Freeway Segment Safety Performance Function (SPF) Development

Prepared by:

Scott Himes*, *Kendra Schenk*, *Ian Hamilton*, *and Frank Gross

Prepared for:

The Ohio Department of Transportation,

Office of Statewide Planning & Research

Project ID Number: *109462*

June 2021

Final Report

Technical Report Documentation Page

<table>
<tr><td>1. Report No.
FHWA/OH-2021-20</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2">4. Title and Subtitle
Freeway Segment Safety Performance Function (SPF) Development</td><td>5. Report Date
June 2021
6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s)
Scott Himes, Kendra Schenk, Ian Hamilton, and Frank Gross</td><td>8. Performing Organization Report No.</td></tr>
<tr><td colspan="2">9. Performing Organization Name and Address
VHB
Venture 1
940 Main Campus Drive, Suite 500
Raleigh, NC 27606-5217

Burgess & Niple, Inc.</td><td>10. Work Unit No. (TRAIS)
11. Contract or Grant No.
33808</td></tr>
<tr><td colspan="2">12. Sponsoring Agency Name and Address
Ohio Department of Transportation
1980 West Broad Street
Columbus, Ohio 43223</td><td>13. Type of Report and Period Covered
Final Report
14. Sponsoring Agency Code</td></tr>
<tr><td colspan="3">15. Supplementary Notes</td></tr>
<tr><td colspan="3">16. Abstract
This report documents the Ohio Department of Transportation's effort to improve the reliability of safety decision-making through the development of planning-level and project design-level safety performance functions (SPFs). Planning-level SPFs will replace calibrated versions of national SPFs in Safety Analyst to support network screening. Project design-level SPFs will replace previously calibrated versions of the Highway Safety Manual (HSM) Chapter 18 SPFs through estimation of Ohio-specific SPFs, adjustment factors (AFs), and a one-directional approach to freeway safety analysis. This research evaluated the performance of the calibrated HSM predictive method against an Ohio-specific bi-directional predictive method, Ohio bi-directional SPFs for base conditions used in combination with HSM AFs, and a one-direction predictive method. The results indicated the one-direction predictive method provided reliable predictions for all crash types and severities when compared to the baseline of the calibrated version of the HSM predictive method. The one-direction method is easier to implement on complex alignments, provides a broad set of AFs, but does exclude some factors that may be relevant to practitioners. Additional factors may be considered along with the one-direction predictive method, as needed, as external crash modification factors to the predictive method.</td></tr>
<tr><td colspan="2">17. Keywords</td><td>18. Distribution Statement
No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161</td></tr>
<tr><td>19. Security Classification (of this report)
Unclassified</td><td>20. Security Classification (of this page)
Unclassified</td><td>21. No. of Pages
153
22. Price</td></tr>
<tr><td colspan="2">Form DOT F 1700.7 (8-72)</td><td>Reproduction of completed pages authorized</td></tr>
</table>

Credits and Acknowledgments

Prepared in cooperation with the Ohio Department of Transportation
and the U.S. Department of Transportation, Federal Highway Administration

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Ohio Department of Transportation or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

The authors acknowledge the support and insights from Mr. Derek Troyer, Ms. Kathy King, Mr. Mike Cronebach, Mr. Mike McNeill, and Ms. Caraline Griffith from the Ohio Department of Transportation. Mr. Derek Troyer, Highway Safety Engineer with the Office of Program Management, provided supporting data and technical oversight. Ms. Kelly Nye and Ms. Vicky Fout served as the research liaison and provided contractual support on the project.

Executive Summary

High-quality data and reliable analytical methods are the foundation of data-driven decision-making. There are various data-driven safety analysis methods for identifying sites with promise and for predicting crash frequency for project design-level analysis. Using more reliable methods, agencies such as the Ohio Department of Transportation (ODOT) maximize the opportunity to reduce crashes or crash severity outcomes by identifying sites with the greatest potential for improvement and allocating resources to achieve the greatest return on investment.

Safety performance functions (SPFs) play a critical role in more reliable safety management methods. SPFs are mathematical equations that predict average crash frequency for a facility based on traffic volume, segment length, and other roadway characteristics. Planning-level SPFs typically include traffic volume and segment length for segments where traffic volume can be included in a variety of manners. Planning-level SPFs have been shown to be more reliable than historical crash counts alone for identifying sites with potential for safety improvement. Project design-level SPFs help to quantify and compare the safety performance of alternative geometric design and traffic operations characteristics.

ODOT engaged in developing Ohio-specific planning-level and project design-level SPFs to improve reliability for network screening and prediction of safety performance for project design alternatives on freeway segments. The planning-level SPFs consist of 10 SPFs for basic freeway segments, broken down by area type and number of bi-directional lanes, as well as 10 SPFs for interchange segments, also broken down by area type and number of bi-directional lanes. The definitions and functional forms used are consistent with and can be incorporated into Safety Analyst to support existing network screening processes.

The project design-level SPFs consisted of a series of base SPFs, applicable adjustment factors (AFs), and severity distribution functions. The project design-level SPFs will be incorporated into ODOT's Economic Crash Analysis Tool (ECAT) as a new one-direction approach to predicting safety performance. The one-direction predictive method replaces the existing predictive method from Chapter 18 of the HSM, including both the SPFs and AFs, formerly known as crash modification factors in the first edition of the Highway Safety Manual. The project documentation includes a Freeway Analysis User Guide, submitted as a standalone document separate from this report. The User Guide provides details needed to complete a one-direction analysis using the new methodology and definitions specifically needed to utilize the methodology which will be incorporated into ECAT.

Table of Contents

List of Figures

List of Tables

1. Problem Statement

High-quality data and reliable analytical methods are the foundation of data-driven decision-making. There are various data-driven safety analysis methods for identifying sites with promise and for predicting crash frequency for project design-level analysis. Using more reliable methods, agencies such as the Ohio Department of Transportation (ODOT) maximize the opportunity to reduce crashes or crash severity outcomes by identifying sites with the greatest potential for improvement and allocating resources to achieve the greatest return on investment.

Safety performance functions (SPFs) play a critical role in more reliable safety management methods. SPFs are mathematical equations that predict average crash frequency for a facility based on traffic volume, segment length, and other roadway characteristics. Planning-level SPFs typically include traffic volume and segment length for segments where traffic volume can be included in a variety of manners. Planning-level SPFs have been shown to be more reliable than historical crash counts alone for identifying sites with potential for safety improvement. Project design-level SPFs help to quantify and compare the safety performance of alternative geometric design and traffic operations characteristics.

Agencies have two options for obtaining SPFs. The first option is to use existing SPFs from another jurisdiction (e.g., national SPFs found in the Highway Safety Manual (HSM) and the American Association of State Highway and Transportation Officials' (AASHTOWare) Safety Analyst). These SPFs have been developed using data from one or more States or regions and need to be calibrated to work in Ohio. The second option is to develop new SPFs using jurisdiction-specific data. Comparatively, SPF calibration is cheaper and simpler, but it produces potentially less reliable results than SPF development. The decision to calibrate existing SPFs or to develop new SPFs generally starts with calibration. If the calibrated SPFs do not perform well, then the next step is to invest in the development of new SPFs. This has been the process in Ohio where ODOT has calibrated the SPFs from the HSM and determined the SPFs for freeways are not performing well enough after calibration.

ODOT engaged in developing Ohio-specific planning-level and project design-level SPFs to improve reliability for network screening and prediction of safety performance for project design alternatives on freeway segments. The planning-level SPFs will be incorporated into Safety Analyst to support existing network screening processes. The project design-level SPFs will be incorporated into ODOT's Economic Crash Analysis Tool (ECAT) as a new one-direction approach to predicting safety performance. The one-direction predictive method replaces the existing predictive method from Chapter 18 of the HSM, including both the SPF and adjustment factors (AFs), formerly known as crash modification factors (CMFs) in the first edition of the HSM.

2. Research Background

2.1 Objectives

The goal of this project is to develop and implement reliable SPFs for freeway segments. This is accomplished through the following objectives.

- Develop planning-level SPFs based on annual average daily traffic (AADT) and segment length for basic freeway segments and interchange segments. Planning-level SPFs are differentiated by number of lanes and area type.
- Develop bi-directional and directional project design-level SPFs based on AADT, segment length, and geometric and operational characteristics. The underlying methodology, SPF crash types and severities, and functional form are consistent with those found in Chapter 18 of the HSM.
- Evaluate whether SPFs for base conditions, in combination with HSM Chapter 18 AFs, provide a reliable fit relative to full models developed within this effort.
- Update ODOT's existing ECAT to incorporate the project design-level SPFs for predictive analysis. The planning-level SPFs are designed to be incorporated into Safety Analyst to support network screening.

2.2 Tasks

To accomplish the research objectives, the VHB team completed the following seven tasks:

- Task 1: Project Management. This included a project startup meeting and monthly status calls.
- Task 2: Literature Review. This included a review of State practices related to calibration and estimation of planning-level and project design-level SPFs.
- Task 3: Data Collection. This included collecting separate datasets for network screening analysis, bi-directional project design-level analysis, and one-directional project design-level analysis.
- Task 4: SPF Development. This included developing SPFs for each database consistent with the national models for each SPF type.
- Task 5: SPF Validation. This included obtaining validation sites for project design-level SPFs and assessing goodness-of-fit measures for planning-level and project-level SPFs.
- Task 6: Tool Development. This included providing a summary of updates necessary to implement the models in ECAT and developing a User Guide for freeway analysis using the revised methodology.
- Task 7: Final Report and Presentation. This included developing the research report and fact sheet.

2.3 Literature Review Findings

The VHB team reviewed ODOT's existing methods and processes for using and incorporating crash prediction models. The team also completed a literature review to identify guiding principles or protocols for developing SPFs, summarize similar efforts undertaken by other State Departments of Transportation (DOTs), identify statistical methods used for developing planning-level and project design-level SPFs, and determine potential geometric or operational characteristics known to impact freeway safety performance. Appendix A includes the details of the full literature review and the following is a summary of the important aspects.

According to data provided on the Federal Highway Administration's (FHWA's) CMF Clearinghouse website (2019), 19 States have developed calibration factors for HSM Part C predictive models or developed network screening-level SPFs. None of the agencies had developed SPFs and associated AFs to update the HSM Part C methodology related to freeways. Key agencies focusing efforts on freeway facilities included the following:

- Florida (Lu et al., 2014). Florida-specific models had superior fit to local data relative to calibrated models for planning-level SPFs. The calibrated models had a constrained impact of AADT; the Florida-specific models had new parameters allowing for adjustments to better fit the data.
- Illinois (Tegge et al., 2010). The researchers developed SPFs for fatal crashes, type-A injuries, type-B injuries, and fatal and injury crashes (all per mile per five-year period). However, the researchers did not differentiate between base and interchange segments.
- Missouri (Sun et al., 2013, 2018). The researchers calibrated the crash prediction models from the HSM Part C for all facility types using FHWA's Interactive Highway Safety Design Model. The researchers excluded segments near interchanges to remove the possibility of including crashes that did not actually occur on the freeway segment.
- North Carolina (Srinivasan and Carter, 2010). Researchers developed planning-level SPFs for freeway segments, including both interchange segments (defined as within 0.5 miles of an interchange) and non-interchange segments. The researchers developed SPFs for total crashes, fatal and injury crashes, fatal and serious injury crashes, property damage only crashes, lane departure crashes, single-vehicle crashes, multiple-vehicle crashes, wet pavement crashes, and nighttime-related crashes.
- Rhode Island (Himes and Le, unpublished). The researchers developed planning-level SPFs that were not differentiated between interchange and non-interchange segments (due to a lack of available information). The SPFs were based on directional segments and included total and fatal and injury crashes.
- South Carolina (Ogle and Rajabi, 2018). The researchers calibrated project design-level SPFs for rural and urban four-lane freeways and urban six-lane

freeways. Since ramp data were not available, segments within 0.5 miles of interchanges were excluded from the calibration procedure.

- Virginia (Kweon and Lim, 2014). The researchers developed planning-level SPFs for Virginia freeway and interchange segments. Interchange segments included the distance between gores and did not include speed-change lane areas or distances upstream or downstream of the gores. Virginia initially considered calibrating the freeway SPFs from Safety Analyst but found that the shape of the SPFs differed from the base models. The researchers also developed SPFs for directional segments, which were incompatible with Safety Analyst.

Based on the literature review, the VHB team discovered little research had been completed comparing the results between a one-direction approach to freeway safety prediction relative to a bi-directional approach, particularly for project design-level SPFs. However, as noted under each State, researchers have generally split freeway segments into interchange (or interchange-related) segments versus basic freeway segment for SPF implementation. All research used generalized linear models, utilizing the negative binomial structure to predict crash frequency on freeway segments, and consistent with the HSM, allowed for a variable dispersion parameter based on segment length.

3. Research Approach

The project team used two tracks to complete the objectives of this research:

1. Develop planning-level SPFs.
2. Develop project design-level SPFs.

Both approaches included data collection, SPF development through regression modeling, and SPF validation. The following subsections discuss the research approach for each track. Complete details on data collection and SPF estimation are provided in Appendix B for the planning-level SPFs and details on the validation procedures are provided in Appendix C. Complete details on data collection and SPF estimation are provided in Appendix D for project design-level SPFs and details on the validation procedures are provided in Appendix E. Prior to each research track, Section 3.1 provides a brief overview of the crash prediction modeling approach used throughout.

3.1 Crash Prediction Model Development Methodology

For crash frequency modeling, count models are traditionally used to quantify the relationship between crash frequency and traffic volumes, design elements, and traffic control features. Negative binomial regression has most commonly been applied to account for the overdispersion inherently found in crash data. The overdispersion parameter estimated from the modeling process is used in the development of the weight factor in the empirical Bayes (EB) analysis method.

Recently, researchers have applied more sophisticated versions of count models to account for temporal and spatial correlations. Depending on the assumption of the correlation between unobserved effects and right-hand side variables, fixed- and random-effects models have been applied to account for temporal and spatial correlations. These more sophisticated models help to reduce potential bias and inconsistency and improve the transferability of the model. However, more sophisticated models can be more time consuming to estimate, limiting the number of models that can be estimated during the modeling process. Additionally, more sophisticated models can prove to be more difficult to reach convergence with a larger number of predictor variables, limiting the number of geometric and operational features that can be included in the model specification.

The project team considered fixed effects, random effects, and mixed effects models to account for unobserved correlations inherent in the data. To balance the reliability and practicality of the models, the project team used negative binomial count models, which is consistent with the current state-of-the-practice.

The project team used a multinomial logit model to estimate severity distribution functions (SDFs) for each of the SPFs. The multinomial logit model allows for some variables to be constrained to have the same effect on each severity level while allowing other variables to have a variable effect among levels. For SDFs, the database was restructured such that the observed unit was the crash instead of the road segment. The SPF for total crashes can be combined with the SDF to estimate the number of crashes of different severity levels.

The following sections describe the results of the Planning-Level SPF Development and Project-Level SPF Development, respectively.

3.2 Research Approach for Planning-Level SPFs

The project team developed bi-directional planning-level SPFs for rural and urban freeways based on data for the entire freeway network. ODOT provided Safety Analyst files separately for roadway inventory, freeway segment volumes, ramp inventory, ramp volumes, and crash data. The roadway inventory data are considered to be a snapshot that remains constant through the study period; ODOT provided crash and traffic volume data for 2014 through 2018.

Table 1 provides an overview of the number of segments, total mileage, and crash sample size for freeway facilities by area type, number of lanes, and segment type.

Based on Table 1, the project team found that crash sample sizes for both total crashes and fatal and injury (FI) crashes were sufficient for estimating separate SPFs for each subtype (as defined by area type, number of lanes, and segment type). However, separate SPFs would not be estimable for every site subtype; some should be excluded, and some combined. Based on Table 1, the following site subtypes were recommended for SPF development:

- Freeway segments:
    - Rural four-lane segments.
    - Rural six-lane segments.
    - Urban four-lane segments.
    - Urban six-lane segments.
    - Urban eight or more-lane segments.
- Interchange segments:
    - Rural four-lane segments.
    - Rural six-lane segments.
    - Urban four-lane segments.
    - Urban five or six-lane segments.
    - Urban seven or more-lane segments.

The project team considered several functional forms for AADT. In each case, segment length was considered as an offset variable (i.e., the parameter for segment length is equal to 1.0).

The following functional forms were considered:

- Exponential Function = $Crashes = L \times e^{(\alpha + \beta_1 \times AADT + D_i)}$
- Power Function = $Crashes = L \times AADT^{\beta_1} \times e^{(\alpha + D_i)}$
- Hoerl Function = $Crashes = L \times AADT^{\beta_1} \times e^{(\alpha + \beta_2 \times AADT + D_i)}$
- Polynomial Function = $Crashes = L \times e^{(\alpha + \beta_1 \times AADT + \beta_2 \times AADT^2 + D_i)}$

Table 1. Overall Study Sample Sizes by Area Type, Number of Lanes, and Segment Type.

Area Type	Lanes	Segment Type	Number of Segments	Total Mileage	Total Crashes	FI Crashes
Rural	4	Base	293	592.44	16,403	3,359
		Interchange	331	103.05	4,112	829
	5	Base*	3	1.42	44	13
		Interchange*	2	0.47	27	2
	6	Base	74	210.25	11,435	2,284
		Interchange	92	21.84	2,056	422
	7	Base*	1	1.72	103	21
		Interchange*	0	0	0	0
Urban	2	Base*	1	0.17	13	4
		Interchange*	6	0.80	129	42
	3	Base*	0	0	0	0
		Interchange*	12	2.44	783	180
	4	Base	521	410.15	17,537	3,968
		Interchange	1,355	359.83	33,718	8.012
	5	Base*	15	9.43	424	90
		Interchange	90	18.08	3,618	865
	6	Base	259	215.41	17,982	4,230
		Interchange	840	255.82	56,218	13,652
	7	Base*	10	3.05	615	169
		Interchange	75	16.97	5,081	1,274
	8	Base	70	41.11	6,915	1,716
		Interchange	316	97.65	30,606	7,808
	9	Base	4	0.67	178	41
		Interchange	10	2.16	682	181
	10	Base	7	5.18	908	316
		Interchange	36	12.00	3,034	885

*Indicates segments were dropped from dataset due to small sample size.

Each equation includes a coefficient for an ODOT District (1 through 12). If the segment is located within that district, the coefficient is used within the prediction as shown. If no coefficient is provided for a specific district for any crash type, then that district is assumed to be a part of the baseline for that crash type. Factors are only applied when a value is present based on the District number.

Table 2 presents the estimated coefficients for all base segments and for urban 7+-lane interchange segments. For these facility types, the project team found the Hoerl

Model to best fit the planning-level data. For these site subtypes, only freeway AADT is included. For urban 7+-lane interchange segments, ramp AADT was not found to be associated with crash outcomes. District indicators are also included in some cases, providing for an increase or decrease in predicted crashes due to underlying factors specific to the district. For those indicators, the baseline is the district(s) not included in the SPF. The indicators do not change the shape of the SPF, but shift it up or down slightly, depending on the direction of effect.

For each SPF, the project team estimated the SPFs considering a variable dispersion parameter that is a function of the segment length. Hauer (2001) identified the possibility that dispersion can logically be related to the segment length; shorter segments will tend of have a higher dispersion than longer segments. Further research by Cafiso et al. (2010) used the logarithm of segment length in the log-dispersion model, resulting in the functional form shown in Equation 1.

$$Dispersion = Constant * L^{\beta_3} \qquad \text{Equation 1}$$

The value of the constant is the exponent of the coefficient from the model of the logarithm of dispersion. If the coefficient of the segment length is negative, the model shows that dispersion is inversely related to the length of the segment, which is consistent with the logic established by Hauer (2001).

Table 2. Hoerl Model SPFs for Base Freeway Segments and Urban 7+-Lane Interchange Segments.

Site Subtype	α	B_1	B_2*	D2	D3	D4	D5	D6	D7	D8	D12	Dispersion	
												Const	B_3
Base Freeway Segments													
Rural 4-Lane Total	-2.147	0.341	0.019	N/A	N/A	N/A	-0.128	-0.297	N/A	-0.262	N/A	0.091	-0.597
Rural 4-Lane FI	-2.777	0.198	0.036	N/A	N/A	N/A	N/A	-0.287	N/A	-0.301	N/A	0.135	-0.649
Rural 6-Lane Total	0.592	0.092	0.021	N/A	-0.220	-0.185	N/A	-0.319	-0.237	N/A	N/A	0.104	-0.906
Rural 6-Lane FI	18.772	-1.964	0.070	N/A	-0.143	N/A	N/A	-0.348	-0.421	N/A	N/A	0.140	-0.994
Urban 4-Lane Total	-1.475	0.282	0.021	-0.194	N/A	N/A	N/A	-0.452	N/A	N/A	-0.192	0.109	-0.819
Urban 4-Lane FI	-4.353	0.420	0.019	N/A	N/A	N/A	N/A	-0.219	N/A	N/A	N/A	0.154	-0.656
Urban 6-Lane Total	-1.134	0.280	0.014	N/A	N/A	N/A	N/A	-0.411	-0.151	N/A	-0.384	0.084	-0.826
Urban 6-Lane FI	-0.916	0.103	0.018	N/A	N/A	N/A	N/A	-0.218	N/A	N/A	N/A	0.107	-0.608
Urban 8+ Lane Total	30.573	-2.816	0.047	0.546	N/A	N/A	N/A	N/A	N/A	0.549	N/A	0.039	-1.129
Urban 8+ Lane FI	21.505	-2.082	0.041	0.424	N/A	N/A	N/A	N/A	N/A	0.339	N/A	0.051	-0.898
Interchange Freeway Segments													
Urban 7+ Lane Total	-14.771	1.659	-0.006	0.186	N/A	0.277	N/A	N/A	N/A	0.210	N/A	0.106	-0.673
Urban 7+ Lane FI	-17.919	1.830	-0.007	N/A	N/A	0.247	N/A	N/A	N/A	N/A	N/A	0.128	-0.442

Note: *Indicates in thousands AADT (AADT/1,000)

Table 3 presents the estimated coefficients for all interchange segments (other than urban 7+ lane segments). For these segments, the project team found the power function to have the best fit to the data, and the SPF includes both freeway segment AADT and the total AADT from all associated interchange ramps (i.e., the sum of entering and exiting vehicles in both travel directions). Equation 2 presents the final model functional form.

$$Crashes = L \times Freeway\ AADT^{\beta_4} \times Total\ Ramp\ AADT^{\beta_5} \times e^{(\alpha + D_i)} \qquad \text{Equation 2}$$

As with base segments, the project team included District indicators to capture unobserved adjustments and used a variable dispersion parameter that is a function of segment length.

Table 3. Power Model SPFs for Interchange Segments.

Site Subtype	α	B_4	B_5	D5	D7	D8	Dispersion	
							Const	B_3
Rural 4-Lane Total	-5.106	0.595	0.136	-0.288	-0.212	-0.171	0.052	-0.855
Rural 4-Lane FI	-6.970	0.565	0.196	N/A	-0.471	N/A	0.012	-2.105
Rural 6-Lane Total	-8.942	0.983	0.107	N/A	0.292	N/A	0.014	-1.592
Rural 6-Lane FI	-15.621	1.385	0.179	N/A	0.399	N/A	0.002	-2.169
Urban 4-Lane Total	-7.086	0.721	0.235	-0.173	N/A	N/A	0.109	-0.818
Urban 4-Lane FI	-8.815	0.752	0.230	-0.181	N/A	N/A	0.144	-0.710
Urban 5/6-Lane Total	-6.637	0.767	0.135	N/A	N/A	0.282	0.063	-0.870
Urban 5/6-Lane FI	-8.683	0.841	0.130	N/A	N/A	N/A	0.185	-0.342

For each SPF site subtype, the project team assessed the reliability of the SPF prediction through goodness-of-fit measures. Appendix C provides more details on the measures used and comparisons for each SPF functional form. In general, the models provide an adequate fit for planning-level crash predictions.

3.3 Research Approach for Project Design-Level SPFs.

The VHB project team also developed bi-directional and one-direction project design-level SPFs for rural and urban freeways based on a sample of sites for which the project team incorporated additional data elements. Figure 1 provides an overview of the freeway segments included in the project-level SPF development dataset, including the following freeways:

- Interstate 70.
- Interstate 71.
- Interstate 75.
- Interstate 77.
- Interstate 80.
- Interstate 90.
- State Route 2.
- State Route 11.

Figure 1. Study Sample Corridors for Project-Level SPFs.

ODOT provided supplemental crash data identifying if crashes occurred in the cardinal or non-cardinal direction of the freeway. The project team used this information to classify the freeway direction for developing directional project-level SPFs.

Additionally, ODOT provided horizontal curve data for select curves on the freeway network and a barrier inventory for supplementing median type (which included the presence of median barrier) and for inclusion of outside barrier. The project team further collected data identifying gore points and end of tapers for entrance and exit speed change lanes. The project team classified whether gore points were for speed change lanes, for a lane add or lane drop, and for whether the ramp is on the left or right. Figure 2 provides an example of how the project team marked gore and taper points in both directions of a freeway segment. The project team additionally used the gore point information to identify the distance from study segments to upstream entrance ramps and to downstream exit ramps. For bi-directional project-level SPFs, the project team used the extents of concurrent freeway segments (without speed-change lanes) to identify bi-directional freeway segments. The project team only estimated speed-change lane SPFs for directional segments (consistent with the approach currently included in the HSM).

Figure 2. Example Ramp Points Collected at Interchange for Both Directions.

Consistent with the first edition of the HSM, the project team evaluated four crash outcomes for bi-directional and one-direction freeway segment models:

- Fatal and injury multi-vehicle crashes (fi_mv). These crashes involve more than one vehicle and have at least one injury in the crash.
- Fatal and injury single vehicle crashes (fi_sv). These crashes involve no more than one vehicle and have at least one injury in the crash.
- Property damage only multi-vehicle crashes (pdo_mv). These crashes involve more than one vehicle and have no reported injuries.
- Property damage only single vehicle crashes (pdo_sv). These crashes involve no more than one vehicle and have no reported injuries.

The project team developed separate bi-directional and one-directional crash prediction models for freeway segments, entry speed-change lanes, and exit speed-change lanes. Equation 3 serves as the foundation for the freeway segment crash prediction model.

$$N_{tot,fs} = L \times \exp(b_0 \ln[AADT_{fr}]) \times (AF_1 \times \ldots \times AF_n) \qquad \text{Equation 3}$$

where:

$N_{y,tot,fs}$ =predicted average total crashes for freeway segment site y, crashes.

L = segment length.

b_0 =regression coefficient freeway AADT.

$AADT_{fr}$ = freeway AADT.

AF_i = adjustment factor for freeway geometric design element, or traffic control feature i.

While the crash prediction model predicts crash frequency (by combined severity categories and crash types), the project team developed SDFs to predict the proportion of K, A, B, C, and O crash severity categories. The probability of each severity category is predicted as a function of traffic volume, geometry, and other roadway characteristics. The proportion is multiplied by the predicted crash frequency to obtain an estimate of the crash frequency for the corresponding severity category.

For entry speed-change lanes and exit speed-change lanes, the foundational equation was modified to Equation 4.

$$N_{tot,sc} = L \times \exp(b_0 \ln[AADT_{fr}] + b_1 \ln[AADT_r]) \times (AF_1 \times \ldots \times AF_n) \qquad \text{Equation 4}$$

where:

$N_{y,tot,sc}$ = predicted average total crashes for speed-change lane site y, crashes.

b_1 =regression coefficient ramp AADT.

$AADT_r$ = ramp AADT.

One version of this equation was developed for entry speed-change lanes and a second version was developed for exit ramp speed-change lanes. This equation includes the entry or exit ramp AADT and its associated regression parameter.

Table 4 provides an overview of the data elements the project team collected and considered in project design-level SPFs. Appendix D provides details on the estimation of project design-level SPFs, including summary statistics for each variable, for the following SPFs:

- Bi-directional freeway segments, with separate SPFs for the following:
  - Multiple-vehicle fatal and injury crashes.
  - Multiple-vehicle property damage only crashes.
  - Single vehicle fatal and injury crashes.
  - Single vehicle property damage only crashes.
- One-direction freeway segments, with separate SPFs for the following:
  - Multiple-vehicle fatal and injury crashes.
  - Multiple-vehicle property damage only crashes.
  - Single vehicle fatal and injury crashes.
  - Single vehicle property damage only crashes.
- One-direction entrance speed-change lanes, with separate SPFs for the following:
  - Fatal and injury crashes.
  - Property damage only crashes.
- One-direction exit speed-change lanes, with separate SPFs for the following:
  - Fatal and injury crashes.
  - Property damage only crashes.

