. The odds of winning in a lottery, for example, do not depend on the payout. Put another way, if the odds for observing a certain crash count at a location (A) are one to a million and at another location (B) the odds for the same crash count are one in one hundred thousand then this crash count for location (A) is a rarer event for this road segment with the specific features and exposure level than for location (B). Thus, the odds can be used to compare locations, regardless of what engineering features and level of exposure these locations have.

Note that this modeling does not use a regression model because the researchers do not have enough data to build a comprehensive model for the local road system. Once the data exist for enough parishes, a comprehensive model may be developed and used to compute expected crash counts and odds based on these crash counts. At the present time, the researchers propose to use a negative binomial model for each class of locations. For the computations of the EB and the tail probability, the researchers follow the steps outlined in Section 5.

The procedure for determining the top p% of locations that warrant further analysis can be summarized as follows:

1. Obtain observed average crash frequency for each location for 3 to 5 years. The researchers should be aware that the more years they choose, the more likely it is that a trend in the crash counts is present, which violates the model assumptions. This implies that neither the simple average nor the EB average will be a good estimate of the current true mean crash count.

2. Select a certain class of road segments, i.e., those with the same exposure and geometric features. For instance, 12 feet lane, 6 feet shoulder, approximately 5000 AADT, straight road, no driveways.

3. Determine the expected crash frequency for each road segment in a class using the Empirical Bayes method outlined in Section 5.

4. For all locations, compute tail probabilities or odds for each expected crash count based on the distribution of the EB estimate from the NB model for given AADT and geometric features as outlined in Section 5. Note that the tail probability will vary with section length and all the covariates in the model.

5. Select a percentage p of highest expected crash locations to be investigated based on the available resources. In most cases, a p of 5% may be appropriate. Note that the top p% is not based on the crash count itself but the computed tail probability. If the tail probability for the EB estimate is less than p%, the location is selected. Odds are often preferred to tail probabilities may be used instead.

6. Select top p% locations from each class based on the tail probability.

7. Investigate the top p% crash locations and determine inexpensive countermeasures and their cost.

8. Use DEA to rank p% from all classes of crash locations with respect to odds, costs, and benefits. Instead of DEA, the cost benefits ratio can be used since it is only slightly less efficient than the DEA.

The previously listed steps provide a straightforward procedure that lends itself to clear step-by-step instruction documentation, including examples and "walk-through" guidance to help local agencies analyze their crash data and evaluate the safety of local roads.

9. Data

The study attempted to build models for predicting crash counts based on the road characteristics and traffic volume. Thus, data for use in the models must be available. The staff of the Highway Safety Research Group (HSRG) contacted several local agencies without success in securing information about AADT and geometric features of road segments. It seems that local agencies lack the resources to maintain a comprehensive database that has both AADT and geometric features listed. Thus the research group decided to develop a database for two pilot parishes and demonstrate the feasibility of the approach.

The two pilot areas are Terrebonne Parish and Lafourche Parish. Houma is located in Terrebonne Parish and Thibodaux is located in Lafourche Parish. The two areas are suitable for a pilot study because they contain a balance of locally maintained roads in urban and suburban areas, as well as locally maintained rural roads.

The first part of this document described the need for specific data when attempting to identify black spots. In this part, the researchers discuss how those needs were addressed in the two chosen pilot parishes. The researchers developed a general approach that can be applied to any location in Louisiana. The first subsection describes how the researchers determined the location of all local roads in the state and how an index to uniquely identify each of those segments was developed. Next, the researchers discuss how the geometric road characteristics were associated with those defined road segments, as well as how the AADT is identified on each of the road segments.

Road Grid

The road grid was taken from the Tiger Shape Files of the US Census Bureau. From all the line segments, the researchers subset those line segments identified as roads. At this point, the researchers could not make the distinction between local roads and other types of roads. While the names of most roads are also included in the raw Tiger Shape file, it was not immediately clear whether a named road is state-maintained or a local responsibility.

The roads in the raw data were stored as a collection of latitude and longitude coordinates. A collection of such pairs had a unique identifier and when connecting the coordinate spots with a line, the road was traced. Each road segment was uniquely identified between each intersection, or from the point where it starts. For example, Highland Road in Baton Rouge was not identified as one single road, but rather was divided into sub-sections defined by the intersection points along the road.

The researchers calculated a new road grid for use in this study, dividing all the road segments into 500 ft. sub-segments whenever possible. A large number of sub segments remained with a length of less than 500 ft. This was unavoidable due to the remainder left over when dividing into 500 ft. segments and the overall length is not evenly divisible by 500. For example, a 1250 ft segment would be divided into three parts: two sub-segments with length 500, and one sub-segment with a length of 250ft. This is important for the calculation of the Empirical Bayes estimate because the risk exposure should be equal between locations to produce a correct estimate of safety performance. The researchers continued to treat intersections separately from road segments; when dividing the road grid into 500 ft. segments, those segments never traversed intersections. The details of the calculation of the Road Grid and the algorithm used can be found in the Technical Appendix.

Crashes

The researchers used five years of data for all crashes occurring within the two parishes from 2005-2009. A large number of crashes reported in the HSRG's LACrash system had geographic identifiers associated with the records. Many records, however, were missing geographic location information, or the researchers determined that the provided latitude and longitude information to be inaccurate. Since the precise location of the crashes is highly important in the identification of black spots, and the only way to match crashes to the calculated road grid is by using geographic location indicators, the researchers had to ensure that every single crash was spotted correctly.

For this reason, the researchers used the crash report, narratives, and diagrams provided by the investigating officer of the crash to determine the exact location of the crash. The researchers did this for all crashes within the noted five-year period indicated to have occurred on local roads. The researchers spotted the crashes, regardless of whether or not the report contained latitude and longitude information, to ensure that the locations were determined accurately. A total of 5,760 crashes were identified in Terrebonne and Lafourche Parishes. The researchers attempted to match all of those crashes in this study.

Matching Crashes with Road Segments

The researchers programmatically matched crash locations (points) with road segments (lines) on the newly calculated road grid. The researchers identified matches using an algorithm that minimizes the cross track distance of the crash point from the vector that describes the road segment. The precise procedure is described in the Technical Appendix. The researchers were able to match 5,597 out of 5,760 with road segments of the calculated grid. This means 97.2% of all crashes could be matched to a road location.

Intersections

Intersections and Road Segments should be treated separately. This is easily accomplished in the data by examining the end points of the individual road segments; however, the definition of an intersection should be broader than just including strictly those line segment ends that touch. For example, when considering roads that are divided by a median, where the opposing directions of travel appear as two separate segments, it should not be considered as two different intersections. For this reason, the researchers used a tolerance of 75 feet to programmatically identify intersections.

Figure 2
Example of intersection identification

Identifying Curves

One important characteristic of road segments to note is the curvature of the road. This is especially important for rural road segments with a relatively high speed limit. In this case, a curve may be the underlying cause of accidents, so it makes sense that the researchers would want to take into account whether or not a location is located in a curve in determining the safety performance of that location.

The researchers developed a way to identify programmatically whether a road segment is part of a curve or whether it is a straight line. The researchers did not identify the degree of the curvature or the radius of the curve, but rather a dummy variable indicator that takes the value of 1 when a road segment is part of a curve and 0 when otherwise, which would also be useful for later analysis. The researchers calculated the directional heading (bearing) for every sub-segment of the road grid. Next, moving along the sub-segments of a road from one intersection to the next intersection point, the researchers calculated the moving standard deviation of the calculated heading using a sliding window of 5 sub-segments.

The researchers experimented with a variety of sizes of sliding windows and determined that a 5 window moving standard deviation provided the best performance. The researchers chose a moving standard deviation because there is a lot of variation in the number of sub-segments that make up a road segment. Individual sub-segments are straight lines themselves, and so many small sub-segments with varying headings are indicative of a curve. Using a moving standard deviation identifies multiple collections of sub-segments within a road segment that may be considered curves. If the road is perfectly straight, then the sub-segments that make up the road will all have the same heading and the standard deviation will be zero. On the other hand, if there is a curve in the road, then the sub-segments that make up the road will have different headings as the researchers move along the road. This means that the standard deviation will be positive.

There is no direct interpretation or units of measurement that are associated with the calculated moving standard deviation of the heading. After examining the calculated moving standard deviations, the researchers observed that the distribution of that statistic was bimodal. There were many segments that were straight and had a very small standard deviation. After a careful calibration exercise, the researchers were confident that if the natural logarithm of the moving standard deviation of a particular segment exceeded the 50^{th} percentile of the distribution of all log moving standard deviations, then this indicates a curve.

This method works better for rural areas than city streets. Curves in the road are much more important on rural roads, as well as a determining crash factor, particularly on higher-speed rural roads, so the method is still suitable for the purpose of this study. On city streets, the individual line segments tend to be much shorter and the variation in the heading much higher in spots that would normally not be identified as curves. See Figure 3 below for an example of how the researchers identified curves. The black line segments are regular road segments that are not considered to be part of a curve. The pink line segments in the figure below indicate road sub-segments that the algorithm determined to be a part of a curve.

Figure 3
Example of curve identification

Other Road Characteristics and Design Features

Collecting the characteristics of the individual roads was the most time consuming task of this project. Since site visits to each of the locations would have been too expensive and time intensive, the researchers made use of Google Earth satellite images, as well as Google Street View.

First is a list of features that the previous literature has identified as being important factors to consider when estimating the safety performance of road segments and intersections. At the same time, the researchers considered which characteristics were able to be observed or measured using Google Earth or Google StreetView. The researchers developed the following list of items based on the information collected:

For Road Segments (non-intersections):

- Number of Lanes
- Type of Center Striping
- Divided Street / One Way Indicator
- Narrow or Regular Lane Width indicator
- Shoulder type (Full, half, no shoulder)
- Speed Limit
- On-street Parking indicator
- Number of Driveways along segment
- Ditch indicator

In addition to the above, the researchers collected the following information for intersections:

- Signaling type (Traffic light, stop sign, none, etc.)
- Protected left turn indicator
- Turning restrictions (No left turn allowed, no right turn allowed)
- "No Right on Red" restriction
- Presence of designated turn lanes

Using the calculated road grid described above, the researchers developed a database application that allowed data collectors to navigate to the locations of the road segments using a browser window running Google Earth. Using information from the satellite pictures, as well as the street view feature, the data collectors filled in forms with the data required for analysis. See Figure 4 for a screenshot of the program used.

The researchers were able to collect most of the above features for all segments on the road grid. For some rural locations in Lafourche Parish, Google Street View was unavailable, making the collection of information on intersections difficult. While the researchers were sometimes able to determine the presence of turn lanes or similar features from the satellite picture alone, it was not possible to collect all pieces of information. In this case, the data collector indicated that the Google data provided insufficient detail to complete all fields.

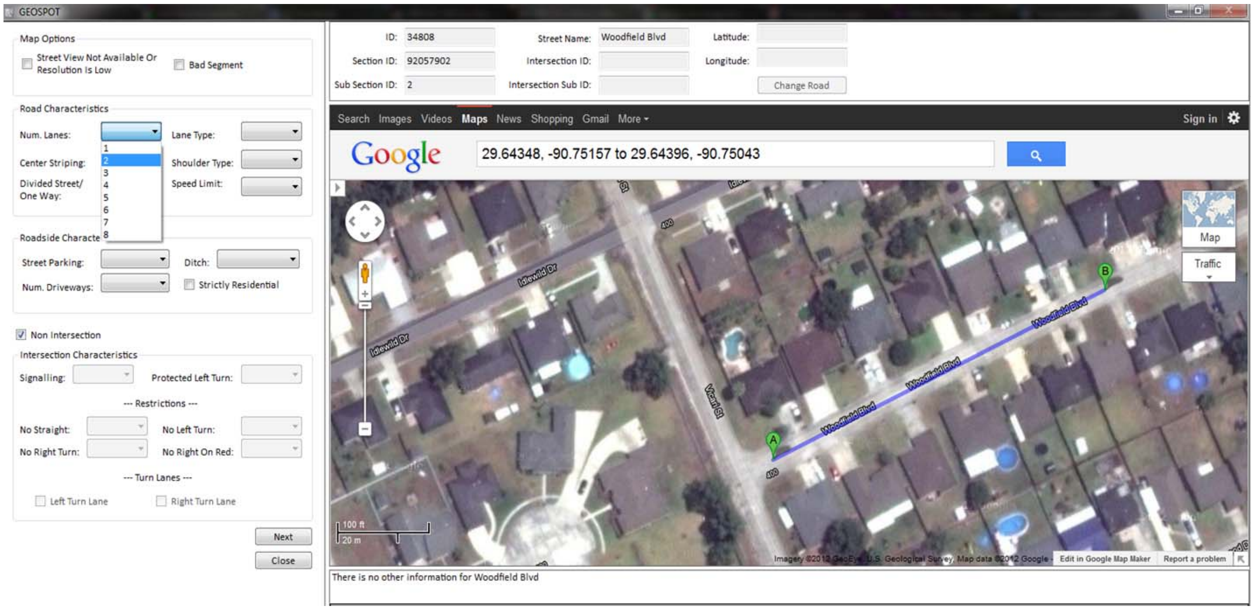


Figure 4
Example of road characteristics spotting program

Average Annual daily Travel (AADT)

The Highway Safety Manual identifies AADT as one of the most important factors to consider when estimating the safety performance of a road segment or intersection. Note, however, that the predicted crash frequency as a function of AADT is a fairly flat function. Predicted crash frequencies do not change very much across wide ranges of AADT.

The researchers obtained Traffic Counts from the South Central Planning Development Commission (SCPDC). The SCPDC referred the researchers to their website to access to traffic counts that are maintained by a third party vendor. The researchers attempted to contact the vendor to discuss access to the traffic count data, but received no response to any inquiries. Therefore, the researchers had a student worker manually extract traffic count data from the website, a time-intensive exercise. Local parish or city officials attempting to identify black spots are likely to have better access to traffic count data.

