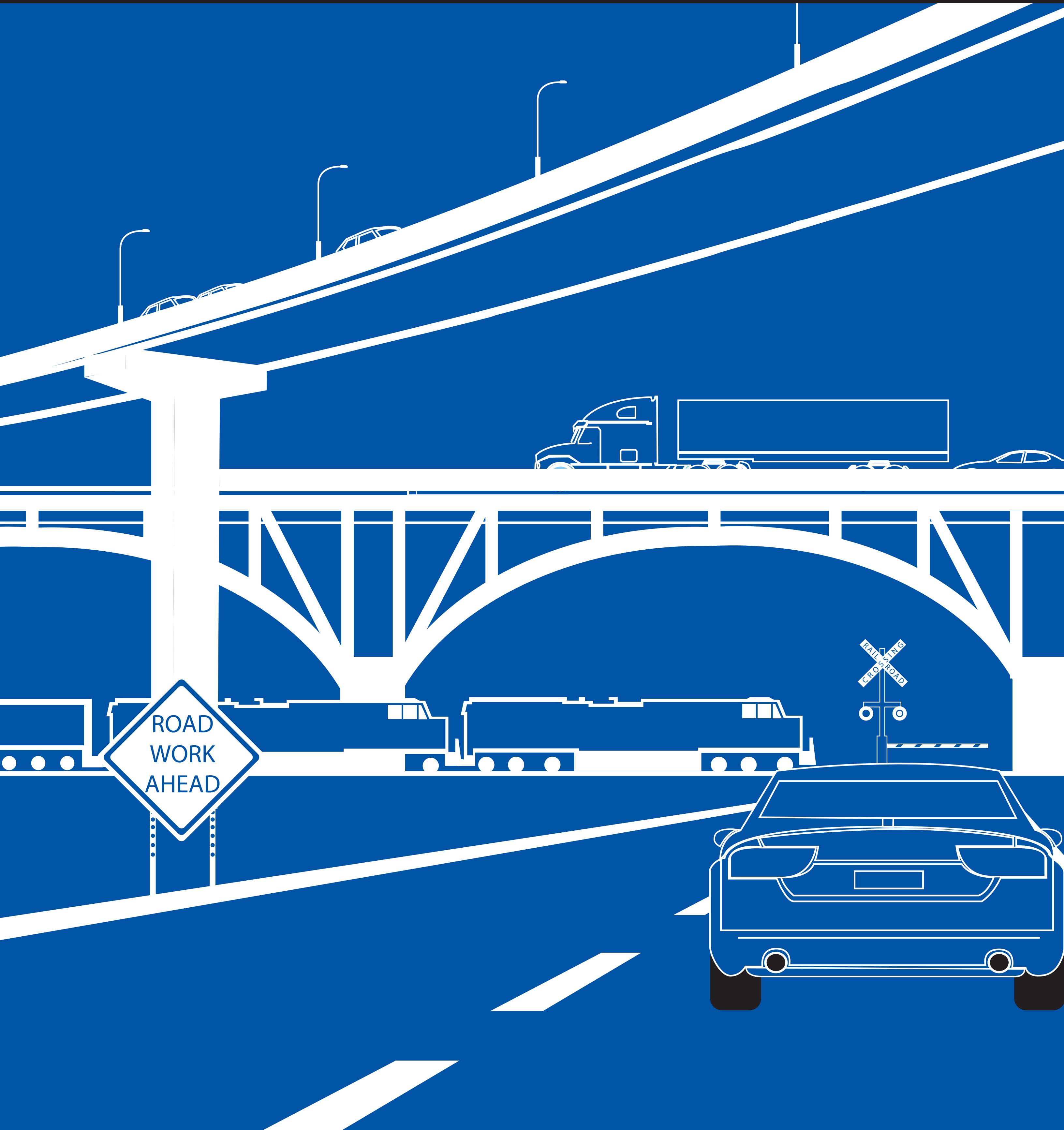
Safety Analysis for SHIFT Implementation

Report Number: KTC-19-23/PL31-1F

DOI: https://doi.org/10.13023/ktc.rr.2019.23



Kentucky Transportation Center
College of Engineering, University of Kentucky, Lexington, Kentucky

in cooperation with
Kentucky Transportation Cabinet
Commonwealth of Kentucky

The Kentucky Transportation Center is committed to a policy of providing equal opportunities for al persons in recruitment, appointment, promotion, payment, training, and other employment and education practices without regard for economic, or social status and will not discriminate on the basis of race, color, ethnic origin, national origin, creed, religion, political belief, sex, sexual orientation, marital status or age.

Kentucky Transportation Center
College of Engineering, University of Kentucky, Lexington, Kentucky

in cooperation with
Kentucky Transportation Cabinet
Commonwealth of Kentucky

KENTUCKY
Transporation Center

Kentucky Transportation Center • University of Kentucky
176 Raymond Building • Lexington, KY 40506 • 859.257.6898 • www.ktc.uky.edu

University of Kentucky

Research Report
KTC-19-23/PL31-1F

Safety Analysis for SHIFT Implementation

Reginald Souleyrette, P.E., Ph.D.
Principal Investigator

Riana Tanzen
Research Assistant

William Staats, P.E., MSCE
Research Engineer

Eric Green, P.E., Ph.D.
Program Manager

and

Mei Chen, Ph.D.
Research Engineer

Kentucky Transportation Center
College of Engineering
University of Kentucky
Lexington, Kentucky

In Cooperation With
Kentucky Transportation Cabinet
Commonwealth of Kentucky

July 2019

1. Report No. KTC-19-23/PL31-1F	**2. Government Accession No.**	**3. Recipient's Catalog No**
4. Title and Subtitle Safety Analysis for SHIFT Implementation		**5. Report Date** July 2019 **6. Performing Organization Code**
7. Author(s): Reginald Souleyrette, Riana Tanzen, William Staats, Eric Green, Mei Chen		**8. Performing Organization Report No.** KTC-19-23/PL31-1F
9. Performing Organization Name and Address Kentucky Transportation Center College of Engineering University of Kentucky Lexington, Kentucky 40506-0281		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** PL-31
12. Sponsoring Agency Name and Address Kentucky Transportation Cabinet 200 Mero Street Frankfort, Kentucky 40622		**13. Type of Report and Period Covered** Final Year **14. Sponsoring Agency Code**
15. Supplementary Notes		

16. Abstract

The Strategic Highway Investment Formula for Tomorrow (SHIFT) program research team evaluated the program's 2018 safety component and developed a new methodology to rank the safety needs of 2020 projects. For the current year, the research team suggests replacing the SHIFT 2018 formulas (based on three naive crash measures) with a new metric, Excess Expected Crashes, formulated from the most current safety analysis guidelines available in the Highway Safety Manual. In addition, the research team developed a user-friendly network screening tool serving as a method to prioritize projects. Another automation tool was developed that will help safety professionals evaluate and prioritize projects in an efficient manner, replacing the laborious manual calculation needed for each project. This report details the new methodology and tools.