Table 4. Data Elements Collected for Project-Level SPF Development.

Variable	Definition
Routename	Provides the combined route type and route number for the segment
County	Provides the county in which the segment is located
District	Provides the ODOT district in which the segment is located
Areatype	Indicates whether the study segment is urban or rural
Thru_lanes	Indicates the number of thru lanes on study segment
Lane_width	Average lane width in feet
Median_type	Indicates the type of median (1: rigid barrier, 2: semi-rigid barrier, 3: flexible barrier, 4: raised median with curb, 5: depressed median, 6: flush paved median, 7: HOV lanes, 8: railroad or other rapid transit, 9: other divided)
Median_width	Median width in feet
Shoulder_out	Outside shoulder width in feet
Shoulder_in	Inside shoulder width in feet
Posted_speed	Posted speed limit in mph
Lane_add	Indicator for a lane add within the study segment
Lane_add_aadt	Ramp AADT for lane added by ramp
Lane_drop	Indicator for a lane drop within the study segment
Lane_drop_aadt	Ramp AADT for lane dropped by ramp
Weave_type	Indicates the type of weaving section (0: no weave, 1: Type A weave, 2: Type B weave, 3: Type C weave)
Freeway_AADT	Directional or bidirectional AADT in vehicles per day
Down_ex_length	Length to downstream exit in miles
Up_en_length	Length to upstream entrance in miles
Down_ex_AADT	Downstream exit AADT in vehicles per day
Up_en_AADT	Upstream entrance AADT in vehicles per day
Seg_length	Study segment length in miles
Length_ex	Segment length within 0.5 miles of downstream exit
Length_en	Segment length within 0.5 miles of upstream entrance
Barrier_in	Length of inside barrier in miles
Barrier_out	Length of outside barrier in miles
Lighting	Indicator for presence of highway or interchange lighting
SCL_type	Speed change lane type (1: entrance, 2: exit)
Ramp_AADT	Exit or entrance ramp AADT
FI	Number of fatal and injury crashes
PDO	Number of property damage only crashes
Fi_mv	Number of fatal and injury multivehicle crashes
Fi_sv	Number of fatal and injury single vehicle crashes
PDO_mv	Number of property damage only multivehicle crashes
PDO_sv	Number of property damage only single vehicle crashes

The project team collected data for a set of validation sites and for 28 sample project sites to compare predictions to observed crash frequency for the following scenarios:

- Application of the uncalibrated HSM predictive method.
- Application of the ODOT calibrated HSM predictive method.
- Application of the base bi-directional SPF with HSM AFs.
- Application of the bi-directional SPF with associated AFs.
- Application of the one-direction SPF with associated AFs.

For the sample project sites, the project team further compared the predicted and observed crash frequency by severity level (K, A, B, C, and O on the KABCO scale), to determine how well the predictive methods performed. Details on the validation procedure and individual results can be found in Appendix E. Overall, the validation results indicated similar predictive performance from the calibrated HSM predictive method and the one-direction predictive method. However, due to improvement in prediction reliability on freeway segments, and ease of implementation, the one-direction predictive method was selected for further application for predicted crash frequency on ODOT freeway segments. Table 5 provides an overview of the recommended SPFs for project design-level applications.

Table 5. One-Direction Project Design-Level SPF Coefficients

Variable Name	Freeway Segments				Entrance Speed-Change Lanes		Exit Speed-Change Lanes	
	MV FI	MV PDO	SV FI	SV PDO	FI	PDO	FI	PDO
Constant	-15.62	-15.87	-7.48	-3.52	-13.90	-8.18	-12.82	-9.00
Urban_area	0.284	0.296	0.206	0.135	N/A	0.258	N/A	N/A
Three_plus_lanes	N/A	-0.137	0.099	0.116	-0.151	N/A	N/A	N/A
Segment Length	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Freeway_AADT	1.565	1.656	0.710	0.347	1.452	0.983	1.359	1.137
Ramp_AADT	N/A	N/A	N/A	N/A	0.176	0.040	0.041	0.026
Left_ramp	N/A	N/A	N/A	N/A	0.605	0.996	0.822	0.732
Shoulder_in	-0.030	-0.015	-0.035	N/A	-0.032	-0.015	-0.032	-0.029
Shoulder_out	-0.026	N/A	N/A	N/A	N/A	N/A	N/A	-0.029
Depressed_med_wid	-0.002	-0.001	-0.003	-0.001	-0.005	-0.002	-0.004	-0.004
Prop_down_lane_change	0.001	0.002	N/A	N/A	N/A	N/A	N/A	N/A
Prop_up_lane_change	N/A	0.001	N/A	N/A	N/A	N/A	N/A	N/A
Lane_add_AADT	0.005	0.005	N/A	N/A	N/A	N/A	N/A	N/A
Lane_down_AADT	0.007	0.011	N/A	N/A	N/A	N/A	N/A	N/A
Weave_A (or B)	0.159	0.230	N/A	N/A	N/A	N/A	N/A	N/A
Ave_degree	0.082	0.139	0.120	0.162	N/A	0.149	N/A	0.234
Curve_prop	N/A	N/A	N/A	N/A	0.166	N/A	0.133	N/A
Median_barrier	N/A	0.046	-0.129	0.043	-0.116	N/A	N/A	N/A
Outside_barrier	N/A	N/A	-0.027	N/A	N/A	0.308	N/A	N/A
Posted_speed	N/A	N/A	N/A	0.012	N/A	N/A	N/A	N/A

4. Research Findings and Conclusions

4.1 Research Findings for Planning-Level SPFs

The planning-level SPFs include data from all freeways in Ohio and are implementable for predicting average total crash frequency or average fatal and injury crash frequency for bi-directional freeway segments. The SPFs provide an indication of the predicted average crash frequency for a freeway site subtype, based on the area type, presence of an interchange, and number of lanes by AADT and segment length.

The project team developed the planning-level SPFs for compatibility with Safety Analyst, which may be directly entered into the software program. In the case of Hoerl model SPFs, an additional factor will need to be added to the SPF, as the functional form differs from the assumed functional form in the software. Figure 3 provides an example of a Hoerl model SPF. The extra factor in the SPF allows the curve to bend but does not overfit the data, which may undermine the purpose of planning-level SPFs.

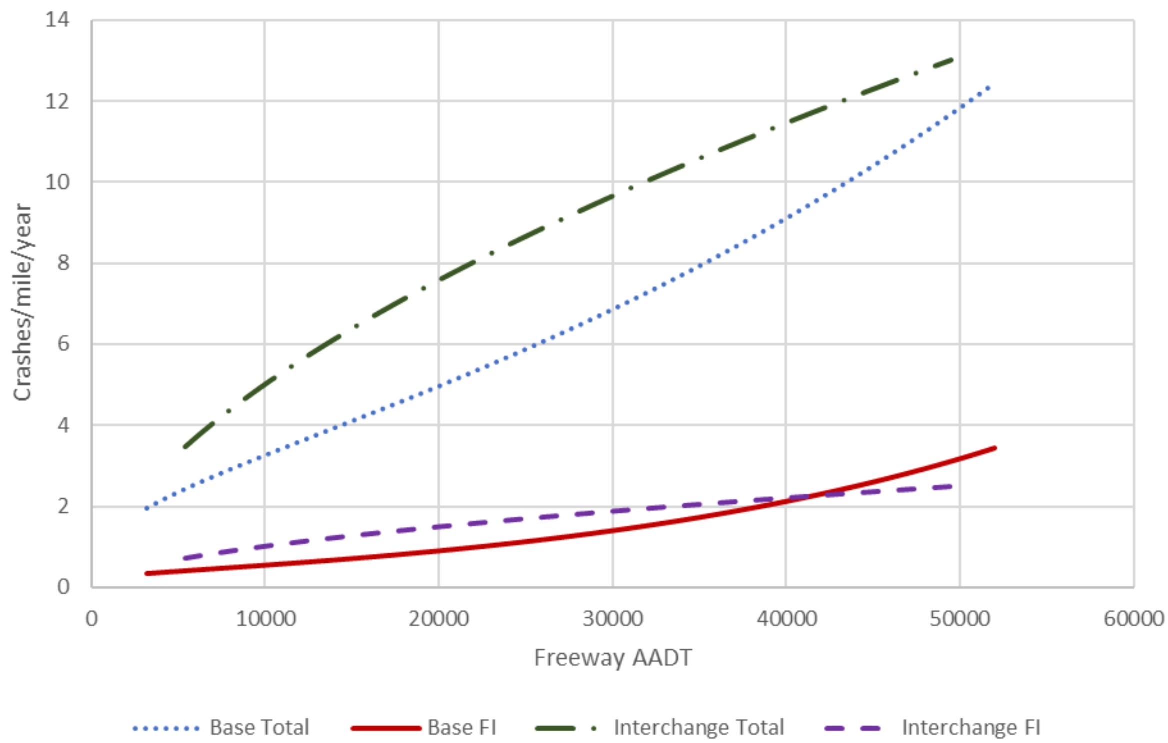

Figure 3. Hoerl Model SPF Implementation.

Additionally, as indicated in Tables 2 and 3, the SPFs include indicators for specific ODOT districts. The district indicators can be used in the planning-level SPF as an additive factor, where the factor takes a value of 1.0 if the segment is located within the district (and the coefficient is added to the constant value). Alternatively, ODOT can implement the SPFs at the District level, adjusting the value of the constant for each district as necessary. Finally, planning-level SPFs include variable dispersion parameters, for which the overdispersion parameter is calculated using Equation 1. The overdispersion parameter is used to calculate expected crash frequency using the EB adjustment for more reliable measures in the network screening process. Overall, these adjustments to the base SPF functional form and implementation will improve the reliability of network screening performance measures for determining sites with a potential for safety improvement.

4.2 Research Findings for Project Design-Level SPFs

The project team evaluated the merits of the following project design-level SPF implementation options:

- Existing bi-directional HSM Chapter 18 predictive method with ODOT-specific calibration factors.
- Bi-directional base SPFs developed for ODOT in this research in combination with HSM Chapter 18 AFs.
- Bi-directional predictive method developed within this research.
- One-direction predictive method developed within this research.

The validation procedure indicated the one-direction predictive method would be the most appropriate for project design-level analysis on Ohio freeway segments. Table 6 provides an overview of the available one-direction SPFs from the new predictive method. Table 7 provides an overview of the availability and applicability of AFs for SPFs. Combined, the SPFs and AFs can be used to predict crash frequency for freeway segments, entrance speed-change lanes, and exit speed-change lanes.

Table 6. Available One-Direction SPFs.

Site Type	Crash Type	Crash Severity	Area type	Lanes
Freeway segments (fs)	Multiple vehicle	Fatal and injury	Rural and urban	All lanes
		Property damage only	Rural and urban	2 lanes
			Rural and urban	3+ lanes
	Single vehicle	Fatal and injury	Rural and urban	2 lanes
			Rural and urban	3+ lanes
		Property damage only	Rural and urban	2 lanes
			Rural and urban	3+lanes
Entrance speed-change lanes (EN)	Combined	Fatal and injury	Combined	2 lanes
				3+ lanes
	Combined	Property damage only	Rural and urban	Combined
Exit speed-change lanes (EX)	Combined	Fatal and injury	Combined	Combined
	Combined	Property damage only	Combined	Combined

Table 7. Prediction Model AFs and Applicability.

Adjustment Factor	Freeway Segments				Entrance Speed-Change Lanes		Exit Speed-Change Lanes	
	MV FI	MV PDO	SV FI	SV PDO	FI	PDO	FI	PDO
Inside shoulder width	✓	✓	✓	N/A	✓	✓	✓	✓
Outside shoulder width	✓	N/A	N/A	N/A	N/A	N/A	N/A	✓
Depressed median width	✓	✓	✓	✓	✓	✓	✓	✓
Degree of curvature	✓	✓	✓	✓	N/A	✓	N/A	✓
Median barrier	N/A	✓	✓	✓	✓	N/A	N/A	N/A
Outside barrier	N/A	N/A	✓	N/A	N/A	✓	N/A	N/A
Downstream exit lane change	✓	✓	N/A	N/A	N/A	N/A	N/A	N/A
Upstream entrance lane change	N/A	✓	N/A	N/A	N/A	N/A	N/A	N/A
Lane addition by ramp	✓	✓	N/A	N/A	N/A	N/A	N/A	N/A
Lane drop by ramp	✓	✓	N/A	N/A	N/A	N/A	N/A	N/A
Type A/B weaving section	✓	✓	N/A	N/A	N/A	N/A	N/A	N/A
Posted speed limit	N/A	N/A	N/A	✓	N/A	N/A	N/A	N/A
Left-side ramp	N/A	N/A	N/A	N/A	✓	✓	✓	✓
Proportion curve	N/A	N/A	N/A	N/A	✓	N/A	✓	N/A

Table 8 provides an overview of factors associated with crash severity on freeway segments and speed-change lanes and whether an increase or presence of the factor is associated with an increase or decrease in that severity level.

Table 8. Factors Associated with Crash Severity Distribution.

Severity Factor	Freeway Segments		Speed-Change Lanes	
	KA Severity	B Severity	KA Severity	B Severity
Posted speed	Increase	Increase	Increase	Increase
Outside shoulder width	Increase	Increase	Decrease	N/A
Inside shoulder width	Decrease	N/A	N/A	N/A
Median width	Increase	Increase	N/A	Decrease
Proportion outside barrier	Decrease	N/A	N/A	N/A
Average degree of curve	N/A	Increase	Increase	N/A
Urban area type	Decrease	Decrease	Decrease	Decrease
Proportion median barrier	Increase	Increase	N/A	Decrease
Presence of lighting	Decrease	Decrease	N/A	N/A
Ramp AADT	N/A	N/A	Decrease	Decrease

The Freeway Safety Analysis User Guide is provided as a companion document for implementing the one-direction predictive method. Additionally, the project team has provided ODOT with a list of changes (and a sample spreadsheet) for incorporating the one-direction predictive method in the ECAT. The guide is intended to provide users with background on collecting necessary data inputs, from where measurements should be taken, and how the calculations are performed using the predictive method. The Guide includes the following details:

- Overview of guide. The guide overview includes an introduction and purpose of the guide, the target audience, and the overall structure of the guide.
- Overview of project design-level safety analysis. This section introduces the HSM Part C project design-level safety analysis method, the reasons and segmentation process for one-direction segmentation, and a comparison of the HSM freeway safety analysis method and the ODOT one-direction safety analysis method.
- Data needs and segmentation process. This section provides an overview of what data are necessary to implement the method and how/when segments should be subdivided for analysis.
- Predictive method for freeway segments and speed-change lanes. This section provides the necessary details to implement the applicable SPFs, AFs, and SDFs. Further this section provides additional information on calibration and guidance on the predictive method.
- Example project applications. This section provides four example problem applications for predicting crash frequency on a basic freeway segment with speed-change lanes, predicting crash frequency on a basic freeway segment with a Type A weave, using SDFs to predict crash frequency by severity level, and using the EB approach for calculating expected crash frequency.

5. Recommendations for Implementation

5.1 Recommendations

Based on the findings and conclusions of this research project, the project team presents the following recommendations:

- Recommendation 1. ODOT should incorporate the planning-level SPFs into Safety Analyst to increase the reliability of network screening measures for freeway facilities. Planning-level SPFs should be calibrated annually to account for yearly trends in average crash frequency and average fatal and injury crash frequency. Further, ODOT should employ cumulative residual (CURE) plots in the calibration process to assess the predictive reliability across the range of AADT and predicted crashes. If and when indicated by CURE plots, ODOT should consider re-estimating planning-level SPFs to support estimation of network screening performance measures.
- Recommendation 2. ODOT should consider a similar process for planning-level SPFs for other facility types as needed. Annual calibration is important to account for annual fluctuations in crash frequency and CURE plots provide a data-driven approach for determining when SPF calibration is insufficient.
- Recommendation 3. ODOT should incorporate the project design-level SPFs and AFs into the planned updates for ECAT. The one-direction approach will provide users with a simpler approach than directional analysis for freeway facilities, including freeway segments, entrance speed-change lanes, and exit speed-change lanes. The one-direction predictive method includes updated data on Ohio-specific freeway severity and crash type distributions.
- Recommendation 4. The one-direction predictive method should be periodically calibrated to account for annual trends in freeway safety performance. This includes calibration of both SPFs and SDFs.
- Recommendation 5. ODOT should consider further refining one-directional predictive models to include additional AFs not included in this research. This research focused on data directly available from existing ODOT databases and did not include supplemental data collection. Inclusion of supplemental data may improve the predictive model's reliability and further improve decision-making on ODOT projects, but this will come with an added cost for data collection and maintenance.

5.2 Implementation Plan

The following is a plan for implementing the research recommendations described in the previous section.

5.2.1 Recommendations for Implementation

- Recommendation 1. To implement Recommendation 1, the Safety Analyst administrator should use the admin tool to update the default SPFs to include an additional multiplicative term supporting implementation of the Hoerl functional form. Additionally, the administrator should update the constant term and coefficients to include the final values provided in this report.
- Recommendation 2. To implement Recommendation 2 (and the second part of Recommendation 1), the ODOT administrator can use Safety Analyst to directly calibrate the models using the most recent statewide crash data. However, to determine if calibration is sufficient, ODOT should consider utilizing tools and resources for developing CURE plots. FHWA's 'The Calibrator' tool can be used to calibrate SPFs and examine CURE plots and other SPF goodness-of-fit (GOF) measures. Additionally, The Kentucky Transportation Center has developed SPF-R, which can be used to develop and assess SPF GOF measures and CURE plots to support calibration and SPF development.
- Recommendation 3. To implement Recommendation 3, the project team has provided a sample mock-up of the one-direction predictive method in ECAT as well as documentation of SPF, AF, SDF, and default distributions for incorporation into future editions of the ECAT. The values for all aspects of predictive analysis are documented in the User Guide.
- Recommendation 4. To further implement Recommendation 3, ODOT should consider promoting the new predictive method to users and providing training on the revised method.
- Recommendation 5. To implement Recommendation 4, ODOT should follow a similar process to that for calibrating planning-level SPFs. ODOT should consider periodically using available tools to calibrate the project design-level SPFs. The Interactive Highway Safety Design Model (IHSDM) provides an automated calibration tool ODOT can use to provide annual calibration. However, to more thoroughly analyze the performance of calibrated models, FHWA's Calibrator or Kentucky's SPF-R can provide additional GOF measures, including CURE plots.
- Recommendation 6. To implement Recommendation 5, ODOT should use the project design-level SPF database as a starting point for future data collection. These locations can serve as an efficient starting point for expanding with supplemental data collection. Most data attributes can be collected virtually, potentially by interns or through contracting with a University to have students provide data collection at a low cost. Additional supplemental data collection may require use of permanent traffic count station data or through other data sources to determine congestion, or heavy volume, related data elements.

5.2.2 Analysis of Benefits and Risks

- Recommendation 1. The expected benefit of implementing Recommendation 1 is that conducting network screening using Ohio-specific freeway SPFs, with the most reliable functional form will improve the selection of sites for targeted safety improvement. Using more reliable measures has been shown to improve overall safety performance and provide for the most cost-effective implementation of safety programs.
- Recommendation 2. The expected benefit of implementing Recommendation 2 is similar to the benefit for recommendation 1. Annual calibration can be done with little cost and promotes estimating the most reliable performance measures for selecting sites for safety improvement. Periodic re-estimation of SPFs can have a higher cost but improves performance measure reliability. There is a risk associated with not recalibrating SPFs, as annual trends in total and fatal and injury crashes may result in segments of a certain type (or in a certain jurisdiction) having a higher expected crash frequency. For example, if a winter has a particularly harsh lake effect snowfall season, freeway segments in Districts 12 and 4 may have a higher crash frequency (thereby elevating expected crash frequency relative to predicted crash frequency for freeway segments in those particular districts).
- Recommendation 3. There are several expected benefits for implementation Recommendation 3. These include the following:
    - Using a one-direction predictive method reduces time and complexity for freeway segment analysis. The current methodology assumes freeway segments are symmetrical, which is often not the case. Analysts spend more time trying to determine how to troubleshoot these situations and often make mistakes in trying to overcome differences between opposing travel directions. Using a one-directional approach will result in fewer calculation errors and reduce analytical time.
    - The new one-direction predictive method requires fewer data elements to be collected and simplifies those calculations that remain in the methodology. As with the previous bullet, this will result in fewer mistakes (resulting in improved decision-making reliability) and a reduced level of effort for data collection and analysis.
    - Implementation in ECAT will help to automate the calculations when using the predictive method. Providing defined inputs and automating calculations will further help to reduce potential errors and time spent on calculations.
    - The reduction in errors, ease in application, and reduction in data elements will also help to reduce the level of effort and cost in reviewing predictive analyses.
- Recommendation 4. The benefits and risks for project design-level SPF calibration are consistent with those in Recommendation 2.

- Recommendation 5. The expected benefit of implementing Recommendation 5 serves to reduce potential risks associated with Recommendation 3. The expected benefit is to reduce the risk of the following:
    - Not accounting for features that may be changed that influence safety performance. The project team provided a method in Appendix F for overcoming this through inclusion of additional CMFs. However, the benefit from Recommendation 5 is that these factors could further be incorporated through Ohio-specific AFs.
    - Improving the ability for designers and analysts to account for potential safety impacts in decision-making. By including more design and operational features, decision-makers will better understand the benefits and risks of making changes to features that currently cannot be accounted for (e.g., clear zone width or median barrier offset).
    - Reducing omitted variable bias for AFs included in the current predictive method. By ignoring some factors for which data are not available, it is possible that their impacts are included in current AFs, overstating the true effect of those AFs.

5.2.3 Agency Coordination

Implementing both the planning-level and project design-level SPFs begins with the ODOT safety program for which the results will be disseminated to end users, including roadway engineering, and consultants. Implementation of the recommendations of this research study will take place mostly within the safety program, but coordination will be required for disseminating the necessary information and for providing training as necessary.

5.2.4 Estimated Costs and Time Frame

- Recommendation 1. The estimated cost of recommendation 1 is negligible. The cost for implementing recommendation 1 includes the time for the Safety Program Manager to update the coefficients and functional forms within the existing Safety Analyst program. The underlying data supporting implementation will not change.
- Recommendation 2. The cost for annual calibration in Safety Analyst is negligible. Based on the updated data, Safety Analyst can compute the annual calibration automatically. There is, however, a cost associated with evaluating SPF performance for determining the need to re-estimate versus calibrate. The software packages for estimating GOF measures (including CURE plots) are free and publicly available; therefore, there is no software related cost. The cost is associated with the time to input the data and run the analytical procedures within both software programs. The cost may range from the time for the Safety Program Manager to conduct the analysis to the cost of hiring a contractor to assess the need for SPF re-estimation. Overall, this cost would not be large, but may expand based the need to include the cost for re-estimation of SPFs by the contractor if dictated by the GOF measures during

calibration. This would be scaled by the number of facility types for which SPF estimation would be required.

- Recommendation 3. The estimated cost for recommendation 3 is unknown at this time. Since ODOT is currently in the process of updating ECAT, the cost to include the new one-direction methodology may be offset if this can be folded into the ongoing effort. However, as this is an existing spreadsheet implementation tool, the amount of programming necessary for implementation should be negligible. The VHB project team anticipates this would be a low-cost implementation in ECAT. The User Guide has been prepared as part of this research effort and would provide the information necessary for software updates. Additionally, the User Guide can serve as a basis for training materials, reducing the cost to develop training for implementing the new one-direction method. There would be some cost to provide training on the new methodology to ECAT users for freeway safety prediction.
- Recommendation 4. The estimated costs for implementing Recommendation 4 are similar to Recommendation 2. The level of cost would be dictated by the resulting need to re-estimate SPFs, SDFs, or crash type distributions periodically.
- Recommendation 5. The estimated cost for implementing Recommendation 5 would be high. Cost may be reduced through the use of interns to collect supplemental data for sample freeway segments provided from this research or through using University students to support data collection. However, this research would include collecting supplemental data (through virtual desktop collection and assessing traffic data as needed), integrating the data with the existing research database, estimating expanded one-direction SPFs, validating the new models, and updating ECAT and the associated User Guide.

Bibliography

American Association of State Highway and Transportation Officials (2010). *Highway Safety Manual, 1st Edition*. American Association of State Highway and Transportation Officials, Washington, D.C.

Anastasopoulos, P. and F. Mannering. (2009). "A note on modeling vehicle accident frequencies with random-parameters count models." *Accident Analysis and Prevention, 41*.

Avelar, R., K. Dixon, and S. Ashraf. (2018). "A Comparative Analysis on Performance of Severe Crash Prediction Methods." Presented at the 97th Annual Meeting of the Transportation Research Board, Washington, DC.

Babyak, M. (2004). "What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models." *Psychosomatic Medicine, 66*, pp. 411-421.

Bhat, C. (2003). "Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences." *Transportation Research Part B, 36*(6).

Bonneson, J., S. Geedipally, M. Pratt, and D. Lord. (2012). *Safety Prediction Methodology and Analysis Tool for Freeways and Interchanges: Final Report*. National Cooperative Highway Research Program, Transportation Research Board, Washington, DC.

Butsick, A. (2014). *Modeling Safety Effects of Geometric Design Consistency on Two-Lane Rural Roadways Using Mixed-Effects Negative Binomial Regression*. Master Thesis, Pennsylvania State University.

Carter, D., R., Srinivasan, F. Gross, and F. Council. (2012). *Recommended Protocols for Developing Crash Modification Factors*. National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C.

Cafiso, S., G. Silvestro, B. Persaud, and M. Begum. (2010). "Revisiting Variability in Dispersion of Safety Performance for Two-Lane Rural Roads." *Transportation Research Record 2148,* TRB, National Research Council, Washington, D.C.

Dixon, D., K. Fitzpatrick, R. Avelar, M. Perez, S. Ranft, R. Stevens, S. Venglar, and T. Voigt. (2015). *Reducing Lane and Shoulder Width to Permit an Additional Lane on a Freeway*. Technical Report (No. FHWA/TX-15/0-6811-1). Texas Department of Transportation, Austin, TX.

Duncan, C., A. Khattak, and F. Council. (1998). "Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions." *Transportation Research Record 1635*, Transportation Research Board, Washington, DC.

Gooch, J., V. Gayah, and E. Donnell. (2016). "Quantifying the safety effects of horizontal curves on two-way, two-lane rural roads." *Accident Analysis and Prevention*, *92*.

Graham, J., D. Harwood, K. Richard, M. O-Laughlin, E. Donnell, and S. Brennan. *NCHRP Report 794: Median Cross-Section Design for Rural Divided Highways*. Transportation Research Board, National Research Council, Washington, D.C., 2014.

Gross, F., B. Persaud, and C. Lyon. (2010). *A Guide to Developing Quality Crash Modification Factors*. Report No. FHWA-SA-10-032. Federal Highway Administration, Washington, D.C.

Gujarati, D. (2003). *Basic Econometrics: Fourth Edition*. McGraw-Hill/Irwin, New York.

Hauer, E. (2001). "Overdispersion in Modelling Accidents on Road Sections and in Empirical Bayes Estimation." *Accident Analysis and Prevention, 33*(6).

Hauer, E. (2006). "The Frequency-Severity Indeterminacy." *Accident Analysis and Prevention*. Vol. 38.

Hauer, E. (2015). *The Art of Regression Modeling in Road Safety*. Springer, New York.

Jovanis, P. and H. Chang. (1986). "Modeling the relationship of accident to miles traveled." *Transportation Research Record 1068*, Transportation Research Board, Washington, DC.

Kweon, Y. and I. Lim. (2014). *Development of Safety Performance Functions for Multilane Highway and Freeway Segments Maintained by the Virginia Department of Transportation*. Report No. FHWA/VCTIR 14-R14. Virginia Department of Transportation. Richmond, VA.

Le, T. and R. Porter. (2012). "Safety Evaluation of Geometric Design Criteria for Spacing of Entrance-Exit Ramp Sequence and Use of Auxiliary Lanes." *Transportation Research Record 2309*, Transportation Research Board, Washington, DC.