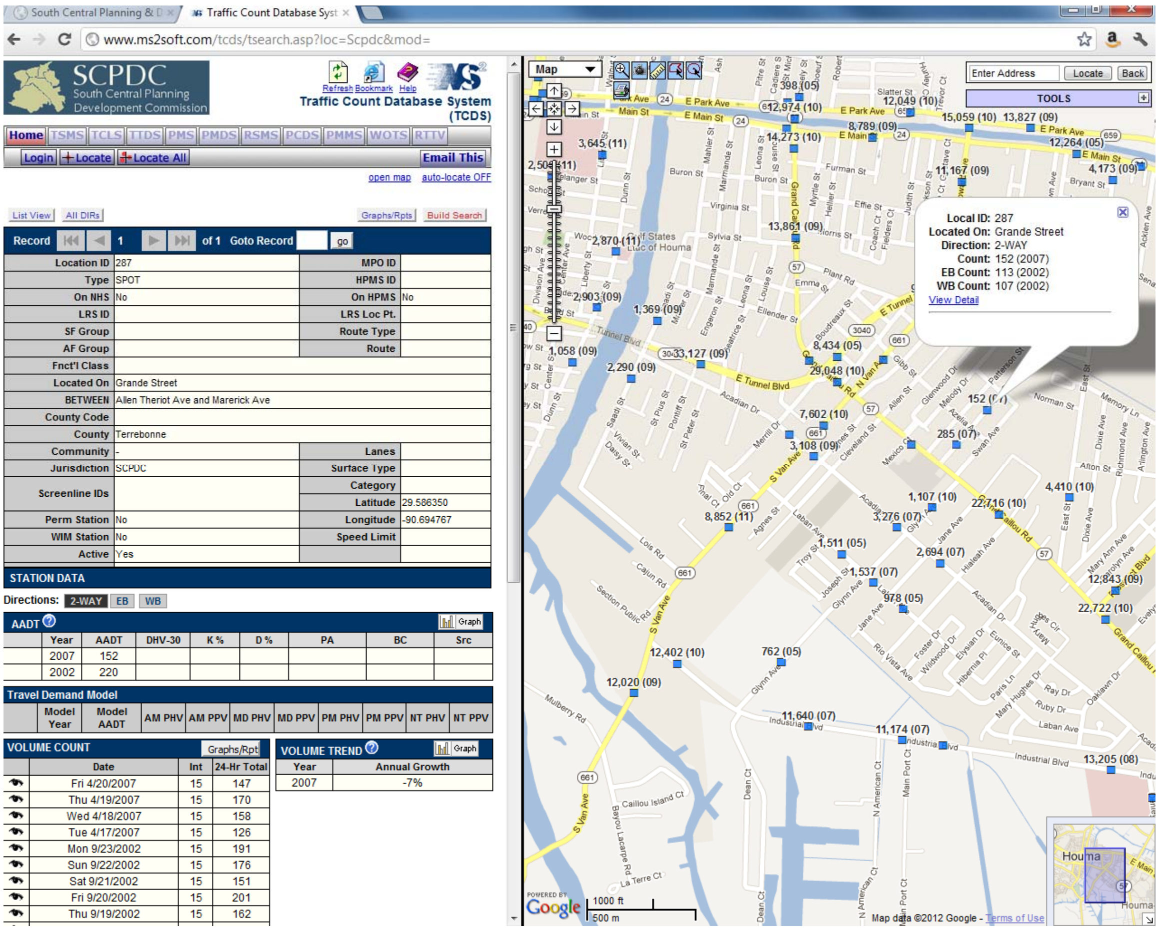


Figure 5
Example of AADT measures

Traffic counts are collected at locations identified using latitude and longitude, which is a single point. As the researchers were particularly interested in the entire segment on the road grid with a particular AADT, they had to calculate the closest road segment to a particular count location and assigned the observed AADT to that segment. This was done in the same way that crashes were assigned to a particular segment (described above). This gave AADT for only a few segments on the grid and it was necessary to manually match additional segments.

Since Predicted Crash Frequencies do not vary a great deal for wide ranges of AADT, AADT was divided into 4 different categories:

- Category 1: AADT = 0—1000
- Category 2: AADT = 1001—2500

- Category 3: AADT = 2501—5000
- Category 4: AADT = 5001+

The manual matching and estimating of AADT was performed by mapping the observed traffic counts, categories, and road segments and judging whether the AADT on adjacent road segments would likely be in the same category as a nearby count. Since the researchers were not using the continuous measure of AADT, but rather the categorical measure, this method of estimating AADT categories is not likely to cause any significant measurement error.

10. Using Microsoft Excel Application for Site Ranking

This section presents an application developed using Microsoft Excel 2010 as the development platform aimed at providing an analyst or highway engineer with a tool for identifying and investigating potentially hazardous sections of highway. This application provides a quick and easy way to select a type of highway section, see all similar sections along with the number of crashes and mean estimates, rank the sections, and finally provide the user with a way to further investigate particular sections. One over-arching theme of this project was to provide effective and inexpensive tools for analysts. Thus, this application was developed to be used on the desktop and utilizes only Excel 2010 and Google Earth (free edition) on the user's end. What follows is a very brief and non-technical description of the underlying data architecture and a discussion on how the application uses this data and interacts with Google Earth to provide visual data.

Underlying Data Architecture

The data used by the application is exactly the same data that is used for developing regression models. How the researchers prepare it is slightly different. The goal is to provide users with a flexible method for choosing highway section attributes when analyzing the crash counts. To facilitate this, the researchers utilize the features of online analytical processing (OLAP). OLAP are basically a collection of indexes that help perform some pre-aggregation of data and enable quick recall. OLAP cubes, the result of applying a set of dimensions to a collection of measures, provide very fast aggregation and slicing of data. Simply put, the researchers can quickly identify highway sections and their associated measures while providing plenty of flexibility for selecting highway section attributes. The researchers developed an OLAP cube including the available highway section attributes as dimensions and the crash count, Empirical Bayes Estimate, and tail probability as measures. In some cases, the researchers also performed some bucketing of continuous variables to make selection easier. An example of this is providing classes of shoulder width constructed by looking at two foot increments. The researchers justify this grouping and subsequent "loss of information" by noting that even in the regression models provided by the HSM, there is little to no change in the estimated crash count for such small changes in shoulder width. The same holds true for many of the other variables, like pavement width and AADT. The classes were all selected based on careful consideration of how the grouping would affect the mean estimates. The cube used by the application resides on a server at the HSRG in an effort to minimize the required resources on the client machine.

While the OLAP cube provides a method for high incident highway section identification, it is not very well suited to returning relational data. The cube is optimized for aggregation, so the researchers must also query the underlying relational database that holds complete crash information in order to provide the analyst with crash details, once a section has been identified for further analysis. A situation may arise where the analyst has identified a section with a high number of crashes relative to similar sections. Then, drilling down to the crashes on this section, he or she may find that the overwhelming majority of these are side-impact crashes. This information may point to a problem with the traffic control configuration or other roadside hazards that are contributing to these crashes, so providing granular crash-level data is also important. This data has already been collected at the HSRG so its inclusion is rather easy to implement.

Application Overview

When implementing this analysis tool, the researchers selected Microsoft Excel for several reasons:

- Excel has the ability to connect directly to OLAP cubes and relational sources and provides an easy to use drag-and-drop cube browser interface. This eliminates the need to develop custom software to interact with the cube.
- Excel supports custom programming using macros and Visual Basic for Applications.
- Single licensing costs for Excel are relatively low and most users are likely to have it on their computer.

Google Earth was chosen because it is a free application that provides some basic tools for analyzing spatial data.

The steps can be summarized as follows:

1. Determine a collection of highway attribute settings to be investigated.
2. Select the settings in the application. This will provide a list of sections with these settings along with their respective crash frequencies and Empirical Bayes estimate.
3. Use the built-in capabilities of the application to map the crashes in Google Earth.
4. Use Google Earth to further investigate the section and identify potentially hazardous aspects of the section that can be addressed.

Opening the application also starts Excel, as shown in Figure 6. There are a few main regions. Along the top of the program is a custom ribbon. Ribbons were introduced to Microsoft Office applications in the 2007 release. They replace the standard drop down menus typically found across the top of an application. Using a Custom Ribbon UI Editor, the researchers are able to create a new ribbon and add custom elements such as buttons, text fields, etc.

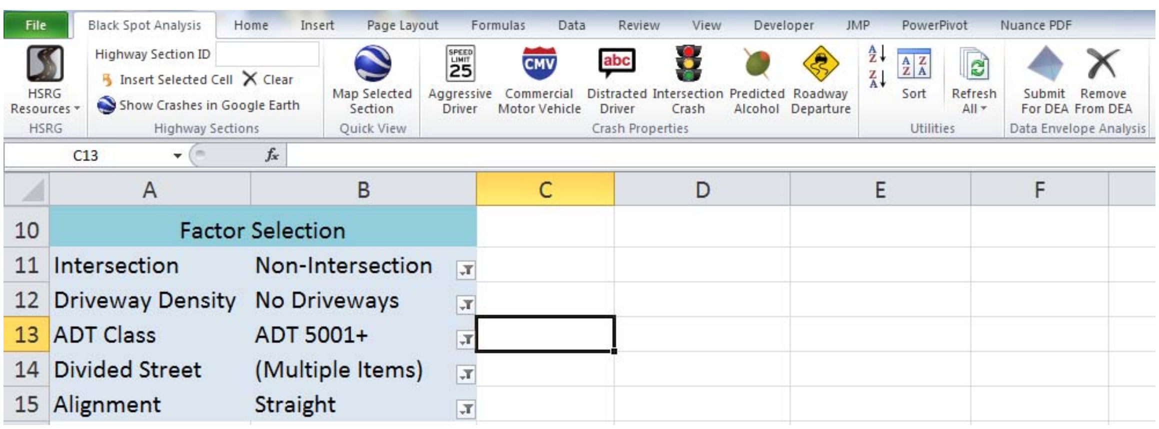
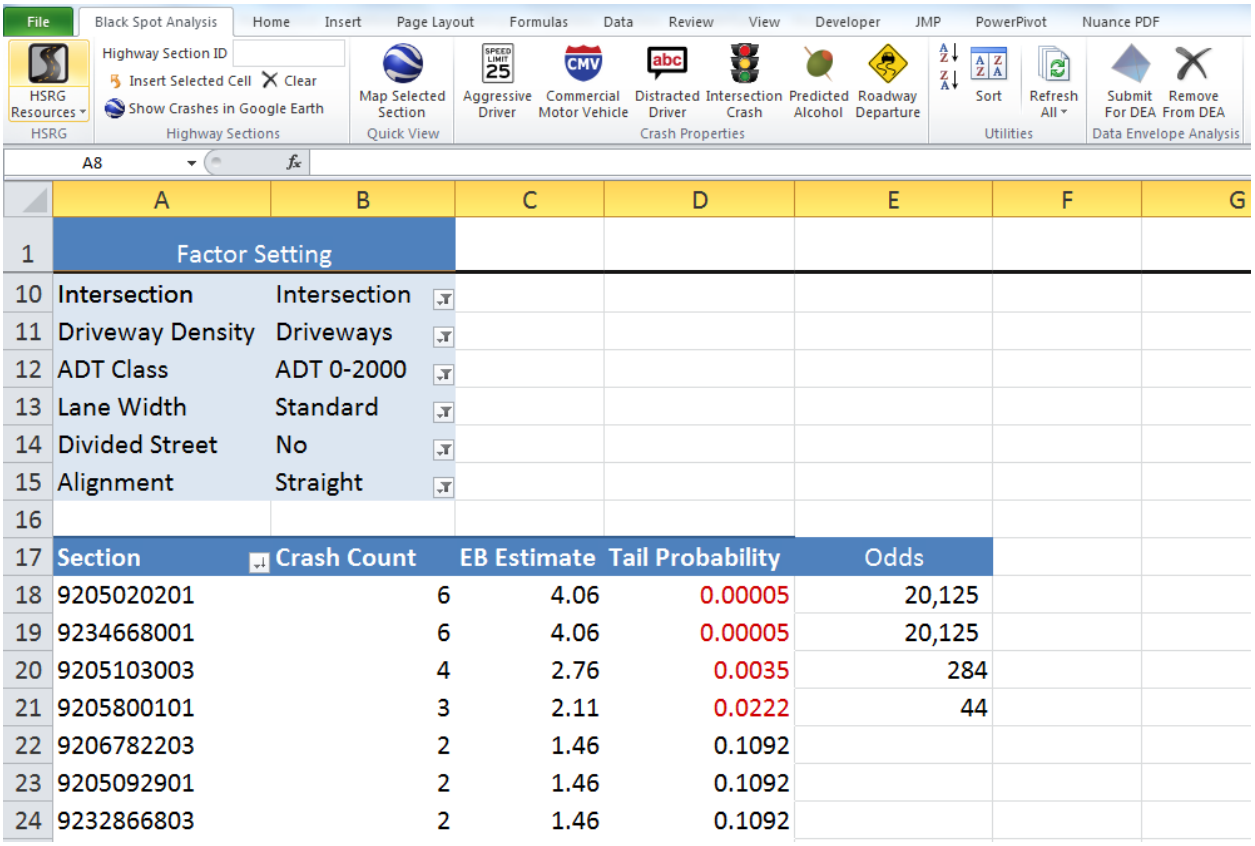


Figure 6
Microsoft Excel black spot analysis application ribbons

In the spreadsheet field area under the “Factor Selection Heading” five factors are listed: Intersection, Driveway Density, AADT Class, Divided Street, and Alignment. To the right in column B, there is the filter that permits the selection of various settings. Both the filters and the filter values are chosen by the user. In the example shown on Figure 7, the following road features are selected: Intersection, Driveways, AADT equal or below 2000, lane width standard 12 ft, No Divided Street and Straight for alignment. The pivot table data are located under the filter area. For this use, the highway section ID numbers, their associated crash counts, the EB estimate, the tail probability p, and associated odds are listed. The odds are computed as (1-p)/p. For instance, consider the intersection with ID=9205020201. The odds of 20,125 indicate that there is a one in 20,125 chance to observe an expected crash count of 4.06 among intersections with straight undivided 12 ft lanes where driveways are present and the AADT is equal or below 2000.