17. Key Words SHIFT, Highway Safety Manual, data driven safety analysis, safety performance function, empirical Bayes, excess expected crashes		**18. Distribution Statement** Unlimited, with approval of the Kentucky Transportation Cabinet	
19. Security Classification (report) Unclassified	**20. Security Classification (this page)** Unclassified	**21. No. of Pages** 17	**22. Price**

Table of Contents

Executive Summary 1
1. Introduction 2
  1.1 Background 2
  1.2 Problem Statement 2
  1.3 Objectives 2
2. Literature Review 3
  2.1 Analysis of SHIFT 2018 Component 3
  2.2 Safety Performance Functions 3
  2.3 Excess Expected Crashes 4
3. Methodology 6
  3.1 Approach 6
  3.2 Development of SPFs 6
  3.3 Network Screening 7
  3.4 Estimation of EEC for projects 8
    3.4.1 EEC for Segment Level 8
    3.4.2 Adjusting EEC for Project Level 10
4. Results 12
5. Conclusion 13
References 14
Appendix A CURE Plots of Safety Performance Functions 15
Appendix B: Regression Parameters and Over-Dispersion Parameters for SPFs 17

List of Figures

Figure 1 Graphical representation of EEC....................5
Figure 2 ArcGIS Map Package for Network Screening....................8
Figure 3 Visualization of Four Cases....................9
Figure 4 Two Segments Having a Common Intersection....................11
Figure 5 Three Segments Having Several Common Intersections....................11

List of Tables

Table 1 Base Conditions for Each Roadway Type....................7
Table 2 EEC Equations for Different Cases....................9
Table 3 Descriptive Statistics of EEC....................12

Acknowledgements

The following individuals contributed to the successful completion of this project through their participation on the Study Advisory Committee:

Eileen Vaughn, Chair .... KYTC Division of Planning
Mike Vaughn.. KYTC Division of Traffic Operations, HSIP
Nathan Ridgway... KYTC Division of Traffic Operations, HSIP
Jill Asher .... KYTC Division of Highway Design
John Moore.. KYTC Asst. State Highway Engineer
Jason Blackburn .... KYTC District 10
Amanda Spencer .... KYTC Division of Planning
Mikael Pelfrey.. KYTC Division of Planning
Travis Thompson .... KYTC Division of Planning

Executive Summary

The Strategic Highway Investment Formula for Tomorrow (SHIFT) program research team evaluated the program's 2018 safety component and developed a new methodology to rank the safety needs of 2020 projects. For the current year, the research team suggests replacing the SHIFT 2018 formulas (based on three naive crash measures) with a new metric formulated from the most current safety analysis guidelines available in the Highway Safety Manual. The new metric is known as Excess Expected Crashes (EEC).

EEC is derived from a crash prediction model or Safety Performance Function (SPF). The function estimates the number of crashes one might expect on a road segment with a given traffic volume and length. The research team has developed state-specific crash prediction models for various roadway types with similar geometrics based on traffic volume and roadway characteristics. EEC is a value that represents the difference between the road segment's current crashes and its expected crashes for the segment type and length. In short, EEC suggests the number of excess crashes a segment should experience compared to others of its type. The new prediction models and EEC are not biased by length or traffic volume. Therefore, the research team proposes that using EEC for an individual segment is a more meaningful method for prioritizing SHIFT projects. All upcoming SHIFT 2020 projects will be ranked based on EEC of the individual projects. The project with the highest EEC receives the maximum score for the SHIFT safety component, with all successive projects receiving incrementally lower scores.

In addition, the research team developed a user-friendly network screening tool to aid in identifying roads with high EECs, serving as a method to prioritize projects. Another automation tool was developed that will help safety professionals evaluate and prioritize projects in an efficient manner, replacing the laborious manual calculation needed for each project.

1. Introduction

1.1 Background

Implemented in 2018, the Strategic Highway Investment Formula for Tomorrow (SHIFT) is KYTC's planning tool for comparing capital improvement projects and prioritizing transportation funding [1]. SHIFT has five basic components: Safety, Asset Management, Congestion, Economic Growth and Benefit/Cost. For SHIFT 2020, the SHIFT safety team desired an improved methodology to rank all projects based on the most current and nationally accepted data-driven safety evaluation methods. This research relies on guidance in the Highway Safety Manual (HSM) to develop the safety component for SHIFT 2020.

1.2 Problem Statement

For the SHIFT 2018 cycle, the safety component was calculated using a combination of three safety measures: Critical Rate Factor (CRF), Crash Frequency (CF), and Crash Density over a segment length (CD*L). Recent research has shown that CRF is not the most accurate or reliable method to compare a segment's crash performance to segments of a similar type. CRF relies on the assumption that crashes and traffic volume have a linear relationship, which is not always true. Crash Frequency (CF) produces a length bias because longer segments will have more space available to accumulate crashes. Evaluations of these two measures are typically affected by regression-to-the-mean bias because both CRF and CF do not address this. The measure CD*L also does not accurately reflect the number of crashes that should be expected on a roadway because factors other than roadway type and length influence crash occurrence, such as roadway geometry and traffic volume. In addition, the contribution of each of these measures to the final safety ranking formulas was arbitrary.

For SHIFT 2020, the safety research team suggested replacing the SHIFT 2018 formulas with a new metric based on the most current safety analysis guidelines available. The new metric is known as Excess Expected Crashes (EEC), derived from a crash prediction model. As the name suggests, EEC represents the number of excess crashes a segment is experiencing compared to others of its type. The project with the highest EEC receives the maximum score for the safety component of SHIFT, with all successive projects receiving incrementally lower scores.

1.3 Objectives

The objectives for this project are:

- Evaluate the methodology for ranking safety projects in the SHIFT 2018 cycle.
- Propose a new method of evaluating safety for the SHIFT 2020 cycle.
- Develop Safety Performance Functions for Kentucky's entire state-maintained roadway network.
- Test the new method on SHIFT 2018 projects and use this method to score SHIFT 2020 projects.
- Develop a tool for Network Screening for safety in Kentucky.

2. Literature Review

The research team first evaluated the project prioritizing method used in SHIFT 2018 and found the method to be often inaccurate and unreliable. Next, the team reviewed the literature related to development of Safety Performance Functions, eventually leading the team to use Excess Expected Crashes (EEC) as the chosen safety performance metric.