Lehmann, E.R. (1990). "Model Specification: the Views of Fisher and Neyman, and Later Developments." *Statistical Science*, *5*(2), pp. 160-168.

Lord, D. and F. Mannering. (2010). "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives." *Transportation Research Part A*, *44*(5).

Lu, J., K. Haleem, A. Gan, and P. Alluri. (2014). "Safety Performance Functions for Florida's Freeways." *Second Transportation and Development Congress*.

Mannering, F., V. Shankar, and C. Bhat. (2016). "Unobserved heterogeneity and the statistical analysis of highway accident data." *Analytical Methods in Accident Research*, *11*.

Milton, J., V. Shankar, and F. Mannering. (2008). "Highway accident severities and the mixed logit model: an exploratory empirical analysis." *Accident Analysis and Prevention*, *40*(1).

Ogle, J. and M. Rajabi.(2018). *Development of South Carolina Databases and Calibration Factors for the Highway Safety Manual: Final Report*. Report No. FHWA-SC-18-05. South Carolina Department of Transportation, Columbia, SC.

Ohio Department of Transportation Office of Roadway Engineering. (2019). *Location and Design Manual, Volume I*. Office of Roadway Engineering, Columbus, OH.

Ohio Department of Transportation Office of Program Management. (2018). *Safety Analysis Guidelines*. Office of Program Management, Columbus, OH.

Park, M., D. Lee, and J. Jeon. (2016). "Random Parameter Negative Binomial Model of Signalized Intersections." *Mathematical Problems in Engineering*.

Shankar, V., R. Albin, J. Milton, and F. Mannering. (1998). "Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model." *Transportation Research Record 1635*, Transportation Research Board, Washington, DC.

Shankar, V. and F. Mannering. (1996). "An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity." *Journal of Safety Research*, *27*(3).

Shankar, V., F. Mannering, and W. Barfield. (1996). "Statistical Analysis of Accident Severity on Rural Freeways." *Accident Analysis and Prevention*, *28*(3).

Shmueli, G. (2010). "To Explain or to Predict?" *Statistical Science*, *25*(3), pp. 289-310.

Srinivasan, R. and D. Carter. (2011). *Development of Safety Performance Functions for North Carolina*. Report No. FHWA/NC/2010-09, North Carolina Department of Transportation, Raleigh, NC.

Srinivasan, R., D. Carter, and K. Bauer. (2013a). *Safety Performance Function Decision Guide: SPF Calibration vs. SPF Development*. Report No. FHWA-SA-14-004. Federal Highway Administration, Washington, D.C.

Srinivasan, R., and K. Bauer. (2013b). *Safety Performance Function Development Guide: Developing Jurisdiction-Specific SPFs*. Report No. FHWA-SA-14-005. Federal Highway Administration, Washington, D.C.

Sun, C., P. Edara, H. Brown, B. Claros, and K. Nam. (2013). *Calibration of the Highway Safety Manual for Missouri*. Report No. cmr14-007. Missouri Department of Transportation, Jefferson City, MO.

Sun, C., P. Edara, H. Brown, J. Berry, B. Claros, and X. Yu. (2018). *Missouri Highway Safety Manual Recalibration*. Report No. cmr18-001. Missouri Department of Transportation, Jefferson City, MO.

Tegge, R., J. Jo, and Y. Ouyang. (2010). *Development and Application of Safety Performance Functions for Illinois*. FHWA-ICT-10-066. Illinois Department of Transportation, Springfield, IL.

Troyer, D., K. Bradbury, and C. Juliano. (2015). "Strength of the Variable: Calculating and Evaluating Safety Performance Function Calibration Factors for the State of Ohio." *Transportation Research Record 2515*, Transportation Research Board, Washington D.C.

Venkataraman, N. (2014). *Random Parameter Analysis of Geometric Effects on Freeway Crash Occurrence*. Ph.D. Disseratation, University of Iceland.

Wang, C., M. Quddus, and S. Ison. (2011). “Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model.” *Accident Analysis and Prevention, 43*.

Washington, S. M. Karlaftis, and F. Mannering. (2003). *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall/CRC.

Wood, J., E. Donnell, and R. Porter. (2015). “Comparison of safety effect estimates obtained from empirical Bayes before-after study, propensity scores-potential outcomes framework, and regression model with cross-sectional data.” *Accident Analysis and Prevention, 75*.

Appendix A: Literature Review

This appendix presents the results of a review ODOT's methods and processes for using and incorporating crash prediction models (CPMs), including both SPFs and AFs. It also provides a summary of related literature on applicable guiding principles or protocols for developing SPFs, similar efforts undertaken by other State DOTs, potential geometric or operational characteristics known to impact freeway safety performance, and statistical methods used for developing network screening-level and project development-level SPFs and associated SDFs. The results of this literature review informed the data elements collected, analytical methods used, and development of implementation tools and guidance for ODOT. This literature review is organized as follows:

- Review of ODOT's existing methods and processes. This section provides an overview of how ODOT uses SPFs in network screening, safety analysis, safety studies, performance-based project development, and the ECAT. Furthermore, this section provides a brief review of previous efforts to calibrate national models for use in Ohio.
- Review of similar efforts undertaken by other State agencies. This section provides an overview of similar development and calibration of State-specific SPFs for freeway facilities. Additionally, this section provides an overview of other States enhancing the HSM Part C implementation spreadsheets. Finally, this section provides a review of other tools used for conducting the roadway safety management process and for project-level design analysis.
- Review of applicable guiding principles or protocols for developing SPFs. This section provides a review of guiding principles and protocols for developing network screening level and project-design level CPMs based on the most current research. This section references protocols for developing HSM-level models as well as sample size requirements necessary for consistent CPM development.
- Review of Statistical Methods for SPF Development. This section provides a review of the statistical analysis considerations and techniques for developing CPMs. Potential biases and confounders are discussed and appropriate methods for overcoming them are provided. Crash frequency and severity distribution modeling techniques are presented separately.
- Review of research on safety impacts of geometric or operational characteristics on freeway facilities. This section briefly reviews freeway-related crash-based studies conducted since National Cooperative Highway Research Program (NCHRP) Project 17-45.

Review of ODOT's Existing Methods and Processes

Network Screening

Annually, ODOT uses AASHTOWare's Safety Analyst to prioritize locations for safety study or review (ODOT Office of Program Management, 2018). ODOT uses the system to identify segments and spot locations that have higher-than-predicted crash frequencies (including injury and fatality frequencies) based on the methods in the HSM Part B. The results are annually illustrated on Safety Integrated Project (SIP) Maps for each county in Ohio and includes both the local and State systems. Segments or spot locations with the highway priority (i.e., have the most potential for safety improvement (PSI)) are indicated in red (rural segments that have 5 or more PSI for all crashes and urban locations that have 5 or more PSI for fatal and injury crashes). Segments or spot locations that have a crash problem, but do not meet the above thresholds are indicated in blue. These segment or spot locations should have lower-cost countermeasures explored. High priority locations may be eligible for supplemental project funding by the ODOT Safety Program to implement "reasonable and practical countermeasures (ODOT Office of Program Management 2018)."

Safety Analysis

The Location and Design (L&D) Manual, Volume 1 recommends that a minimum data-driven safety assessment (DDSA) should be performed in the early phases of project development, allowing schedule, scope, and budget considerations to be accounted for when reasonable and practical countermeasures are to be included in the project (ODOT Office of Roadway Engineering 2019). The Safety Analysis Guidelines indicate that a minimum safety assessment should include the following steps:

- Obtain applicable studies for the project area.
- Determine if the location is on the ODOT SIP map.
- Determine the ranking on ODOT or local safety priority list or within Local Road Safety Plan.
- Analyze historical/observed crash data.

DDSA is applicable to all ODOT Let projects, except for maintenance, pavement surface treatment, spot repair, and slot paving projects. Further, while Local Let projects are exempt from performing DDSA, analysis is strongly recommended to understand the impacts of the project on crash frequency and severity (ODOT Office of Roadway Engineering 2019). A non-complex project assessment should be conducted for projects that do not require an alternatives analysis as part of the project development process. The primary purpose of assessing safety in this case is to identify any cost-effective safety improvements that should be included in the project to mitigate a current safety problem, and/or assist with the development of design exception justifications. This type of analysis considers whether crash type percentages are above statewide averages, whether safety countermeasures can be included in the current project, and if the location is on State or local priority lists.

The second analysis type is complex projects assessment with alternative analysis without "safety" in the purpose and need statement. Safety impacts are considered along with other impacts, but safety funding is not requested. This analysis considers whether the alternatives will result in the use of a different SPF from the existing conditions to determine whether predicted or expected crashes should be evaluated. The results of the safety analysis should be considered along other metrics; however, the preferred alternative cannot result in significantly more crashes than the existing conditions or an increase in fatal or serious injury crash frequencies.

The final analysis type is complex project assessments with alternatives analysis and a safety component. Since safety is a component of the purpose and need, this assessment includes additional steps to verify a safety benefit is being achieved. This method requires that both the predicted and expected crash frequencies as well as the PSI for the existing conditions should be calculated. This analysis then considers whether the alternatives will result in the use of a different SPF from the existing conditions to determine whether predicted or expected crashes should be evaluated. The analyst should then determine if at least one alternative reduces crashes or crash severity. This step is repeated until at least one alternative improves safety. Further, this assessment considers whether safety funds are being sought, and if so, the benefit cost ratio is obtained to verify a value above 1.0.

The L&D Manual also notes that safety should be considered and evaluated for every project, but there is no need to include safety countermeasures for projects without safety included in the purpose and need. Projects should be evaluated to determine if there is a reasonable and practical countermeasure or countermeasures that can be incorporated into the project without expanding the scope. Decisions should be documented on the appropriate DDSA documentation form.

Safety Studies

Highway safety studies provide the basis for good decision making and facilitate the timely implementation of necessary improvements for ODOT's dedicated annual $102 million for engineering improvements at high-crash or severe crash locations (ODOT Office of Program Management 2018). ODOT's safety study process consists of the following steps:

1. Collect data and diagnose crash patterns.
2. Identify potential for site safety improvements and possible countermeasures.
3. Perform relevant traffic studies.
4. Evaluate countermeasures.
5. Develop plan and finalize report.

Steps 2 through 4 are repeated as necessary to facilitate the identification and evaluation of countermeasures that best address the safety needs of the site. Step 1 requires data to be collected for consistent use with the HSM and ECAT. Step 2 uses the ECAT to calculate the predicted crash frequency for peer sites and the expected crash frequency for the actual site (including historical crash data). The difference between the predicted and expected crashes is the potential for site safety improvement. Based on the results, the predicted and actual crash history should be reviewed to identify potential safety countermeasures. The ECAT tool incorporates Part D countermeasures to help with this process. This process can be completed for formal safety studies and abbreviated safety studies.

Performance Based Project Development

ODOT has adopted the performance-based project development (PBPD) philosophy where proposed improvements should be targeted and right-sized based on project specific needs (ODOT Office of Roadway Engineering 2019). PBPD can happen throughout the project development process, including planning, preliminary engineering, and design (more specific details can be found in the L&D Manual. The L&D Manual emphasizes that the level of effort should be commensurate with the complexity of the situation, and that the HSM and its predictive performance methods should be used to quantify safety impacts. However, the results of HSM analyses should not normally be the sole basis of making decisions. The ECAT should be used, when appropriate, to compare expected crashes between alternatives. It is also of note that a PBPD alternative can accept an increase in crashes on proposed alternatives; however, the magnitude and severities of increases should be further considered. If increased crash frequency or severity is expected on an alternative, users should consider further mitigation with application of appropriate safety countermeasures.

Previous Efforts to Calibrate

In 2014, ODOT collected data, aggregated historic observed crashes, and used specialized spreadsheet tools to calibrate the HSM Part C predictive models (Troyer et al. 2015). ODOT noted that calibration factors may not be sufficient in some cases and would not perform as well as agency-specific SPFs. However, using CURE plots, ODOT found that calibrations were sufficient, and prioritized developing Ohio-specific SPFs for future efforts. Table 9 provides an overview of the calibration factors rural two-lane highways, rural multilane highways, and urban and suburban arterials segments and intersections. The calibration factors in Table 9 indicate that the uncalibrated models in the HSM underpredict crashes on rural highways and over-predict crashes on urban highways. In all cases, the uncalibrated models underpredict crashes at intersections. The calibration factors are largest for urban and suburban signalized intersections (over 3.0 in both cases). Note that the calibration factors are for total crashes. Separate calibration factors were developed for fatal and injury and for property damage only crashes. All calibration factors are provided in the Ohio HSM SPFs Calibration Factors spreadsheet, and several of these are updated since those provided in Table 9 were developed. The calibration factors for freeway segments generally indicate higher calibration factors when fewer lanes are present and decrease as the number of lanes increases. Many ramp terminal types were not analyzed and do not have Ohio-specific calibration factors.

Table 9. Ohio 2014 Calibration Factors.

Segment Type	Calibration Factor	Intersection Type	Calibration Factor
R2U	1.20	R23ST	1.51
RMD	1.31	R24ST	1.50
RMU	1.61	R24SG	1.86
U2U	1.02	RM3ST	1.66
U3T	0.45	RM4ST	1.73
U4U	0.24	RM4SG	1.33
U4D	0.79	U3ST	1.34
U5T	0.36	U3SG	3.35
N/A		U4ST	1.60
		U4SG	3.71

ECAT

ODOT enhanced the NCHRP 17-38 implementation spreadsheets based on Ohio-specific needs. ODOT combined the three spreadsheets, plus the enhanced version of the Interchange Safety Analysis Tool (ISATe), to create a single spreadsheet application tool for completing the calculations. ODOT's tool, the ECAT combines the ability to perform predictive and EB analysis along with the ability to conduct alternatives analysis and complete a benefit-cost analysis. The ECAT incorporates the ODOT-specific calibration factors provided above and crash type distributions specific to ODOT. Additionally, the ECAT incorporates HSM Part D CMFs (up to ten may be identified but only four can be applied for a single element).

Crash data are entered using the CAM tool and the toolbox can be used to automatically assign crashes to segments and intersections based on information provided in the project elements description table. If the tool is unable to assign a crash to a project element it requires users to assign the crash to an element. Once the crashes are assigned, the user creates analysis sheets for each element for the study site based on what was entered in the project information worksheet. Data for existing and proposed alternatives may be entered on the same worksheet, with input boxes provided on different parts of the worksheet (each with boxes labeling existing versus proposed). The verify analysis tool is used to determine if all data have been entered correctly. The project summary report provides a summary of anticipated safety performance for existing conditions predicted average crash frequency, expected average crash frequency, potential for safety improvement, and proposed conditions expected average crash frequency. The tool also provides an observed crash report, a design exception report, or can complete a benefit cost analysis. The benefit cost tool allows the user to provide a service life, initial cost, annual maintenance and energy costs, and salvage value for countermeasures selected or input.

Other countermeasures are added to this table automatically when specified in the data entry worksheets.

Review of Similar Efforts Undertaken by Other State DOTs

Development and Calibration of State-Specific SPFs for Freeways

The FHWA's CMF Clearinghouse (www.cmfclearinghouse.org) provides a summary of research developing State-specific SPFs and/or calibration factors. While the website is not comprehensive to date, it served as a starting reference point for identifying applicable research. As of June 30, 2016, 19 States had developed calibration factors for Part C predictive models or developed network screening-level SPFs. None of the research summarized developed SPFs for CMFs for revision or update to the HSM Part C predictive method. However, while 19 States had completed research, only 5 of those States had developed calibration factors or network screening SPFs. A summary of these efforts is provided in alphabetical order.

Florida

Lu et al. (2014) developed SPFs for comparison to the existing calibrated SPFs used for network screening. The authors used four years of data to develop SPFs for within and outside of interchange areas for rural and urban segments. Interchange areas were defined consistently with the SafetyAnalyst definition (extending 0.3 miles upstream and 0.3 miles downstream of gore points for the interchange). The modeling results indicated that the Florida-specific models had superior fit to the data relative to the calibrated models. Calibrated models are constrained to have the same functional slopes as the base models, whereas new parameters allow for adjustments to better fit the data. The researchers compared the fits using the Freeman-Tukey R-square, the mean absolute deviation (MAD), and the mean square prediction error.

Illinois

Tegge et al. (2010) developed network screening SPFs for application in Illinois. The researchers developed SPFs consistent with the facility type definitions and functional forms used in SafetyAnalyst to replace the base values. The researchers developed a Visual Basic for Applications software tool for the Illinois DOT to conduct future updates to SPFs. They developed SPFs separately for fatal crashes, type-A injuries, type-B injuries, and fatal and injury crashes (all per mile per five-year period). The researchers developed separate models for rural four and six-plus lane freeways and urban four, six, and eight-plus lane freeways. They did not differentiate between base and interchange segments. The authors developed the SPFs assuming the negative binomial distribution.

Missouri

Sun et al. (2013) calibrated the CPMs from the HSM Part C predictive method for all applicable facility types in Missouri. Sun et al. (2018) recalibrated the CPMs for all applicable facility types in 2018. In both cases, the researchers used FHWA's IHSDM to calibrate the HSM models included in Chapter 18 of the HSM Supplement. In 2013, the researchers identified a random sample of freeway segments from the roadway inventory for calibration. In 2018, the researchers used the same segments, with updated data. In both cases, segments near interchange areas were excluded to remove the possibility of including crashes that did not actually occur on the freeway segment. Calibration factors were developed separately for each SPF included in the predictive method for rural four-lane and urban four and six lane freeway segments. Calibration factors were separated by multi-versus single-vehicle crashes and fatal and injury versus property damage only crashes. In 2018, the property damage only crashes were generally greater than 1.0 for rural and urban freeways and were between 0.5 and 0.85 for fatal and injury crashes on all four-lane freeways. The calibration factors for fatal and injury crashes were nearly 1.0 for six lane freeways. These findings were consistent with the calibration factors developed in 2013.

North Carolina

Srinivasan and Carter (2010) developed network screening-level SPFs for North Carolina freeway segments. Separate SPFs were developed by area type, number of lanes, and location in proximity to interchange areas (consistent with SafetyAnalyst). Rural segments were evaluated as four-lane and six-plus lanes. Urban segments were analyzed as four-lane, six-lane, and eight-plus lanes. North Carolina defined segments with 0.5 miles of an interchange as interchange segments. Separate SPFs were estimated for total crashes, fatal and injury crashes, fatal and serious injury crashes, property damage only crashes, lane departure crashes, single-vehicle crashes, multi-vehicle crashes, wet pavement crashes, and nighttime-related crashes.

Rhode Island

Himes and Le (unpublished) developed network screening-level SPFs for Rhode Island. Due to the unique character of Rhode Island, the SPFs were not developed separately by area type, area type was included in the model specifications due to smaller samples of rural freeways. Additionally, there was no data on base versus interchange segments; therefore, all segments were combined for model development. However, segments were entered into the database directionally, meaning that SPFs developed were for directional segments. The researchers developed SPFs separately for four-lane freeways and six to eight lane freeways. Separate SPFs were developed for total crashes and for fatal and injury crashes.

South Carolina

Ogle and Rajabi (2018) recently calibrated the project-level SPFs for South Carolina using three years of data for freeway facilities. The researchers developed calibration factors for rural four-lane freeways (2.59), urban four-lane freeways (2.69), and urban six-lane freeways (3.66). Since ramp data were not available, segments within 0.5 miles of interchanges were excluded from the calibration procedure.

Virginia

Kweon and Lim (2014) developed network screening-level SPFs for Virginia freeway segments. Separate SPFs were developed by area type, number of lanes, and location in proximity to interchange areas. Rural segments were evaluated as four-lane or six-plus lanes. Urban segments were evaluated as four-lane, six-lane, and eight-plus lanes. Virginia initially considered calibrating the freeway SPFs from Safety Analyst but found that the shape of the SPFs differed from the base models. Virginia DOT (VDOT) developed separate SPFs for total crashes and fatal and injury crashes. Moreover, VDOT developed SPFs for directional segments rather than bi-directional segments. The research report noted this incompatibility with Safety Analyst.

States Enhancing Part C Spreadsheets

While the IHSDM directly implements the HSM Part C predictive method for project-level evaluation, there have several spreadsheets developed to assist with the implementation of the HSM methods directly. As part of NCHRP Project 17-38, the researchers developed implementation spreadsheets for rural, two-lane highways, rural multilane highways, and urban and suburban arterials. Further, as part of NCHRP Project 17-45, the researchers developed an enhanced version of the ISATe. Since the development of the base spreadsheets, other agencies have been updating or revising the NCHRP 17-38 spreadsheets to fit their needs. To this point, there have not been any updates/revisions to the ISATe.

For the most part, agencies have been directly using the NCHRP 17-38 spreadsheets as they currently exist or have made minor modifications to update the base calibration factors and crash severity and type distributions based on agency-specific data. Massachusetts and Louisiana serve as two example States that have updated the NCHRP 17-38 spreadsheets with their own data distributions and calibration factors.

The Alabama DOT and VDOT pooled funds to develop extended versions of the NCHRP 17-38 spreadsheets. The pooled funds were used to automate the spreadsheets, allow for automated report creation, and eliminate the need for user manipulation of the Site Total worksheet to perform the site-specific EB method. The Washington State DOT has since provided updated versions of the extended spreadsheets. These revised versions have more fully automated the worksheets through the use of macros and have expanded their capability to allow for multi-year analysis assuming linear growth rates for traffic volumes.

The Illinois DOT has combined the spreadsheets into one workbook with a front-end Graphic User Interface (GUI). The Illinois-specific workbooks include Illinois-specific calibration factors and default crash type and severity distributions. Illinois has also developed a separate benefit/cost tool that uses GUIs to allow for data entry for economic analysis. The predictive method; however, is not directly incorporated into this tool. This tool computes checks to ensure AADT and other data fit within HSM parameters and an extensive dropdown menu is provided to allow users to select countermeasures (some based on the predictive method) for economic analysis. Crash data used for economic analysis must be entered by hand in whole numbers. The CMFs from the selected countermeasures are used to calculate the safety benefit.

The Pennsylvania DOT has also adapted the NCHRP 17-38 spreadsheets to have a GUI and to exist as two separate tools (Tool A and Tool B). Tool A focuses on developing a base or existing alignment into the predictive method. The user can also select to use models for a county, district, or from the HSM. The spreadsheets are applicable for rural, two-lane highways, rural, multilane highways, and urban and suburban arterials.

Once the data are entered, the tool automatically develops a printable report on predicted and expected crash frequencies for each of the facility types. Multiple facility types and segments can be analyzed at once.

Tool B is known as the alternative analysis tool. In this tool, the existing alignment can be imported and reviewed. The user can make changes to the alignment for analysis of multiple alternatives. Additionally, the alternative analysis tool allows for other countermeasures (up to three) to be selected (from an existing menu) or generated.

Finally, once the alternatives have been entered, the user can enter project lifecycle cost information. Note that separate entry forms are provided for predictive method-based costs and additional countermeasure costs. The tool automatically provides the boxes that need data based on changes from the existing alignment to the alternative alignment being analyzed. Once the data are entered, the spreadsheet automatically generates a report summarizing the predicted crash performance for each of the alternatives (up to three) as well as the economic performance summary.

Review of Related Tools and their Capabilities

There are several tools that have been developed to help implement Parts B (roadway safety management) and C (predictive method) of the HSM. The purpose of safety management tools is to assist with planning-level safety analyses including those in the HSM Part B. Safety management tools that most States use are focused on automating and improving the delivery of the Highway Safety Improvement Program (HSIP). Most support hotspot projects, some support systemic projects; however, very few tools include features for all steps of the roadway safety management process. Often, States use multiple tools to support their HSIP. The following provides a brief overview of some of the tools used for roadway safety management. The purpose of predictive method tools are to assist in the application of predictive methods for project design-level decision-making.

AASHTOWare Safety Analyst

AASHTOWare Safety Analyst is a suite of tools to be used by State and local highway agencies for the implementation of the HSM safety management process. It requires upfront work to structure and import safety data but can then be used as a management tool for housing data, prioritizing project work, maintaining records of updated conditions, and evaluating countermeasure safety effectiveness. Ultimately, AASHTOWare Safety Analyst™ supports the decision-making process through state-of-the-art safety management approaches. This tool can help users to:

- Improve the programming of site-specific highway safety improvements.
- Develop and evaluate a roadway safety program of individual safety projects.
- Apply predictive methods to estimate crash frequency and severity.
- Estimate the expected effectiveness of infrastructure countermeasures using CMFs.

Features:

- Can export results for visualization.
- Crash analysis (other than network screening).
- Data entry/editing.
- All steps of Safety Management Process (Network Screening, Diagnosis, Countermeasure Selection, Economic Appraisal, Project Prioritization, Safety Effectiveness Evaluation).
- SPF functional form is flexible; however, SPFs must have a multiplicative, exponential form consisting of one or more terms of the following forms:
    - C – constant term.
    - e^{C} – expontential term with a constant exponent.
    - e^{CV} – exponential term with a variable exponent.
    - V^{C} – variable power term.

Benefits:

- Rigorous and comprehensive analysis capabilities, many other tools use this software as their basis for safety analysis.
- Implements nearly everything in HSM 1st Edition Part B.
- Developers are working on an API that may fix interface concerns in the longer term.
- Largest user base of advanced safety management tools.
- Includes an extract, transform, and load (ETL) tool to help import data.

- AASHTO has a lower risk of discontinuing support for their products than many vendors.
- Reports can be exported to geographic information system (GIS) fairly easily (require simple edits in excel) for visualization on top of the linear referencing system.
- Options for user management or easy login without credentials.

Drawbacks/Limitations:

- User interface not fully streamlined or user friendly. However, much of the graphical design is to accommodate accessibility to disabled and partially-blind persons.
- The documentation is not easily consumable (i.e., very long and difficult to understand) and limited training is readily available beyond in-person training.
- Could be viewed as data intensive compared to other tools. However, the minimum data requirements do not exceed those required to use SPFs in network screening. The software is flexible to omit sites from without enough data, but it can be cumbersome to fix erroneous data.
- Requires knowledge of databases, data transfers, and ETL to support tool and data importing.
- Training and standard analysis procedures are recommended to guide most users.
- Relatively steep learning curve and requires knowledge of HSM fundamentals.

AgileAssets Safety Analyst

AgileAssets Safety Analyst is developed by AgileAssets, which was originally developed using the core of the AASHTOWare Safety Analyst tool. It is a browser-based safety management system that integrates a GIS and linear referencing system (LRS). The tool stores safety-related data, allows for visualization of traffic and crash data, integrates GIS/LRS for network screening and analysis, applies user-defined decision trees to determine the most effective treatments, applies HSM methods (including EB), optimizes the application of limited resources, and provides reporting tools such as the HSIP reporting requirements. This tool integrates the HSM methods with crash data visualization reports, including tabular, graphical, and GIS map-based formats. The system can be configured to meet agency requirements and customized to existing data formats. This software requires crash, roadway, and traffic volume data and can help users to:

- Improve the programming of site-specific highway safety improvements.
- Develop a roadway safety program or individual safety projects.
- Apply predictive methods to estimate crash frequency and severity.
- Screen the network for sites with potential for safety improvement.

- Diagnose site-specific and system-wide safety issues.
- Select potential countermeasures to address specific safety issues.
- Estimate the cost-effectiveness of potential countermeasures.
- Prioritize projects within an overall safety program.

Features:

- Visualization capabilities.
- Crash analysis.
- All steps of Safety Management Process (Network Screening, Diagnosis, Countermeasure Selection, Economic Appraisal, Project Prioritization, Safety Effectiveness Evaluation).

Benefits:

- Rigorous and comprehensive analysis capabilities; can be customized to existing data formats; compatible with HSM.
- The tool was built on the algorithms from initial versions of AASHTOWare Safety Analyst with some development and a new interface.

Drawbacks/Limitations:

- Unsure about drawbacks—more research is needed.

FHWA GIS Safety Analysis Tools v4.0

The user can evaluate crashes at designated spots or intersections, along specific roadway segments or strips, clustered around a specific roadway feature, or within a defined corridor. All the programs allow the user to produce results for all crashes or for a given subset of crashes that can be defined using any of the variables contained in the crash and roadway inventory files. The software requires ESRI ArcGIS software suite. Data requirements include roadway data, crash data, and traffic data to perform crash analyses. The user can also perform additional ad hoc queries using the analysis and mapping tools in ArcView. Custom reports can be generated using the Crystal Reports engine bundled with ArcGIS.