	A	B	C	D	E
1	Factor Setting				
10	Intersection	Intersection			
11	Driveway Density	Driveways			
12	ADT Class	ADT 0-2000			
13	Lane Width	Standard			
14	Divided Street	No			
15	Alignment	Straight			
16					
17	Section	Crash Count	EB Estimate	Tail Probability	Odds
18	9205020201	6	4.06	0.00005	20,125
19	9234668001	6	4.06	0.00005	20,125
20	9205103003	4	2.76	0.0035	284
21	9205800101	3	2.11	0.0222	44
22	9206782203	2	1.46	0.1092	
23	9205092901	2	1.46	0.1092	
24	9232866803	2	1.46	0.1092	

Figure 7
Microsoft Excel black spot analysis example

Note that the sections are ranked in descending order according to their EB estimate. Those sections at the top of the list may be good candidates for further investigation provided the tail probability is smaller than the cutoff value. As the factor settings, e.g. AADT, lane width etc., are changed, the sheet updates to reflect the new highway sections and their crash counts, EB estimates, tail probabilities and odds. This allows the analyst to first choose the type of highway sections he or she wants to investigate.

Note that the factors in the factor setting grid can be changed using the pivot table field shown in Figure 8. The pivot field list is used to control which elements are included in the factor settings. Currently in the filter area, only Intersection, Driveway Density, AADT Class, Divided Street, Alignment are being considered and so those are listed factor selection area mentioned above. But the user may add or delete any fields in the list.

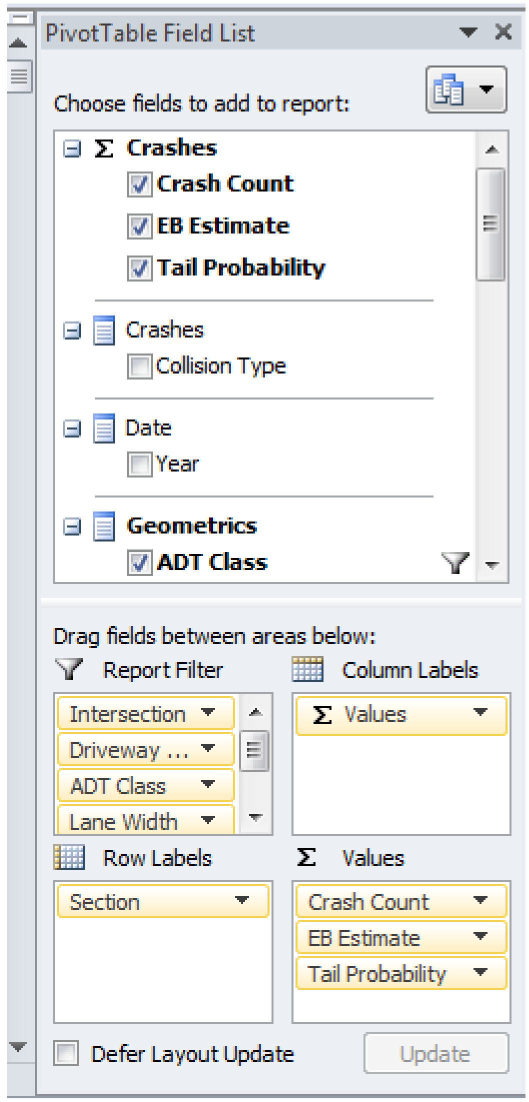


Figure 8
Pivot table fields

Figure 9 shows some of the fields that could be selected, such as left turn lane, no left turn, no right on red, etc. To facilitate the evaluation of this crash location, the researchers provide an easy way to show the crashes for a given section in Google Earth. This functionality resides in the "Black Spot Analysis" ribbon.

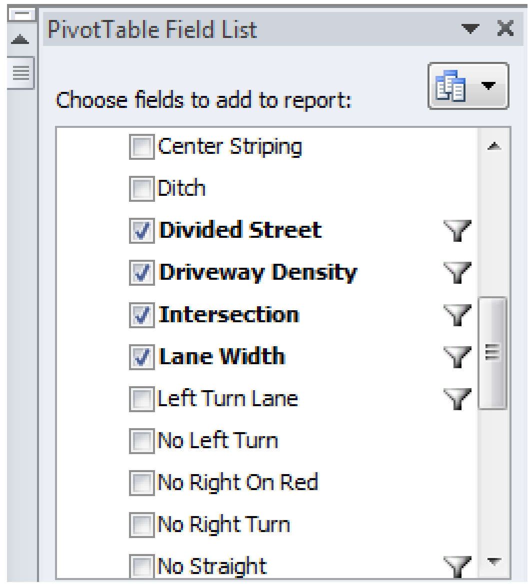


Figure 9
Optional fields

The quickest way to map the crashes is to click on a highway section ID and press the "Map Selected Section" button in the ribbon as shown in Figure 10. When the button is pressed, a macro written into the worksheet is initialized. This macro uses the selected highway section ID to retrieve all crashes associated with it from the underlying relational database. The macro first creates a kml file (i.e., a file that Google Earth uses to store information about points). As the crashes are read from the database, information about them is saved to the kml file. Saving this data in the kml file allows it to be viewed once the locations are loaded into Google Earth. Some examples of data elements include the date, day of week, time of day, collision type, etc. Similarly, crash related elements are written to a second worksheet in the excel application labeled 'Crash Details.' This gives a list and attributes of the crashes so that the analyst can see the crash data in a tabular form. The macro finishes by making a call to the operating system to open the kml file. This loads the kml file in Google Earth (or the client's computer's default kml file viewer) and zooms in to display the mapped crashes.

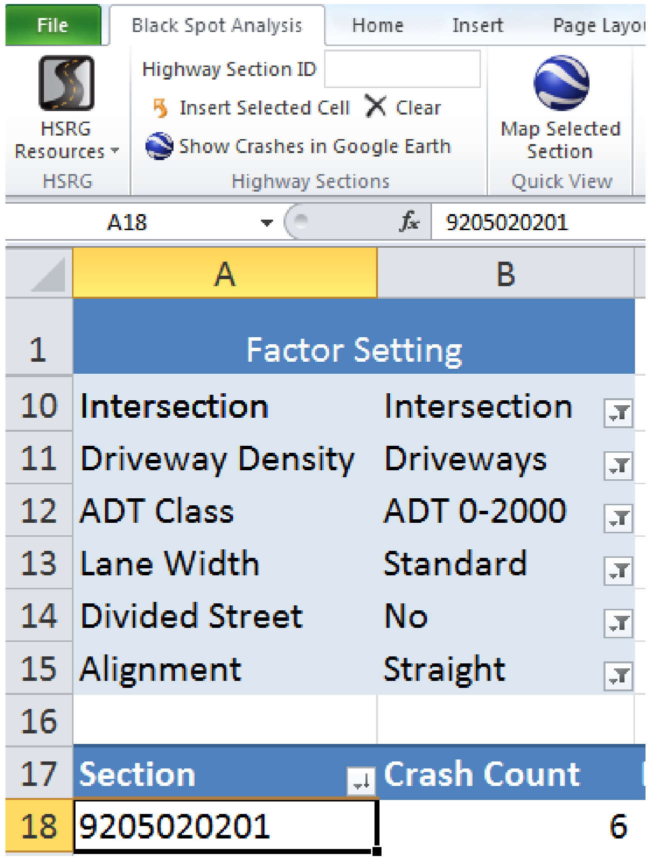


Figure 10
Selected road segment/intersection

Figure 11 below shows the crashes displayed in Google Earth and the information associated with a selected crash.

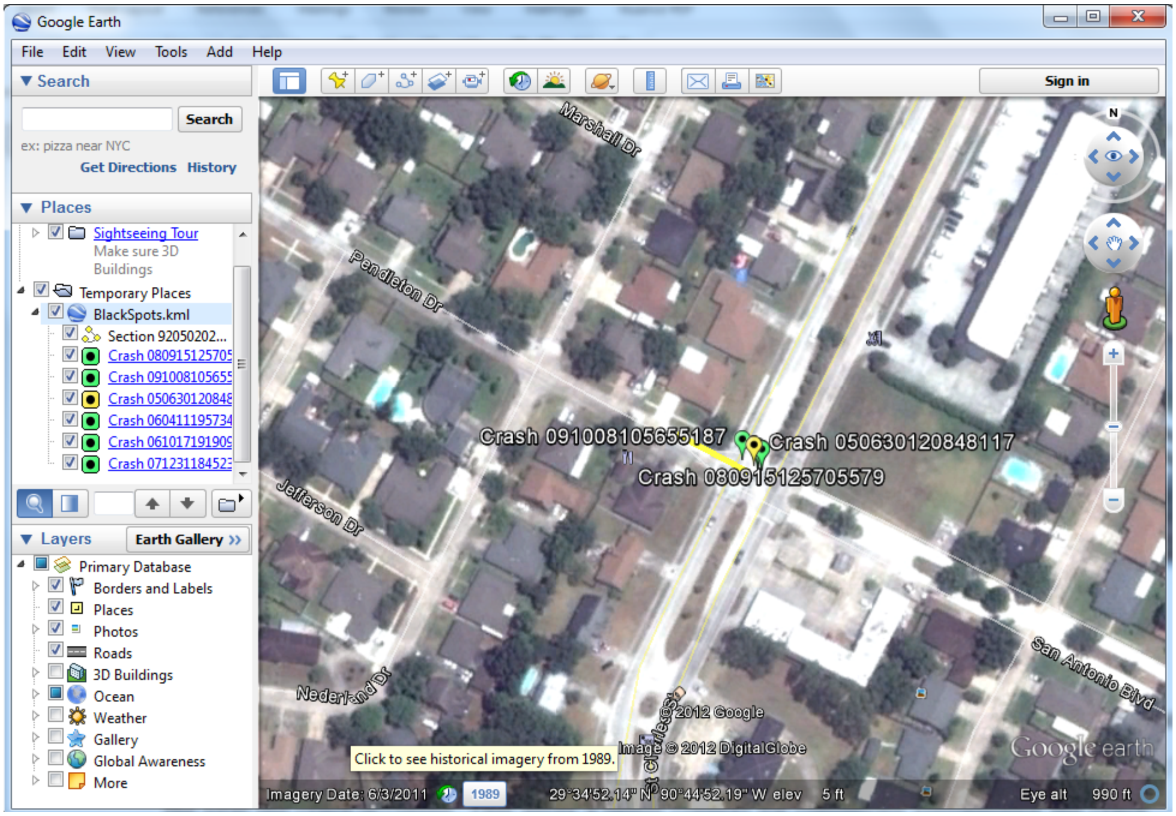


Figure 11
Google Earth picture for selected intersection

With the crashes shown in Google Earth, the user can see properties related to the highway. Selecting any crash in the upper left panel pulls up its properties. Google Earth also allows the user to view aerial images through time. Usually, there will be an image corresponding to the year the crash occurred. This gives a better context for conditions surrounding the crash, particularly if some aspects of the section have been changed over time. Clicking on one of the pins that identify the crashes provides short description of the crash, such as aggressive driving and alcohol, as shown in Figure 12.

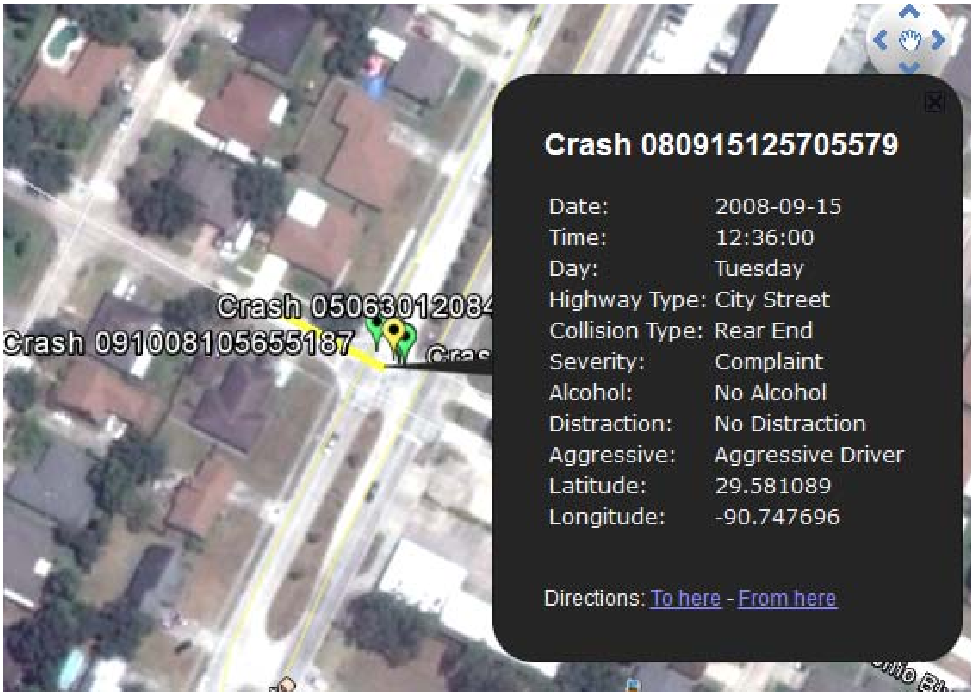


Figure 12
Detailed information for selected crash

Google Street View allows the analyst to obtain a view of the intersection (shown in Figure 13). The view can be used to assess any road hazards or suggest any particular road improvements. Note that Google Earth allows to view the intersection from different directions.

Figure 13
Intersection view

There are additional buttons in the ribbon (Figure 14) to control which crashes are mapped. These give the user the option to only see crashes that involved factors like aggressive, distracted, drivers under the influence, commercial motor vehicles, drivers that left the roadway (departure), or crashes that occurred at intersections. It should be noted here that any combinations of these properties can be used as well.

Figure 14
Options for selecting types of crashes

Hence with only a few clicks an analyst can quickly compare crash counts on similar highway sections and identify those that stand out. He or she can also employ Google Earth's capabilities to get specific characteristics of the highway section and possible identify factors that contribute to the abnormal crash count.

DISCUSSION OF RESULTS

Selection of Factor Level Settings

This analysis included Terrebonne and Lafourche parishes. All crashes were map spotted and all factor levels (such as driveways, lane width, etc.) of the roads segments and intersections were recorded in a database using Google Earth. The AADT was obtained from websites. Table 1 shows the factors and levels for each factor that were collected. However, for many of the factors, the levels could not be identified for every road segment. For instance, the speed limit was only identified when a sign was visible on Google Earth.