2.1 Analysis of SHIFT 2018 Component

For the 2018 segments or intersections, the safety score was calculated using a combination of Critical Rate Factor (CRF), Crash Frequency (CF), and Crash Density over a segment length (CD*L). CRF is a measure that compares a segment's crash rate to a crash rate that is considered critical, or much greater than the average crash rate for a segment of that roadway type. CF is simply the total number of crashes a location experiences in five years. CD*L is an attempt to distinguish each SHIFT project based on its roadway type. It represents the average crash density (crashes per mile) for each roadway type (interstate, parkway, rural and urban multilane, rural and urban two lanes, etc.). The following equations show how the three components were weighted to create a combined safety score for segments and intersections:

Segment (L>0.2): $= 0.25*((CD*L)_{\dagger scaled}) + 0.25*(CRF_{\dagger\ scaled}) + 0.50*(CF_{\dagger scaled})$
Intersection (L<=0.2): $= 0.5*(CF_{\dagger scaled}) + 0.5*(CRF_{\dagger scaled})$

The three components for each project were scaled from 0-100 based on how their magnitudes ranked in comparison to all other SHIFT projects. The scaled values of the three components were combined for each SHIFT project to create a single safety score. The scaled components were weighted differently based on the length of a project. If the project was less than or equal to 0.2 miles, the project was considered an intersection. If the project was greater than 0.2 miles in length, the project was considered a segment.

One of the main shortcomings of this method is that it does not account for the potential non-linear relationship between traffic volume and crashes. CRF assumes that more traffic volume will produce more crashes, which is not always accurate. It is possible that a low-volume road has more crashes than a high-volume road due to other factors (i.e., roadway's geometric attributes). Another issue is that CF and CD*L have bias towards segment length, however, a longer segment will not always have more crashes just because it has more space to accumulate crashes. Therefore, with the previous method, projects with higher traffic volume and longer length received higher scores whether there were additional crashes occurring or not. Regression to the mean bias is also not addressed with any of the components, which means they do not account for temporal fluctuation in crashes. These biases can produce misleading results, and when used for project prioritization, there was a possibility that worthy potential projects were not chosen.

Another issue with this method is that the weighting of each of the three components shown in the equations above is arbitrary and contributes to a length bias. For example, in both the segment and intersection equations, CF contributes 50% of a project's score. As discussed, CF is influence by the length of a project, and longer projects tend to have higher crash totals.

2.2 Safety Performance Functions

Safety performance functions (SPF) are one of the fundamental building blocks of the crash predicting methods in the HSM. The development of a SPF requires a database of roadway segments (or ramps or intersections) that contains information on segment length, traffic volume for that site, and the number of crashes. Negative Binomial regression is used to create an equation that relates predicted crashes to traffic volume and length [2]. Where the traditional method assumes that crashes have a linear relationship with traffic volume, most SPFs exhibit an exponential relationship between crashes and volume. Segment length is treated as an offset in that it is directly proportional to the crash prediction. If a road segment does not

identically match the base conditions of the homogenous roadway segments, Adjustment Factors (AF) are used to account for the difference between site conditions and specified base conditions [3].
The generalized functional form of an SPF for a roadway (not ramp or intersection) and a ramp is defined as follows:

$$\textbf{SPF Predicted Crashes} = L * e^{a} * AADT^{b} * AF$$

Where,

L = Length of segment
AADT = Annual Average Daily Traffic
a = regression parameter for intercept
b = regression parameter for AADT
AF = adjustment factor (if needed)

The formula for the SPF of an intersection is expressed as follows:

$$\textbf{SPF Predicted Crashes} = L * e^{a} * AADT_{Major}^{b1} * AADT_{Minor}^{b2} * AF$$

Where,

$AADT_{Major}$ = Annual Average Daily Traffic of the major road
$AADT_{Minor}$ = Annual Average Daily Traffic of the minor road
a, b_1, b_2= regression parameters

Based on the roadway type used in the regression model, the model form varies and the regression coefficients change. Statistical packages with built-in tools such as SPSS, SAS, and RStudio can execute this regression model. SPFs can also be developed in Microsoft Excel using solver or custom functions. Though these tools can generate SPF manually, sometimes improving the model development becomes cumbersome since it requires several iterations and filtering of the roadway dataset. An automation tool named “SPF-R” in RStudio proved to be more efficient and customizable to a variety of potential uses [4].

The source code is available on GitHub at: http://github.com/irkgreen/SPF-R. As an input, the code requires a CSV-format file with a specific set of attributes of roadway segments, intersections, or ramps. The SPF-R tool itself must be configured for a specific project with its own paths for input and output, as well as any additional model specifications. An excel file with model parameters and data, a Cumulative Residual (CURE) plot, a scatter plot, and four box plots (length, crashes, crashes per mile and AADT) are the outputs of the tool. A CURE plot is a measure of goodness-of-fit of the crash model. It is a graph of the cumulative residuals versus an independent variable (i.e., traffic volume) [5]. At a given site, residuals are the difference between actual crashes and the SPF predicted crashes.

2.3 Excess Expected Crashes

The estimation of Excess Expected Crashes (EEC) requires an Empirical Bayes (EB) estimate of crashes along with the SPF for predicted crashes. For safety analysis, the EB method addresses two problems: it increases the precision of estimates when the usual estimate is too imprecise to be useful, and it addresses the regression-to-mean bias [6]. The SPF crash prediction and the historical crash data are balanced using a weight parameter that is a function of how well the SPF model represents the dataset from which it was correlated. If the SPF has poor correlation, the weight parameter places more emphasis on the historic crash data, and vice versa. The EB method uses the following formula:

$$\textbf{EB Expected Crashes} = w * SPF\ Crashes + (1 - w) * Historic\ Crashes$$

$$w = \frac{1}{1 + (SPF\ Crashes/Segment\ Length)/\theta}$$

Where,

w = weight (based on over-dispersion parameter from calibrated SPF)
SPF Crashes = predicted crashes on a segment from SPF
Historic Crashes = total historic crashes on a segment
θ = Over-dispersion parameter

The difference between EB expected crashes and SPF predicted crashes is a measure termed as Excess Expected Crashes (EEC). EEC quantifies the number of crashes occurring at a location more than what would be expected. EEC is positive when more crashes are occurring than expected, and EEC is negative when fewer crashes are occurring than expected. Figure 1 shows a visual representation of the relationship between SPF predicted crashes, historic crashes, EB expected crashes, and EEC.