Features:

- Visualization capabilities.
- Crash analysis.
    - Spot/Intersection Analysis.
    - Strip Analysis.
    - Cluster Analysis.
    - Sliding-Scale Analysis.
    - Corridor Analysis.
- Only applies to Network Screening.

Benefits:

- Develop summary statistics of crashes (e.g., number of crashes by injury severity and crash cost) associated with spot locations, corridors, or a given roadway feature (e.g., bridge, railroad crossing, or traffic signal).
- Develop crash maps.
- Identify high-crash concentrations within a corridor.
- Screen the network for high-crash locations based on visual inspection of crash maps.
- Screen the network for high-crash locations based on crash frequency and the sliding scale method.
- Generate reports of summary data.

Drawbacks/Limitations:

- The tool does not account for the non-linear relationship between crash frequency and traffic volume. The EB and full Bayes methods are not included in the tool, so it does not account for regression-to-the-mean in network screening.
- This tool may not be updated in the future as the use of GIS among transportation agencies has grown since this tool was first developed. The version of this tool may lose compatibility with newer computers and software. For example, this tool will not work with ESRI ArcGIS Version 10, which is the latest version of the software.

Numetric

Numetric is built by transportation agency engineers. “Numetric will connect to your crash and asset databases, pull in all data, make data formatting consistent, and ensure all future data stays consistent.” The software provides crash data query, ETL, user management, and safety management functionality through economic analysis phase.

Features:

- Crash report narrative is searchable
- Incorporates agency-specific SPF models.
- Queries can be exported as a .csv file
- Data is housed online (“the cloud”) after being pulled from DOT databases, so there is no hardware/software to install. This also means when new data appears on DOT’s databases, it will appear on Numetric as well.
- On the fly visualization capabilities built into the software.
- Network screening is available, but capabilities are unclear since it is proprietary.
- Some other steps of Safety Management Process may be available or supported but need further research.
- Numetric can be integrated with ESRI Roads and Highways.

Benefits:

- Scalable in analysis size (network to project level).
- Customizable based on available data.
- Wide range of analysis capabilities from network screening to economic analysis, including basic query.

Drawbacks/Limitations:

- While Numetric covers economic analysis, it does not have any capability to assist in prioritizing proposed projects or evaluating the effectiveness of proposed projects.
- The network screening tools use methods and measures of performance inconsistent with the HSM.
- Potentially limited analysis capabilities to calculate new data or run new reports on the fly (except simply querying data).

United States Road Assessment Program (usRAP) / ViDA Software

usRAP has three major functions, (1) risk mapping, (2) performance tracking, and (3) star ratings. Risk maps are used to document the risk of fatal and serious injury crashes and show where risk is high and low. Five color-coded risk levels are used for the development of maps for crash density, crash rate, crash rate ratio, and potential crash savings. The categories include highest risk (5 percent of system), medium-high risk (10 percent of system), medium risk (20 percent of system), medium-low risk (25 percent of system), and lowest risk (40 percent of system). The maps are presented based on crash type (e.g., roadway departure crashes). Star ratings are based on inspection of roads to examine how well they protect users from crashes and from deaths and serious injuries when crashes occur. Road protection scores (RPSs) are used to derive star ratings to identify differences in road design or management which are likely to lead to different probabilities of fatal or serious crashes and RPSs can be determined relatively quickly, with low cost. Roadway data can be used without detailed crash data to estimate star ratings. Performance tracking compares the safety performance of highways over time and relates those changes to ongoing safety improvement programs.

Features:

- Visualization capabilities.
- Some steps of Safety Management Process (Network Screening, Economic Appraisal, Safety Effectiveness Evaluation).
- Detailed crash data not needed for analysis.

Benefits:

- GIS based with visualization tools. Can conduct network screening on a larger scale.

Drawbacks/Limitations:

- Less rigorous analysis capabilities, risk mapping only available for fatal and severe injury crashes.

SPF Tool

The SPF Tool is web-based software that uses SPFs to assist with managing transportation systems. The SPF Tool provides network screening, diagnostic, economic analysis, and effectiveness evaluation tools with data visualization and reporting capabilities. The Tool manages and maintains SPFs and data and assesses how well CMFs fit local conditions. The SPF Tool web-based platform also allows for security through user management tools.

Features:

- Visualization capabilities.
- Some steps of Safety Management Process (Network Screening, Economic Appraisal, Safety Effectiveness Evaluation).
- SPF and CMF management.

Benefits:

- The software permits flexible jurisdiction-specific SPF format entry and provides support for existing SPF databases.
- The safety effectiveness evaluation allows users to create jurisdiction-specific CMFs using EB before-after analyses.

Drawbacks/Limitations:

- The tool is proprietary, and therefore, a black box.
- Small company, potentially limited support and software not yet widely proven by other agencies.

Vision Zero Suite (VZS)

VZS is a suite of analytical tools developed by DiExSys LLC. It is designed to provide decision support analysis for solving road safety problems at the system, corridor, and project levels. VZS is designed to be customized to meet the needs of individual agencies. The capabilities of VZS include 1) manage crash and roadway data; 2) conduct network screening to identify sites with potential for safety improvement; 3) diagnose safety issues at specific locations; 4) select appropriate countermeasures based on the safety issues identified in diagnosis; 5) estimate the benefit-cost ratio for a given project, and 6) prioritize treatments at a given location.

Features:

- Visualization capabilities.
- Data entry/editing.
- Most steps of Safety Management Process (Network Screening, Diagnosis, Countermeasure Selection, Economic Appraisal, Project Prioritization).

Benefits:

- Scalable in analysis size (network to project level).
- Customizable based on available data.
- Wide range of analysis capabilities.

Drawbacks/Limitations:

- This is a propriety tool, unsure of exact safety analysis capabilities or methodologies.

FHWA Highway Safety Benefit-Cost Analysis (BCA) Tool

This tool supports the implementation of the methods described and demonstrated in the Highway Safety BCA Guide. Specifically, the Tool provides a method for preparing a simple economic analysis of infrastructure projects, helping users to quantify projects costs as well as direct and indirect safety-related benefits of project alternatives. Direct safety benefits include the expected change in crash frequency and severity. Indirect benefits include the operational and environmental benefits that result from a reduction in crashes (i.e., reduced travel time, improved travel time reliability, reduced fuel use, and reduced emissions). The Tool is intended for project-level analysis of single or multiple improvements at a given location. It can also support network-level economic analysis for projects that include multiple locations (e.g., systemic improvements).

Features:

- Economic Appraisal.

Benefits:

- Manage economic analysis data.
- Develop alternative strategies for improving and managing highway facilities.
- Evaluate and compare the benefits and costs of alternative strategies.
- Determine the optimal timing of projects.
- Develop reports based on the analysis results.
- Support systemic improvements through network-level economic analysis.

Drawbacks/Limitations:

- This is an online tool that requires Internet Explorer Version 6 or newer.

Other States' Custom Tools

Connecticut

The Connecticut DOT is working with the University of Connecticut—through the Connecticut Transportation Safety Research Center—to develop a safety management tool specific to Connecticut. The Tool is still under development but will have corresponding modules for each step in the Safety Management Process as outlined in the HSM. The Tool will also use analysis methods from the HSM and is heavily based on the capabilities of AASHTOWare Safety Analyst, with some workflow and interface improvements. The Tool will also feature a map-based component to visualize results of network screening to identify sites with potential for safety improvement. SPFs were developed specifically for this tool; it has the flexibility for the functional form to change as required.

Alabama

Critical Analysis Reporting Environment (CARE) is data analysis software package developed by University of Alabama. CARE analyzes collision, exposure, vehicle, driver/population, and road infrastructure data to identify safety issues in a road network. Also, it has applications to formulate emphasis areas, objectives, and strategies to address the identified causal and contributing factors. Furthermore, it has applicability in prioritizing specific locations for treatments targeted to the emphasis area and estimating the number of collisions that could be reduced for a given emphasis area. The network screening capabilities are based on crash frequency and severity and does not account for the non-linear relationship between crashes and traffic volume. The EB and Full Bayes methods are not included. The software includes the following functions: narrative searching, hotspot determination, GIS integration, Open Road (street view). CARE can be downloaded free of charge.

Iowa

The Crash Mapping Analysis Tool (CMAT) identifies common vehicular collision sites, diagnoses the safety issues at these sites, and provides an evaluation of potential safety improvement. CMAT was developed by the Center for Transportation Research and Education at Iowa State University under the direction of the Iowa DOT's office of Traffic and Safety. CMAT can be downloaded free of charge.

The Safety, Analysis, Visualization and Exploration Resource (SAVER) is a GIS-based software that allows queries of collision data, an in-depth analysis of collision sites, and an analysis of the causes/contributing factors to the corresponding incidents. SAVER also reads supplemental report data, including traffic citations, crime incidents, operating while intoxicated from the existing National Model/TraCS software.

New York State

New York State DOT is in the early stages of developing the Crash Location Engineering and Analysis Repository (CLEAR) software. CLEAR will replace three existing systems (ALIS, SIMS and PIES) used to manage and analyze crash data. The CLEAR platform will provide New York State DOT with a spatially enabled, web-based crash geocoding, analysis and management solution that has been deployed within a secure cloud infrastructure using Esri ArcGIS technology. The solution will include a Crash Geocoding Engine to determine the location of crashes from information on crash reports using a composite, multi-tiered geocoding algorithm; the Interactive Crash Editor allowing users to interactively geocode and edit crashes; an Automated Crash Geocoder will perform geocoding of all new crashes nightly; a CLEAR Safety application which will provide the widely accepted six-step safety management process outlined in the AASHTO HSM; a CLEAR Data Viewer for visualization, query and analysis of crash and safety data, an Intersection Inventory Manager; and a Mobile Site Investigation application allowing Safety Engineers to visualize crashes and perform site analysis from the field in a map-centric application.

Massachusetts

The Interactive Mapping Portal for Analysis and Crash Tracking (IMPACT) is in the final stages of development. IMPACT provides MassDOT with a spatially enabled, web-based crash geocoding, analysis and management solution that has been deployed within a secure cloud infrastructure using Esri ArcGIS technology. The solution includes an Interactive Crash Locator (ICL) allowing users to process crash records for interactive geocoding and crash location; a robust Geocoding Engine which provides a multi-tiered algorithm to determine the location of crashes from locational information captured from crash reports received from the Registry of Motor Vehicles; a Crash Data Portal for users to explore, analyze, and download the crash data using interactive dashboards, a sophisticated crash data query and visualization tool, a robust reporting module based on SQL Server Reporting Services; an Administrator

Tool allowing the system to be highly configurable; and a Screening Engine to incorporate SPFs for collectors and arterials.

There are relatively few options for project-level, design-related tools. Most States use the excel spreadsheets or IHSDM depending on the type of analysis they're doing and the stage of project development that the project is in.

HSM Part C Spreadsheets

NCHRP Project 17-38 developed implementation spreadsheets and NCHRP Project 17-45 developed the Enhanced ISATe and User's Manual to help new users understand how to apply the predictive method included in Part C of the HSM (including the supplemental chapters). The spreadsheets implement the crash prediction procedure for rural two-lane two-way roads (HSM Chapter 10), rural multilane highways (HSM Chapter 11), and urban and suburban arterials (HSM Chapter 12). The ISATe can be used to evaluate freeway and interchange safety. Several States (discussed above) have modified the defaults values to fit their needs.

Features:

- Implements Part C predictive method of the HSM.

Benefits:

- Analyzes the safety performance of multiple alternatives.
- Considers impacts of geometric features associated with CMFs.
- Spreadsheets were developed and vetted by researchers that developed the analysis methods, or by researchers involved in developing training tools for those methods.

Drawbacks/Limitations:

- Cannot easily integrate analysis results from segments and intersections or of segments and intersections of various facility types.
- Somewhat cumbersome to coordinate analysis for complex projects.
- Does not include a visualization component.
- SPF and CMF formatting must be maintained when developing agency-specific SPFs or CMFs.
- Base spreadsheets do not allow additional countermeasures to be considered along with predictive method.

IHSDM

IHSDM is a suite of software analysis tools for evaluating safety and operational effects of geometric design decisions on highways. It is a decision-support tool for individual locations and is not intended to be used as a safety management tool for an entire network. Intended users include highway project managers, designers, and traffic and safety reviewers in State and local highway agencies and engineering consulting firms.

IHSDM currently includes six evaluation modules:

- Crash prediction module:
    - Estimates the expected frequency of crashes on a highway using geometric design and traffic characteristics. IHSDM implements the predictive methods in the HSM.
- Design consistency module:
    - Estimates the magnitude of potential speed inconsistencies to help identify and diagnose safety issues at horizontal curves of existing highways or proposed designs.
- Intersection review module:
    - Performs a diagnostic review to systematically evaluate an intersection design for typical safety concerns.
- Policy review module:
    - Checks highway segment design elements for compliance with relevant highway geometric design policy at several stages during the highway design process.
- Traffic analysis module:
    - Estimates operational quality-of-service measures for an existing or proposed design under current or projected future traffic flows.
- Driver/Vehicle module:
    - Estimates a driver's speed and path along a highway and corresponding measures of vehicle dynamics.

Features:

- Implements Part C predictive method of the HSM.

Benefits:

- Analyzes the safety performance of multiple alternatives.
- Considers impacts of geometric features associated with CMFs.
- Allows additional features to be included in predictive analyses.
- Allows multiple facility types to be analyzed together.
- Updated to include the latest predictive methods approved for inclusion in the HSM (even before the next edition is printed).
- Spreadsheet data entry format available for simple analyses.
- Includes a simplistic visualization component.
- Developers are working on an API that may fix interface concerns long-term.

Drawbacks/Limitations:

- The Crash Prediction Module is the only module that works on most types of roads, others are extremely limited or only apply to rural two-lane roads.
- There is a learning curve for entering data and conducting analyses for the first time.
- If station-based data entry used, requires building multiple highways and developing connections to analyze intersections and ramp terminals.
- SPF and CMF formatting must be maintained when developing agency-specific SPFs or CMFs.

Summary

Most tools implementing the roadway safety management process only perform a subset of the overall process, not all steps. Moreover, off-the-shelf products tend to follow the same methods as the original version of AASHTOWare's Safety Analyst, meaning that SPF functional form is constrained. However, this is not the case with the most recent version of Safety Analyst. All of the tools that implement the Part C predictive methodology are a faithful implementation of the method and are therefore constrained to the same functional form and need for similar CMFs. More details are provided below on developing CPMs for consistent use with these applications.

Review of Applicable Guiding Principles or Protocols for Developing SPFs

Introduction

The HSM provides analytic tools and techniques for quantifying the potential safety effects of decisions made regarding the location, classification, design, or operation of a roadway facility (HSM 2010). These safety effects are quantified in terms of a calculated frequency of crashes associated with the facility. The analysis scenario could be an existing condition or for a proposed alternative design condition. The users of the HSM include transportation planners, highway designers, traffic engineers, and other transportation professionals who make discretionary road planning, design and operational decisions.

The analytic tools and techniques are documented in Parts B, C, and D of the HSM. Those tools and techniques in Part B support the roadway safety management process. The tools and techniques in Part C support the design process and some operational decisions. Part D is focused on cataloging the safety effect of a wide range of design and operational treatments that can be used with the Part C tools and techniques.

Part C of the HSM provides one predictive method for each of the following facility types: rural two-lane roads, rural multilane highways, urban and suburban arterials, freeways, and ramps. Each method can be used to estimate the average crash frequency for an entire facility or selected sites (i.e., segments, intersections, or speed-change lanes) that comprise the facility.

The predictive method is applied to a given one-year time period, traffic volume, and constant geometric design characteristics of the roadway.

An objective of this research is to develop predictive methods that can be used to estimate crash frequency for freeway facilities. This includes developing planning level SPFs based on AADT and segment length for basic freeway segments and interchange segments. Additionally, this includes developing bi-directional and directional project design-level SPFs based on AADT, segment length, and geometric and operational characteristics. It is envisioned that this predictive method will be developed as consistent with the existing predictive methods in Part C as possible.

This section documents the findings from a review research protocols appropriate for predictive model development.

Guidance on Research Protocols

This section describes guidance provided in the HSM and other sources for developing crash prediction models. The first subsection describes the components of a crash prediction model. The second subsection summarizes guidelines in the HSM for the development of SPFs. The third subsection summarizes considerations for the development of crash prediction models (and their component SPFs and CMFs).

HSM Crash Prediction Model Components

Part C of the HSM provides one predictive method for each of the following facility types: rural two-lane roads, rural multilane highways, urban and suburban arterials, freeways, and ramps. The predictive method for each facility type is described in a separate chapter within Part C. A predictive method consists of (1) one or more CPMs and (2) guidelines for using these CPMs and interpreting the results.

A CPM is used to estimate the predicted average crash frequency of a specific type of site (e.g., segment, signalized intersection) with specific geometric design elements and traffic control features. With one exception, all sites in the HSM are defined to include both directions of travel (when the road of interest supports travel in both directions). The exception is freeway speed-change lane sites in Chapter 18, which focus on one direction of travel.

Each CPM in the HSM has the following general form:

$$N_p = C \times N_{SPF} \times (CMF_1 \times \ldots \times CMF_n) \qquad \text{Equation 5}$$

where:

N_p = predicted average crash frequency, crashes/yr;

C = local calibration factor;

N_{SPF} = predicted crash frequency for site with base conditions;

CMF_i = crash modification factor for geometric design element, or traffic control feature i (i = 1 to n); and

n = total number of CMFs.

Each CPM includes a SPF, one or more CMFs, and a local calibration factor (C). The SPF is used to predict the crash frequency N_{SPF} for a site having characteristics that match a specified set of "base conditions." These conditions describe the typical site's design elements and control features (e.g., 12 ft lane width). The set of CMFs are used to adjust N_{SPF} such that the CPM can provide reliable estimates of the predicted crash frequency N_p for sites that do not match all base conditions.

One base condition value is specified for each variable represented in a CMF. These variables are referred to herein as "base variables." The set of specified values are referred to as "base condition values" for the SPF. Each Part C chapter lists the base variables and the base condition values associated with the CPMs in that chapter.

The SPF is developed to include AADT volume and segment length as variables. These variables are not included in the CMFs so they are not considered (or used to define) the SPF base conditions. SPFs are used for network screening without inclusion of any CMFs.

The CPM can be used to evaluate any given site having known values for the geometric design elements and traffic control features recognized by its CMFs. When a site has an element or a feature whose value equals the base condition value, the corresponding CMF has a value of 1.0. When an element or feature has a value that is different from the base condition value, the corresponding CMF has a value that is different from 1.0.

Typically, there is one CMF for each base variable. If the variable is continuous (e.g., lane width), the associated CMF includes the variable. If the variable is discrete (e.g., add lighting), then the associated CMF is a constant (e.g., 0.90 for "add lighting") and the base condition is inferred from the CMF description (e.g., base condition is "no lighting present").

The predictive methods in Part C use a site-based approach for safety evaluation. With this approach, the road facility of interest is separated into homogenous road segment sites and intersection sites. The predictive method is used to estimate the average crash frequency for each site. The estimates for all sites are combined to obtain an estimate of the average crash frequency for the facility. A homogenous segment is one that has key safety-influential characteristics (e.g., AADT, lane width) that are relatively constant for the length of the segment.

HSM Guidance for Development of SPFs

The appendix to Part C of the HSM provides guidelines for the development of jurisdiction-specific SPFs for use in the Part C predictive methods. Key elements of these guidelines are summarized in the following list:

- In preparing the crash data to be used for development of SPFs, crashes are assigned to roadway segments and intersections following the definitions explained in Section A.2.3 and illustrated in Exhibit A-4 (note: the section and exhibit referenced here are in the HSM).
- The jurisdiction-specific SPF should be developed with a statistical technique such as negative binomial regression that accounts for the overdispersion typically found in crash data and quantifies an overdispersion parameter so that the model's predictions can be combined with observed crash frequency data using the EB Method.
- The jurisdiction-specific SPF should include the effects of the following traffic volumes: AADT volume for roadway segments, and major- and minor-road AADT volumes for intersections.
- The jurisdiction-specific SPF for any roadway segment facility type should have a functional form in which predicted average crash frequency is directly proportional to segment length.

The first bullet is a reminder that the HSM uses a site-based approach for safety evaluation. When the EB Method is used, observed crashes need to be associated with each site being evaluated. Rules are provided in the referenced HSM sections for this purpose.

The second bullet specifies the need to estimate an overdispersion parameter for each CPM so that it can be used with the EB Method. The analytic elements of this method are based on the assumption that the distribution of crashes is adequately described by the negative binomial distribution. If a different distribution is used, then the EB Method will need to be modified to use the measure of overdispersion produced by the new distribution.

The third bullet states the important association between crash frequency and the volume of traffic served by the site. For intersections or freeway speed-change lanes, the volume of both intersecting flows has an important association with site crash frequency. Thus, ramp volumes are important for the development of a predictive model for speed-change lane sites.

The last bullet acknowledges the important feature of the HSM site-based approach. The predicted crash frequency for a facility consisting of Y consecutive segment sites should not change with a change in the length of two or more of the Y segments. That is, all other factors being equal, a site that is X miles in length should have a predicted crash frequency that is twice that for a site that is X/2 miles in length. This logical model characteristic is achieved when the SPF includes the segment length variable without a regression coefficient (i.e., it is an offset variable).

CPM Development Considerations

This section describes protocols for the development of a CPM. These protocols were synthesized from various reports that describe the challenges associated with SPF and CMF development. Also described are study design techniques that can mitigate or eliminate these challenges. The presentation in this section is subdivided into four subsections. These subsections are identified in the following list:

- SPF from Base-Condition Database.
- SPF from Multiple-Variable Database.
- SPF and Inferred CMFs from Multiple-Variable Model.
- CMF from Before-After Study.

The protocols described in these four subsections provide useful state-of-the-practice considerations associated with the preparation of a sound study design for CPM development. The focus of the discussion is on protocols to guide the development of the models for ODOT Freeways. The discussion herein is not intended to be a comprehensive description of the model development process or of alternative statistical modeling approaches.

SPF from Base-Condition Database

This subsection describes protocols for the development of a SPF suitable for use in a CPM. There is no discussion in this subsection of protocols for the development of CMFs that would be used with this SPF. However, the last subsection provides some protocols for the development of CMFs from before-after studies.

The HSM Appendix to Part C describes two methods for developing SPFs for use with the CPMs described in Part C. One method is based on the use of a “base-condition database.” This type of database includes only the AADT and segment length for sites whose geometric design elements and traffic control features match the base condition values for the base variables specified for the CPM. The SPF coefficients are estimated with the base-condition database using regression analysis.

The objective of the SPF development described in this subsection is to produce a model that provides a reliable prediction of the average crash frequency for sites like the one of interest. This type of model is called “empirical” by Lehmann (1990) and “predictive” by Shmueli (2010). It is described as having an “application” focus by Hauer (2015) because it can be used with the EB method (and the site’s observed crash frequency) to estimate a specific site’s expected average crash frequency.

The protocols for developing an SPF using a base-condition database are listed in Table 10. Protocol 1 recognizes the need for equal sample size for each unique combination of independent variables. This protocol is explained using an example. Consider the development of an SPF for a speed change segment. The SPF has two variables, they are: freeway ADT and ramp ADT. Because these are continuous variables, they are converted into scalar categories that divide the range into three to five intervals. The freeway ADT ranges from 15,000 to 60,000 veh/d. Using three intervals, the categories are 15,000 to 30,000, 30,000 to 45,000, and 45,000 to 60,000. The ramp ADT ranges from 500 to 1,500 veh/d. Using three intervals, the categories are 0 to 500, 500 to 1,000, and 1,000 to 1,500. The factorial representation of these two variable categories is shown in the middle section of Table 11. There are nine combinations shown representing the full factorial of two variables, each with three levels (i.e., 9 = 3 × 3). For this example, it was determined that a minimum site sample size of 270 intersections was needed. These 270 sites are shown to be equally distributed to each of nine ADT combinations.

If the SPF is applicable to a segment and segment length is an offset in the SPF regression model, then it is excluded from the factorial development. More generally, any variable in the model that is not associated with a regression coefficient is not used to determine the factorial combinations.

If one or more additional variables are included in the SPF, then the number of factorial combinations is increased accordingly. For example, if two districts are represented in the database and an indicator variable for "District A" is included in the SPF, then the number of combinations shown in Table 11 doubles (18 = 2 × 9) and the sample size for each cell is cut in half (i.e., to 15). In this manner, 135 sites are obtained from District A and 135 sites are obtained from District B.

Table 10. Protocols for obtaining an SPF from a Base-Condition Database.

Study Design Element	Protocol	Benefit
Site selection and data assembly	1. Strive for an equal sample size for each unique combination of independent variables in the factorial representation of these variables.	Minimizes correlation among the independent variables.
	2. Avoid selecting sites that are relatively near each other.[2, 3]	Minimizes spatial correlation among observations.
	3. Avoid low sample size. At least 100 to 200 intersections or 100 to 200 miles; At least 300 crashes for the crash type and severity category of interest.[1]	Adherence to minimum number of sites criterion promotes transferability of model. Adherence to minimum number of crashes criterion promotes model goodness of fit.
Regression model form selection	4. The model form should allow for a non-linear relationship between traffic volume and crash frequency.[2] The model should predict a crash frequency that approaches zero as the traffic volume approaches zero.	Minimizes bias in the predicted value due to incorrect functional form. Adherence to logical boundary conditions reduces burden on data to empirically justify model form.
	5. If multiple jurisdictions are represented in the data, determine if the regression coefficients vary by jurisdiction and, if they do, include one or more "jurisdiction" variables in the model to account for this variation.[3]	Avoids confounding the influence of "jurisdiction" with that of the other independent variables or the intercept coefficient.
Statistical Modeling Approach	6. If the database includes multiple years of data for one or more sites, then combine the years of data for each site into one observation and use "number of years" in the model without a regression coefficient (i.e., include "number of years" as an offset variable).	Eliminates temporal correlation among observations.

Notes:

1. Source: Srinivasan et al. 2013a

2. Source: Srinivasan and Bauer 2013b

3. Source: Carter et al. 2012

Table 11. Illustrative factorial design for a two-variable SPF.

Ramp ADT Category	Site Sample Size by Freeway ADT Category			Total Sites
	15,000 to 30,000	30,000 to 45,000	45,000 to 60,000	
0 to 500	30	30	30	90
500 to 1,000	30	30	30	90
1,000 to 1,500	30	30	30	90
Total sites:	90	90	90	270

Protocol 2 addresses possible spatial correlation among observations in the database when the distance between some sites is small and that for other sites is large. For this discussion, one observation represents one site. Sites that are near to one another often have similar characteristics (e.g., similar design, similar drivers) and thus, they can overemphasize the influence of these characteristics in a regression analysis.

Spatial correlation can lead to incorrect estimates of the standard errors of the coefficients.

Protocol 3 addresses the need for a reasonably large sample size to offset the random variation inherent in crash data. The minimum sample size criterion stated in Table 10 considers both the number of sites and the number of crashes at the collective set of sites. It may be necessary to include multiple years of data (up to 5 years) at each site to satisfy the minimum crash criterion.

Protocol 6 addresses possible temporal correlation among observations in the database when these observations represent many consecutive years for some sites and fewer years for other sites. For this discussion, one observation represents one year at one site. At a given site, crash frequency in one year is likely to be similar to the next year because most site characteristics do not change in this time period. If some sites are represented in the database by multiple annual observations (one observation for each of several consecutive years) and other sites are represented by fewer annual observations, then the safety influence of the characteristics of the sites with multiple annual observations can be overemphasized. Temporal correlation can lead to incorrect estimates of the standard errors of the coefficients. This correlation can be minimized by aggregating the consecutive years of data for each site with multiple years of data such that each observation in the database represents one site. The crash count for the aggregated observation represents the total for all consecutive years. An average value is used for traffic volume and any other independent variables that change during the consecutive years. A variable for “number of years” is included with each observation and is used as an offset in the regression model.

SPF from Multiple-Variable Database

Like the first subsection, this subsection describes protocols for the development of a SPF suitable for use in a CPM. There is no discussion in this subsection of protocols for the development of CMFs that would be used with this SPF. However, the last subsection provides some protocols for the development of CMFs from before-after studies.