Table 1
Factors and factor levels

Factors	Levels	Number of Levels	Intersection levels*	Non Intersections Levels*
AADT	0-2000, 2001-5000, 5000+, unknown	4	3	3
Alignment	Straight, curve	2	X	2
Center Striping	Yes, No	2	X	X
Ditch	Yes, No	2	X	X
Divided Street	Yes, No	2	2	2
Driveway Density	Driveway, No Driveway	2	X	2
Lane Width	12ft+, <12ft	2	X	X
Shoulder Width	Yes, No	2	X	X
Speed Limits	35, 45, 55, 65, 70	5	X	X
Street Parking	Yes, No	2	X	X
Strictly Residential	Yes, No	2	X	X
Left Turn lane	Yes, No	2	2	
Protected Left Turn	Yes, No	2	2	
No Left Turn	Yes, No	2	X	
No Right on Red	Yes, No	2	X	
No Right Turn	Yes, No	2	X	
No Straight	Yes, No	2	X	
Right Turn Lane	Yes, No	2	2	
Signaling	Yes, No	2	X	
Total Number of Combinations			48	24

**The X indicates possible level and a number indicates the actual levels analyzed.*

The more factors that are chosen, the more combinations there are that will have to be evaluated. Using all factors would lead to many combinations to consider. Many of these combinations of factor levels would not have enough data to allow an analysis. For instance, not all road segments have speed limit information. Table 2 shows the total number of combinations of levels for intersections and for road segments analyzed.

Table 2
Number of combinations investigated

	Number of levels selected for analysis
Intersection	48
Roadway Segment	24
Total	72

Selection of Road Segments and Intersections

The choices of levels included in the analysis were based on the available information for the factor levels. There were very few road segments and intersections with 0-2000 AADT. Thus this group was combined with 2000-5000. Table 6 of Appendix B gives the 48 combinations chosen for analysis. For instance, Section Level 1 combination of the intersections includes AADT of 0-5000, no left-turn lane, undivided street, no protected left turn, and no right turn lane. For all other factors, all levels were included. For instance, a specific lane was not chosen because there was no difference between the levels of this factor. There were 89 intersections for the chosen combination. Twenty-two had 1-5 crashes and three intersections had 6-10 crashes. Six of these sections had tail probabilities below 5% and thus were analyzed on Google Earth. Out of the 4,821 intersections, there were 30 intersections that had probabilities below 5% and thus were analyzed through Google Earth to identify countermeasures.

Table 7 of Appendix B shows similar information for non-intersection road segments. The factors included are AADT, alignment, divided street, and driveway density. These 24 combinations had 13,382 road segments, of which 42 road segments had tail probabilities below 5%. Six of the selected sections do not have a clear Google Street view.

Selecting Countermeasures

The potential countermeasures for each of these 30 intersections and 36 roadway segments are selected based on the analysis of Google Street view snapshots. One drawback in this process is that most of the Google Street views were captured in 2008. Thus, some of the issues identified may have already been corrected. The selected road segment/intersection should be visited to identify the correct status of the road segment.

Crash Modification Factor (CMF)

The Federal Highway Administration's guideline ' Low-cost Safety Enhancements for Stop Controlled and Signalized Intersections' is followed to select the inexpensive safety countermeasures for Louisiana. From the recently published Highway Safety Manual (HSM), the countermeasures are divided into three major groups:

1. Countermeasures with CMF value
2. Countermeasures with no CMF value
3. Treatments with unknown crash effects *[61], [68]*.

Some of the potential countermeasures used in the study do have associated CMF values. These countermeasures are marked as countermeasures with ‘No CMF value.’ The potential countermeasures for both intersections and roadway segments are shown in Appendix B.

Cost Estimation

The evaluation of the cost of the countermeasure is important to enable a ranking of projects. The local cost data are taken from several local agencies. Some package prices of the countermeasure installation are considered from the Louisiana Road Safety Program (LRSP) data. The estimated range of the costs is shown in Appendix B along the available CMF values.

Ranking of Projects

After completion of selection of countermeasures, costs, and CMF’s, the following measures are considered as criteria for ranking the countermeasure projects:

- Cost measure is the expected cost of implementing a possible low-cost countermeasure at the road segment or intersection.

- Benefit measure is the expected reduction in crashes estimated as (1 – CMF) * EB, where CMF is the Crash Modification Factor of the countermeasures and EB is the Empirical Bayes estimate of the expected number of crashes at a site.

- Risk measure (hazard) is based on the tail probability estimate at the site. The odds for the expected value is (1-p)/p where p is the tail probability. To smooth the large differences of the odds, the researchers applied a logarithmic transformation $\log_2$(odds).

The ranking should take into consideration for all three measures the cost, benefits, and risk. If only the costs of the countermeasures were selected for ranking, then high benefits of countermeasures would be neglected and sites with high expected crash counts and benefits could be ranked low. If the benefit were the only ranking criterion, the cost efficiency would suffer. The reason for including a risk measure is to avoid high risk road segments ranking low if only costs and benefits were to be considered. Taking into consideration only costs and monetary benefits could bias road improvement project selection toward sites with a high volume of traffic. While this seems sensible from a cost perspective, it is unacceptable from a standpoint of safety. There are many situations where the public does not accept decisions based solely on a cost-benefit analysis due to a perception of an unacceptable risk. Public projects such as road safety improvements require careful consideration of other objectives besides cost and benefits. Therefore, the researchers propose that a measure of risk (hazard) for a road segment be included into the cost benefit analysis.

By combining cost, benefit, and risk measures, a multi-criteria ranking method based on the Data Envelopment Analysis (DEA) procedure is suggested, which applies weights to the different criteria without user input. The preference weights are calculated by linear programming, a method which can easily be applied to evaluate and rank thousands of sites. The DEA is an extreme point method; i.e., it compares each location to all other locations with weights calculated to be the most favorable for the particular location being evaluated.

In DEA, the efficiency of the project is the weighted output over weighted input. The objective of the DEA is to identify the projects that produce the largest values of outputs by consuming the least amount of inputs. In this case, the input is the cost of the project. The researchers select two output measures. One is the benefit that is measured by the expected crash reduction. The other output is the risk that is measured by the proposed risk (hazard) measure for a road segment. Overall, the suggested ranking provides a good compromise between obtaining a large benefit in crash cost reduction and including more projects on high risk road segments.

The result of applying the DEA multi-criteria ranking method is compared with the other ranking criteria in the summary tables and graphs below. The researchers compare the results of the five alternative ranking methods. For the comparison, the researchers selected different available budgets (arbitrarily selected values for illustration) and serve as the total cost limit for that particular budget. The researchers evaluate each ranking for each budget limit according to the following measures:

- Cumulated Benefit (CB) expresses the total improvement in Benefit measure that is the expected reduction in crashes estimated as (1 – CMF) * EB, defined earlier.

The higher the CB the better is the selection for benefit of crash reduction.

- Cumulated risk avoided (CR) expresses the total hazard avoided, the odd values the researchers applied to a logarithmic transformation, and the risk measure is $\log_2$(odds) as discussed earlier. The higher the CR, the better the selection for risk improvement.

The results of the DEA ranking for the 30 intersections and 36 road segments are summarized in Table 3 and Table 4. For each measure, Figure 15 and 16 show a comparison of the different rankings methods. The Cumulated Risk avoided (CR) and the Cumulated Benefit (CB) for each ranking method is expressed as a percentage of the optimal DEA ranking. The higher the percent for a ranking, the closer is it is to the best performing DEA ranking. The researchers note that the ranking based on benefits/costs is very close to the optimal ranking obtained using DEA. Although the DEA provides a slightly better ranking, the ranking based on the benefit/cost ratio may be used in most practical cases because it is easier to obtain.

Table 3
Comparison of the benefit and risk improvement
for the different ranking methods for 30 intersection sites

	Cost criterion		**Benefit criterion**		**Crash Count**		**Risk criterion**		**DEA criterion**		**Benefit/Cost crit.**	
Total cost limit ($)	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided
5,000	14	134	14	172	14	172	14	172	25	176	25	177
10,000	23	237	33	321	28	248	17	272	42	390	39	348
15,000	28	283	42	416	31	348	22	345	54	528	50	465
20,000	42	346	53	534	40	403	29	414	61	622	61	622
25,000	50	406	56	634	50	543	42	608	68	679	66	655
30,000	57	507	61	702	59	671	47	670	73	723	72	704
35,000	64	605	64	733	65	750	58	753	77	766	77	766

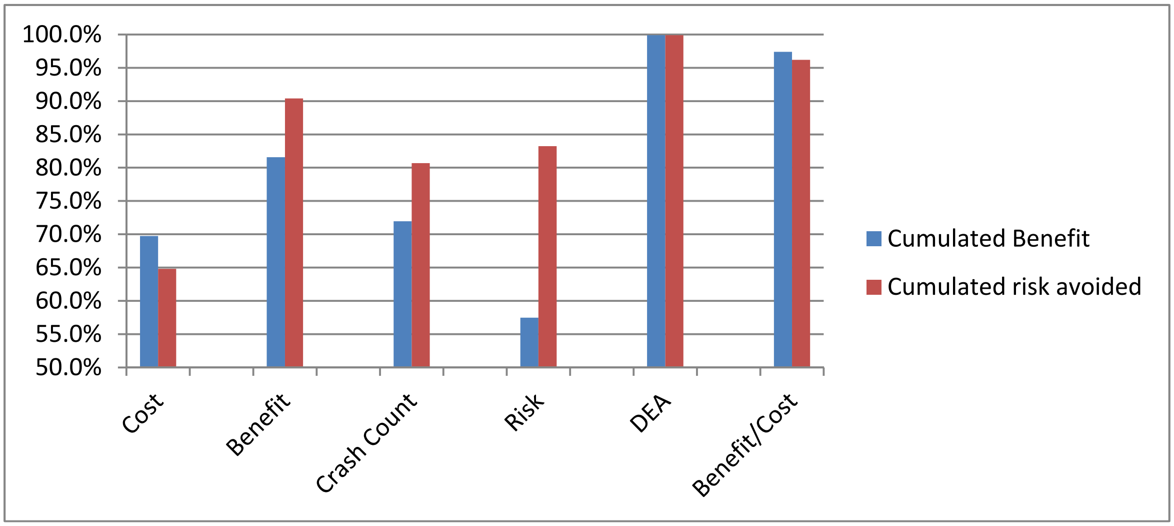


Figure 15
Comparison of benefit and risk improvement

Table 4
Comparison of the benefit and risk improvement

	Cost criterion		**Benefit criterion**		**Crash Count**		**Risk criterion**		**DEA criterion**		**Benefit/Cost crit.**	
Total cost limit ($)	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided	Cumul. Benefit	Cumul. risk avoided
20,000	13	188	28	296	28	296	28	296	31	304	31	304
40,000	18	257	43	441	44	478	46	513	50	557	50	536
60,000	35	498	54	563	54	612	53	633	58	647	58	647
80,000	58	673	59	632	60	717	61	762	68	794	68	794
100,000	64	732	70	768	68	827	69	862	73	860	73	860
120,000	77	928	78	918	77	931	76	934	77	928	77	918

It is apparent from Figures 15 and 16 that the DEA provides the optimal ranking and the benefits/costs is the second best ranking. The figures also show how the crash count alone does not provide an optimal reduction in risk and increase in benefits.

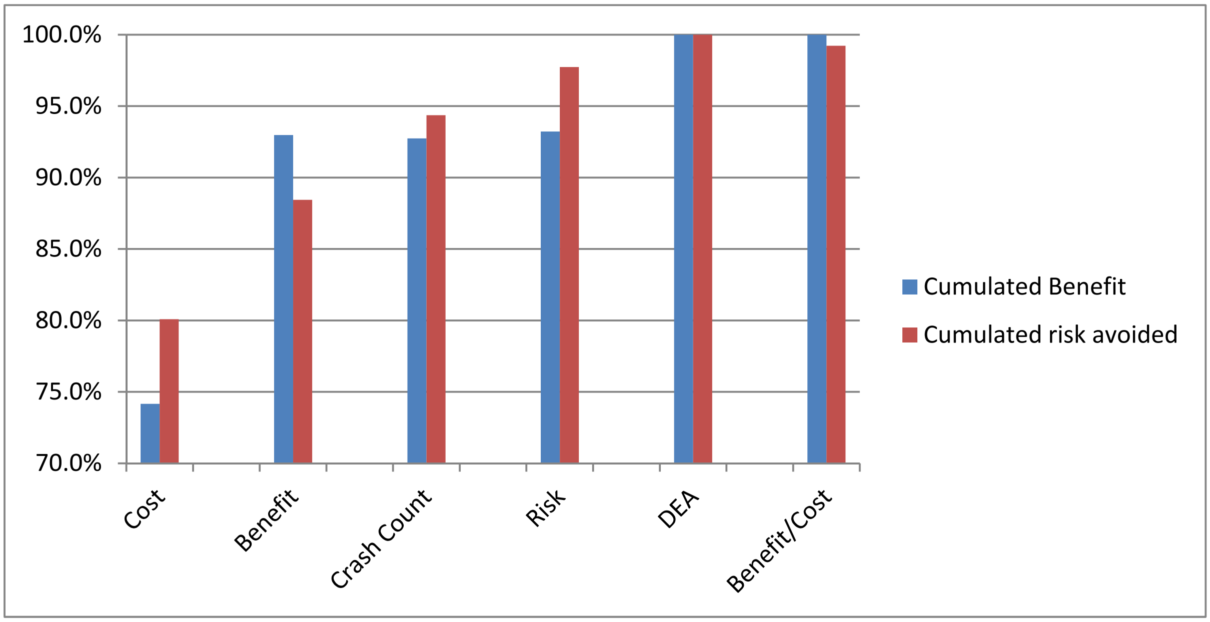


Figure 16
Comparison of benefit and risk improvement for different ranking methods for road segments

The complete DEA ranking of the 66 sites is provided in Appendix B (Tables 6 and 7). These tables also provide the crash counts, EB estimate, benefits computed as (1-CMF)*EB, costs for the improvements, tail probability, log odds, DEA efficiencies, road name, countermeasures, and the CMF's for the individual and combined countermeasures. It should be noted that some of the road segments and intersections with high crash counts and high risk are ranked lower than other road segments/intersection with lower crash counts and risks because of the differences in costs. If costs are not an issue, then the road segments/intersections may be ranked according to risks and benefits only. The researchers also note that the benefits were calculated using a fixed cost for all crashes not differentiating between fatal, injury and PDO crashes. Weighting benefits by crash severity would require separate EB estimates for each severity. Also, the benefits would be more sensitive to the occurrence of fatalities at a crash site and would thus result in ranking highest all the sites with fatal crashes. Since fatalities are relatively rare among all crashes (0.4%) the ranking would be greatly affected by random single fatality events which may not be related to road hazard but instead driver behavior, such as alcohol use and not wearing a seat belt. Thus the researchers opted against weighting the benefits by crash costs of the severity of a crash.