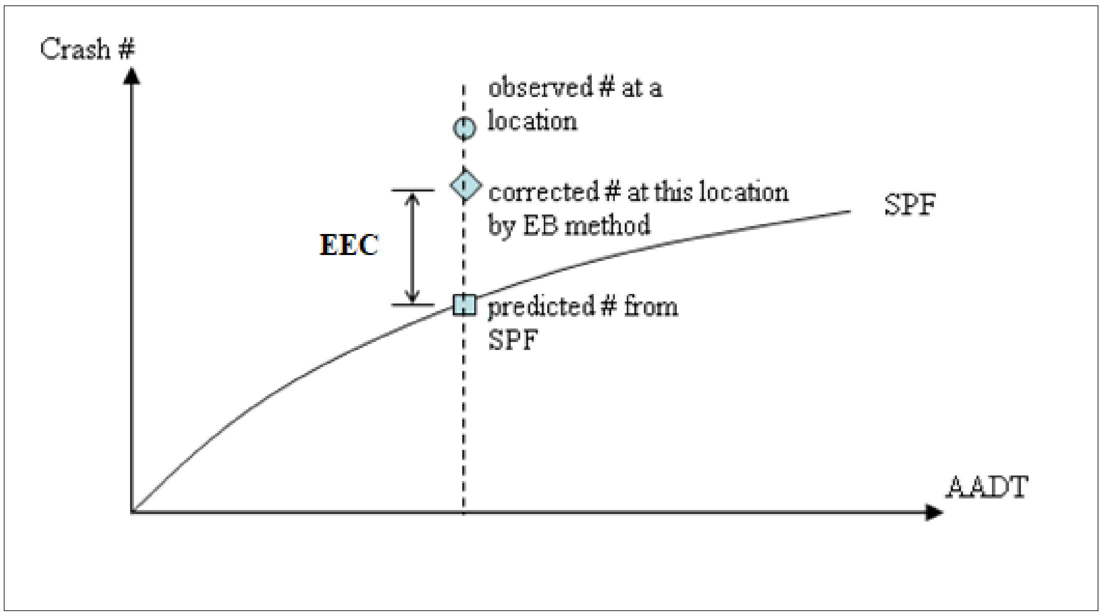


Figure 1 Graphical representation of EEC

3. Methodology

3.1 Approach

The HSM promotes the use of Safety Performance Functions (SPFs) to model crash frequency based on traffic volume and length of homogeneous roadway segments. SPFs are typically modeled using negative binomial regression, which is a more accurate representation of the relationship between crashes and traffic volumes than the assumed linear relationship modeled by CRF. The estimated number of crashes calculated by an SPF represents the number of crashes one might expect on an average length of road with a given traffic volume. If a road segment does not identically match the base conditions of the homogenous roadway segments used to calibrate the SPF, then an adjustment factor (AF) must be applied to the SPF's crash prediction to account for the difference in roadway characteristics. Furthermore, the HSM recommends the use of the Empirical Bayes (EB) method, which combats regression to the mean by combining the SPF crash prediction for a segment with the historical crash data of that segment.

The difference between EB expected crashes and SPF predicted crashes is defined as Excess Expected Crashes (EEC). EEC better quantifies the number of crashes occurring at a location. For the SHIFT 2020 cycle, EEC will be used as a standalone measure to replace the three measures used to evaluate safety in the SHIFT 2018 cycle.

Instead of using the CD*L measure to distinguish between crash patterns on different roadway types, the safety team developed a new SPF for each roadway type for the SHIFT 2020 cycle. Individualized SPFs for each roadway type were used to calculate crash predictions, EB estimates, and EECs for projects on each roadway type. This method more accurately captures the differences in crash patterns on differing roadway types than a simple crash density average (CD*L). The Safety Team developed SPFs for the following roadway types: ramps, intersections, rural two-lanes, rural interstates/parkways, rural multilane divided highways, rural multilane undivided highways, urban two-lanes, urban interstates/parkways, urban multilane divided highways, and urban multilane undivided highways.

All SHIFT 2020 projects will be ranked based on the EEC of each project. The project with the highest EEC receives the maximum number of safety points toward the overall SHIFT score based on the weight of the safety component. Each successive project receives a lower score, with the amount of score reduction being linear and based on the total number of projects in the SHIFT 2020 cycle. Due to the linear nature of the scoring process, a project may have an EEC an order of magnitude higher than the next highest ranked project, even though their SHIFT safety scores may close in magnitude. In this case, the safety score for the higher-ranked project will come with a warning that the EEC is much greater than the next highest SHIFT project.

3.2 Development of SPFs

State-specific Safety Performance Functions were developed for ten principal Kentucky state-maintained roadway types. These include rural two-lane, rural divided multilane, rural undivided multilane, urban two-lane, urban divided multilane, urban undivided multilane, rural interstates and parkways, urban interstates and parkways, ramps and intersections. The development of SPFs for roadways required a comprehensive dataset with information on segment length, traffic volume, and five years of crash data. For the safety component of SHIFT 2020, crash data from the years 2013 to 2017 were used to develop the models. When developing SPFs for intersections, length was omitted, and separate traffic volumes were used for major and minor routes [7]. Along with this information, the process needed detailed geometric attributes of each segment, i.e., lane width, shoulder width, number of lanes, median width and type, curve class, and grade class.

The SPFs were developed using the single source code in the program R-Studio named “SPF-R”. The model required several iterations and filtering of the dataset, thus this program consolidated SPF development and assessment into one streamlined process. One of the major parameters to assess the SPFs and measure the goodness-of-fit was the CURE plot, which required several steps to generate. SPF-R readily generated a CURE plot along with each SPF [7]. The values of the regression parameters of SPFs were different for each roadway type and SPF-R generated an SPF based on the exclusive input dataset using the default model form shown in Equation 1 (roadways and ramps) and Equation 2 (intersections). To avoid errors in the SPF function, two filters were pre-set for every model. It filtered out segments with zero AADT and lengths less than 0.01 miles.

For each roadway type, several SPFs were modeled as different combinations of geometric attributes for base conditions. Based on goodness-of-fit measures, the best models were chosen. An important criterion for choosing the most useful model was the availability of Adjustment Factors. If the attributes of a roadway did not match with the base conditions, Adjustment Factors (AF) adjusted the predicted crashes. HSM and CMF Clearinghouse are the sources for AF and if AF is not available for any base condition, that attribute is eliminated from the model. Table 1 shows the base conditions used for each roadway’s SPF development and Appendix A includes the CURE plots of SPFs. Appendix B includes corresponding regression parameters and over-dispersion parameters.