As noted in the first subsection, the HSM Appendix to Part C describes two methods for developing SPFs for use with the CPMs described in Part C. The second method is based on the use of a “multiple-variable database.” This database includes AADT and segment length as well as all base variables. Initially, a regression model is developed to include all significant database variables. Next, the SPF is made applicable to the base conditions by (1) substituting values in the regression model variables that correspond to the base condition values and (2) mathematically reducing the model to include only AADT and segment length variables.

The objective of the SPF development described in this subsection is the same as that described in the first section. That is, the objective is to produce an “empirical” (i.e., predictive) model that provides a reliable prediction of the average crash frequency for sites like the one of interest.

The multiple-variable database includes sites with a diversity of characteristics, as opposed to the sites in a base-condition database. A statistical modeling approach that controls for these differences is needed to ensure the model prediction is reliable when applied to a given site. (note that this control is provided through the site selection process when the SPF is developed using a base-condition database).

The SPF development process is focused on building a parsimonious model. This process is intended to produce a model that provides reliable predicted values while using as few variables as possible. The Akaike Information Criterion (AIC) is one statistic that can be used for this purpose. This criterion is used in a forward model building process where one variable is added to the model at a time. This model's AIC is compared to that of another model with fewer variables. The model form having the smallest AIC is retained as the basis for comparison with any new models being considered. All database variables are evaluated in this manner.

The protocols for developing an SPF using a multiple-variable database are listed in Table 10 and Table 12.

Table 12. Protocols for obtaining an SPF from a Multiple-Variable Database.

Study Design Element	Protocol[1]	Benefit
Site selection and data assembly	7. Strive to include variables in the database that are (1) available to users of the SPF and (2) describe the presence, location, or dimension of roadway features having a likely influence on crash frequency or severity. Use a factorial representation of the most influential variables (see Protocol 1).	Inclusion of influential variables in the model improves reliability of the model prediction.
Statistical Modeling Approach	8. Avoid including many independent variables such that the model is over-fit to the data (see also Protocol 3). [2, 3]	Promotes transferability of model.

Notes:

1. Include Protocols 1 to 6 from Table 9.

2. Source: Srinivasan and Bauer 2013b.

3. Source: Carter et al. 2012.

Protocol 8 is intended to avoid the development of models with more independent variables than can be justified by the data. This type of model is considered to "over-fit" to the data. In an over-fit model, one or more model variables may be explaining a trend in the data that is, in fact, just an artifact of underlying random processes (i.e., the variables are not explaining the effect of systematic influences on the dependent variable). Overfitting is most likely to occur when the ratio of observations to model variables is relatively small (say less than 10 to 15 observations per variable; Babyak 2004).

The scale parameter φ is a statistic that can be used to assess the potential for over- or under-fitting. It is computed using the following equation:

$$\varphi^2 = \sum_{i=1}^{n} \frac{(X_i - y_i\ N_i)^2}{(n-p)\ V[X_i]} \qquad \text{Equation 6}$$

where:

φ = scale parameter;

n = number of observations (i.e., segments, intersections, or roundabouts in database);

$V[X_i]$ = crash frequency variance for a group of similar locations, crashes^2;

N_i = predicted average crash frequency for observation i, crashes/yr;

X_i = reported crash count for y_i years for observation i, crashes;

p = number of model variables; and

y_i = time interval during which X_i crashes were reported for observation i, yr.

When the distribution of the dependent variable is specified by the negative binomial distribution, the following equation can be used to estimate the crash frequency variance $V[X_i]$:

$$V[X_i] = y_i\ N_i + \frac{(y_i\ N_i)^2}{K} \qquad \text{Equation 7}$$

where K is the inverse dispersion parameter (= $1/k$, where k = overdispersion parameter).

The potential for an over-fit model increases as the scale parameter value decreases below 1.0. Similarly, the potential for an under-fit model increases as the scale parameter increases above 1.0. Note that deviation from 1.0 could also be an indication that the assumed distribution of the dependent variable is not accurately describing the data (and that the data may be following a different distribution).

Cross-validation is the most widely-accepted method for assessing whether a model is over fit. With this procedure, the data is split into two parts. One part is used to estimate the model and the other part is used to validate the estimated model. Model goodness-of-fit statistics are computed for model when fit to the estimation data and again for this model when applied to the validation data. Each pair of fit statistics is compared. The likelihood of the model being over-fit increases as the difference between the pair of fit statistics increases.

SPF and Inferred CMFs from Multiple-Variable Model

Unlike the prior two subsections, this subsection describes protocols for the development of a CPM (i.e., both an SPF and its associated CMFs). The CMFs obtained from this development are referred to as “inferred CMFs” because they are obtained using regression analysis of cross section data. Regression results of this type rely on statistical control of differences among sites (as opposed to the more rigorous site-selection control that is available in the traditional before-after study).

The method for developing the CPM is based on the use of a multiple-variable database. The model form will include an SPF and one or more CMFs. Each CMF is associated with one or more model variables that collectively describe a road characteristic. A CMF will be a constant value when it is associated with the presence of a characteristic (e.g., lighting presence). A CMF is likely to have a functional form when it is associated with a characteristic that has a continuous dimension (e.g., lane width).

The objective of the SPF development described in this subsection is to produce a model that provides (1) a reliable prediction of the average crash frequency for the site of interest and (2) one or more CMFs each of which provide a reliable prediction of the effect of a change in a specific road characteristic on the average crash frequency. This type of model is called “explanatory” by Lehmann (1990) and by Shmueli (2010). It is described as having a “research” focus by Hauer (2015) because it is used to describe the effect of one or more road characteristics on safety. Like the empirical model described in the previous sections, the explanatory model can also be used with the EB method (and the site’s observed crash frequency) to estimate the site’s expected average crash frequency.

The CPM development process is focused on building an explanatory model. It is *not* based on the principle of parsimony (as described in the previous subsection). For explanatory model development, an independent variable is retained in the model if (1) its associated CMF produces a value whose direction and magnitude are logical and consistent with the findings of previous research on similar facilities and (2) its regression coefficient has sufficient statistical certainty to produce a t-statistic having an absolute value of about 1.0 or more. These variable retention rules are applicable regardless of whether the independent variable of interest is correlated with another model variable. They are intended to minimize the potential for omitted variable bias.

The aforementioned rules are used in a backward model building process where a full model is estimated initially. The one variable that is most contrary with the rules is selected for removal, the reduced model is re-fit to the data, and the rules are reassessed to determine if any additional variables should be removed.

If one or more of the roadway characteristic variables is associated with a well-established CMF, this CMF is incorporated directly in the regression model (i.e., without a regression coefficient). If the CMF has a medium-to-large level of uncertainty associated with its value, Bayesian regression analysis can be used to explicitly incorporate the CMF value and its uncertainty. If the CMF has a small level of uncertainty, then the CMF value can be incorporated in the model and the traditional (frequentist) regression analysis can be used.

The protocols for developing an SPF using a multiple-variable database are listed in Table 10, Table 12, and Table 13. For Protocol 10, the objective is to define road segment sites along which the presence, location, and dimension of key roadway characteristics is constant. The key roadway characteristics of interest are typically those that have a measurable influence on safety. The various chapters in Part C of the HSM identify key roadway characteristics used for roadway segmentation.

Table 13. Protocols for obtaining SPFs and Inferred CMFs from a Multiple-Variable Database.

Study Design Element	Protocol[1]	Benefit
Site selection and data assembly	9. For key roadway characteristics likely to be changed as a result of a safety improvement, use propensity scores to identify sites that have a similar likelihood of having the characteristic (e.g., turn bay, warning sign, wide lane).[2, 3] Include only the identified sites in the database.	Minimizes endogeneity bias (i.e., bias that can occur in regression results when the characteristic of interest is often implemented to improve safety).
	10. Strive to identify homogenous segments.[2]	Improves ability to quantify safety effect of independent variables.
Regression model form selection	11. Give preference to using segment length in the model without a regression coefficient (i.e., include segment length in the model as an offset variable).	Ensures that predicted crash frequency increases in direct proportion to the increase in segment length.

Notes:

1. Include Protocols 1 to 6 from Table 9 and 7 to 8 from Table 11.

2. Source: Srinivasan and Bauer 2013b.

3. Source: Carter et al. 2012.

The HSM also indicates that segments less than 0.10 mile may be too short for accurate crash location and may require significant effort to manage for safety evaluation purposes. For these reasons, a practical minimum segment length of 0.10 miles is encouraged.

In some situations (e.g., urban areas), it is difficult to find road segments that are at least 0.10 miles long and in which the key characteristics are constant. In these circumstances, it may be necessary to establish segmentation criterion using only those few variables that have the strongest influence on safety. With this approach, the remaining variables would be allowed to vary along the segment length. If the characteristic changes in presence, location, or dimension along the length of the segment, then a length-weighted average value is used for the variable value.

CMF from Before-After Study

This subsection describes protocols for the development of a CMF based on a before-after study. The discussion herein is based largely on the CMF-development protocols described by Carter et al. (2012).

The objective of the CMF development described in this subsection is to produce a CMF that provides a reliable indication of the change in safety caused by a change in a specified road characteristic. There are several before-after study methods that can achieve this objective. The three methods that have acceptable statistical rigor and are most commonly used are identified in the following list:

- Before-after with comparison group.
- Before-after with EB method.
- Before-after with full Bayes method.

Gross et al. (2010) provides a detailed description of each of the CMF development process associated with each of aforementioned study methods.

The protocols for developing a CMF based on a before-after study are listed in Table 14. Protocol A is intended to minimize site-selection bias due to regression-to-the-mean tendencies. The EB and full Bayes methods can minimize this bias if it is present in the treatment sites.

Table 14. Protocols for obtaining CMFs from a before-after study.

Protocol	Benefit
A. If the candidate study sites were selected for treatment (by the operating agency) because of their safety history, then use a study method that accounts for regression-to-the-mean.[1]	Minimizes the potential for bias in the estimate of treatment effectiveness.
B. Account for changes in traffic volume during the study period.[1]	Minimizes the potential for bias in the estimate of treatment effectiveness.
C. Account for changes in the road environment during the study period (e.g., weather, demographics, economy, vehicle fleet composition)[1]	Minimizes the potential for bias in the estimate of treatment effectiveness.
D. Verify no extraneous changes occur during the study period (e.g., other treatments installed, shift in crash reporting threshold).[1]	Maximizes the reliability of the estimate of treatment effectiveness.
E. Confirm suitability of comparison group or reference group.[1]	Maximizes the reliability of the estimate of treatment effectiveness.
F. If EB or full Bayes method is used, develop the SPF in conformance with the protocols in Table 14.	Maximizes the reliability of the estimate of treatment effectiveness.

Note:

1. Source: Carter et al. 2012.

Protocol B recognizes the need to adjust the observed change in safety between the before and after periods to extract the effect of a change in traffic volume between periods. This adjustment can be implemented in any of the three study methods.

Protocol C recognizes the need to statistically control for changes in the road environment in the before-after analysis. For the EB and full Bayes methods, this control can be provided by including yearly multipliers in the SPF for the reference group. For the comparison-group method, this control is indirectly provided in the set of comparison sites.

Protocol D is a due-diligence consideration during study site selection. The analyst should confirm that there were no extraneous changes at the treatment sites. Such changes would likely be confounded with the treatment effect and bias the treatment's estimated safety effect.

Protocol E is most applicable to the comparison-group method given its relatively small size relative the number of treatment sites. Gross et al. (2010, pp. 14-16) provide guidelines for evaluating the suitability of a comparison group for the evaluation of a given set of treatment sites.

Protocol F addresses the development of an SPF for use in the EB or full Bayes methods. For either method, the database should include reference sites that have traffic and physical characteristics that are similar to those of the treated sites. This database assembled is considered an "average-condition database" because it includes sites that are considered representative of the region of interest. Each observation in the database represents one year at one site. This database should include the following variables: calendar year, AADT, and segment length. The calendar year indicates the year associated with the AADT and crash data. The SPF coefficients are estimated with the average-condition database using regression analysis. The protocols for developing this SPF are listed in Table 15.

Protocol G addresses the need for a reasonably large sample size to offset the random variation inherent in crash data. The minimum sample size criterion stated in Table 15 considers both the number of sites and the number of crashes at the collective set of sites. The minimum number of crashes (i.e., 300 crashes) should be satisfied for each of at least three years.

Table 15. Protocols for obtaining an SPF from an average-condition database.

Study Design Element	Protocol[3]	Benefit
Site selection and data assembly	G. Avoid low sample size. At least 100 to 200 intersections or 100 to 200 miles; At least 300 crashes per year for the crash type and severity category of interest. At least 3 years of data are recommended. [1]	Adherence to minimum number of sites criterion promotes transferability of model. Adherence to minimum number of crashes criterion promotes model goodness of fit.
Statistical Modeling Approach	H. If the number of years varies by site, and yearly multipliers (i.e., an indicator variable for each year) are to be quantified, then minimize the effects of temporal correlation by using special regression techniques (e.g., generalized estimating equations, random effects model, negative multinomial model).[2]	Minimizes temporal correlation among observations.

Notes:

1. Source: Srinivasan et al. 2013a.

2. Source: Srinivasan and Bauer 2013b.

3. Include Protocols 2, 4, and 5 from Table 9.

Protocol H addresses possible temporal correlation among observations in the database when the number of consecutive years for each site varies among sites. For this discussion, one observation represents one year at one site. At a given site, crash frequency in one year is likely to be similar to the next year because most site characteristics do not change in this time period. If some sites are represented in the database by multiple observations (one observation for each of several consecutive years) and other sites are represented by fewer yearly observations, then the safety influence of the characteristics of the sites with multiple yearly observations can be overemphasized. Temporal correlation can lead to incorrect estimates of the standard errors of the coefficients.

Network Screening

The previous sections focused on the protocols for developing CPMs from SPFs and CMFs consistent with the HSM Part C. However, the protocols for SPF development for network screening models are consistent with those for Part C CPMs. The major difference is the use of CMFs in the CPM when compared to network screening SPFs. Network screening SPFs can be developed from the multiple variable database (there is no need to assume base conditions and/or use a base-condition database). The network screening SPFs still follow Protocols 1 through 8 provided in Tables 10 and 12 but consist only of segment length and ADT.

Review of Statistical Methods for SPF Development

Modeling Techniques for Crash Frequency

Poisson Regression

Poisson regression was introduced in the 1980s to estimate crash frequency as a function of independent variables (Jovanis and Chang 1986). The Poisson distribution was introduced as an alternative to the normal distribution to model non-negative count data. In the case of segments, the probability of segment i have y_i crashes per time period is given as:

$$P(y_i) = \frac{e(-\lambda_i)\lambda_i^{y_i}}{y_i!} \quad \text{Equation 8}$$

where $P(y_i)$ is the probability of segment i having y_i crashes per time period and λ_i is the Poisson parameter for entity i (Lord and Mannering 2010). It is assumed that the Poisson parameter is equal to the segment's expected number of crashes per year. Poisson regression models estimate the Poisson parameter as a function of explanatory variables with the form $\lambda_i = e(\beta X_i)$ where X_i is a vector of explanatory variables and β is a vector of estimable parameters. However, the Poisson distribution assumes that crash events are independent, and it cannot handle over- or under-dispersion. Therefore, crash analysis has been extended to the negative binomial (or Poisson-gamma) model.

Negative Binomial Regression

When the variance of the data distribution is larger than the mean value, the data are noted to be overdispersed. The negative binomial model assumes the Poisson parameter follows a gamma probability distribution. The Poisson parameter is rewritten as $\lambda_i = e(\beta X_i + \varepsilon_i)$, where the added term is a gamma-distributed error term with a mean of 1 and a variance of α. The parameter α is referred to as the overdispersion parameter. If the overdispersion parameter approaches zero, then the model approaches a Poisson model. The negative binomial model is the most commonly used in transportation safety research, but it does not handle excess zeros well, nor does it handle underdispersed date well (Lord and Mannering 2010). Moreover, the negative binomial model does not account for heterogeneity or correlations among observations. The following extensions can be considered for spatial and temporal correlations.

Random and Fixed Effects

Panel models typically specify either fixed effects or random effects as ways to allow different intercept terms for each cross-sectional unit of interest (typically spatial). The differing intercepts for each spatial unit address cross-sectional heterogeneity, or unmeasured variables that determine crash frequency or severity. The fixed effects model can address situations where the unobserved effects are correlated with right-hand-side variables in the model. The main limitation of the fixed effects model is its inability to estimate regression parameters associated with variables that vary within cross-sectional units (e.g., a county is considered as a fixed effect, but the effect of county is found to change by year within counties). Additionally, the standard errors of the parameters estimated by fixed effects can be larger than standard errors from random effects if there is little variation in the predictor variables across the cross-sectional units.

Fixed effects models are typically estimated by including cross-sectional dummy variables for each spatial unit (minus one) included in the analysis. For an example, see Le and Porter (2012). These variables are used to capture the difference between locations where the study segments or intersections are located. For a fixed effects model, the intercepts for each spatial unit are allowed to vary, but they are constrained (i.e., fixed) over time. Additionally, the effects of each predictor variable are the same for each unit and over time. Time series variables (i.e., dummy variables for years) may be included to capture the shared unobserved effects for each year (for example if crashes are higher in one year due to a pattern of more snow and/or ice). Moreover, interaction terms may be applied to account for differences in effects of predictor variables and individual units; however, those effects are still fixed for each unit over time.

According to Gujarati (2003), if too many dummy variables are introduced, the model can overfit the data. Additionally, the addition of a wide range of dummy variables can introduce multicollinearity; which, if present, can reduce the precision of parameter estimates. As noted previously, fixed-effect models cannot handle variables that are time invariant, meaning we cannot estimate their impacts. Additionally, fixed-effect models can lead to necessary assumptions for the model's error term that may or may not hold true. Some of these problems may be alleviated through the use of random effects models.

The main assumption with the random effects model is that the unobserved effects are uncorrelated with the independent/right-hand-side variables (i.e., the unobserved effects are expressed through the model's error term). If this assumption holds, the random effects estimator is consistent and estimated parameters will have smaller standard errors than the fixed effects estimator. Random effects models account for unobserved heterogeneity and serial correlation by treating the data in a time-series cross-sectional panel. Note that serial correlation can be eliminated by aggregating data for multiple years. The Poisson parameter is rewritten as $\lambda_{ij} = e(\beta X_{ij} + \eta_{ij})$ where j refers to the group each segment belongs (the group can be temporal or spatial), where segments in the group share unobserved effects. η_{ij} is a random effect for observation group j. The model assumes η_{ij} is randomly distributed across groups such that $e(\eta_j)$ is gamma distributed with a mean of 1 and a variance of α. What we are assuming with this methodology is that each unit analyzed (e.g., county) is drawn from a larger universe of units and they have a common mean value for the intercept and the individual differences in intercept values are reflected in the error term (Gujarati 2003). Shankar et al. (1998) concluded that random effects negative binomial models provide benefits when exogenous variables relating to spatial and temporal effects are absent. When spatial and temporal effects are explicitly included, the negative binomial model is adequate and the random effects negative binomial model loses its distributional advantage.

The choice to select the fixed versus random effects model hinges on the assumption of correlation between the error component and the independent variables. If the error term is assumed to be uncorrelated with the independent variables, random effects models may be appropriate. If the error term is assumed to be correlated with the independent variables, fixed effects models may be appropriate. If the number of time series data is large and number of spatial units is small, there will be little difference between the two model types. If the number of spatial units is large and number of time series data is small, the two methods can differ significantly. A Hausman test can be used to help choose between fixed effects and random effects models. If the null hypothesis (the fixed effects and random effects estimators do not differ substantially) is rejected, the fixed effects model is suggested to be more appropriate. Mixed effects models have also been proposed, incorporating both fixed and random effects to account for shared unobserved effects (Butsick 2014, Wood et al. 2015, Gooch et al. 2016).

Random Parameters

Mannering et al. (2016) summarized the statistical consequences of ignoring heterogeneity in crash data for human elements, vehicle characteristics, roadway characteristics, traffic characteristics, and environmental characteristics. The authors suggested that if unobserved heterogeneity is ignored, and the effects of observable variables is restricted to be the same across all observations, the resulting model will be misspecified and parameter estimates will be both biased and inefficient. Traditional fixed parameter methods do not account for heterogeneity of segments, thus underestimating the standard errors (and thereby overinflating t-ratios). Additionally, Mannering et al. (2016) point out that fixed and random effects models necessitate panel data, while the random parameters model can be estimated with cross-sectional or panel data. The random parameters modeling methodology was developed allowing each estimated model parameter to vary across individual segments in the dataset, accounting for unobserved heterogeneity from one segment to another. Venkatraman (2014) adds that since independent variables Xs are constrained to the observed dataset, the modeling objective is to maximize the information available from the Xs. Therefore, the insights from the model are limited to the range of values observed in the Xs (which is consistent with other type of regression models). In order to maximize the objective, β is allowed to vary across segments and years rather than held constant.

The estimable parameters are expressed as:

$$\beta_i = \beta + \varphi_i \qquad \text{Equation 9}$$

where φ_i is a randomly distributed error term (e.g., normally distributed term with mean zero and variance σ^2. Anastasopolous and Mannering (2009) found the normal distribution to provide the best fit for all parameters found to be random. Here the Poisson parameter becomes $\lambda_i \mid \varphi_i = e(\beta_i X_i + \varepsilon_i)$ in the negative binomial regression model. The log-likelihood with this random parameter is written as:

$$LL = \sum_{\forall i} ln \int_{\varphi_i} g(\varphi_i) P(n_i | \varphi_i) d\varphi_i \qquad \text{Equation 10}$$

where g(.) refers to the probability density function of φ_i.

Since the numerical integration of the negative binomial models with random parameter distribution is computationally cumbersome (meaning that the time taken to estimate the model can be an order of magnitude higher for each estimated model), a simulation-based maximum likelihood is used to maximize the simulated log-likelihood function. A quasi-random sequence such as Halton draws is used since it provides a more efficient distribution of draws for numerical integration than random draws (Washington et al. 2003). Several researchers have suggested that 200 draws is an appropriate number when using this method (Milton et al. 2008, Bhat 2003, Anastasopoulos and Mannering 2009).

For traditional fixed parameter models, the elasticity and marginal effect for variables is directly computed from the fixed parameter estimate and is used in the development of safety prediction methods. For example, an elasticity may report that for a one percent increase in an independent variable, the dependent variable decreases by X percent. For random parameters models, average marginal effects and elasticities can be computed, and as shown by Anastasopolous and Mannering (2009) and Park et al. (2016), the marginal effects can be quite different (when comparing the fixed parameter estimate to the averages for the roadway segment population from the random parameter) from fixed parameter models, having accounted for heterogeneity. Additionally, more information is provided from the resulting model.

In addition to the elasticity and the marginal effect, the random parameter can be used to suggest that in Y percent of segments an increase in the independent variable would be associated with lower crash frequency and for 100 – Y percent of segments, it would be associated with higher crash frequency. The standard deviation of the parameter estimate is an important part of the modeling result, and it is unclear how useful or confusing this would be to practitioners. For the random parameters methodology, both the mean parameter estimate and standard deviation have an associated standard error. If the standard error of the standard deviation is significant, then the variable is included for a random parameter estimate. If the standard error is insignificant, then the variable is included for a fixed parameter estimate.

An important aspect of the development of safety prediction models is the transferability of the model to users in other spatial units. Lord and Mannering (2010) note that random parameter models are very complex to estimate, may not necessarily improve predictive capability, and model results may not be transferable to other data sets because the results are observation specific. Mannering et al. (2016) add that while random parameter models are often criticized due to a lack of transferability to different locations, there are two important considerations. The first is that that finding significant random parameters means that unobserved heterogeneity is present in the dataset (which as noted above can lead to biased and inconsistent fixed parameter estimates). The second is that transferability would still be a problem due to the fixed parameter estimate having bias that is a function of unobserved heterogeneity. However, it is important to note that random parameters models can pose convergence problems in the presence of some variables.

Summary

In summary, count models are traditionally used to quantify the relationship between crash frequency and traffic volumes, design elements, and traffic control features. Negative binomial regression has most commonly been applied to account for the overdispersion inherently found in crash data. The overdispersion parameter estimated from the modeling process is used in the development of the weight factor in the EB analysis method.

Relatively recently, researchers have been applying more sophisticated versions of count models to account for temporal and spatial correlations. Depending on the assumption of the correlation between unobserved effects and right-hand side variables, fixed- and random-effects models have been applied to account for temporal and spatial correlations. However, recent research has shown that ignoring heterogeneity in crash data will result in estimate of effects that are biased and inconsistent (statistical significance may be overinflated). Random parameters models allow each model parameter to vary across individual segments, maximizing the information available for each right-hand side variable included in the model.

As models become more sophisticated, bias and inconsistency are generally lessened, improving the transferability of the model. However, more sophisticated models can be more time consuming to estimate, limiting the number of models that can be estimated during the modeling process. Additionally, more sophisticated models can prove to be more difficult to reach convergence with a larger number of predictor variables, limiting the number of geometric and operational features that can be included in the model specification.

Modeling Techniques for Severity Distribution

The crash severity distribution may change significantly with traffic volumes, design elements, traffic control features, and other characteristics. Estimating a SDF using logit models is one possibility for quantifying the relationship between crash severity and various roadway characteristics. In this regard, the logit model can be used to predict probabilities (or proportions) of crash severity outcomes as a function of traffic volume, geometry, and other roadway characteristics. The multinomial logit (Shankar and Mannering 1996), nested logit, (Shankar et al. 1996) and ordered outcome models (Duncan et al. 1998) are possible model alternatives. The databases used to estimate the severity models will consist of the same crashes and segments as the frequency model databases but will be restructured so that the basic observation unit (i.e., database row) is the crash instead of the road segment. The SPFs and SDFs can be combined to estimate the number of crashes of different severity levels (see Wang et al. 2011) for example.

The approach was successfully applied in NCHRP Project 17-45 by Bonneson et al. (2012), which created content for the freeway and ramp chapters of the HSM. This project predicted the proportion of crashes for fatals, incapacitating injuries, non-incapacitating injuries, and possible injuries. SDFs were developed to be combined with the safety prediction models to minimize the frequency-severity indeterminacy problem described by Hauer (2006). The SDF model considers all severity levels together and can be used to predict the shift in in crashes among levels due to a change in roadway conditions (Bonneson et al. 2012). The researchers used a multinomial logit model to analyze crash severity. This model allows for some variables to be constrained to have the same effect on each severity level while allowing other variables to have a variable effect among levels. Possible injuries were used as the baseline and variables were included in the model to predict the proportion of other crash severities. The proportion of barrier and proportion volume during high-volume hours were consistently found to reduce the proportion of fatals, incapacitating injuries, and non-incapacitating injuries. Lane width was associated with reduced fatality and non-incapacitating injury proportions. The proportion of rumble strips, proportion of horizontal curves, and rural areas were associated with a higher proportion of fatals, incapacitating injuries, and non-incapacitating injuries.

A recent investigation of the reliability of this approach has found that it tends to outperform the direction calibration of SPFs for each severity level (Avelar et al. 2018). The researchers compared the direct estimation of severity-level SPFs (i.e., KAB crashes) and a combination of total crashes SPFs with SDFs (proportion of KAB crashes). The authors used data from Oregon intersections and Texas turnaround sites. In both cases, the two methods produced similar results. The authors further compared the sensitivity by degree of dispersion and observed correlation levels of total and severe injury crashes with potential explanatory variables. These analyses favored the SDFs in combination with the total crashes SPF. However, the authors did note that the KAB SPF outperformed the combination method when KAB crashes and non-KAB crashes have a common predictor with an opposite direction of effect.

Review of Research on Safety Impacts of Geometric or Operational Characteristics of Freeway Facilities

The project team conducted a literature review detailing geometric and operational characteristics that have been shown to be significantly associated with freeway safety performance in previous studies. This literature review is important for identifying factors for inclusion in project design-level SPFs. The project team focused the literature review on recent research completed since the publication of NCHRP Project 17-45, as the NCHRP Project provided a detailed review at that time. This literature review focuses on typical freeway facilities and excludes literature related to identifying the safety effects of managed lanes. The project team conducted a literature review focused on the safety effects of managed lanes for

NCHRP Project 17-89A and is familiar with geometrics and operations related to managed lanes.