CONCLUSIONS

This report developed a procedure for identifying candidates for crash countermeasures, selecting inexpensive crash countermeasures, providing costs and benefits of countermeasures, and ranking the projects using costs, benefits, and crash risks for local Louisiana roads. The deliverables include an Excel application that uses OLAP to obtain a ranking of candidates for road improvements. This application makes use of crash data, engineering features, and AADT to compute EB estimates and tail probabilities for each road segment and intersection. Road segments and intersections with a tail probability below 5% are selected as candidates for countermeasures. These candidates are evaluated using Google Earth, countermeasures are suggested, and costs and benefits of the countermeasures are obtained using published information. The resulting road improvement projects are then ranked using multi criteria DEA including costs, benefits and crash risks.

While the procedure was able to identify 30 intersections and 36 road segments in two Louisiana Parishes that are candidates for improvement and for which countermeasures were identified, several issues were encountered that impacted the development of the procedure.

1. The procedure relies on data that were not readily available.
    a. There is no database of a road inventory for local roads that could be used to obtain engineering features of the roads. Thus, the time consuming task included identifying road features on Google Earth for 36,000 road segments. These road features included lane and shoulder width, curves, driveway density, and intersection features such as turn lane and traffic controls.
    b. There is no complete database for AADT on local roads available. AADT is published sporadically on websites. While the AADT may be collected by many local agencies, it was not made available to the researchers. The AADT used in this project was obtained from a variety of websites and AADT was estimated for roads without available AADT.
    c. While using Google Earth has the advantage of saving time by not having to travel to the sites, the disadvantage is that the views may be a year old or older depending on the area. Thus the candidates for road improvement may include road segments and intersection that have been changed already.
    d. The crash data used were from 2005 to 2009 and thus crash patterns may have changed already in the past three years. While the crash data were readily available, the exact locations for crashes on local roads are often not available.

Thus all crash locations were map spotted to obtain reliable GPS information of the locations.

2. Statistical modeling of hazardous locations.
    a. Because of the lack of data, an SPF was not developed. To develop an SPF, exposure data and engineering features are needed. The AADT were not sufficient for estimating an SPF for local roads.
    b. Ranking of the top p% requires the computation of the tail probability. The Negative Binomial distribution used in this report requires an estimate of the mean and the variance. While the mean may be obtained from an SPF multiplied by known CMFs, the variance of this product is not readily available. Therefore, the approach used in this project was to discretize all factors (road features) and AADT and create classes of road segments/intersections with similar engineering features and exposure levels (AADT).

3. Countermeasures
    a. The countermeasures chosen were limited to engineering. Many engineering countermeasures have costs and crash reduction factors associated with them. Therefore, these costs and crash reduction factors can be used in an initial ranking of projects.
    b. Education and enforcement countermeasures should also play an important role in the evaluation of sites. Thus the engineering countermeasures alone may not lead to a desired reduction in crashes. Specifically, sites with crash reductions that are judged unsatisfactory should be evaluated with respect to human factors. For instance, the location of bars or other places frequented by young drivers may be affecting crashes.
    c. The countermeasures suggested through the proposed procedure should not be used as the final decision, but only as an initial guideline to obtain a ranking of many sites. As mentioned earlier, the Google Earth street view may not reflect the current features.

Although the project had to overcome considerable data issues which accounted for much of the time and resources used in this project, the procedure and Excel application is rather robust with respect to identifying candidates for countermeasures. Although the researchers suggest training for engineers to use this application, the procedures are easy to follow and the results are straightforward.

The process developed in the project serves as the basis of a local road safety improvement program which allows local agencies with guidelines and procedures for a systematic system-wide road improvement methodology. Such a safety improvement program includes the following steps:

1. Prepare local road inventory for the agency. Divide roads into segments of nearly equal length and obtain road engineering features including the following elements:

Table 5
Engineering features and settings

Factor	Road segment of 500 feet	Intersection
Lane Width	<12ft. , 12ft. >12ft.	<12ft. , 12ft. >12ft.
Shoulder	none, <6ft., >=6ft.	none, <6ft., >=6ft.
Alignment	Curve or straight	Curve or straight
Divided street	yes, no	yes, no
Driveways	yes, no	yes, no
Left turn lane		yes, no
Protected left turn		yes, no
Right turn lane		yes, no
Signal		light, type of sign
Speed	25, 30, 35, 45 55, 60	25, 30, 35, 45 55, 60

2. Obtain an inventory of AADT. AADT is necessary to determine exposure levels. AADT may be classified as <2000, 2000-5000, and >5000.
3. Obtain crash information and assure correct GPS location information is available for each crash.
4. Select road features and AADT class to obtain ranking of crash sites.
5. Select all locations with less than 5% tail probability and create a list of road segments and intersections that are candidates for countermeasures.
6. Obtain information on the selected sites from Google Earth to identify potential road hazards.
7. Identify initial countermeasures, costs of countermeasures, and benefits using available CMFs.
8. Rank sites using DEA. The benefits/cost ratio may be used as an approximate measure.
9. This ranked list serves as a baseline for the road improvement program. The program should include a detailed investigation of the sites through a site visit and the determination of other factors that could affect the road safety of the sites but is not

easily obtained from Google Earth. Such factors may include, but are not limited to, presence of schools, bars, restaurants, entertainment establishments, etc. Speed related issues may be obtained from crash data or police agencies that do speed enforcement.

10. Specifically, sites where the benefits of initial countermeasures are not large enough to reduce the number of crashes to an acceptable level should be investigated for other factors not obtained from Google Earth.

The above steps serve as a guideline to institute a systematic system-wide road improvement program. This program should also include resources available to implement countermeasures, collaboration with enforcement agencies, the Louisiana DOTD, and the Louisiana Highway Safety Commission to determine a plan for implementing the countermeasures in engineering, education and enforcement.

RECOMMENDATIONS

There are several recommendations that can be derived from this project. While we recommend that the project be continued by including other parishes, there are prerequisites that need to be addressed first. The main prerequisites include:

1. Create a database for local road sections and features in parishes.

Before a system-wide local road improvement program can be implemented, a road inventory must be available. Many larger agencies may have information on road features but this information is often not readily available for analysis because it is not stored in one accessible database. The research team contacted many agencies to obtain road inventories without success. Either the data are not electronically stored by the agencies or they are stored in different formats that cannot be easily retrieved. It was also difficult to determine who has the responsibility of storing and the authority over these data. Although it is difficult to imagine how road improvements could be managed without knowledge of the engineering features at a local level, the researchers were not able to obtain these road features for local roads. Efforts should be made to obtain an inventory for local roads on a state-wide basis. This inventory should be made easily accessible to engineers and researchers.

2. Create a database of AADT for local roads of parishes and cities.

The AADT for local roads is not available on a system-wide level. There are sporadic AADT values published on websites, but these cover only main routes. Local agencies may have additional AADT numbers but they are not readily available. Effort should be made to create system-wide AADT counts for local roads. Without AADT, no safety program can be established that takes into consideration exposure levels. The missing AADT can then be estimated through various algorithms. This database should be readily available to all engineers and researchers.

3. Increase number of local agencies that provide electronic crash records with GPS information.

The third element for a successful local safety program is the availability of crash records with GPS information. This will allow the identification of hazardous locations. Without the location information, no effective road safety program can be established.

Once these three databases have been established, the application provided with this report can be used to rank crash sites and determine countermeasures.

ACRONYMS, ABBREVIATIONS, AND SYMBOLS

AASHTO	American Association of State Highway and Transportation Officials
cm	centimeter(s)
FHWA	Federal Highway Administration
ft.	foot (feet)
in.	inch(es)
LADOTD	Louisiana Department of Transportation and Development
LTRC	Louisiana Transportation Research Center
lb.	pound(s)
m	meter(s)
AADT	Average Annual Daily Travel
OLAP	Online analytical processing
EB	Empirical Bayes Estimate
PDO	Property damage only
DEA	Data Envelope Analysis
SPF	Safety Performance Function
NegBin	Negative Binomial
CMF	Crash Modification Factor
LSU	Louisiana State University

REFERENCES

1. Jo, J., Lee, J., Ouyang, Y., and Li, Z. 2009. *Crash Data Analysis and Engineering Solutions for Local Agencies*. Urbana, IL: Illinois Center for Transportation, September.
2. Vachal, K., and Johnson, K. 2010. *Systems Analysis to Improve Local Road Safety: Phase I*. Research Proposal. North Dakota State University.
3. Sohn, S. 1999. Quality function deployment applied to local traffic accident reduction. Accident Analysis & Prevention 31, no. 6: 751–761. doi:10.1016/S0001-4575(99)00038-X.
4. Lambert, J., Peterson, K., and Joshi , N. 2006. Synthesis of quantitative and qualitative evidence for accident analysis in risk-based highway planning. Accident Analysis & Prevention 38, no. 5: 925–935. doi:10.1016/j.aap.2006.03.003.
5. Sorensen, M. and Elvik, R. 2007. Black Spot Management and Safety Analysis of Road Networks. Olso, Norway: Transport\okonomisk institutt.
6. Cheng, W., and Washington, S. 2005. Experimental evaluation of hotspot identification methods. Accident Analysis & Prevention 37, no. 5: 870–881. doi:10.1016/j.aap.2005.04.015.
7. Hauer, E. 1996. Identification of Sites with Promise. Transportation Research Record 1542, no. 1: 54–60. doi:10.3141/1542-09.
8. Elvik, R. 2006. New Approach to Accident Analysis for Hazardous Road Locations. Transportation Research Record 1953, no. 1: 50–55. doi:10.3141/1953-06.
9. Elvik, R. 2007. State-of-the-art approaches to road accident black spot management and safety analysis of road networks. Transportøkonomisk institutt (Norway).
10. Hauer, E., Kononov, J., Allery, B., and Smith, M. 2002. *Screening the Road Network for Sites with Promise. Transportation Research Record: Journal of the Transportation Research Board* 1784/2002. Safety and Human Performance: 27–32.
11. Persaud, B., Lyon, C., and Nguyen, T. 2007. Empirical Bayes Procedure for Ranking Sites for Safetly Investigation by Potential for Safety Improvement. Transportation Research Record: Journal of the Transportation Research Board 1665 / 1999. Safety and Human Performance: 7–12.
12. Morency, P., and Cloutier, M-S. 2006. From targeted "black spots" to area-wide pedestrian safety. Injury Prevention 12, no. 6: 360–364. doi:10.1136/ip.2006.013326.

13. Geurts, K., Wets, G., Brijs, T., Vanhoof, K., and Karlis, D. 2006. Ranking and selecting dangerous crash locations: Correcting for the number of passengers and Bayesian ranking plots. Journal of Safety Research 37, no. 1: 83–91. doi:10.1016/j.jsr.2005.10.020.

14. Geurts, K., Wets, G., Brijs, T., and Vanhoof, K. 2004. Identification and Ranking of Black Spots: Sensitivity Analysis. Transportation Research Record 1897, no. 1: 34–42. doi:10.3141/1897-05.

15. Hamidi, A., Fontaine, M., and Demetksy, M. 2010. *A Planning-Level Methodology for Identifying High-Crash Sections of Virginia's Primary System*. Research Report. Charlottesville, VA: Virginia Transportation Research Council, August. http://www.virginiadot.org/vtrc/main/online_reports/pdf/11-r4.pdf.

16. Sorensen, M. 2007. *Best Practice Guidelines on Black Spot Management and Safety Analysis of Road Networks.* Oslo, Norway: Transport\okonomisk institutt.

17. National Research Council, American Association of State Highway Transportation Officials, and National Cooperative Highway Research Program. 2010. Highway Safety Manual. 1st ed. Vol. 1. Washington D.C.: American Association of State Highway and Transportation Officials.

18. Kononov, J., and Allery, B. 2003. Level of Service of Safety: Conceptual Blueprint and Analytical Framework. Transportation Research Record 1840, no. 1: 57–66. doi:10.3141/1840-07.

19. Park, E.S. and Lord, D. 2007. Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. Transportation Research Record 2019, no. -1 (December): 1–6. doi:10.3141/2019-01.

20. Aguero-Valverde, J., and Jovanis, P. 2009. Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking. Transportation Research Record: Journal of the Transportation Research Board 2136, no. -1 (December): 82–91. doi:10.3141/2136-10.

21. Hauer, E., Allery, B., Kononov, J., and Griffith, M. 2004. How Best to Rank Sites with Promise. Transportation Research Record 1897, no. 1: 48–54. doi:10.3141/1897-07.

22. Fridstrom, L. 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. Accident Analysis & Prevention 27, no. 1: 1–20. doi:10.1016/0001-4575(94)E0023-E.

23. El-Basyouny, K., and Sayed, T. 2010b. Safety performance functions with measurement errors in traffic volume. Safety Science 48, no. 10 (December): 1339–1344. doi:10.1016/j.ssci.2010.05.005.

24. Abdel-Aty, M., and Pande A. 2005. Identifying Crash Propensity Using Specific Traffic Speed Conditions. Journal of Safety Research 36, No. 1: 97–108. doi:10.1016/j.jsr.2004.11.002.

25. Zhang, Y., and Liu, Y. 2009. Missing Traffic Flow Data Prediction Using Least Squares Support Vector Machines in Urban Arterial Streets. In 2009 IEEE Symposium on Computational Intelligence and Data Mining, 76–83. Nashville, TN, USA. doi:10.1109/CIDM.2009.4938632. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4938632.

26. Hauer, E., and Hakkert, A. 1988. Extent and some implications of incomplete accident reporting. Transportation Research Record: Journal of the Transportation Research Board 1185: 1–10.

27. Langley, J.D. 2003. Missing cyclists. Injury Prevention 9, No. 4: 376–379. doi:10.1136/ip.9.4.376.

28. Stutts, J., and William, H. 1998. Police Reporting of Pedestrians and Bicyclists Treated in Hospital Emergency Rooms. Transportation Research Record 1635, No. 1: 88–92. doi:10.3141/1635-12.