Table 1 Base Conditions for Each Roadway Type

Roadway Type	Base Conditions
Rural 2 Lane	Lane Width = 9 ft Shoulder Width = 3 ft Horizontal Curve = Class A Grade = Class A
Urban 2 Lane	---
Rural Multilane (Divided)	Shoulder Width = 10 ft
Rural Multilane (Undivided)	Lane Width = 12 ft
Urban Multilane (Divided)	Median Width over 20
Urban Multilane (Undivided)	Lane Width = 12 ft
Rural Interstate and Parkways	---
Urban Interstate and Parkways	---

3.3 Network Screening

Safety Performance Functions can identify locations that may benefit the most from a safety treatment, an application defined as Network Screening [5]. After the development of SPFs for each roadway type, intersections and ramps, Network Screening was performed using ArcGIS software. The goal was to create a user-friendly tool that anybody with even a limited knowledge of ArcGIS would be able to use. The network screening approach in the HSM requires that a roadway should be divided into homogeneous segments based on engineering judgment and by certain roadway attributes [8]. For this project, the roads of each roadway type were segmented by the conditions for each roadway type’s SPFs so that each segment was only as short as it needed to be (not longer than 1.5 mile). This segmentation fulfilled the SPF and AF needs for that roadway type. EECs were calculated for all segments of all roadway types, intersections, and ramps, and then plotted in GIS. Finally, all of the statewide EEC data was compiled into a single ArcGIS

Map Package along with Kentucky's Highway District, Metropolitan Planning Organization, and Area Development District boundaries. The EECs were color coded according to their magnitude. This allowed each entity to zoom into their jurisdiction and identify roads with high EECs that would serve as good project candidates. Figure 2 shows the network screening tool.

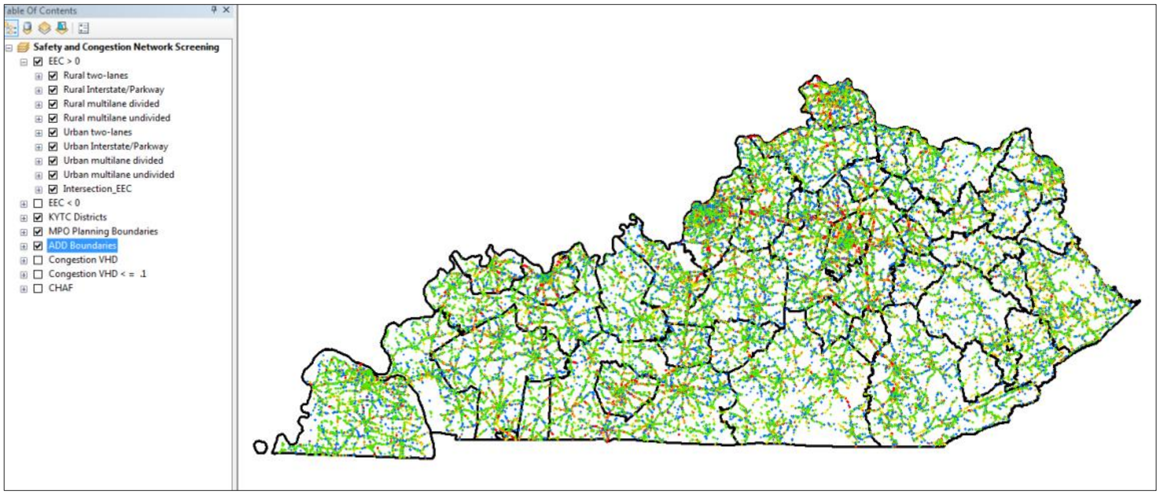


Figure 2 ArcGIS Map Package for Network Screening

3.4 Estimation of EEC for projects

There were 1288 projects to evaluate for SHIFT 2020. Some of the projects were divided into multiple segments that contained three types of elements: roadway section, intersections, and ramps. The EEC of a segment is the summation of all EECs of the roadway segments, intersections and ramps that fall inside a segment. The final EEC of a project was obtained by summing the EECs of each segment included under that project. The task was done in two phases:

- Calculating the EEC of individual segments
- Calculating the EEC of the projects

3.4.1 EEC for Segment Level

This is straight forward if the beginning and ending mile points of a project coincide with those of the roadway segments. It becomes more complex when either the starting point, ending point or both mile points do not coincide with the segment mile points. In these cases, calculating the weighted EEC over length of the roadway becomes a requirement.

From the list of projects to be prioritized, four cases are summarized below. Figure 3 is a schematic representation of the four cases, where red dots refer to the beginning and ending mile points of a project. Table 2 shows the equations used to calculate the EECs for each case.

- Case 1- The beginning and ending mile points coincide with those of the segments
- Case 2 – The ending mile point coincides with the mile point of the segments, but the beginning mile point does not
- Case 3 – The beginning mile point coincides with the mile point of the segments, but the ending mile point does not
- Case 4 - Neither the beginning nor the ending mile point coincides with those of the segments

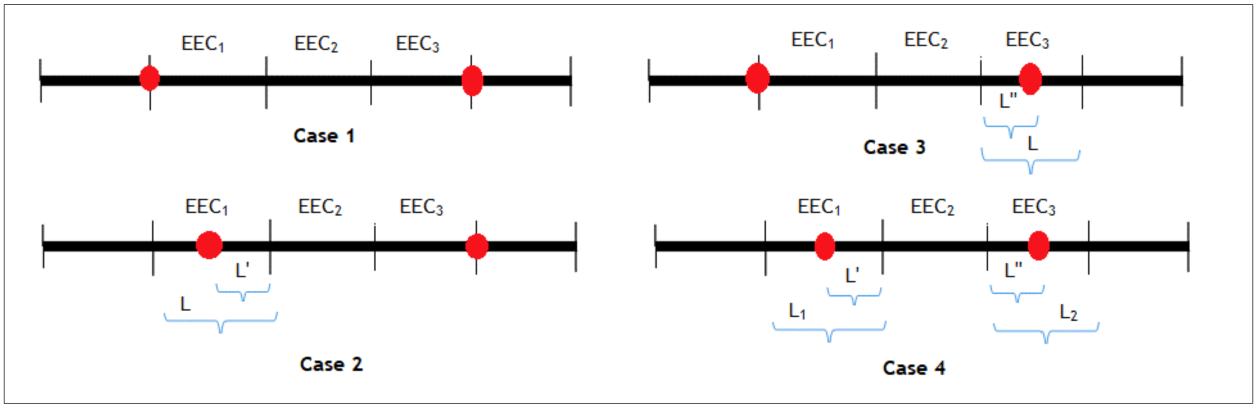


Figure 3 Visualization of Four Cases

Table 2 EEC Equations for Different Cases

Case	**Final EEC**
Case 1	$EEC_1 + EEC_2 + EEC_3$
Case 2	$\frac{L'}{L} * EEC_1 + EEC_2 + EEC_3$
Case 3	$EEC_1 + EEC_2 + \frac{L''}{L} * EEC_3$
Case 4	$\frac{L'}{L1} * EEC_1 + EEC_2 + \frac{L''}{L2} * EEC_3$

In calculating EEC, the four cases need to be addressed differently and doing an individual calculation for each project is time-consuming and laborious. The team developed an automation tool to perform this process quickly and efficiently for any number of projects at once.

3.4.1.1 Data Preparation

The data preparation process requires four files in CSV/SHP format.