Bonneson et al. (2012) developed a predictive method for freeway facilities and ramps for inclusion in the HSM. The predictive method developed SPFs separately for freeway segments and speed change lanes. Additionally, SPFs were developed separately for freeway segments by number of lanes for urban freeways from four to ten lanes, by number of vehicles involved in the crash (i.e., multiple-vehicle versus single-vehicle), and by crash severity (i.e., fatal and injury versus property damage only). For speed change lanes, separate SPFs were developed for ramp entrances versus ramp exits, crash severity, and number of lanes.

CMFs were developed for freeway segments and speed change lanes for the following characteristics:

- Horizontal curvature. CMFs assume a tangent base condition and incorporates information on whether curves are on one or both roadbeds, proportion of segment length with curvature, and equivalent radius of curve. Separate coefficients were estimated by multiple-vehicle or single-vehicle crashes and fatal and injury or property damage only crashes.
- Lane width. The CMF assumes a 12-foot lane width as the base condition. Separate coefficients were estimated by multiple-vehicle or single-vehicle crashes and fatal and injury or property damage only crashes.
- Inside shoulder width. The CMF assumes a 6-foot inside shoulder width as the base condition. Separate coefficients were estimated by multiple-vehicle or single-vehicle crashes and fatal and injury or property damage only crashes.
- Median width. The CMF assumes the base condition is a 60-foot median, 6-foot inside shoulder, and no barrier present in the median. The CMF can be adjusted for proportion of segment length with barrier present in the median, median width, and distance from edge of inside shoulder to barrier face. Separate coefficients were estimated by multiple-vehicle or single-vehicle crashes and fatal and injury or property damage only crashes.
- Median barrier. The CMF assumes the base condition is no barrier present in the median. The CMF can be adjusted for proportion of segment length with barrier present and distance from the edge of inside shoulder to barrier face. Separate coefficients were estimated by multiple-vehicle or single-vehicle crashes and fatal and injury or property damage only crashes.
- High volume. The CMF assumes a base condition of no hours having a volume that exceeds 1,000 veh/hr/lane and can be adjusted by proportion of AADT during hours where volume exceeds 1,000 veh/hr/lane. Separate coefficients were estimated by multiple-vehicle or single-vehicle crashes and fatal and injury or property damage only crashes.

CMFs were also developed for multiple-vehicle crashes on freeway segments and single-vehicle crashes on freeway segments. The CMF for multiple-vehicle crashes on freeway segments is for lane changes. The base condition is no significant lane changing due to ramp entry or exit (i.e., no ramp entrance within 0.5 miles of the segment). The CMF can be adjusted by the proportion of segment length within a Type B weave, weaving section length, distance from segment to upstream entrance ramp or downstream exit ramp, and entrance or exit ramp volumes. Separate coefficients were estimated by crash severity.

The CMFs for single-vehicle crashes on freeway segments include the following:

- Outside shoulder width. The CMF assumes a base condition of a 10-foot outside shoulder width. The CMF can be adjusted by the proportion of total segment length with horizontal curvature and outside shoulder width. Separate coefficients were estimated by crash severity.
- Shoulder rumble strips. The CMF assumes a base condition of no rumble strips present. The CMF can be adjusted by the proportion of total segment length with horizontal curvature and the proportion of segment length with rumble strips present on the inside and/or outside shoulders. The CMF only applies to fatal and injury crashes.
- Outside clearance. The CMF assumes a base condition of a 30-foot clear zone, a 10-foot outside shoulder, and no barrier present in the clear zone. The CMF can be adjusted by the proportion of segment length with a barrier present on the roadside, clear zone width, and distance from edge of outside shoulder to barrier face. The CMF only applies to fatal and injury crashes.
- Outside barrier. The CMF assumes a base condition of no barrier present in the clear zone. The CMF can be adjusted by the proportion of segment length with a barrier present on the roadside and the distance from the edge of the outside shoulder to the barrier face. Separate coefficients were estimated by crash severity.

Finally, ramp entrance and ramp exit CMFs were developed for ramp entrances and ramp exits, respectively. The ramp entrance CMF can be adjusted by the length of ramp entrance, the ramp side (i.e., left or right), and the ramp AADT. Separate coefficients were estimated by crash severity. The ramp exit CMF can be adjusted by the length of ramp exit and the ramp side. Separate coefficients were estimated by crash severity.

Le and Porter (2012) quantified the relationship between ramp spacing and freeway safety for total crashes, fatal and injury crashes, and multivehicle crashes. Ramp spacing was included as an inverse variable, accounting for the non-linear relationship between ramp spacing and safety. The safety impacts of ramp spacing are greatest when ramp spacing is smaller and loses effectiveness for larger values of spacing (the safety effects level-off for spacing greater than 4,000 feet). The parameters were significant for total crashes and multivehicle crashes and were nearly significant for fatal and injury crashes. The results indicated that multivehicle crashes are more sensitive to changes in ramp spacing than other crash types. Additionally, while the frequency of severe injury crashes increases as ramp spacing decreases, the proportion of crashes resulting in severe injuries decreases as spacing decreases. Further, the results supported the Green Book recommendation that auxiliary lane should be used when ramp spacing is less than 1,500 feet. The safety effects of ramp spacing level-off when auxiliary lanes are present at around 2500 feet. The dataset included segments with HOV lanes; these segments were found to have increased crash frequency over those without HOV lanes.

Graham et al. (2014) developed guidelines for designing median typical cross-sections through crash analysis and simulation of existing median designs and median barrier effectiveness. The results of the crash-based analysis indicated that for rural four-lane freeways, cross-median crashes decrease with wider medians, but rollover crashes generally increase with wider medians. These effects were noted to be almost equal in magnitude but opposite in direction. The crash analysis indicated both effects are continuous in nature when compared to median width. The crash analysis further indicated opposite effects for cross-median crashes and rollovers based on median slope ratio but are opposite to the results for median width. Flatter slope ratios are associated with increased cross-median crashes and fewer rollover crashes. Flatter slopes on rural four lane freeways are also associated with fewer fixed object crashes.

Graham et al. (2014) also developed models for traversable and barrier medians. The models can be used to estimate the safety differences between median types along with various geometric characteristics and barrier types. Overall, the results indicate that flexible median barriers may be cost-effective at lower traffic volumes than shown in current AASHTO median barrier warrants.

Dixon et al. (2015) examined the safety impacts of reducing lane and shoulder with to permit an additional lane on urban freeways in Texas. They found that the impacts of reducing shoulders outweighs the safety benefits of adding lanes when the total paved width is not changed. However, they found that if total paved width is increased when adding a travel lane, it is possible to identify lane and shoulder width combinations such that the number of crashes remains constant. They found that crashes reduce 9 percent for each additional foot of right shoulder. Furthermore, they found a five percent decrease in total crashes on basic freeway segments for each additional foot of left-shoulder. The authors also found a safety benefit for increasing 11-foot travel lanes to 12-foot travel lanes on basic freeway segments. When there are two directional lanes, the benefit is approximately a five percent decrease in KAB crashes. This percentage increases up to a five-directional-lane freeway, where there is a 12 percent decrease in KAB crashes.

Dixon et al. (2015) found that expected crash frequency decreases as the distance to the closest upstream ramp increases for basic freeway segments. For study segments, the closest downstream ramp varied from 17 feet to nearly 7,000 feet, with an average of 1,860 feet. The closest upstream ramp varied from 320 feet to 7,170 feet, with an average of 1,739 feet. They found that expected crash frequency decreases as the distance to the closest downstream ramp increases for basic freeway segments.

The effects for upstream and downstream ramps were similar for total crashes, while the effect size was slightly greater for upstream ramps for fatal and injury crashes.

Conclusions and Recommendations

This overall literature review will be an important resource for guiding the development and implementation of freeway SPFs throughout the remainder of this project. The first part of this review will guide the implementation of SPFs into specific tools and methods already in place for the department. Additionally, the first part of this review provides insights on how models and methods should be developed for integration into planning, safety studies, PBPD, and project-level decision-making. The review of how other agencies have developed or calibrated SPFs and implemented those in existing tools, as well as the review of existing tools, provides insights into the functional forms and specifications our team should consider throughout the model development process. The guiding principles and modeling considerations further build off the review of tools and provide best practices for modeling methods, potential biases, and maximizing the use of existing data. Finally, the review of freeway-related research provides further insight into factors that should be considered and functional forms that should be considered in statistical models developed.

Appendix B: Planning-Level SPF Development

Introduction

This appendix presents the process for developing planning-level SPFs for network screening. Network screening SPFs include bi-directional segments. The project team developed separate analysis databases for the following SPF categories:

1. Network screening databases:
    a. Base segments.
    b. Interchange segments.
2. Bi-directional project design-level database:
    a. Base segments.
3. Directional project design-level databases:
    a. Base segments.
    b. Entry speed change segments.
    c. Exit speed change segments.

This Appendix focuses on data collection and SPF development for the network screening databases.

Network Screening SPF Methodology

Network screening SPFs estimate average crash frequency or average fatal and injury (FI) crash frequency for a site based on a population of similar sites. They generally only include AADT and segment length for a set of base conditions. The typical form of a freeway segment network screening SPF (as included as a base model in Safety Analyst) is shown as Equation 11:

$$N = L \times exp^{\alpha} \times AADT^{\beta} \qquad \text{Equation 11}$$

where:

N = predicted number of crashes at a site per year.

L = segment length (in miles).

AADT = annual average daily traffic.

α and *β* are regression parameters estimated during the modeling process.

As discussed in the Network Screening SPF Development section, other functional forms may be considered for SPF development. Further, the project team examined freeway facility mileage and total number of crashes by area type, number of lanes, segment type, and crash type (total or FI) for facility subtype grouping. Network screening using the SPFs can focus on total crashes and/or FI crashes. The project team did not develop SPFs for other crash or severity types.

Crash Prediction Model Development Methodology

For crash frequency modeling, count models are traditionally used to quantify the relationship between crash frequency and traffic volumes, design elements, and traffic control features. Negative binomial regression has most commonly been applied to account for the overdispersion inherently found in crash data. The overdispersion parameter estimated from the modeling process is used in the development of the weight factor in the EB analysis method.

Relatively recently, researchers have been applying more sophisticated versions of count models to account for temporal and spatial correlations. Depending on the assumption of the correlation between unobserved effects and right-hand side variables, fixed- and random-effects models have been applied to account for temporal and spatial correlations. As models become more sophisticated, bias and inconsistency are generally lessened, improving the transferability of the model. However, more sophisticated models can be more time consuming to estimate, limiting the number of models that can be estimated during the modeling process. Additionally, more sophisticated models can prove to be more difficult to reach convergence with a larger number of predictor variables, limiting the number of geometric and operational features that can be included in the model specification.

The project team considered fixed effects, random effects, and mixed effects models to account for unobserved correlations inherent in the data. The project team applied negative binomial count models as is the current state-of-the-practice.

The following sections describe the results of the Network Screening SPF Development and Project-Level SPF Development, respectively.

Network Screening SPF Development

The project team developed network screening SPFs for rural and urban freeways based on data for the entire freeway network. ODOT provided Safety Analyst files separately for roadway inventory, freeway segment volumes, ramp inventory, ramp volumes, and crash data. The roadway inventory data are considered to be a snapshot that remains constant through the study period; ODOT provided crash and volume data for a 2014 through 2018 study period.

The project team combined the original segments from the Safety Analyst data files when no features changed between segments. The purpose was to create longer segments (reducing the impact of short segment-related issues), reduce the potential for outliers to leverage their weight on the results, and to increase the reliability of estimated SPFs. Table 16 provides an overview of number of segments, total mileage, and crash sample size for freeway facilities by area type, number of lanes, and segment type.

Based on Table 16, the project team found that crash sample sizes for both total crashes and FI crashes were sufficient for estimating separate SPFs for each subtype (as defined by area type, number of lanes, and segment type). However, separate

SPFs would not be estimable for every site subtype; some should be excluded, and some combined. Sample sizes were very small for urban segments with fewer than four lanes. Additionally, segments with fewer than four lanes are likely to act more as ramp segments. Based on Table 16, the following site subtypes were recommended for SPF development:

- Freeway segments:
    - Rural four-lane segments.
    - Rural six-lane segments.
    - Urban four-lane segments.
    - Urban six-lane segments.
    - Urban eight or more-lane segments.
- Interchange segments:
    - Rural four-lane segments.
    - Rural six-lane segments.
    - Urban four-lane segments.
    - Urban five/six-lane segments.
    - Urban seven or more-lane segments.

From the individual datasets, the project team combined ramp inventory and volume data for freeway segments coded with interchanges (by interchange number). Individual ramps were retained for each interchange, and interchange ramp volumes were considered for the following:

- Total combined ramp volumes (entrance and exit).
- Maximum ramp volume (of all ramps within interchange area).
- Combined entrance ramp volumes.
- Combined exit ramp volumes.

Table 16. Overall Study Sample Sizes by Area Type, Number of Lanes, and Segment Type.

Area Type	Lanes	Segment Type	Number of Segments	Total Mileage	Total Crashes	FI Crashes
Rural	4	Base	293	592.44	16,403	3,359
		Interchange	331	103.05	4,112	829
	5	Base*	3	1.42	44	13
		Interchange*	2	0.47	27	2
	6	Base	74	210.25	11,435	2,284
		Interchange	92	21.84	2,056	422
	7	Base*	1	1.72	103	21
		Interchange*	0	0	0	0
Urban	2	Base*	1	0.17	13	4
		Interchange*	6	0.80	129	42
	3	Base*	0	0	0	0
		Interchange*	12	2.44	783	180
	4	Base	521	410.15	17,537	3,968
		Interchange	1,355	359.83	33,718	8.012
	5	Base*	15	9.43	424	90
		Interchange	90	18.08	3,618	865
	6	Base	259	215.41	17,982	4,230
		Interchange	840	255.82	56,218	13,652
	7	Base*	10	3.05	615	169
		Interchange	75	16.97	5,081	1,274
	8	Base	70	41.11	6,915	1,716
		Interchange	316	97.65	30,606	7,808
	9	Base	4	0.67	178	41
		Interchange	10	2.16	682	181
	10	Base	7	5.18	908	316
		Interchange	36	12.00	3,034	885

*Indicates segments were dropped from dataset due to small sample size.

The project team considered several functional forms for SPF development. However, while considering functional form, it became readily apparent that the presence of outliers in the data played a role in the decision of the final functional form. The project team first identified outliers in the dataset for each site subtype and removed those during the SPF development process.

Figure 4 provides an example of a CURE plot for an SPF for rural four lane freeway base segments with the presence of outliers. The CURE plot shows the cumulative error in prediction versus observed crashes when plotted against the main variable of interest, in this case, freeway AADT. The CURE plot in Figure 4 shows two main concerns. The first is that the cumulative residuals do not come close to adding up to zero (indicating prediction bias). The second is that there is a systematic over-prediction (resulting in negative residuals) and for most of the plot, the cumulative residuals fall outside the 95 percent confidence interval, indicating the error is not random. We see a consistently decreasing slope across AADT indicating a long, systematic bias in the SPF. We see a large drop at around 21,000 AADT, indicating a potential outlier at this location. Figure 5 shows a CURE plot for the same model functional form for the same dataset, removing the outlier at 21,000 AADT. Comparing Figure 4 and Figure 5, the maximum cumulative residual reduced from -

900 to approximately -500. The overall slope from 10,000 AADT to 40,000 AADT has flattened, indicating the systematic bias in the SPF has been reduced. The cumulative residuals fall within the confidence interval until approximately 37,000 AADT, which is a substantial improvement.

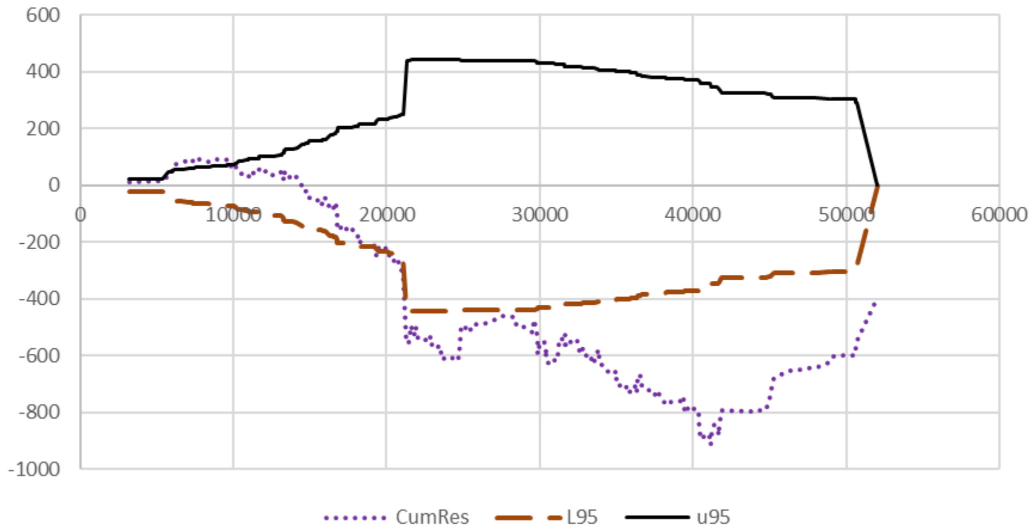

Figure 4. Example CURE Plot for Rural Four Lane SPF with Outliers.

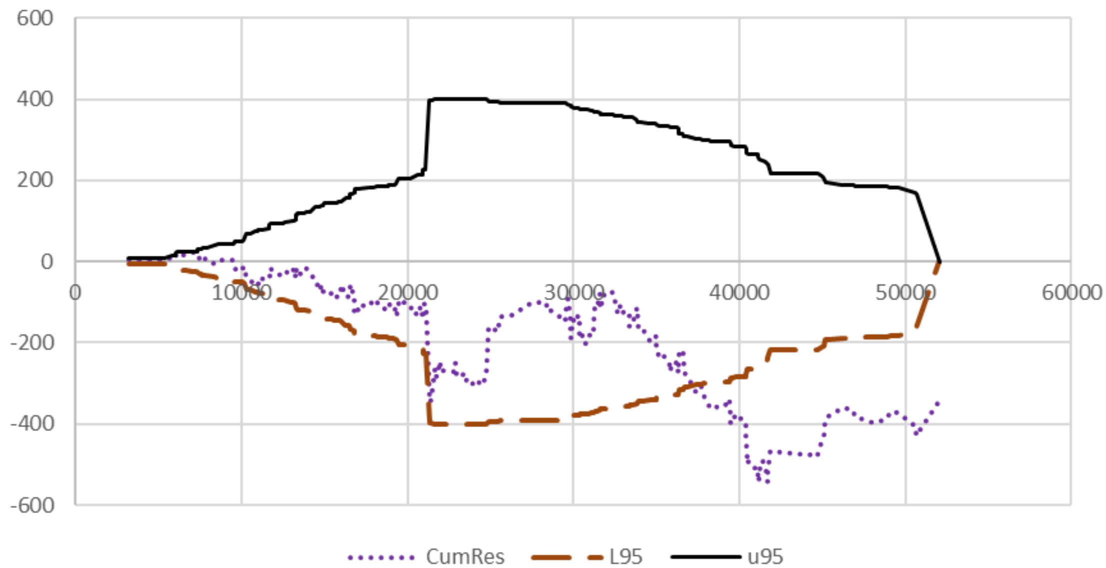

Figure 5. Example CURE Plot for Rural Four Lane SPF with Outlier Removed.

Figure 6 provides the CURE plot for the final rural four lane freeway SPF, with several outliers removed, and the final functional form chosen. The project team removed 12 outliers in total. The resulting model fits the data well, reduces the overall maximum cumulative residual to approximately -200 and there are only a few places where the cumulative residuals approach the 95 percent confidence interval. The same process was used to develop all SPFs by segment subtype and crash type.

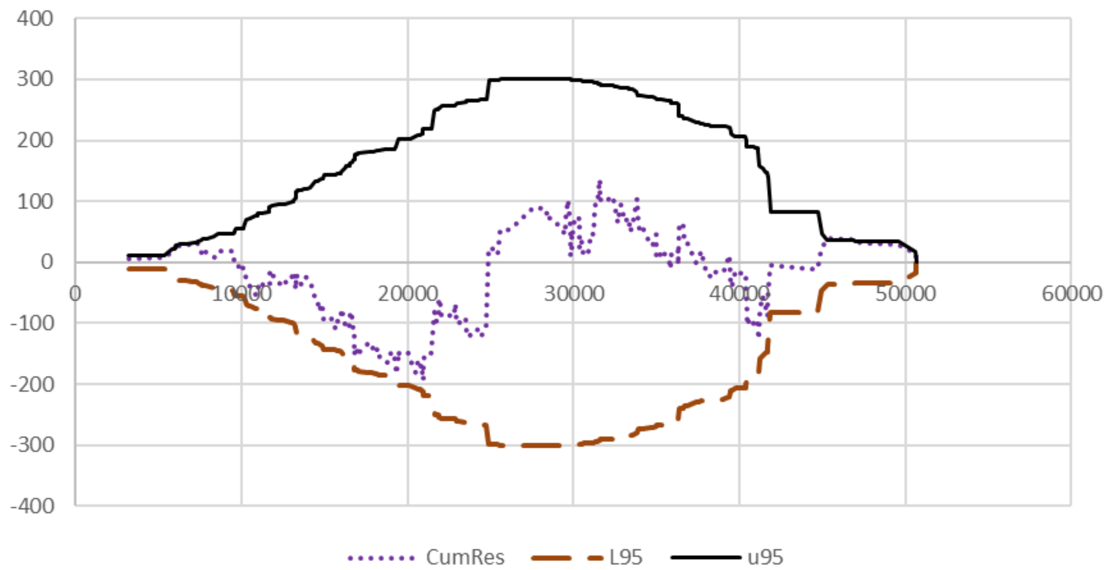

Figure 6. Example CURE Plot for Final Rural Four Lane SPF.

Table 17 provides an overview of the final sample sizes and summary statistics for freeway segment SPF development. Table 18 provides an overview of the final sample sizes and summary statistics for interchange freeway segment SPF development. It is important to note that the summary statistics for crash data are for five-year totals (2014 – 2018 data). Note that AADT are average values for the five years for each segment. The number of years (in this case five) are included as an offset during model development in order for SPFs to predict annual crashes.

The project team considered several functional forms for AADT. In each case, segment length was considered as an offset variable (i.e., the parameter for segment length is equal to 1.0). The following functional forms were considered:

- Exponential Function = $Crashes = L \times e^{(\alpha + \beta_1 \times AADT + D_i)}$
- Power Function = $Crashes = L \times AADT^{\beta_1} \times e^{(\alpha + D_i)}$
- Hoerl Function = $Crashes = L \times AADT^{\beta_1} \times e^{(\alpha + \beta_2 \times AADT + D_i)}$
- Polynomial Function = $Crashes = L \times e^{(\alpha + \beta_1 \times AADT + \beta_2 \times AADT^2 + D_i)}$

Each equation includes a coefficient for an ODOT District (1 through 12). If the segment is located within that district, the coefficient is used within the prediction as shown. If no coefficient is provided for a specific district for any crash type, then that district is assumed to be a part of the baseline for that crash type. Factors are only applied when a value is present based on the District number.

Table 17. Sample Size and Summary Statistics for Base Segments.

Site Subtype	Number of Segments	Total Mileage	Crash Sample Size	Variable	Mean	SD	Min	Max
Rural 4 Lane Total	281	571.05	15,504	AADT	23,769	11,565	3,218	50,679
				Length	2.03	1.93	0.01	10.69
				Tot_Crash	55.17	59.32	0	317
Rural 4 Lane Fatal & Injury	290	579.23	3,330	AADT	24,442	11,965	3,218	52,049
				Length	2.00	1.92	0.01	10.69
				FI_Crash	11.48	13.16	0	99
Rural 6 Lane Total	71	209.54	11,314	AADT	45,738	8,633	32,701	83,544
				Length	2.95	2.87	0.01	13.72
				Tot_Crash	159.35	163.52	0	814
Rural 6 Lane Fatal & Injury	72	209.55	2,240	AADT	45,956	8,770	32,701	83,544
				Length	2.91	2.87	0.01	13.72
				FI_Crash	31.11	32.51	0	136
Urban 4 Lane Total	499	353.20	14,442	AADT	31,666	15,972	4,966	82,241
				Length	0.71	0.74	0.01	6.02
				Tot_Crash	28.94	33.06	0	266
Urban 4 Lane Fatal & Injury	500	363.61	3,361	AADT	31,731	15,962	4,966	82,241
				Length	0.73	0.83	0.01	11.18
				FI_Crash	6.72	8.12	0	61
Urban 6 Lane Total	250	192.31	15,716	AADT	70,726	25,040	22,948	143,661
				Length	0.77	0.93	0.01	6.30
				Tot_Crash	62.86	70.65	0	551
Urban 6 Lane Fatal & Injury	243	171.88	3,510	AADT	70,991	25,640	22,948	143,661
				Length	0.71	0.93	0.01	9.81
				FI_Crash	14.44	14.98	0	110
Urban 8+ Lane Total	76	42.21	6,644	AADT	106,798	29,841	35,200	172,818
				Length	0.56	0.51	0.01	2.52
				Tot_Crash	87.42	87.20	0	390
Urban 8+ Lane Fatal & Injury	76	42.21	1,731	AADT	106,798	29,841	35,200	172,818
				Length	0.56	0.51	0.01	2.52
				FI_Crash	22.78	24.42	0	126

Table 18. Sample Size and Summary Statistics for Interchange Segments.

Site Subtype	Number of Segments	Total Mileage	Crash Sample	Variable	Mean	SD	Min	Max
Rural 4 Lane Int Total	303	95.75	3,427	Freeway AADT	23,021	11,240	5,366	49,573
				Ramp AADT	8,754	9,310	614	70,196
				Length	0.32	0.19	0.01	1.02
				Tot_Crash	11.31	10.08	0	51
Rural 4 Lane Int Fatal & Injury	304	97.56	714	Freeway AADT	23,227	11,361	5,366	49,573
				Ramp AADT	8,819	9,273	614	70,196
				Length	0.32	0.19	0.01	1.02
				FI_Crash	2.35	2.61	0	18
Rural 6 Lane Int Total	76	18.72	1,439	Freeway AADT	48,637	9,784	32,701	83,544
				Ramp AADT	21,900	19,538	1,526	89,947
				Length	0.25	0.20	0.01	1.06
				Tot_Crash	18.93	21.41	0	122
Rural 6 Lane Int Fatal & Injury	76	18.72	292	Freeway AADT	48,637	9,784	32,701	83,544
				Ramp AADT	21,900	19,538	1,526	89,947
				Length	0.25	0.20	0.01	1.06
				FI_Crash	3.84	5.18	0	30
Urban 4 Lane Int Total	1,260	330.56	25,957	Freeway AADT	32,906	18,698	5,126	121,172
				Ramp AADT	31,277	45905	336	424,498
				Length	0.26	0.19	0.01	1.15
				Tot_Crash	20.60	23.87	0	187
Urban 4 Lane Int Fatal & Injury	1,243	328.68	25,157	Freeway AADT	32,595	18,456	5,126	121,172
				Ramp AADT	30,404	44,196	336	424,498
				Length	0.26	0.19	0.01	1.15
				FI_Crash	4.80	6.19	0	45
Urban 5 & 6 Lane Int Total	723	198.55	31,825	Freeway AADT	75,540	28,956	10,553	166,075
				Ramp AADT	69,603	71,080	2,740	424,498
				Length	0.27	0.20	0.01	1.31
				Tot_Crash	44.02	42.33	0	269
Urban 5 & 6 Lane Int Fatal & Injury	819	250.70	11,320	Freeway AADT	77,310	29,261	10,553	166,075
				Ramp AADT	71,188	69,435	2,740	424,498
				Length	0.31	0.22	0.01	1.42
				FI_Crash	13.82	14.89	0	105
Urban 7+ Lane Total	394	114.38	29,465	Freeway AADT	113,086	29,033	40,235	178,276
				Ramp AADT	101,783	79,038	2,384	424,498
				Length	0.29	0.22	0.01	1.71
				Tot_Crash	74.83	66.39	0	478
Urban 7+ Lane Fatal & Injury	393	118.96	8,229	Freeway AADT	113,634	28,939	40,235	178,276
				Ramp AADT	102,205	79,552	2,384	424,498
				Length	0.30	0.21	0.01	1.71
				FI_Crash	20.99	19.19	0	93

In each case, the project team estimated the SPFs considering a variable dispersion parameter that is a function of the segment length. Hauer (2001) identified the possibility that dispersion can logically be related to the segment length; shorter segments will tend of have a higher dispersion than longer segments. Further research by Cafiso et al. (2010) used the logarithm of segment length in the log-dispersion model, resulting in the functional form shown in the following equation:

$$Dispersion = \ Constant * L^{\beta_3} \qquad \text{Equation 12}$$

The value of the constant is the exponent of the coefficient from the model of the logarithm of dispersion. If the coefficient of the segment length is negative, the model shows that dispersion is inversely related to the length of the segment, which is consistent with the logic established by Hauer (2001).