29. Elvik, R., and Mysen, A. 1999. Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. Transportation Research Record 1665, No. 1: 133–140. doi:10.3141/1665-18.

30. Dhillon, P. 2001. Assessment of hospital and police ascertainment of automobile versus childhood pedestrian and bicyclist collisions. Accident Analysis & Prevention 33, No. 4: 529–537. doi:10.1016/S0001-4575(00)00066-X.

31. Alsop, J. 2001. Under-reporting of motor vehicle traffic crash victims in New Zealand. Accident Analysis & Prevention 33, no. 3: 353–359. doi:10.1016/S0001-4575(00)00049-X.

32. Hilbe, J. 2007. Negative binomial regression. Cambridge: Cambridge University Press.

33. Greene, W. 2008. Econometric analysis. 6th ed. Upper Saddle River N.J.: Pearson/Prentice Hall.

34. Cameron, A., and Trivedi, P. 1998. Regression analysis of count data. Cambridge UK: Cambridge University Press.

35. Cameron, A. 1990. Regression-based tests for overdispersion in the Poisson model. Journal of Econometrics 46, no. 3: 347–364. doi:10.1016/0304-4076(90)90014-K.

36. Hardin, J., and Joseph, H. 2001. Generalized Linear Models and Extensions. College Station: Stata Press.

37. Greene, W. 2008. Functional forms for the negative binomial model for count data. Economics Letters 99, no. 3: 585–590. doi:10.1016/j.econlet.2007.10.015.

38. Lord, D., and Park, P. 2008. Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. Accident Analysis & Prevention 40, no. 4 (July): 1441–1457. doi:10.1016/j.aap.2008.03.014.

39. El-Basyouny, K., and Sayed, T. 2006a. Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models. Transportation Research Record 1950, no. 1: 9–16. doi:10.3141/1950-02.

40. Lord, D., Washington, S., and Ivan, J. 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis & Prevention 37, no. 1: 35–46. doi:10.1016/j.aap.2004.02.004.

41. Lord, D., Washington, S., and Ivan, J.N. 2007. Further notes on the application of zero-inflated models in highway safety. Accident Analysis and Prevention 39, No. 1: 53–57.

42. Warton, D. I. 2005. Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics 16, No. 3: 275–289. doi:10.1002/env.702.

43. Kim, D., Washington, S., and Oh, J. 2006. Modeling Crash Types: New Insights into the Effects of Covariates on Crashes at Rural Intersections. *Journal of Transportation Engineering 132*, no. 4: 282. doi:10.1061/(ASCE)0733-947X(2006)132:4(282).

44. Hadi, M., Aruldhas, J., Chow, L., and Wattleworth, J. 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transportation Research Record 1500: 169–177.

45. Zhang, Y., Ye, Z., and Lord, D. 2007. Estimating Dispersion Parameter of Negative Binomial Distribution for Analysis of Crash Data: Bootstrapped Maximum Likelihood Method. Transportation Research Record 2019, no. -1: 15–21. doi:10.3141/2019-03.

46. Chang, L. 2005. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. Safety Science 43, No. 8: 541–557. doi:10.1016/j.ssci.2005.04.004.

47. Ma, J., and Kockelman, K. 2006. Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity. Transportation Research Record 1950, No. 1: 24–34. doi:10.3141/1950-04.

48. Kockelman, K., 2001. A model for time- and budget-constrained activity demand analysis. Transportation Research Part B: Methodological 35, No. 3: 255–269. doi:10.1016/S0191-2615(99)00050-8.

49. Ma, J., Kockelman, K., and Damien, P. 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis & Prevention 40, no. 3 (May): 964–975. doi:10.1016/j.aap.2007.11.002.

50. El-Basyouny, K., and Sayed, T. 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis & Prevention 41, No. 4 (July): 820–828. doi:10.1016/j.aap.2009.04.005.

51. Aitchison, J. 1964. Two papers on the comparison of Bayesian and frequentist approaches to statistical problems of prediction: Bayesian tolerance regions. In Research Methods Meeting. Series B (Methodological). Royal Statistical Society, February.

52. Casella, G. 1985. An Introduction to Empirical Bayes Data Analysis. The American Statistician 39, No. 2: 83–87.

53. Bayarri, M., and Berger, J. 2004. The interplay of Bayesian and Frequentist Analysis. Statistical Science 19, No. 1: 58–80.

54. Gelman, A. 2009. Bayesian data analysis. 2nd ed. Chapman and Hall.

55. Hauer, E. 1997. Observational before-after studies in road safety. Pergamon Press.

56. Carlin, B., and Louis, T. 1998. Bayes and Empirical Bayes Methods for Data Analysis. Monographs on Statistics and Applied probability 69. London: Chapman and Hall.

57. Hauer, E., Harwood, D., Council, F., and Griffith, G. 2002. Estimating Safety by the Empirical Bayes Method: A Tutorial. Transportation Research Record 1784, No. 1: 126–131. doi:10.3141/1784-16.

58. Lord, D. 2006. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention 38, No. 4: 751–766. doi:10.1016/j.aap.2006.02.001.

59. Federal Highway Administration. 2012. FHWA Local Roads Safety Resource CD, 2012. http://safety.fhwa.dot.gov/local_rural/training/resourcecd/.

60. Patterson, T. and Munnich, L. Five-County Minnesota Case Study: Rural Roadway Fatal Crash Characteristics and Select Safety Improvement Programs, Center for Excellent in Rural Safety, University of Minnesota, 2008.

61. Federal Highway Administration (FHWA), Low-cost Safety Enhancements for Stop Controlled and Signalized Intersections, FHWA-SA-09-020, 2009

62. Rakauskas, M., Ward, N., Gerberich, S., and Alexander, B. Rural and Urban Safety Cultures: Human-Centered Interventions Toward Zero Deaths in Rural Minnesota, Report No.: MN/RC 2007-41, University of Minnesota, 2007.

63. McGee, H., and Hanscom, F. Low Cost Treatments for Horizontal Curve Safety-FHWA, Report No.: FHWA-SA-07-002, December, 2006.

64. Isebrands, H., Hallmark, S., Hans, Z., McDonald, T., Preston, H., and Strom, R. Safety Impacts of Street Lighting at Isolated Rural Intersection, Minnesota Department of Transportation and Minnesota Local Road Research Board, December, 2004.

65. Preston, H., Storm, R., Donath, M., and Shankwitz, C. Review of Georgia's Rural Intersection Crashes: Application of Methodology for Identifying Intersections for Intersection Decision Support (IDS), Minnesota Department of Transportation, July, 2007.

66. Golembiewski, G.A., and Chandler, B. Roadway Departure Safety: A Manual for Local Rural Road Owners, Federal Highway Administration, U.S. Department of Transportation, Report No.: FHWA-SA-11-09, January, 2011.

67. Carlson, P., Rose, E., Chrysler, S., and Bischoff, A., Simplifying delineator and chevron applications for horizontal curves, Texas Department of Transportation, March, 2004.

68. American Association of State Highway Transportation Officials. 2010. Highway Safety Manual. 1st ed. Vol. 2. Washington D.C.: American Association of State Highway and Transportation Officials.

69. American Association of State Highway Transportation Officials. 2010. Highway Safety Manual. 1st ed. Vol. 3. Washington D.C.: American Association of State Highway and Transportation Officials.

70. Latham, F., and Trombly, J. Low Cost Traffic Engineering Improvements: A Primer, Federal Highway Administration (FHWA), U.S. Department of Transportation, Report No.: FHWA-OP-03-078, April, 2003.

71. Federal Highway Administration, Two Low-Cost Safety Concepts for Two-Way STOP-Controlled, Rural Intersections on High-Speed Two-Lane, Two-Way Roadways, U.S. Department of Transportation, Report No.: FHWA-HRT-08-063, September 2008.

72. Molino, J., Katz, B., Hermosillo, M., Dagnall, E., and Kennedy, J. Simulator Evaluation of Low-Cost Safety Improvements on Rural Two-Lane Undivided Roads: Nighttime Delineation for Curves and Traffic Calming for Small Towns, Federal Highway Administration, U.S. Department of Transportation, Report No. FHWA-HRT-09-061, February 2010.

73. Mass Highway Traffic Engineering. Low Cost Non-Intersection Safety Improvements, January 2008

74. Farid, F., Johnston, D., Laverde, M., and Chen, C. Application of incremental benefit cost analysis for optimal budget allocation to maintenance, rehabilitation and replacement of bridges. Transportation Research Record 1442, 1994; 88–100.

75. Sinha K., Kaji, T., and Liu, C. Optimal allocation of funds for highway safety improvement projects. Transportation Research Record 808, 1981; 24–30.

76. Skinner R. RRR design standards: Cost effectiveness issues. In: Proceedings of the Conference on the Effectiveness of Highway Safety Improvements, Nashville, TN, 1985; 41–50.

77. Sinha K., and Hu, K. Assessment of safety impacts of highway projects. In: Proceedings of the Conference on the Effectiveness of Highway Safety Improvements, Nashville, TN, 1985; 31–40.

78. Pal, R., and Sinha, K. Optimization approach to highway safety improvement programming. Transportation Research Record 1640, 1998; 1–9.

79. Melachrinoudis, E., and Kozanidis, G. A mixed integer knapsack model for allocating funds to highway safety improvements. Transportation Research Part A: Policy and Practice 2002; 36; 789-803.

80. Brown, B. Use of dynamic programming in optimally allocating funds for highway safety improvements. Transportation Planning and Technology 1980; 6; 131–138.

81. Norese, M., and Viale, S. A multi-profile sorting procedure in the public administration. European Journal of Operational Research, 2002; 138; 365-379.

82. Yedla, S., and Shrestha, R. Multi-criteria approach for the selection of alternative options for environmentally sustainable transport system in Delhi. Transportation Research Part A: Policy and Practice 2003; 37; 717-729.

83. Hinloopen E., **Ni**jkamp P., and Rietveld, P. Integration of ordinal and cardinal information in multi-criteria ranking with imperfect compensation. *European Journal of Operational Research* 2004; 158; 317-338.

84. Odeck, J. Identifying traffic safety best practice: an application of DEA and Malmquist indices. Omega 2006; 34; 28-40.

85. Tudela A, Akiki N, Cisternas R. Comparing the output of cost benefit and multi-criteria analysis: An application to urban transport investments. Transportation Research Part A: Policy and Practice 2006; 40; 414-423.

86. Ghorbani, S., and Rabbani, M. A new multi-objective algorithm for a project selection problem. Advances in Engineering Software 2009; 40; 9-14.

87. Kozanidis, G. Solving the linear multiple choice knapsack problem with two objectives: Profit and equity, Computational Optimization and Applications 2009; 43; 261-294.

88. Zongzhi, L., Madanu, S., Zhou, B., Wang, Y., and Abbas M. A Heuristic Approach for Selecting Highway Investment Alternatives. Computer-Aided Civil and Infrastructure Engineering 2010; 25; 427-439.

89. Teng, J-Y, Huang, W. C., and Lin, M. C. Systematic budget allocation for transportation construction projects: A case in Taiwan. Transportation, 2010; 37, 331-361.

90. Farrell, M.J., 1957. The measurement of productive efficiency, *Journal of the Royal Statistical Society*, Series A 120: 253ety.

91. Charnes, A., Cooper, W.W., and Rhodes, E., 1978. Measuring the efficiency of decision-making units. European Journal of Operations Research 2: 429pean.

92. Cooper W.W., Seiford, L.M., and Tone K. Data Envelopment Analysis, A comprehensive text with models, applications, references and DEA-solver software. Kluwer: Dordrecht; 1999.Amatatsu, H., Ueda, T.,

93. Amatatsu, Y., 2011. Efficiency and returns-to-scale of local governments, J Oper Res Soc 63: 299-305.

94. Zhao, Y., Triantis, K., Murray-Tuite, P., and Edara, P., 2011. Performance measurement of a transportation network with a downtown space reservation system: A network-DEA approach. Transportation Research Part E: Logistics and Transportation Review 47(6): 1140-1159.

95. Golany, B. An interactive MOLP procedure for the extension of data envelopment analysis to effectiveness analysis. *Journal of the Operational Research Society* 1988; 39; 725–734.

96. Stewart, T. Relationships between data envelopment analysis and multicriteria decision analysis. *Journal of the Operational Research Society* 1996; 47; 654–665.

97. Belton, V., and Stewart, J. DEA and MCDA: competing or complementary approaches? In: Meskens N, Roubens M. (Edts), Advances in decision analysis, Kluwer: Dordrecht; 1999; 87–104.

APPENDIX A

Calculating Distances and Matching Points with Line Segments

This appendix describes the algorithms and formulas that were used to implement the procedures described in the report above.

Calculating Distances between Coordinates

It is possible to implement some of the procedures using off-the-shelf GIS programs, such as ArcGIS. The researchers did not use such programs and provide documentation for everything done in the most general terms possible. Analysts will therefore be able to follow the suggestions or alter the procedures to their own specification with a large degree of flexibility.

The researchers use an implementation of the Haversine formula for distance calculations because it remains very well-conditioned and computationally stable at small distances and is therefore suitable for the problem. The researchers calculate distances, d, using the formula[1]

$$\mathrm{d} = r \cdot 2 \cdot \mathrm{atan2}\left(\sqrt{a}, \sqrt{1-a}\right)$$

$$a = \sin^2 \frac{\mathrm{lat}_2 - \mathrm{lat}_1}{2} + \cos \mathrm{lat}_1 \cdot \cos \mathrm{lat}_2 \cdot \sin^2 \frac{\mathrm{lon}_2 - \mathrm{lon}_1}{2}$$

The researchers use a radius of $r = 20{,}902{,}231$ ft. for all of the calculations.

Headings are calculated such that

$$\theta_{12} = \mathrm{atan2}(\sin(\mathrm{lon}_2 - \mathrm{lon}_1) \cdot \cos \mathrm{lat}_2\ ,\ \cos \mathrm{lat}_1 \cdot \sin \mathrm{lat}_2 - \sin \mathrm{lat}_2 \cdot \cos \mathrm{lat}_2 \cdot \cos(\mathrm{lon}_2 - \mathrm{lon}_1))$$

In order to re-segment the road grid into 500 ft. sub-segments, the researchers move along each road from intersection to intersection and, given a set of starting coordinates and initial heading, calculating the coordinates of the end point after moving 500 ft. using the following formulas.