- File 1- A file with segment information. Critical information includes a unique segment identifier (denoted as "SegmentNo", unique route identifier, and beginning and ending mile points (denoted as BEGINNINGMILE and ENDINGMILE respectively).
- File 2 - A file containing the EECs of all eight types of roadways by segment. It must include a unique route identifier, and the beginning and ending mile point of each segment (denoted as BEGIN_MP and END_MP respectively).
- File 3 - A file containing the EECs of all intersections. It needs to include unique route identifiers of the two intersecting roads and the mile point on the major road.
- File 4 - A file containing the EECs of all ramps. It needs to include unique route identifier.

The next step is to spatially join File 1 and 2, File 1 and 3 and File 1 and 4. This is done using the Spatial Join tool in ArcGIS. The first output, "Roadway_EEC" contains all the roadway segments that fall within the beginning and ending mile points of a project. If a project's mile points fall in between a segment, it will include that segment as well. The second and third output files, "Intersection_EEC" and "Ramp_EEC" contain all the EECs of the intersections and ramps that fall within the beginning and ending mile points of each project. The automation tool requires these three files in CSV-format with specific sets of attributes and specific column names.

3.4.1.2 Code Configuration

The python code is available on GitHub: https://github.com/rianatanzen068/SHIFT_Safety. This is an online open source code, which anyone can download and use for safety project prioritization. The code can be modified for improvement and when the changes are committed to the GitHub repository, other safety professionals can access and use the modification. Following is the cell-by-cell explanation of the interactive notebook:

- Cell 1: Imports the necessary packages: pandas and numpy.
- Cell 2: Imports the "Roadway_EEC.csv" file as a data frame.
- Cell 3: Adds a new column "LENGTH"; Segment length = END_MP – BEGIN_MP.
- Cell 4 and 5: Defines a function which determines whether the mile points of the project fall in between any segment (Case 2, 3, 4) and creates a new column "StartEnd".
- Cell 6 and 7: Defines a function to calculate a new length which is necessary to calculate weighted EEC of a segment depending on its position (start or end of a project). Also, creates a column "NewLength".
- Cell 8: Calculates the weighted EEC for the segments.
- Cell 9: Defines a function to replace the weighted EEC with the original EEC.
- Cell 10: Creates a new column "NEW_EEC" with the modified and original EEC.
- Cell 11: Sums up the EEC depending on their "SegmentNo", stores in a column "Roadway_EEC", and names the data frame as "Roadway_EEC_new".
- Cell 12: Replaces the NaN to 0 in the Roadway_EEC" column.
- Cell 13: Imports the "Intersection_EEC.csv" file as a data frame.
- Cell 14: Merges two data frames: "Intersection_EEC" and "Roadway_EEC_new" to "Merged1".
- Cell 15: Merges two data frames: "Merged1" and "Ramp_EEC" to "Final_EEC".
- Cell 16: Sums up the roadway, intersection and ramp EEC for each segment creating a new column "Final_EEC".
- Cell 17: Sorts the EECs in ascending order.
- Cell 18: Saves the resulting data frame to "FinalList.csv".
- Cell 19: Provides a description of the EECs.
- Cell 20, 21 and 22: Defines a function to determine whether the EECs are positive, negative or zero and gives the value counts of each.

3.4.2 Adjusting EEC for Project Level

Each project has a unique CHAF ID and every segment under a project shares the same identifier. Therefore, the final EEC of a project is the summation of EECs for segments having the same CHAF ID. But for some projects, this EEC had to be adjusted. Since intersections are considered as polygons, in some cases, two segments shared the same polygon. Figure 4 and Figure 5 show some examples where blue lines are the segments to be evaluated. This adjustment ensured there were no duplicate intersections in one project.