Table 19 presents the estimated parameters for all base segments and for urban 7+-lane interchange segments. For these facility types, the project team found the Hoerl Model to best fit the network screening-level data. For these site subtypes, only freeway AADT is included. For urban 7+-lane interchange segments, ramp AADT was not found to be associated with crash outcomes. District indicators are also included in some cases, providing for an increase or decrease in predicted crashes due to underlying factors specific to the district. For those indicators, the baseline is the districts not included in the SPF. The indicators do not change the shape of the SPF, but shift it up or down slightly, depending on the direction of effect.

Table 20 presents the estimated parameters for all interchange segments (other than urban 7+ lane segments). For these segments, the project team found the power function to have the best fit to the data, and the SPF includes both freeway segment AADT and the total AADT from all associated interchange ramps (i.e., the sum of entering and exiting vehicles in both travel directions). The following equation presents the final model functional form:

$$Crashes = L \times Freeway\ AADT^{\beta_4} \times Total\ Ramp\ AADT^{\beta_5} \times e^{(\alpha + D_i)} \qquad \text{Equation 13}$$

As with base segments, the project team included District indicators to capture unobserved adjustments. Additionally, the variable dispersion parameter was significant and is calculated in the same manner as for base segments.

When looking at Table 19, the coefficients are unstable in general (across site subtypes) due to multicollinearity when including log-AADT and AADT. However, including both as a Hoerl function improved the CURE plots (and overall model fit) compared to the other functional forms (This is discussed in greater detail in the Network Screening SPF Validation Results section). Due to the multicollinearity, parameter statistical significance is inconsistent from site subtype to site subtype. However, the main effect of log-AADT was significant for all site-subtypes when considered alone. Ultimately, the VHB team felt that the improved predictive reliability warranted using the Hoerl model as the final functional form for base segment SPFs. The Hoerl function allows for a bend in the SPF function and maintains the assumption that no crashes are expected when no traffic is present. The exponential and polynomial functional forms do not hold this assumption.

Table 19. Hoerl Model SPFs for Base Freeway Segments and Urban 7+-Lane Interchange Segments.

Site Subtype	α	B_1	B_2*	D2	D3	D4	D5	D6	D7	D8	D12	Dispersion	
												Const	B_3
Base Freeway Segments													
Rural 4-Lane Total	-2.147	0.341	0.019	N/A	N/A	N/A	-0.128	-0.297	N/A	-0.262	N/A	0.091	-0.597
Rural 4-Lane FI	-2.777	0.198	0.036	N/A	N/A	N/A	N/A	-0.287	N/A	-0.301	N/A	0.135	-0.649
Rural 6-Lane Total	0.592	0.092	0.021	N/A	-0.220	-0.185	N/A	-0.319	-0.237	N/A	N/A	0.104	-0.906
Rural 6-Lane FI	18.772	-1.964	0.070	N/A	-0.143	N/A	N/A	-0.348	-0.421	N/A	N/A	0.140	-0.994
Urban 4-Lane Total	-1.475	0.282	0.021	-0.194	N/A	N/A	N/A	-0.452	N/A	N/A	-0.192	0.109	-0.819
Urban 4-Lane FI	-4.353	0.420	0.019	N/A	N/A	N/A	N/A	-0.219	N/A	N/A	N/A	0.154	-0.656
Urban 6-Lane Total	-1.134	0.280	0.014	N/A	N/A	N/A	N/A	-0.411	-0.151	N/A	-0.384	0.084	-0.826
Urban 6-Lane FI	-0.916	0.103	0.018	N/A	N/A	N/A	N/A	-0.218	N/A	N/A	N/A	0.107	-0.608
Urban 8+ Lane Total	30.573	-2.816	0.047	0.546	N/A	N/A	N/A	N/A	N/A	0.549	N/A	0.039	-1.129
Urban 8+ Lane FI	21.505	-2.082	0.041	0.424	N/A	N/A	N/A	N/A	N/A	0.339	N/A	0.051	-0.898
Interchange Freeway Segments													
Urban 7+ Lane Total	-14.771	1.659	-0.006	0.186	N/A	0.277	N/A	N/A	N/A	0.210	N/A	0.106	-0.673
Urban 7+ Lane FI	-17.919	1.830	-0.007	N/A	N/A	0.247	N/A	N/A	N/A	N/A	N/A	0.128	-0.442

Note: *Indicates in thousands AADT (AADT/1,000).

Figures 7 through 11 provide graphical representations of the SPFs for each site subtype. These graphics include the SPFs for both total and fatal and injury crashes for base and interchange segments. Figure 12 provides the graphical representations of rural freeway SPFs and Figures 13 and 14 provide graphical representations of urban total and fatal and injury crashes, respectively.

Table 20. Power Model SPFs for Interchange Segments.

Site Subtype	α	B_4	B_5	D5	D7	D8	Dispersion	
							Const	B_3
Rural 4-Lane Total	-5.106	0.595	0.136	-0.288	-0.212	-0.171	0.052	-0.855
Rural 4-Lane FI	-6.970	0.565	0.196	N/A	-0.471	N/A	0.012	-2.105
Rural 6-Lane Total	-8.942	0.983	0.107	N/A	0.292	N/A	0.014	-1.592
Rural 6-Lane FI	-15.621	1.385	0.179	N/A	0.399	N/A	0.002	-2.169
Urban 4-Lane Total	-7.086	0.721	0.235	-0.173	N/A	N/A	0.109	-0.818
Urban 4-Lane FI	-8.815	0.752	0.230	-0.181	N/A	N/A	0.144	-0.710
Urban 5/6-Lane Total	-6.637	0.767	0.135	N/A	N/A	0.282	0.063	-0.870
Urban 5/6-Lane FI	-8.683	0.841	0.130	N/A	N/A	N/A	0.185	-0.342

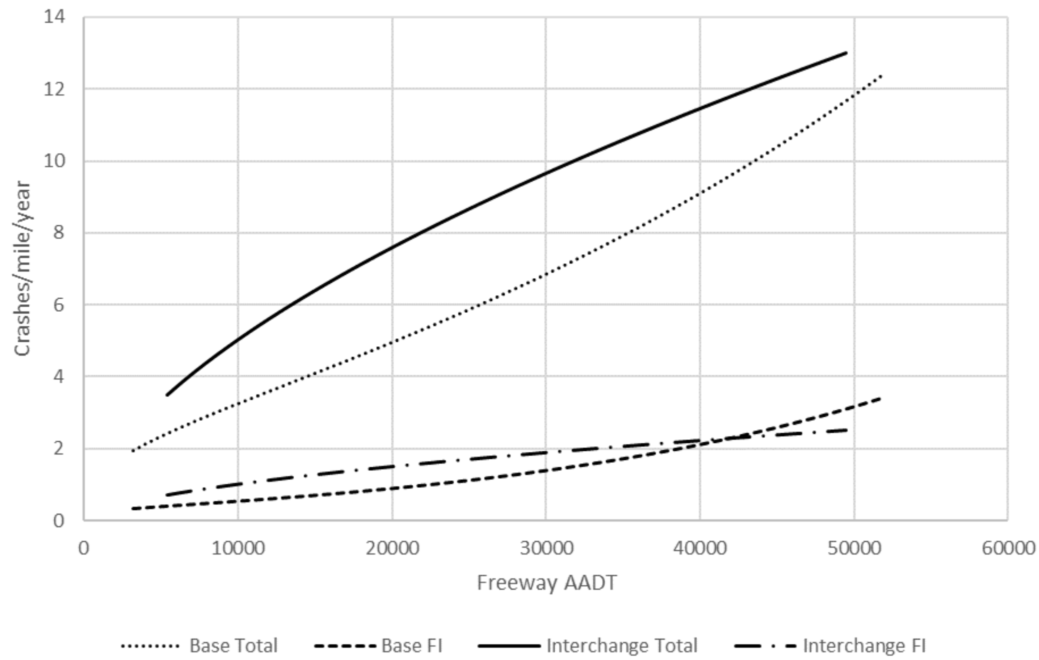

Figure 7. Rural Four Lane Freeway SPFs.

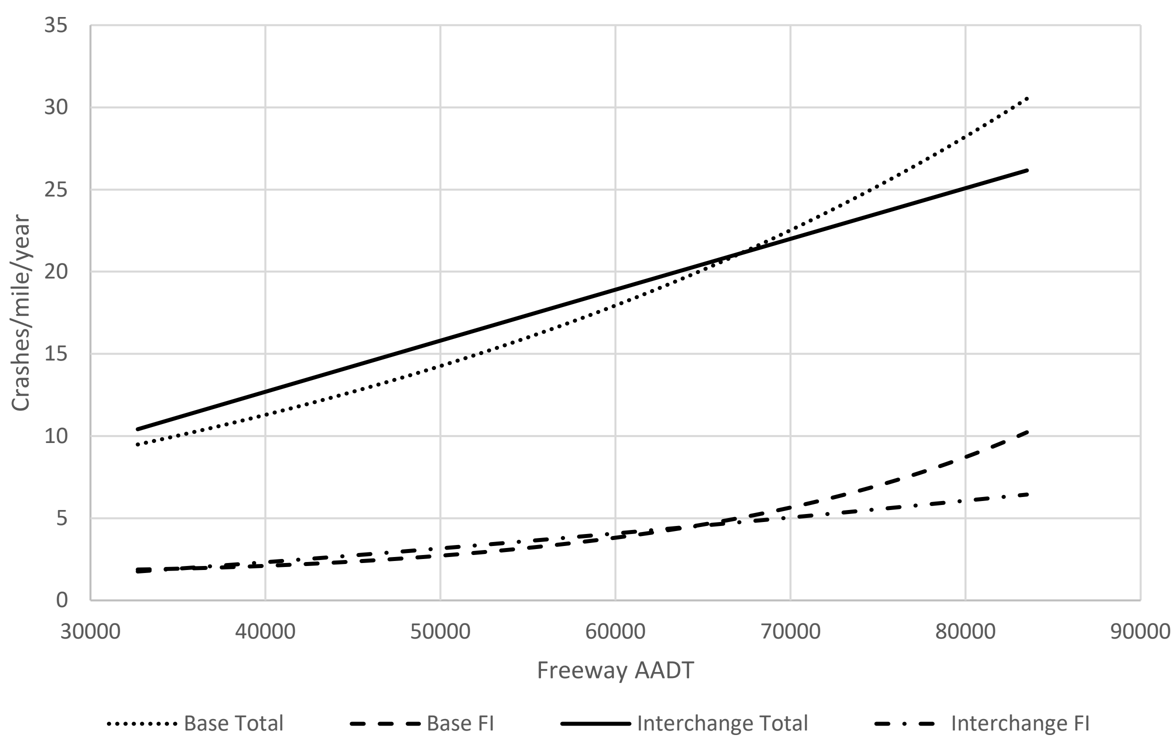

Figure 8. Rural Six Lane Freeway SPFs.

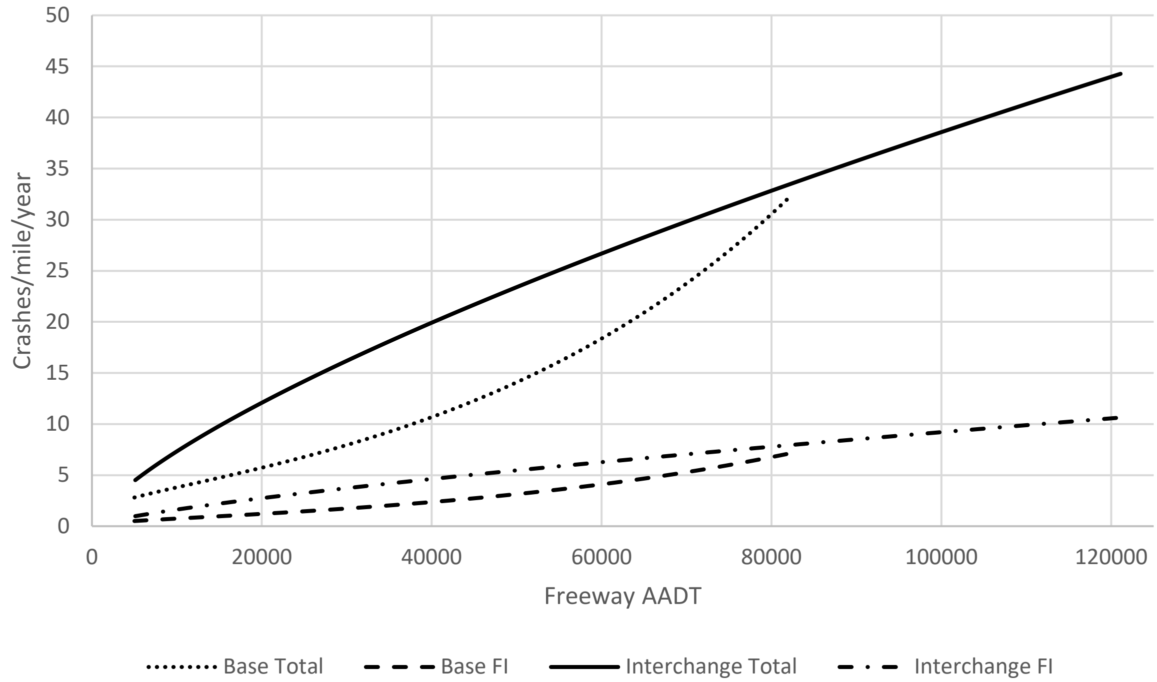

Figure 9. Urban Four Lane Freeway SPFs.

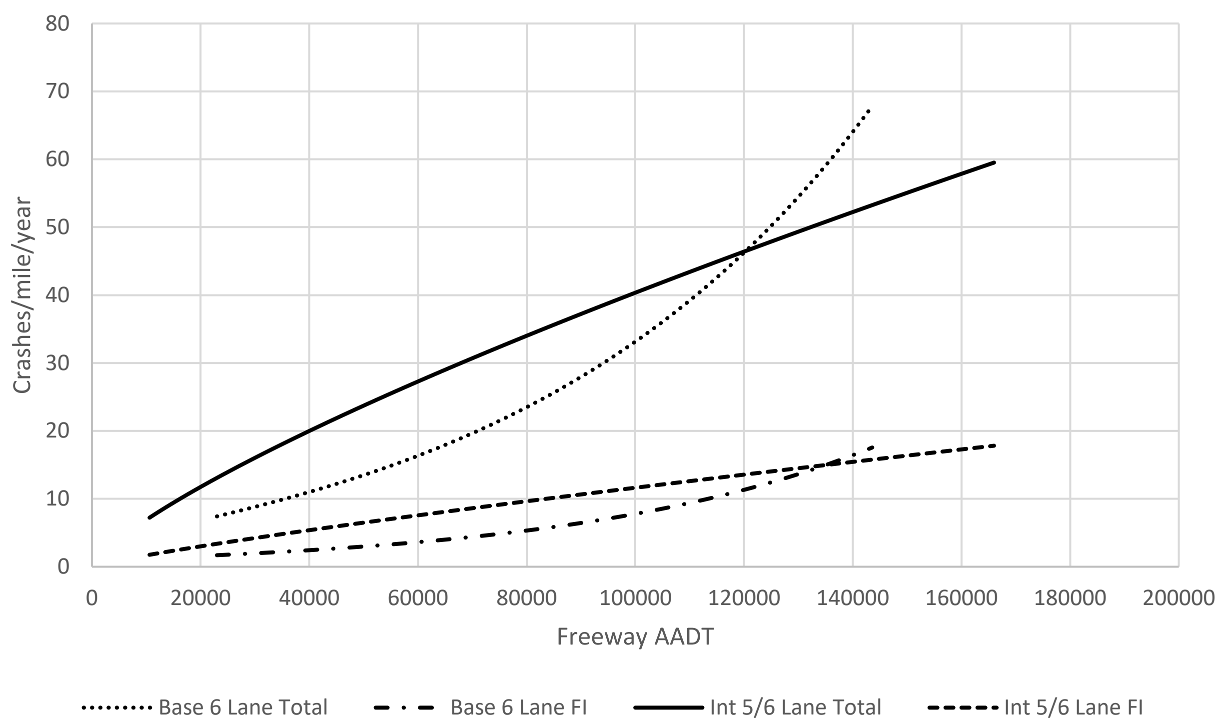

Figure 10. Urban Five/Six Lane Freeway SPFs.

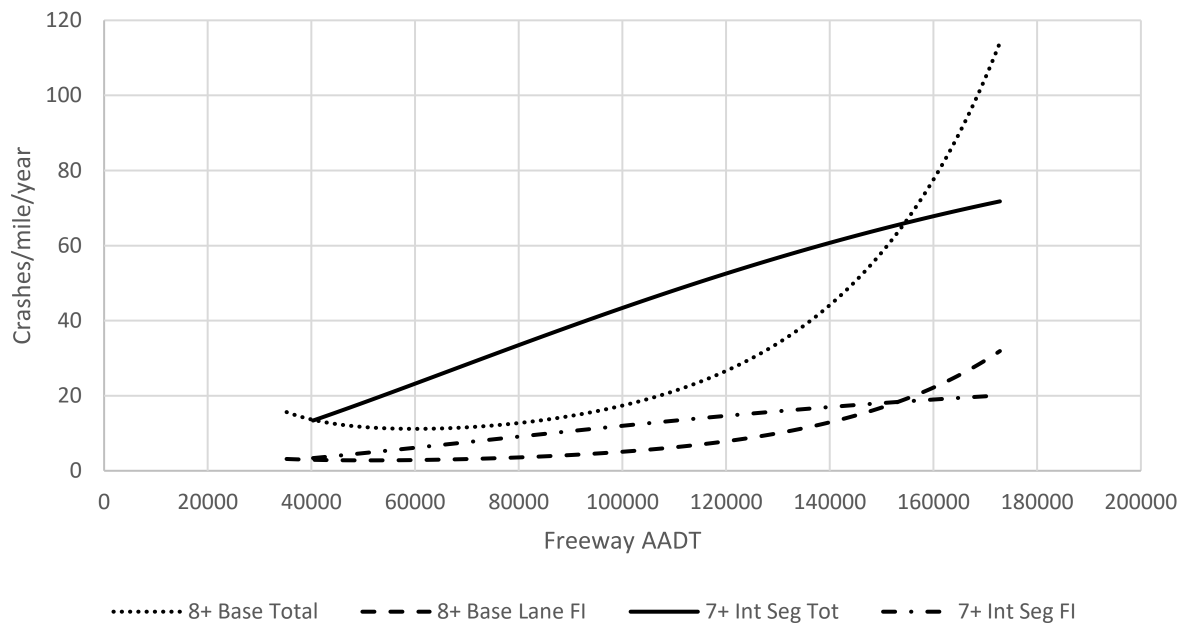

Figure 11. Urban Seven/Eight Lane Freeway SPFs.

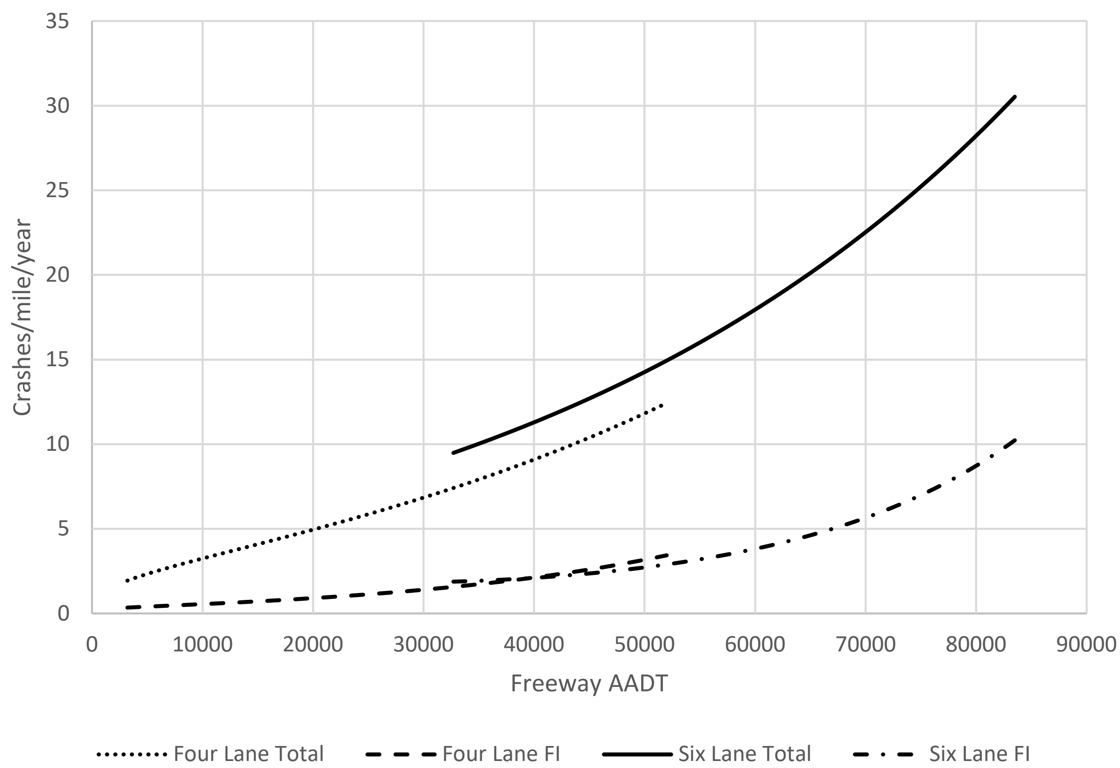

Figure 12. Rural Freeway SPFs.

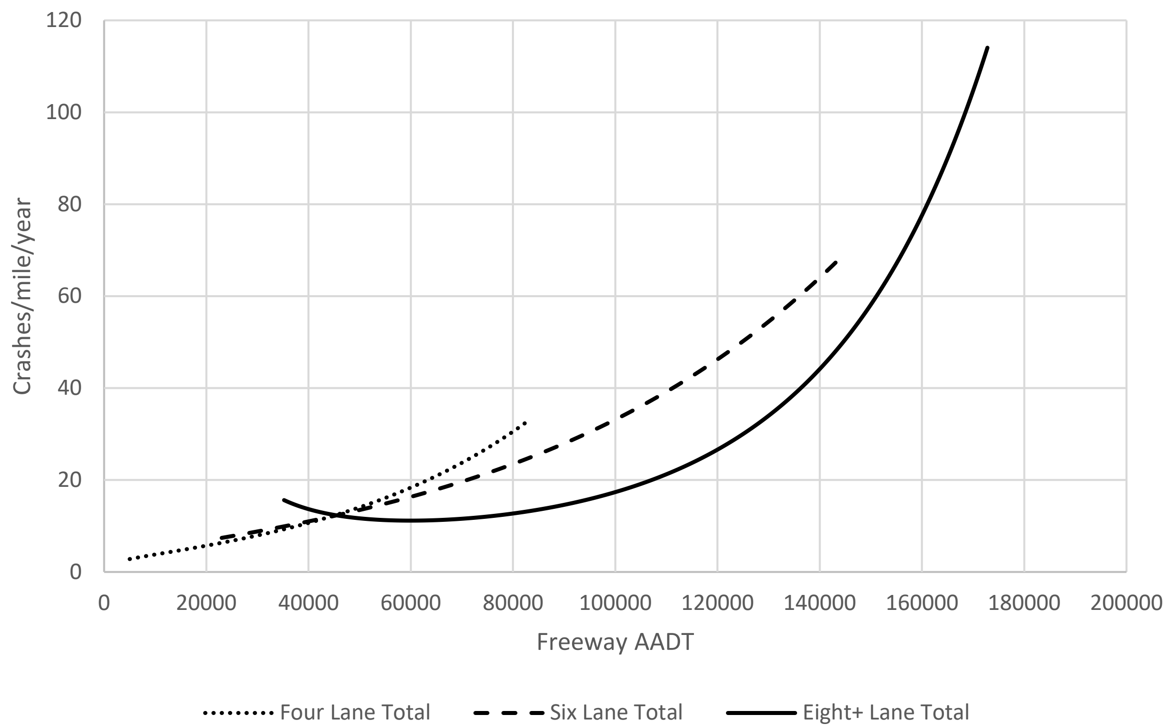

Figure 13. Urban Freeway Total Crash SPFs.

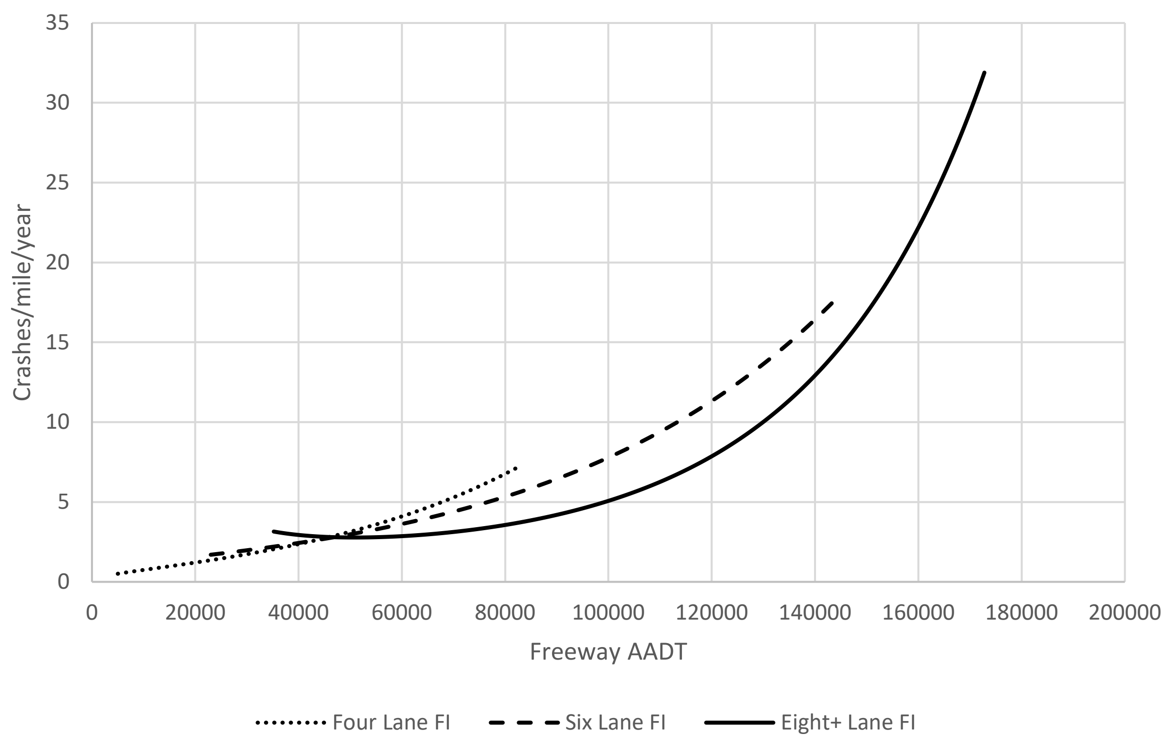

Figure 14. Urban Freeway Fatal and Injury Crash SPFs.