1 See Sinnott, R.W. (1984). "Virtues of the Haversine." *Sky and Telescope, Vol. 68:2* for a derivation of the formula.

$$\mathrm{lat}_2 = \mathrm{lat}_1 + \frac{d}{r}\cos\theta$$

$$\mathrm{lon}_2 = \mathrm{mod}\left(\mathrm{lon1} + \frac{d}{r}\sin\theta/\cos\mathrm{lat}_1 + \pi, 2\pi\right) - \pi$$

where, d is the distance moved (i.e., 500 ft.) and r is again the radius of the earth.

Matching Points with Line Segments

Crash locations and traffic count locations are identified as points with geographic coordinates. The road grid used in this study is described by a set of geographic coordinates that, when connected, form a set of line segments. In order to match crash locations or AADT count locations with correct road segment, proceed as follows:

1. Calculate the distance between the crash location and every end point of all road segments. Keep only those segments for which at least one of the endpoints has a distance that is less than 500 feet.
2. For each of the remaining segments, calculate the cross track distance[2] of the point 0 from the vector described by the segments using the formula

$$CTD_0 = \mathrm{asin}(\sin\frac{d_{10}}{r}\ \sin(\theta_{10} - \theta_{12})) * r$$

3. Where, the subscript 0 indicates the crash location and subscripts 1 and 2 index the end points of the remaining road segments after step 1. Positive numbers indicate that the point lies to the right of the track, negative numbers indicate that the point lies to the left of the track.
4. Calculate the distance along track using the distances calculated in step one and the cross track distances from step two. The formula used is

$$D_{along} = \mathrm{asin}(\frac{\sqrt{\sin^2 D_{10} - \sin^2 CTD_0}}{\cos CTD_0})$$

This is the distance from the endpoint 1 along the course to the other endpoint 2 to the point abeam 0.

2 This is sometimes also called “cross track error” when used to describe how far airplanes are off course, for example.

5. Keep only those segments for which the following is true: (1) the along track distance of point 0 from endpoint 1 is less than the distance of point 0 to endpoint 1. AND (2) the along track distance of point 0 from endpoint 2 is less than the distance of point 0 to endpoint 2.

6. If there is more than one segment remaining from step 4, then out of the resulting segments, pick the one that has the smallest absolute cross track error and use it as the match.

7. If the absolute cross track error is less than 75 ft., then calculate the "meeting coordinates" along the track (the point along 1 toward 2 that is abeam point 0) and keep it.

8. If can't find segment closer than 75 ft., then ignore that crash.

APPENDIX B
Tables for Countermeasures and Estimated Costs

Table 6
Section levels and intersections analyzed

							No. of Sections with crashes for 5 years (2005-2009)			Sections selected based on Tail Prob. Values			No. of Sections analyzed		
Sect. Level	AADT Class	Left Turn Lane	Divided St.	Protected Left Turn	Right Turn Lane	No. of Sections	1-5 Crashes	6-10 Crashes	11 or above crashes	1-5 Crashes	6-10 Crashes	11 or above crashes	1-5 Crashes	6-10 Crashes	11 or above crashes
1	0-5,000	No	No	No	No	89	22	3	0	3	3	0	3	3	0
2	0-5,000	Yes	Yes	Yes	Yes	0	0	0	0	0	0	0	0	0	0
3	0-5,000	No	Yes	No	No	12	2	0	0	0	0	0	0	0	0
4	0-5,000	No	No	No	Yes	0	0	0	0	0	0	0	0	0	0
5	0-5,000	No	No	Yes	No	2	1	0	0	0	0	0	0	0	0
6	0-5,000	Yes	No	No	No	0	0	0	0	0	0	0	0	0	0
7	0-5,000	No	Yes	Yes	Yes	0	0	0	0	0	0	0	0	0	0
8	0-5,000	Yes	Yes	No	Yes	0	0	0	0	0	0	0	0	0	0
9	0-5,000	Yes	Yes	Yes	No	0	0	0	0	0	0	0	0	0	0
10	0-5,000	Yes	No	Yes	Yes	0	0	0	0	0	0	0	0	0	0
11	0-5,000	No	Yes	Yes	No	2	1	0	1	0	0	1	0	0	1
12	0-5,000	Yes	No	No	Yes	0	0	0	0	0	0	0	0	0	0
13	0-5,000	No	No	Yes	Yes	0	0	0	0	0	0	0	0	0	0
14	0-5,000	No	Yes	No	Yes	0	0	0	0	0	0	0	0	0	0
15	0-5,000	Yes	Yes	No	No	0	0	0	0	0	0	0	0	0	0
16	0-5,000	Yes	No	Yes	No	2	1	0	0	0	0	0	0	0	0
17	5,000+	No	No	No	No	25	9	0	0	1	0	0	1	0	0
18	5,000+	Yes	Yes	Yes	Yes	0	0	0	0	0	0	0	0	0	0
19	5,000+	No	Yes	No	No	6	2	0	0	0	0	0	0	0	0
20	5,000+	No	No	No	Yes	3	1	0	0	0	0	0	0	0	0
21	5,000+	No	No	Yes	No	0	0	0	0	0	0	0	0	0	0
22	5,000+	Yes	No	No	No	0	0	0	0	0	0	0	0	0	0
23	5,000+	No	Yes	Yes	Yes	0	0	0	0	0	0	0	0	0	0
24	5,000+	Yes	Yes	No	Yes	0	0	0	0	0	0	0	0	0	0
25	5,000+	Yes	Yes	Yes	No	0	0	0	0	0	0	0	0	0	0
26	5,000+	Yes	No	Yes	Yes	0	0	0	0	0	0	0	0	0	0
27	5,000+	No	Yes	Yes	No	0	0	0	0	0	0	0	0	0	0
28	5,000+	Yes	No	No	Yes	1	1	0	0	0	0	0	0	0	0
29	5,000+	No	No	Yes	Yes	0	0	0	0	0	0	0	0	0	0
30	5,000+	No	Yes	No	Yes	0	0	0	0	0	0	0	0	0	0
31	5,000+	Yes	Yes	No	No	1	1	0	0	0	0	0	0	0	0
32	5,000+	Yes	No	Yes	No	3	0	0	0	0	0	0	0	0	0
33	Unknown	No	No	No	No	3,882	236	10	4	50	10	4	0	5	4
34	Unknown	Yes	Yes	Yes	Yes	9	0	0	0	0	0	0	0	0	0
35	Unknown	No	Yes	No	No	490	23	1	0	6	1	0	0	1	0
36	Unknown	No	No	No	Yes	23	3	0	0	1	0	0	1	0	0
37	Unknown	No	No	Yes	No	55	12	3	0	3	3	0	1	3	0
38	Unknown	Yes	No	No	No	21	3	0	0	0	0	0	0	0	0
39	Unknown	No	Yes	Yes	Yes	2	0	0	0	0	0	0	0	0	0
40	Unknown	Yes	Yes	No	Yes	8	3	0	0	1	0	0	1	0	0
41	Unknown	Yes	Yes	Yes	No	18	0	2	1	0	0	1	0	0	1
42	Unknown	Yes	No	Yes	Yes	31	6	1	0	1	1	0	0	1	0
43	Unknown	No	Yes	Yes	No	15	5	0	0	0	0	0	0	0	0
44	Unknown	Yes	No	No	Yes	9	2	0	1	0	0	1	1	0	1
45	Unknown	No	No	Yes	Yes	1	0	0	0	0	0	0	0	0	0
46	Unknown	No	Yes	No	Yes	17	1	0	0	1	0	0	0	0	0
47	Unknown	Yes	Yes	No	No	29	0	0	0	0	0	0	0	0	0
48	Unknown	Yes	No	Yes	No	65	12	1	0	5	1	0	1	1	0
					Total	4,821	347	21	7	72	19	7	9	14	7

Table 7
Matrix of roadway segment levels and crash information

							No. of Sections with crashes for 5 years (2005-2009)			Sections selected based on Tail Prob. Values			No. of Sections analyzed		
							Crashes			Crashes			Crashes		
Sect. Level	AADT Class	Left Turn Lane	Divided St.	Protected Left Turn	Right Turn Lane	No. of Sections	1-5	6-10	11 or above	1-5	6-10	11 or above	1-5	6-10	11 or above
1	0-5,000	No	No	No	No	89	22	3	0	3	3	0	3	3	0
2	0-5,000	Yes	Yes	Yes	Yes	0	0	0	0	0	0	0	0	0	0
3	0-5,000	No	Yes	No	No	12	2	0	0	0	0	0	0	0	0
4	0-5,000	No	No	No	Yes	0	0	0	0	0	0	0	0	0	0
5	0-5,000	No	No	Yes	No	2	1	0	0	0	0	0	0	0	0
6	0-5,000	Yes	No	No	No	0	0	0	0	0	0	0	0	0	0
7	0-5,000	No	Yes	Yes	Yes	0	0	0	0	0	0	0	0	0	0
8	0-5,000	Yes	Yes	No	Yes	0	0	0	0	0	0	0	0	0	0
9	0-5,000	Yes	Yes	Yes	No	0	0	0	0	0	0	0	0	0	0
10	0-5,000	Yes	No	Yes	Yes	0	0	0	0	0	0	0	0	0	0
11	0-5,000	No	Yes	Yes	No	2	1	0	1	0	0	1	0	0	1
12	0-5,000	Yes	No	No	Yes	0	0	0	0	0	0	0	0	0	0
13	0-5,000	No	No	Yes	Yes	0	0	0	0	0	0	0	0	0	0
14	0-5,000	No	Yes	No	Yes	0	0	0	0	0	0	0	0	0	0
15	0-5,000	Yes	Yes	No	No	0	0	0	0	0	0	0	0	0	0
16	0-5,000	Yes	No	Yes	No	2	1	0	0	0	0	0	0	0	0
17	5,000+	No	No	No	No	25	9	0	0	1	0	0	1	0	0
18	5,000+	Yes	Yes	Yes	Yes	0	0	0	0	0	0	0	0	0	0
19	5,000+	No	Yes	No	No	6	2	0	0	0	0	0	0	0	0
20	5,000+	No	No	No	Yes	3	1	0	0	0	0	0	0	0	0
21	5,000+	No	No	Yes	No	0	0	0	0	0	0	0	0	0	0
22	5,000+	Yes	No	No	No	0	0	0	0	0	0	0	0	0	0
23	5,000+	No	Yes	Yes	Yes	0	0	0	0	0	0	0	0	0	0
24	5,000+	Yes	Yes	No	Yes	0	0	0	0	0	0	0	0	0	0
25	5,000+	Yes	Yes	Yes	No	0	0	0	0	0	0	0	0	0	0
26	5,000+	Yes	No	Yes	Yes	0	0	0	0	0	0	0	0	0	0
27	5,000+	No	Yes	Yes	No	0	0	0	0	0	0	0	0	0	0
28	5,000+	Yes	No	No	Yes	1	1	0	0	0	0	0	0	0	0
29	5,000+	No	No	Yes	Yes	0	0	0	0	0	0	0	0	0	0
30	5,000+	No	Yes	No	Yes	0	0	0	0	0	0	0	0	0	0
31	5,000+	Yes	Yes	No	No	1	1	0	0	0	0	0	0	0	0
32	5,000+	Yes	No	Yes	No	3	0	0	0	0	0	0	0	0	0
33	Unk	No	No	No	No	3882	236	10	4	50	10	4	0	5	4
34	Unk	Yes	Yes	Yes	Yes	9	0	0	0	0	0	0	0	0	0
35	Unk	No	Yes	No	No	490	23	1	0	6	1	0	0	1	0
36	Unk	No	No	No	Yes	23	3	0	0	1	0	0	1	0	0
37	Unk	No	No	Yes	No	55	12	3	0	3	3	0	1	3	0
38	Unk	Yes	No	No	No	21	3	0	0	0	0	0	0	0	0
39	Unk	No	Yes	Yes	Yes	2	0	0	0	0	0	0	0	0	0
40	Unk	Yes	Yes	No	Yes	8	3	0	0	1	0	0	1	0	0
41	Unk	Yes	Yes	Yes	No	18	0	2	1	0	0	1	0	0	1
42	Unk	Yes	No	Yes	Yes	31	6	1	0	1	1	0	0	1	0
43	Unk	No	Yes	Yes	No	15	5	0	0	0	0	0	0	0	0
44	Unk	Yes	No	No	Yes	9	2	0	1	1	0	1	1	0	1
45	Unk	No	No	Yes	Yes	1	0	0	0	0	0	0	0	0	0
46	Unk	No	Yes	No	Yes	17	1	0	0	1	0	0	0	0	0
47	Unk	Yes	Yes	No	No	29	0	0	0	0	0	0	0	0	0
48	Unk	Yes	No	Yes	No	65	12	1	0	5	1	0	1	1	0
				Total		4821	347	21	7	73	19	7	9	14	7

Table 8
Countermeasures, potential CMF values, and estimated costs for intersections