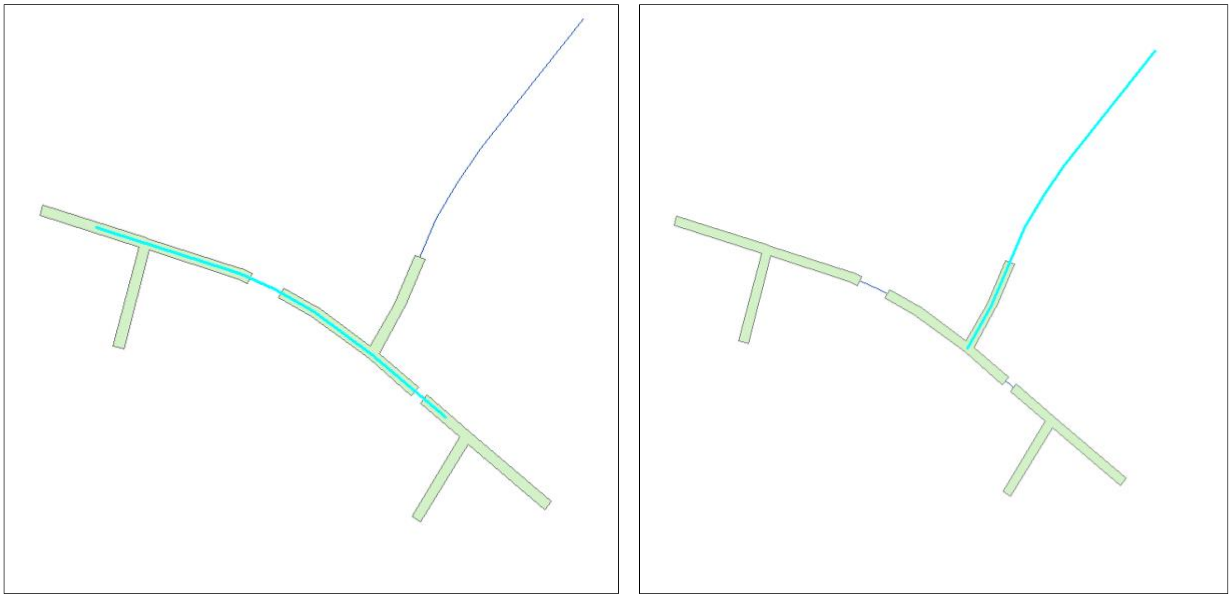


Figure 4 Two Segments Having a Common Intersection

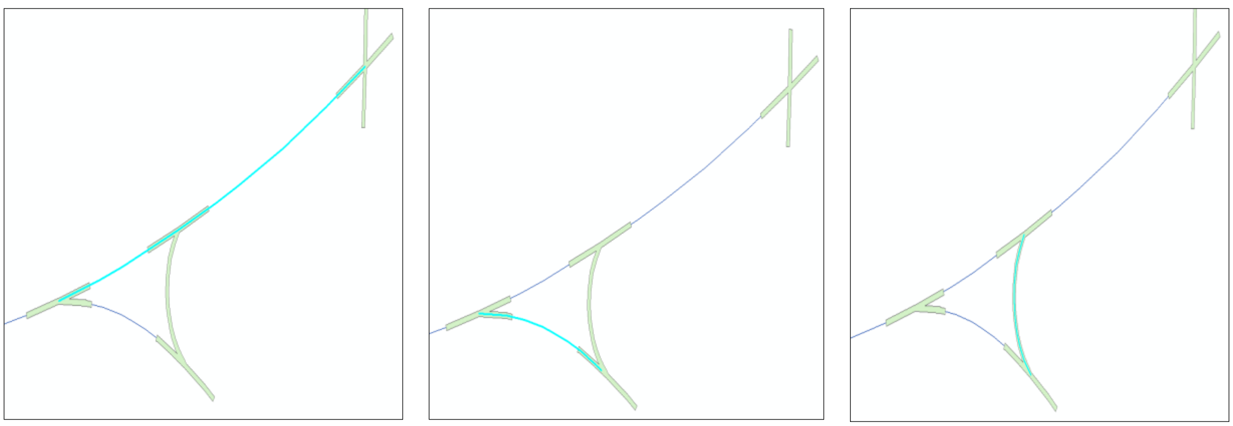


Figure 5 Three Segments Having Several Common Intersections

4. Results

The research team calculated the EEC for a total of 1286 projects that were slated for the SHIFT 2020 cycle. Among those projects, 590 had negative EEC values, meaning that fewer crashes are occurring than expected. The remaining 696 projects had positive EEC values, and thus were identified as having the potential for safety improvements. Table 3 presents the descriptive statistics of the EECs of all projects and also the projects with positive EECs.

Table 3 Descriptive Statistics of EEC

	All Projects	Projects with positive EEC
Count	1286	696
Mean	22.47	58.02
Standard Deviation	92.1	108.07
Minimum	-618.24	0.000048
25%	-7.07	3.90
Median	0.83	17.07
75%	20.19	59.3
Maximum	1123	1123

5. Conclusion

In this study, HSM methodologies were adopted into the current project prioritization process, with the end result of enhancing the decision-making processes. This report explains the drawbacks of the method used in SHIFT 2018 and describes the new methodology to rank the potential of SHIFT 2020 projects. After network screening and developing SPFs for all state-maintained roadways in Kentucky, EEC was estimated for all potential projects so they could be ranked appropriately. This process helps to select appropriate countermeasures, creates the ability to compare safety consequences among various alternatives, and allows identification of cost-effective strategies. In the course of this research, two tools were developed: an ArcGIS tool for network screening and a Python automation tool for estimating project EEC. The tools are user-friendly and customizable to a variety of potential uses.

References

[1] SHIFT | KYTC [Online]. Retrieved May 1, 2019, from https://transportation.ky.gov/SHIFT/Pages/default.aspx.

[2] Srinivasan, Raghavan, Daniel Carter, and Karin Bauer. *Safety Performance Function* 25 *Decision Guide: SPF Calibration vs SPF Development*. Publication FHWA-SA-14-004. 26 FHWA, U.S. Department of Transportation, 2013.

[3] AASHTO. Highway Safety Manual, first edition. *American Association of State Highway* 13 *and Transportation Officials*, Washington, DC, 2010.

[4] Blackden, Chris, Eric Green, Reginald Souleyrette, and William Staats. *Automating Safety Performance Function Development to Improve Regression Models*. No. 18-05202. 2018.

[5] Srinivasan, Raghavan and Karin Bauer. *Safety Performance Function Development Guide: Developing Jurisdiction-Specific SPFs.* U.S. Department of Transportation: Federal Highway Administration, 2013a.

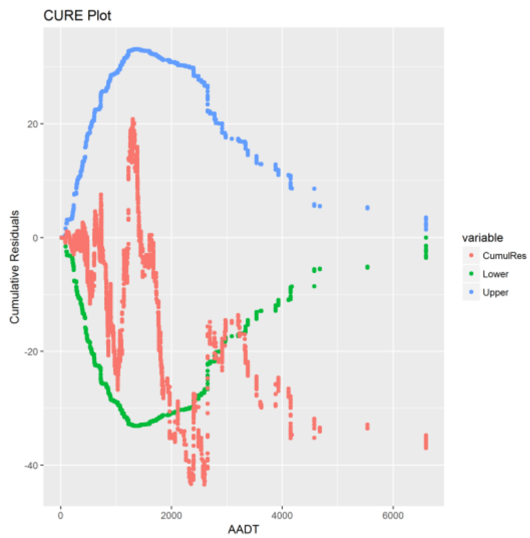
[6] Hauer, E., Harwood, D.W., Council, F.M. and Griffith, M.S., 2002. *Estimating safety by the empirical Bayes method: a tutorial*. Transportation Research Record, *1784*(1), pp.126-131.

[7] Blackden, Chris, Eric Green, Reginald Souleyrette, and William Staats. *Automating Safety Performance Function Development to Improve Regression Models*. No. 18-05202. 2018.

[8] Green, Eric R. *Segmentation Strategies for Road Safety Analysis*. 2018.

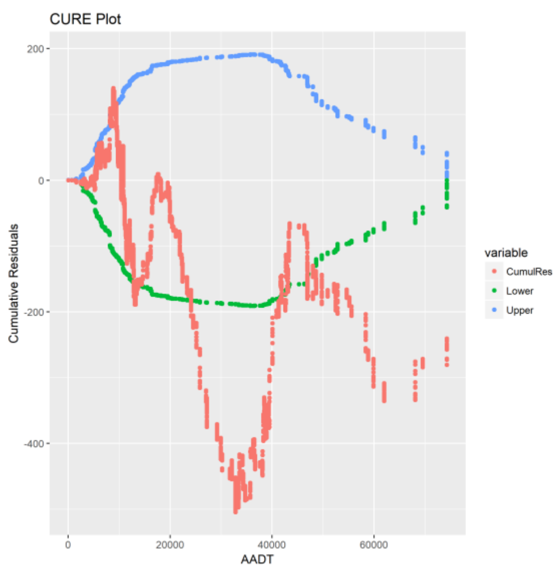
Appendix A CURE Plots of Safety Performance Functions

This is a set of eight Safety Performance Function CURE Plots, used as goodness-of-fit measures.

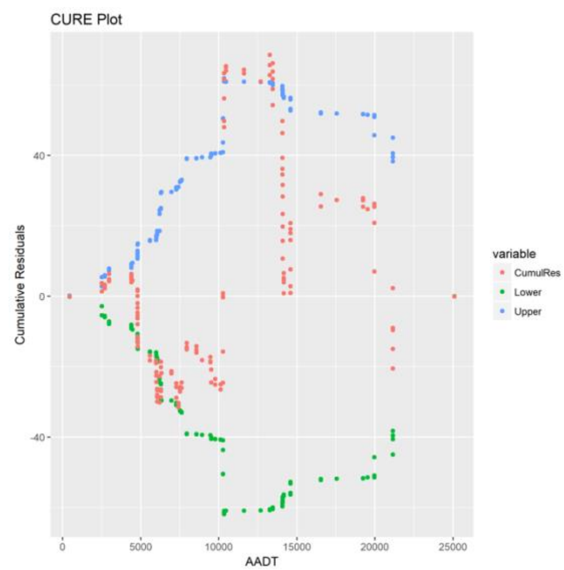


Rural Two Lane

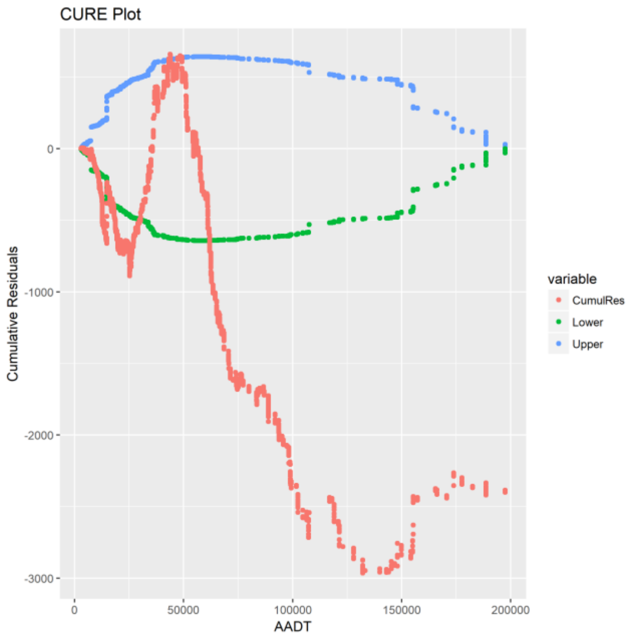
Urban Two Lane



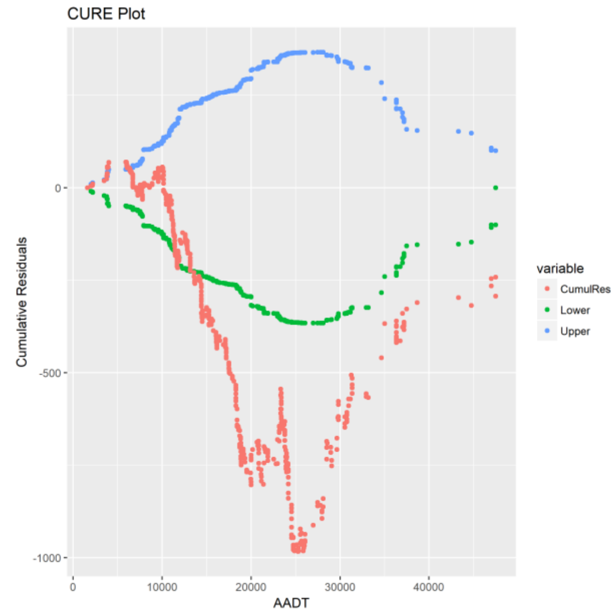
Rural Multilane (Divided)



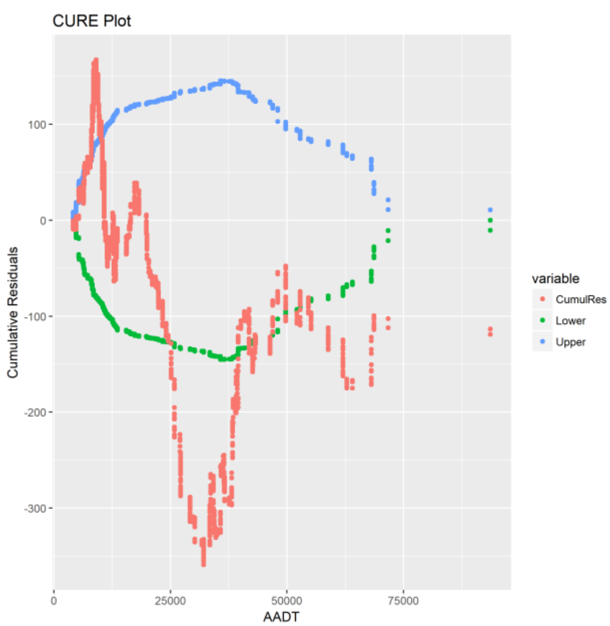
Rural Multilane (Undivided)



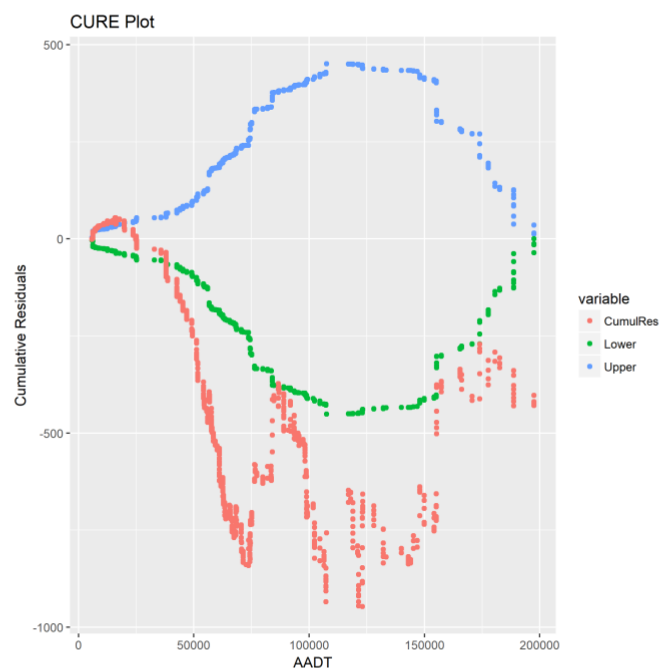
Urban Multilane (Divided)



Urban Multilane (Undivided)



Rural Interstates and Parkways



Urban Interstates and Parkways

Appendix B Regression Parameters and Over-Dispersion Parameters for SPFs

The following table represents the regression parameters for eight roadway types.

Roadway Type	a	b	θ
Rural two Lane	-4.492	0.844	1.532
Urban two Lane	-3.65	0.78	1.126
Rural Multilane (Divided)	-5.337	0.768	1.951
Rural Multilane (Undivided)	-6.962	1.045	0.649
Urban Multilane (Divided)	-4.171	0.761	0.814
Urban Multilane (Undivided)	-6.894	1.15	0.882
Rural Interstate and Parkways	-6.358	0.869	2.448
Urban Interstate and Parkways	-10.595	1.305	1.642