Section	DEA rank	Crash Count	EB	(1-CMF) *EB	Cost avrg.	Tail Prob.	log2 (odd)	DEA Eff.	Road Name	Countermeasure	CMF1	CMF2	CMF3	CMF4	CMF
9206901201	1	4	2.7	0.8	75	0.10%	10	1.0	Prospect Blvd./ LA 24	1) Revise signal timing for yellow and all-red intervals	0.7				0.70
9205225001	2	4	3.0	0.9	75	1.75%	6	1.0	Industrial Blvd./ LA 661	1) Revise signal timing for yellow and all-red intervals	0.7				0.70
9206887203	3	5	2.9	0.9	125	0.34%	8	0.6	Saint Charles St./Valhi Blvd.	1) Revise signal timing for yellow and all-red intervals 2) Regulate minimum spacing of driveway (No CMF value)	0.70				0.70
9205093503	4	10	8.2	3.3	800	0.00%	48	0.5	Suthon Ave./Elizabeth St.	1) Improve visibility of intersections by providing enhanced sign and pavement markings, 2) Regulate minimum spacing of driveway (No CMF value)	0.60				0.60
9206942802	5	25	23.9	7.2	1500	0.00%	69	0.4	Bayou Gardens Blvd./ Alma St.	1) Revise signal timing for yellow and all-red intervals, 2) Advane 'Signal Ahead' sign (No CMF Value), 3) Regulate minimum spacing of driveway (No CMF value)	0.7				0.70
9205337001	6	16	14.9	4.5	900	0.00%	28	0.4	East Main St./East Woodlawn Ranch Rd/Little Caillou Rd.	1) Revise signal timing for yellow and all-red intervals, 2) Add advance signal ahead signage (No CMF value)	0.70				0.70
9234826501	7	25	24.4	7.3	1500	0.60%	7	0.4	Audubon Ave./LA 448	1) Revise signal timing for yellow and all-red intervals, 2) Advane 'Signal Ahead' sign (No CMF Value), 3) Regulate minimum spacing of driveway (No CMF value)	0.7				0.70
9207015404	8	28	22.9	14.0	4000	0.00%	172	0.3	Frank Street/Prospect Blvd.	1) Revise signal timing for yellow and all-red intervals, 2) Enhance roadway illumination at night, 3) Provide left turn lane at major road approach, 4) Regulate minimum spacing of driveway (No CMF value)	0.70	0.62	0.90		0.39
9205089203	9	9	7.4	3.0	1050	0.00%	42	0.3	New Orleans Blvd./6th St.	1) Improve visibility of intersections by providing enhanced sign and pavement markings, 2) Install warning signs (No CMF value)	0.60				0.60
9235046501	10	8	6.6	2.6	1050	0.00%	37	0.3	Williams St./Queen St.	1) Improve visibility of intersections by providing enhanced sign and pavement markings, 2) Install warning signs (No CMF value)	0.60				0.60

Table 8
Countermeasures, potential CMF values, and estimated costs for intersections (cont'd)

Section	DEA rank	Crash Count	EB	(1-CMF) *EB	Cost avrg.	Tail Prob.	log2 (odd)	DEA Eff.	Road Name	Countermeasure	CMF1	CMF2	CMF3	CMF4	CMF
9205337001	11	16	14.9	4.5	1500	0.00%	28	0.2	East Woodlawn Ranch Rd. / LA 56	1) Revise signal timing for yellow and all-red intervals, 2) Advane 'Signal Ahead' sign (No CMF Value), 3) Regulate minimum spacing of driveway (No CMF value)	0.7				0.70
9205088203	12	14	11.5	4.9	2850	0.00%	73	0.2	6 th Street/Williams Avenue	1) Improve visibility of intersections by providing enhanced sign and pavement markings, 2) Install flashing beacons	0.60	0.95			0.57
9206820402	13	13	10.7	4.6	2850	0.00%	67	0.2	Beatrice St./East Tunnel Blvd.	1) Improve visibility of intersections by providing enhanced sign and pavement markings, 2) Install flashing beacons,	0.60	0.95			0.57
9205268701	14	7	5.0	1.5	900	0.00%	17	0.2	Louisiana 57/Prospect Rd.	1) Revise signal timing for yellow and all-red intervals, 2) Add advance signal ahead signage (No CMF value) 3) Regulate minimum spacing of driveway (No CMF value)	0.70				0.70
9207015501	15	6	4.4	1.3	750	0.06%	11	0.2	Prospect Blvd./Frank St.	1) Revise signal timing for yellow and all-red intervals 2) Regulate minimum spacing of driveway (No CMF value)	0.7				0.70
9205103003	16	4	2.9	0.9	550	1.13%	6	0.1	Barrow St./ Belanger St.	1) Revise signal timing for yellow and all-red intervals, 2) Sign for No right-turn on Red (No CMF value), 3) Remove site obstructions (No CMF value)	0.70				0.70
9206830605	17	7	5.5	1.6	1150	0.00%	17	0.1	East Main St/Prospect Blvd.	1) Revise signal timing for yellow and all-red intervals, 2) Sign for No right-turn on Red (No CMF value), 3) Add advance signal ahead signage (No CMF value)	0.70				0.70
9205676005	18	6	4.4	1.3	950	0.06%	11	0.1	Prospect Blvd./ Grand Caillou Rd.	1) Revise signal timing for yellow and all-red intervals 2) Regulate minimum spacing of driveway (No CMF value)	0.7				0.70
9206535901	19	6	4.4	1.6	1250	0.06%	11	0.1	St. Louis Canal Rd./Bayou Gardens Blvd.	1) Revise signal timing for yellow and all-red intervals 2) Provide left turn lane at major road approach,	0.7	0.9			0.63

Table 8
Countermeasures, potential CMF values, and estimated costs for intersections (cont'd)

Section	DEA rank	Crash Count	EB	(1-CMF) *EB	Cost avrg.	Tail Prob.	log2 (odd)	DEA Eff.	Road Name	Countermeasure	CMF1	CMF2	CMF3	CMF4	CMF
9206902301	20	6	4.2	1.7	1400	0.03%	12	0.1	Cummins St./Moffet Rd.	1) Improve visibility of intersections by providing enhanced sign and pavement markings, 2) Install warning signs (No CMF value) 3) Remove site obstructions (No CMF value)	0.60				0.60
60458402302	21	4	2.9	1.1	1000	1.13%	6	0.1	Ridgefield Ave./ Thompson Place	1) Provide Stop bar on minor approaches and Pavement marking improvements	0.60				0.60
9207005303	22	5	3.7	1.1	1000	0.88%	7	0.1	Saint Charles St./ Southdown Blvd.	1) Revise signal timing for yellow and all-red intervals, 2) Sign for No right-turn on Red (No CMF value), 3) Add advance signal ahead signage (No CMF value)	0.70				0.70
62078481006	23	7	5.7	2.5	2850	0.00%	31	0.1	Bowie Rd./Ardoyne Dr.	1) Improve visibility of intersections by providing enhanced sign and pavement markings, 2) Install flashing beacons,	0.60	0.95			0.57
9206807001	24	7	5.7	2.5	2850	0.00%	31	0.1	East St./ Daniel Turner Court	1) Improve visibility of intersections by providing enhanced sign and pavement markings, 2) Install flashing beacons, 3) Regulate minimum spacing of driveway (No CMF value)	0.60	0.95			0.57
9205020201	25	6	4.2	1.3	1400	0.03%	12	0.1	Pendleton Dr./ St. Charles St./ St. Antonio Blvd.	1) Signal and sign improvement, 2) Regulate minimum spacing of driveway (No CMF value), 3) Remove site obstructions (No CMF value)	0.70				0.70
9206901203	26	18	14.7	2.9	10000	0.00%	100	0.1	Gleanmore Ave./ Prospect Blvd.	1) Convert Stop Control to Signal Control, 2) Modify left turn signal phase, 3) Provide left turn lane at major road approach 4) Regulate minimum spacing of driveway (No CMF value)	0.95	0.90	0.94		0.80
9205097501	27	6	3.3	2.1	2750	0.00%	27	0.1	Canal Street/ High St.	1) Provide Stop bar on minor approaches and Pavement marking improvements, 2) Enhance roadway illumination at night	0.60	0.62			0.37
9234668001	28	6	4.2	2.4	3500	0.03%	12	0.1	Bayou Blue Bypass Rd./ Burma Rd.	1) Provide left turn signal phase, 2) Improve roadway illumination at night ,	0.70	0.62			0.43
9207123702	29	4	2.9	1.7	2500	0.48%	8	0.1	East St./ LA 57	1) Revise signal timing for yellow and all-red intervals, 2) Enhance roadway illumination at night 3) Regulate minimum spacing of driveway (No CMF value) 4) Remove site obstructions (No CMF value)	0.70	0.62			0.43
9205743701	30	4	2.9	0.2	1150	1.13%	6	0.0	Alma St./ Bayou Gardens Blvd.	1) Modify left turn signal phase, 2) Remove site obstructions (No CMF value)	0.94				0.94

<0.01% is shown as 0.00%

Table 9
Countermeasures, potential CMF values, and estimated costs for roadway segments

Section	DEA rank	Crash Count	EB	(1-CMF)* EB	Cost ($)	Tail Prob.	log2 (Odds)	DEA Eff.	Road Name	Countermeasure	CMF 1	CMF 2	CMF 3	CMF
9233021601	1	7	3.5	1.22	400	0.00%	35	1.0	Arms St.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9207043602	2	20	17.5	10.70	3330	0.00%	114	1.0	Bayou Gardens Blvd	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
9206781903	3	20	17.9	9.86	5,000	0.00%	61	0.6	Saint Charles St	1) Convert 4 lane undivided to 5 lane undivided with a TWLTL	0.45			0.45
9206887002	4	4	3.5	2.15	1575	0.00%	15	0.4	Saint Charles St.	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
9206884402	5	4	2.0	0.70	600	0.00%	17	0.4	Southdown Mandalay Rd	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9205669702	6	3	1.8	1.08	900	0.01%	14	0.4	Saint Charles St.	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
9207043603	7	10	8.8	5.36	4725	0.00%	48	0.4	Bayou Gardens Blvd	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
9204953102	8	11	9.7	5.89	5300	0.00%	54	0.3	Westside Blvd	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
9234076002	9	7	6.3	2.20	2,250	0.01%	13	0.3	Audubon Ave.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9207015502	10	23	13.5	7.45	9,000	0.00%	121	0.3	Prospect Blvd	1) Convert 4 lane undivided to 5 lane undivided with a TWLTL	0.45			0.45
6045849000[illegible]	11	9	5.3	1.85	3,300	0.00%	37	0.2	Williams St.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9235046503	12	7	4.2	1.44	2,600	0.00%	27	0.2	Williams St.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65

Table 9
Countermeasures, potential CMF values, and estimated costs for roadway segments (cont'd)

Section	DEA rank	Crash Count	EB	(1-CMF)* EB	Cost ($)	Tail Prob.	log2 (Odds)	DEA Eff.	Road Name	Countermeasure	CMF 1	CMF 2	CMF 3	CMF
9206500803	13	4	3.7	1.27	2,250	2.04%	6	0.2	Industrial Blvd	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9205257402	14	5	3.0	1.04	2,100	0.00%	18	0.2	Payne St.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9205102802	15	5	3.0	1.04	2,100	0.00%	18	0.2	School St.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9206901202	16	10	5.9	3.26	7,000	0.00%	43	0.1	Prospect Blvd	1) Convert 4 lane undivided to 5 lane undivided with a TWLTL	0.45			0.45
9206892902	17	3	2.1	1.30	2850	1.37%	6	0.1	Saint Charles St.	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
60458490303	18	8	3.9	2.38	5350	0.00%	34	0.1	Brocato Ln	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
9205226301	19	7	4.2	1.44	3,300	0.00%	27	0.1	Industrial Blvd	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9205676004	20	8	4.8	2.61	6,000	0.00%	32	0.1	Prospect Blvd	1) Convert 4 lane undivided to 5 lane undivided with a TWLTL	0.45			0.45
9232873602	21	5	2.9	1.01	2,550	0.00%	26	0.1	Madewood Dr	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9206916902	22	5	3.0	1.04	2,700	0.00%	18	0.1	Stovall St.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9204953103	23	3	2.6	1.61	4800	0.06%	11	0.1	Westside Blvd	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
60458490202	24	5	3.0	1.04	3,300	0.00%	18	0.1	Brocate Ln	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65

Table 9
Countermeasures, potential CMF values, and estimated costs for roadway segments (cont'd)

Section	DEA rank	Crash Count	EB	(1-CMF)* EB	Cost ($)	Tail Prob.	log2 (Odds)	DEA Eff.	Road Name	Countermeasure	CMF 1	CMF 2	CMF 3	CMF
9207003903	25	4	2.7	1.63	5500	0.00%	15	0.1	Mystic Blvd	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
9207043604	26	3	2.6	1.61	5500	0.06%	11	0.1	Bayou Gardens Blvd	1) Change the width of an existing median 2) Install Combination Horizontal Alignment/ Advisory Speed Signs 3) Install Changeable Accident Ahead Warning Signs	0.8	0.87	0.56	0.39
9205337002	27	5	2.5	0.88	3,300	0.00%	23	0.1	E Woodlawn Ranch Road	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9206178906	28	4	2.0	0.70	2,700	0.00%	17	0.1	E Woodlawn Ranch Road	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9233271602	29	4	2.4	0.83	3,300	0.01%	13	0.1	St Louis St.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9205676003	30	7	4.2	2.29	9,000	0.00%	27	0.1	Prospect Blvd	1) Convert 4 lane undivided to 5 lane undivided with a TWLTL	0.45			0.45
9233248204	31	3	1.5	0.52	2,850	0.10%	10	0.1	Triple Oaks Dr	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9207015403	32	5	3.0	1.65	9,000	0.00%	18	0.1	Prospect Blvd	1) Convert 4 lane undivided to 5 lane undivided with a TWLTL	0.45			0.45
9206756403	33	3	1.5	0.52	3,200	0.10%	10	0.1	Linda Ann Ave	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9205109502	34	4	1.3	0.44	2,700	1.71%	6	0.1	Gabasse St.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9205874805	35	4	1.3	0.44	2,700	1.71%	6	0.1	Corporate Dr.	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65
9206881304	36	3	1.2	0.42	3,000	2.75%	5	0.0	E Woodlawn Ranch Road	1) Centerline and edgeline marking improvement 2) Install rumble strip	0.76	0.86		0.65

<0.01% is shown as 0.00%

APPENDIX C

Example of Data Problems

Crash Coded as Segment Crash but too Close to Intersection

Figure 17
Google view of Mystic Blvd. (Picture Id- 12)

Figure 18
Google view of Arms St. (Picture Id- 37)

The Tiger file used for the development of road segments does often not delineate the exact location of intersections. Thus some road segments were misclassified as non-intersection though the Google map revealed that there was an intersection. Additional quality control review is necessary to reduce misclassifications. However, this would require additional resources for the project.

Roadway Segment with no Google Street View (4 Sites Shown Here)

Some road segments in rural areas do not provide a street view in Google Earth. Some of these road segments have very low traffic counts and are not good candidates for road improvements and others are recent developments that do not have street views yet.

Figure 19
Google view of Sanders Road (Picture Id- 9)

Figure 20
Google view of N J Theriot Road (Picture Id- 36)

Figure 21
Google view of Clendenning Road (Picture Id- 41)

Figure 22
Google view of White Rose Dr. (Picture Id- 38)