Appendix C: Planning-Level SPF Validation

This appendix presents the process for validating planning-level SPFs for network screening. While developing planning-level SPFs, the project team evaluated the fit of the predictive models for each of the functional forms considered, including the following:

- Exponential Function = $Crashes = L \times e^{(\alpha+\beta_1 \times AADT)}$
- Power Function = $Crashes = L \times AADT^{\beta_1} \times e^{\alpha}$
- Hoerl Function = $Crashes = L \times AADT^{\beta_1} \times e^{(\alpha+\beta_2 \times AADT)}$
- Polynomial Function = $Crashes = L \times e^{(\alpha+\beta_1 \times AADT+\beta_2 \times AADT^2)}$

The project team used the following GOF measures to evaluate each SPF functional form:

- MAD: measures the average magnitude of variability of prediction. Smaller values are preferred to larger values in comparing two or more competing SPFs. Note that the results presented in this section are for five-year aggregate crash counts at each segment.
- Modified R^2: measures the amount of systematic variation explained by the SPF. Larger values indicate a better fit to the data in comparing two or more competing SPFs.
- Log-Likelihood: measures the goodness of fit of a statistical model to the sample of data for given values of the unknown parameters. Smaller values of log-likelihood are preferred to larger values in comparing two or more competing SPFs. Traditionally, the number of parameters should be compared when considered two or more models, but since the project team focused on the same variables in each model, the number of estimated parameters is the same in each comparison.
- CURE plots: CURE plots (See Figure 15 for an example) provide a graphical representation of cumulative residuals (which are observed crashes minus predicted crashes for each segment) against a variable of interest sorted in ascending order (e.g., major road traffic volume). CURE plots help to identify the following concerns:
    - Long trends: trends in the CURE plot (increasing or decreasing) indicate regions of bias.
    - Percent exceeding the confidence limits: cumulative residuals outside the confidence limits indicate a poor fit over that range of the variable. Cumulative residuals frequently outside the confidence limits indicate a

notable bias in the SPF. A reasonable upper threshold for the percent of the CURE plot exceeding the 95 percent limits is 5 percent.

- Vertical changes: large vertical changes in the CURE plot are potential indicators of outliers, which require further examination.

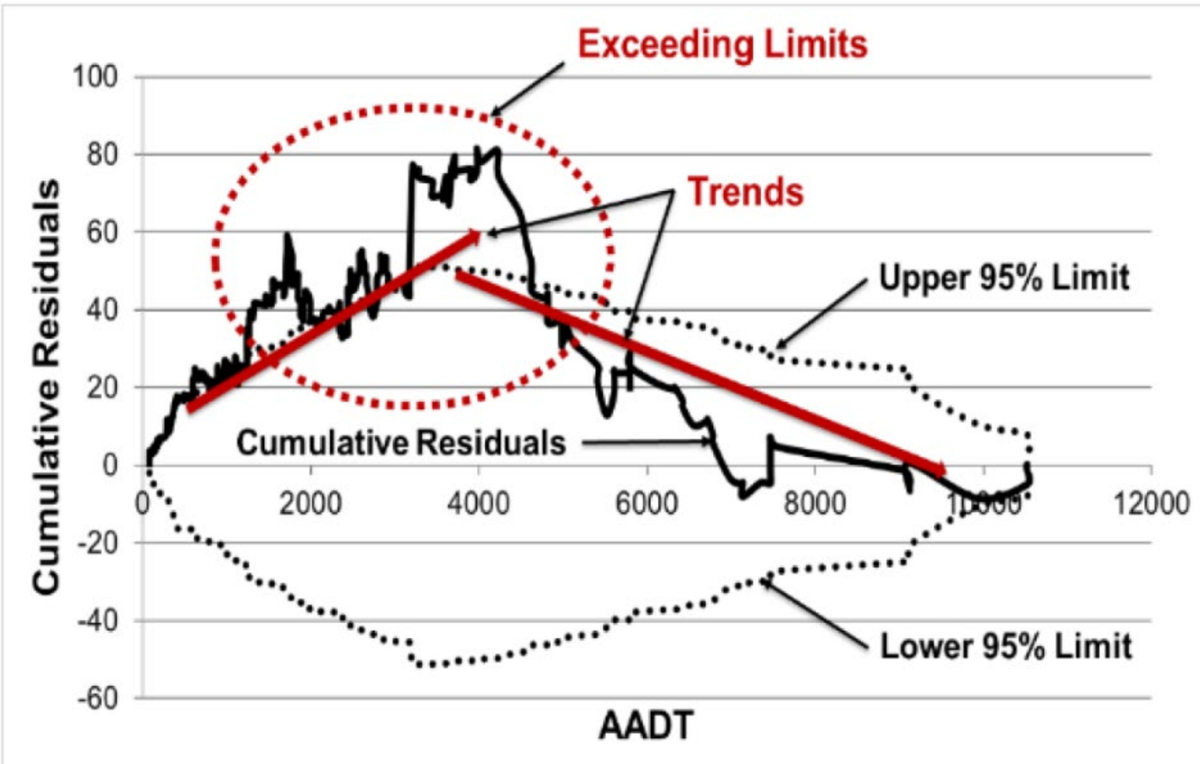

Figure 15. Example CURE Plot.

The project team further considered evaluating the mean absolute percentage error (MAPE) for each model; however, the MAPE (for which a smaller value is better) cannot be used when there are zero values because there would be a division by zero. For this reason, the MAPE was excluded from analysis.

This section provides an overview of the model fit statistics for each SPF functional form. The CURE plots are only provided for the final functional form and model, but each subsection provides the summary statistics for each model estimated. For freeway interchange segments, functional forms were not evaluated and compared, as the power function was used to incorporate both freeway mainline and ramp traffic volumes. This section provides the CURE plots for the final freeway interchange segment models for validation purposes.

Rural Four Lane Freeway Segment Total Crashes

Table 21 provides the GOF measures for each model functional form for rural four-lane freeway segment total crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 21. Rural Four Lane Freeway Segment Total Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	11.62	0.103	-1031.84
Power	11.81	0.102	-1032.63
Hoerl	11.48	0.106	-1028.30
Polynomial	11.37	0.107	-1027.51

The results indicated the Hoerl and polynomial functions provided the best fits of the data. The polynomial model suffers from some foundational issues (e.g., zero traffic volume does not result in zero predicted crashes) and does not provide a substantial benefit over the Hoerl model form. Additionally, as shown in Figure 16, the CURE plot for the Hoerl model indicates a good fit of the data, only 3 percent of the observations exceed the 95 percent confidence intervals, and there is some oscillation about the zero line, indicating only minor systematic bias in predictions against AADT.

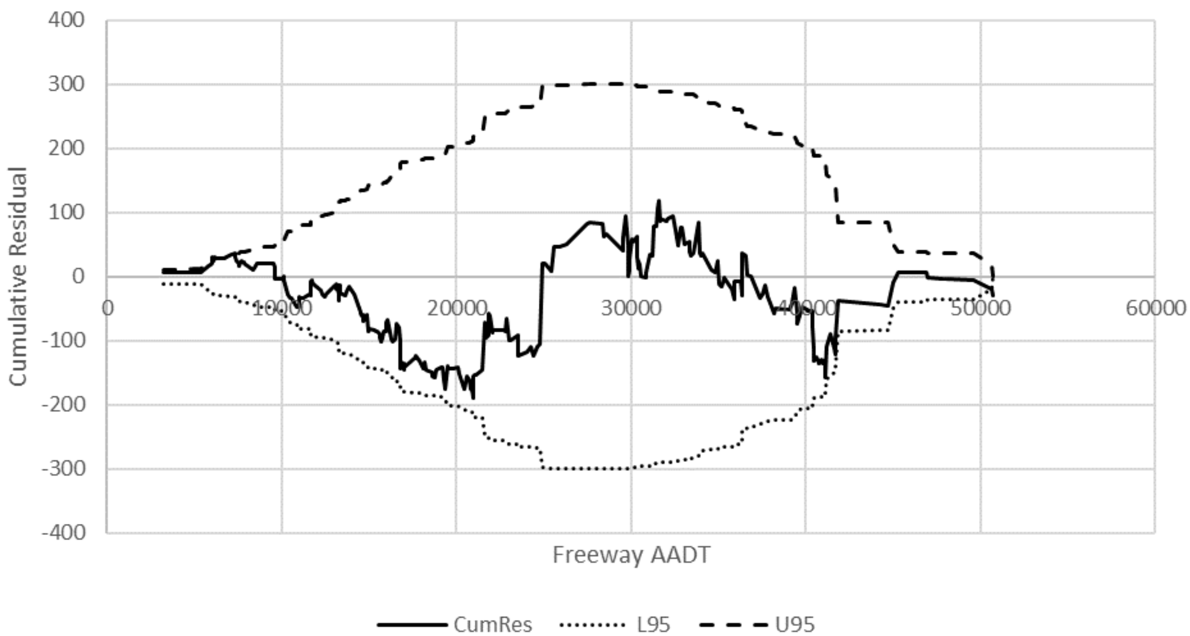

Figure 16. CURE Plot for Rural Four Lane Freeway Segment Total Crashes SPF.

Rural Four Lane Freeway Segment FI Crashes

Table 22 provides the summary statistics for each model functional form for rural four lane freeway segment FI crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 22. Rural Four Lane Freeway Segment FI Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	3.72	0.116	-744.53
Power	3.80	0.114	-745.98
Hoerl	3.53	0.127	-734.65
Polynomial	3.72	0.117	-743.09

The results indicated the Hoerl function provided the best fit of the data. As shown in Figure 17, the CURE plot for the Hoerl model indicates a good fit of the data, only 2 percent of the observations exceed the 95 percent confidence intervals, and there is general oscillation about the zero cumulative residual. The area where the observations exceed the confidence intervals is about 40,000 AADT, for which very few sites were observed.

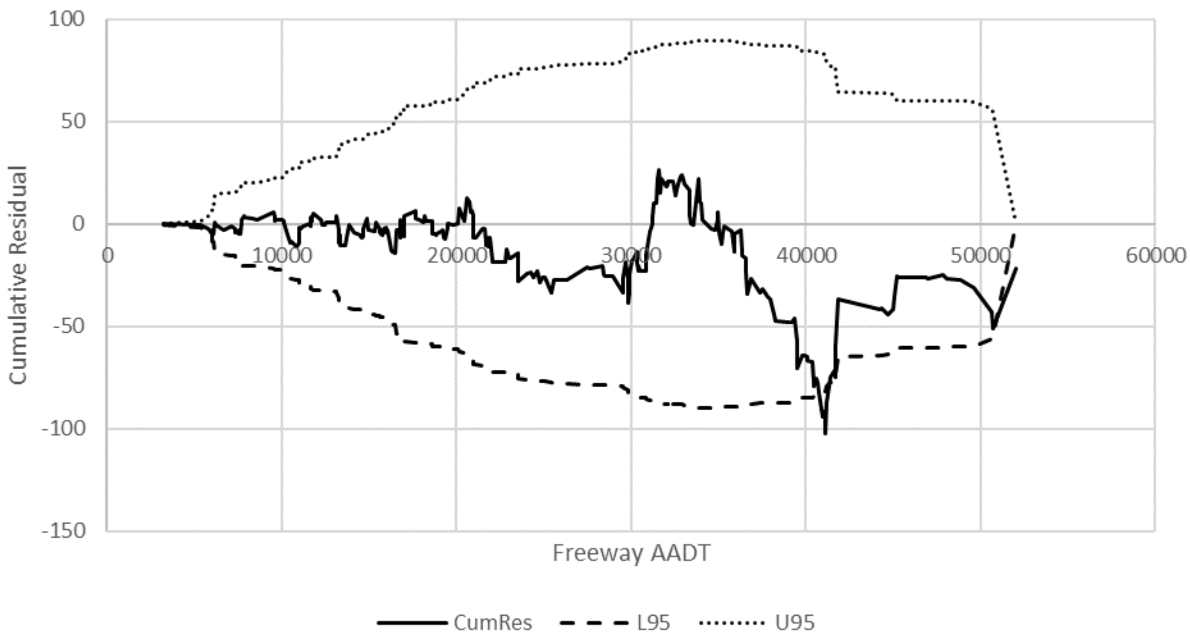

Figure 17. CURE Plot for Rural Four Lane Freeway Segment FI Crashes SPF.

Rural Six Lane Freeway Segment Total Crashes

Table 23 provides the summary statistics for each model functional form for rural six-lane freeway segment total crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 23. Rural Six Lane Freeway Segment Total Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	24.02	0.053	-322.94
Power	23.79	0.052	-323.30
Hoerl	23.75	0.053	-322.94
Polynomial	23.65	0.053	-322.92

The results indicated a consistent fit of the data across all model functional forms. The inclusion of the district indicators seems to have more impact on the results than the functional form in this case. As shown in Figure 18, the CURE plot for the Hoerl model indicates a good fit of the data, only 4 percent of the observations exceed the 95 percent confidence intervals, and there is general oscillation about the zero cumulative residual. The area where the observations exceed the confidence intervals is above 60,000 AADT, for which very few sites were observed.

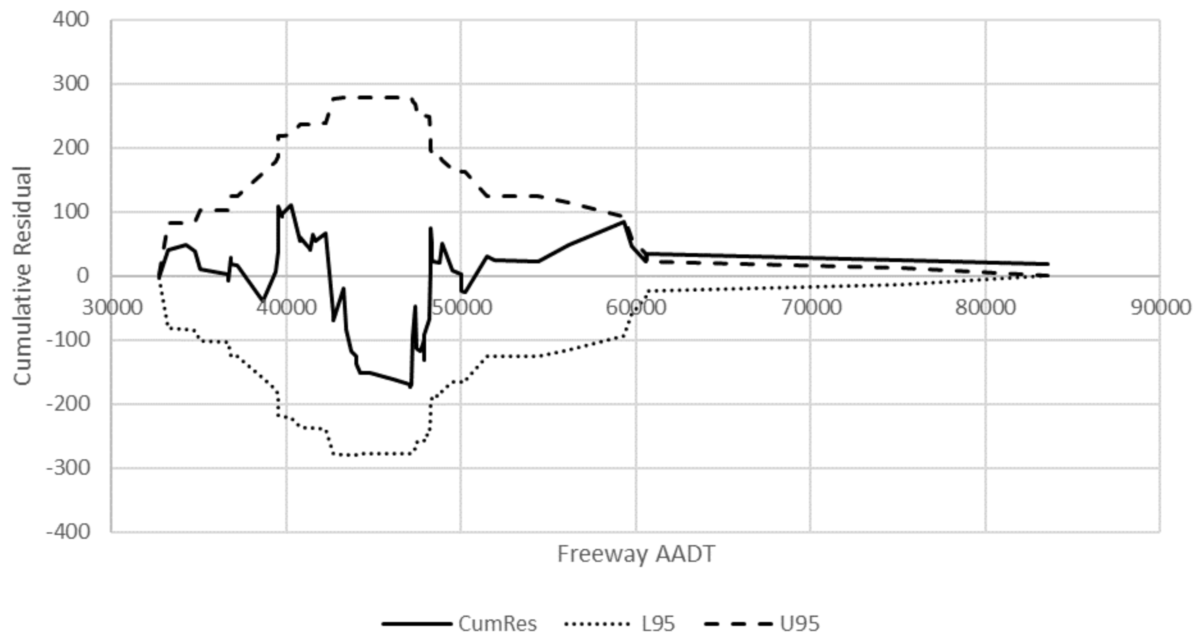

Figure 18. CURE Plot for Rural Six Lane Freeway Segment Total Crashes SPF.

Rural Six Lane Freeway Segment FI Crashes

Table 24 provides the summary statistics for each model functional form for rural six-lane freeway segment FI crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 24. Rural Six Lane Freeway Segment FI Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	6.21	0.058	-230.40
Power	6.57	0.052	-231.86
Hoerl	6.07	0.061	-229.67
Polynomial	6.14	0.060	-229.93

The results indicated a consistent fit of the data across all model functional forms; however, the Hoerl model provided the best overall validation statistics. As shown in Figure 19, the CURE plot for the Hoerl model indicates a good fit of the data, only the final observation exceeds the 95 percent confidence intervals, and there is general oscillation about the zero cumulative residual. The area where the observations exceed the confidence intervals is above 84,000 AADT, for which only one site was observed.

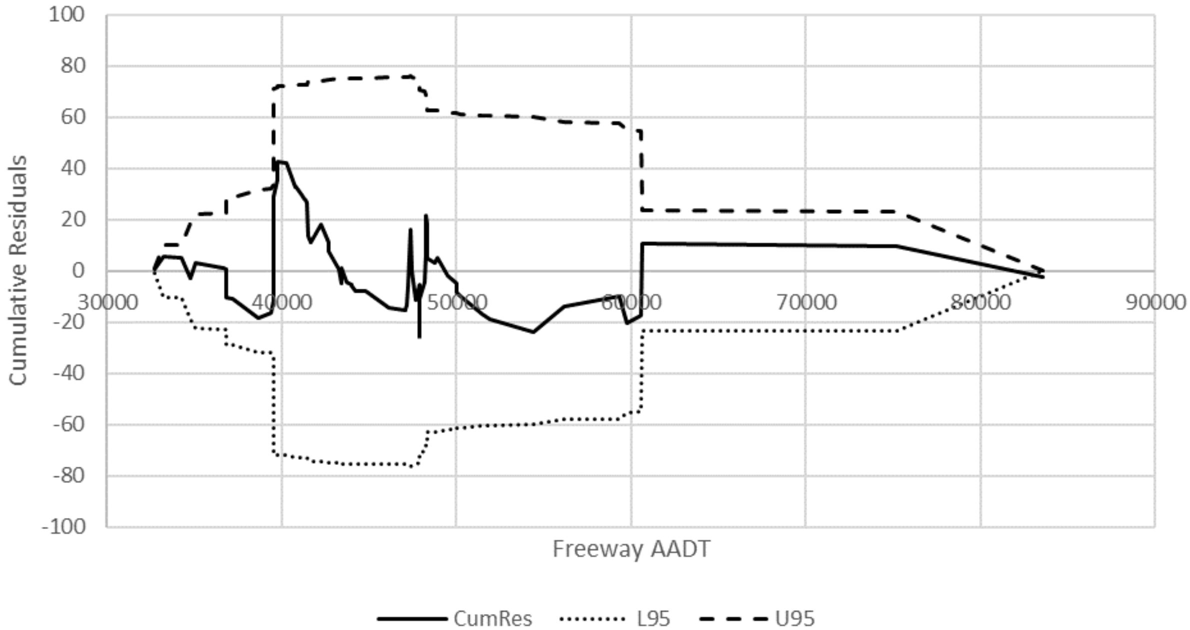

Figure 19. CURE Plot for Rural Six Lane Freeway Segment FI Crashes SPF.

Urban Four Lane Freeway Segment Total Crashes

Table 25 provides the summary statistics for each model functional form for urban four-lane freeway segment total crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 25. Urban Four Lane Freeway Segment Total Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	8.79	0.085	-1770.47
Power	9.14	0.081	-1778.53
Hoerl	8.79	0.086	-1768.45
Polynomial	8.85	0.086	-1768.91

The results indicated a consistent fit of the data across all model functional forms; however, the Hoerl model provided the best overall validation statistics. As shown in Figure 20, the CURE plot for the Hoerl model indicates a relatively good fit of the data, only the final six percent of observations exceed the 95 percent confidence intervals. There does seem to be some systematic bias between 15,000 and 30,000 AADT; however, the observations all fall well within the 95 percent confidence interval. The area where the observations exceed the confidence intervals is above 60,000 AADT, for which there were few sites.

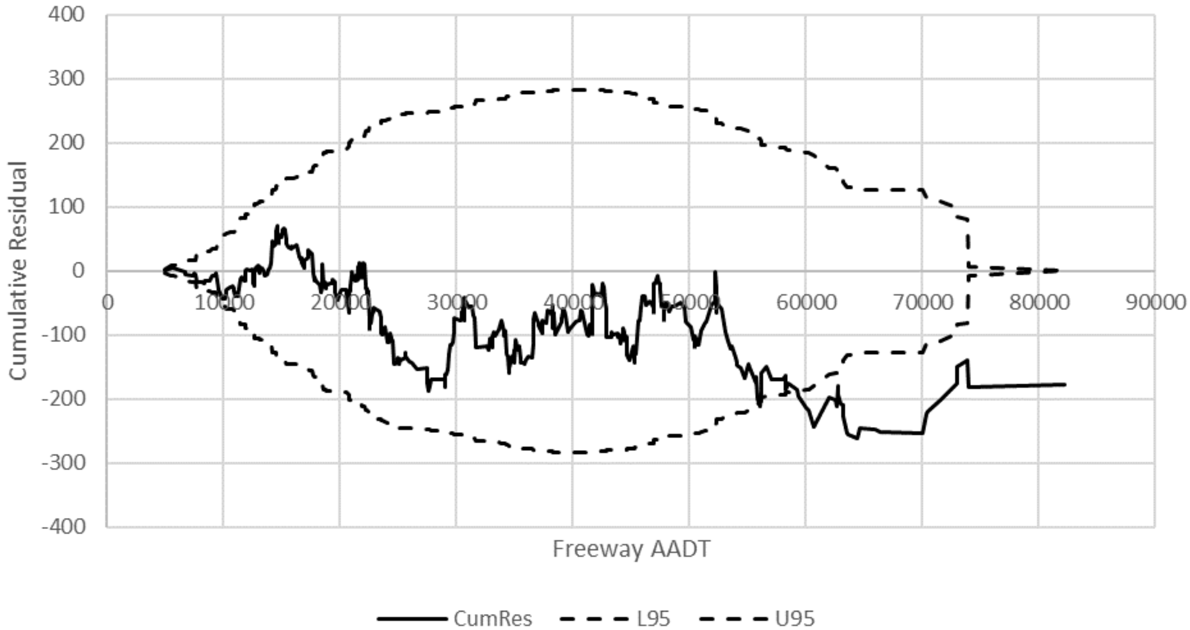

Figure 20. CURE Plot for Urban Four Lane Freeway Segment Total Crashes SPF.

Urban Four Lane Freeway Segment FI Crashes

Table 26 provides the summary statistics for each model functional form for urban four lane freeway segment FI crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 26. Urban Four Lane Freeway Segment FI Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	2.77	0.089	-1180.16
Power	2.77	0.088	-1182.00
Hoerl	2.76	0.091	-1178.01
Polynomial	2.76	0.091	-1178.09

The results indicated a consistent fit of the data across all model functional forms; however, the Hoerl model provided the best overall validation statistics. As shown in Figure 21, the CURE plot for the Hoerl model indicates a relatively good fit of the data, only the final three percent of observations exceed the 95 percent confidence intervals. There does seem to be some systematic bias between 15,000 and 30,000 AADT; however, the observations all fall well within the 95 percent confidence interval. The area where the observations exceed the confidence intervals is above 63,000 AADT, for which there were few sites.

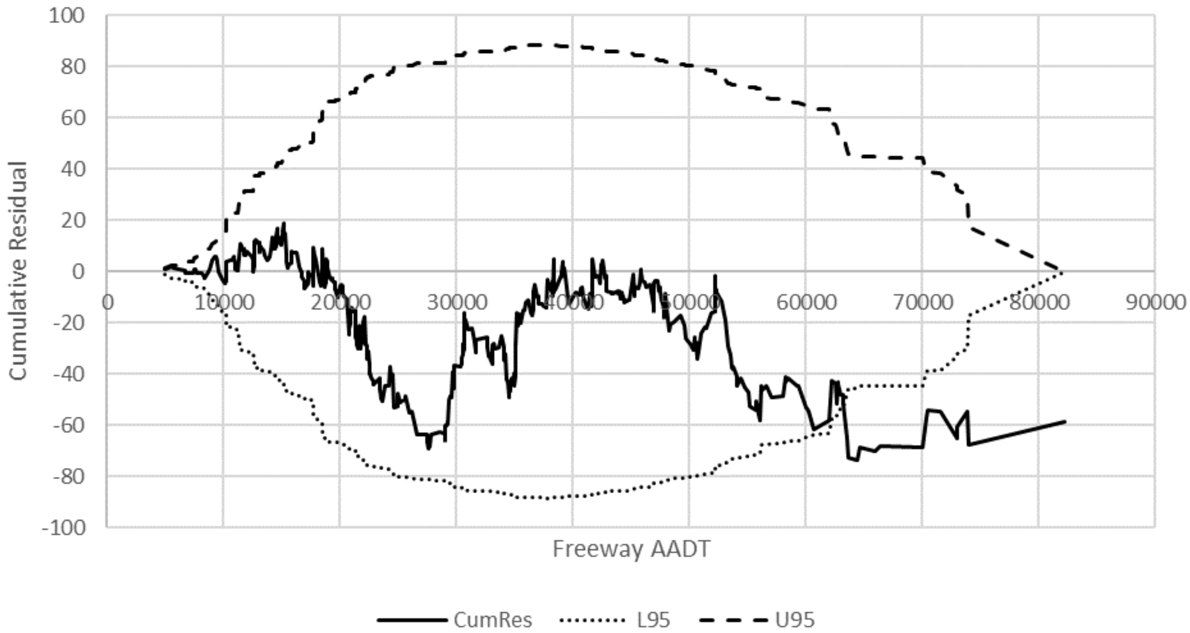

Figure 21. CURE Plot for Urban Four Lane Freeway Segment FI Crashes SPF.

Urban Six Lane Freeway Segment Total Crashes

Table 27 provides the summary statistics for each model functional form for urban six lane freeway segment total crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 27. Urban Six Lane Freeway Segment Total Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	16.02	0.081	-1050.72
Power	16.30	0.078	-1054.82
Hoerl	15.84	0.082	-1050.31
Polynomial	15.81	0.082	-1050.17

The results indicated a consistent fit of the data across all model functional forms, but the polynomial model provided the best overall validation statistics. However, as indicated with rural four-lane freeways, the Hoerl model was selected based on foundational theory and a lack of substantial improvement for the polynomial model. As shown in Figure 22, the CURE plot for the Hoerl model indicates a good fit of the data, only the final two percent of observations exceed the 95 percent confidence intervals. The area where the observations exceed the confidence intervals is above 130,000 AADT, for which there were few sites.

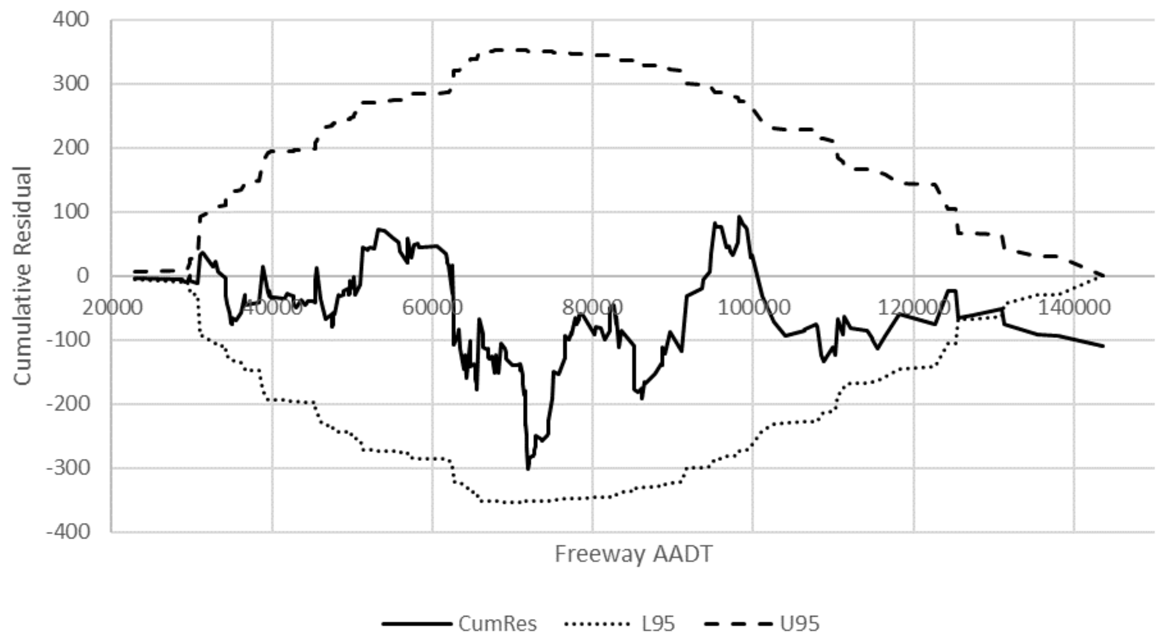

Figure 22. CURE Plot for Urban Six Lane Freeway Segment Total Crashes SPF.

Urban Six Lane Freeway Segment FI Crashes

Table 28 provides the summary statistics for each model functional form for urban six lane freeway segment FI crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 28. Urban Six Lane Freeway Segment FI Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	4.77	0.094	-725.05
Power	4.89	0.088	-729.56
Hoerl	4.76	0.094	-725.01
Polynomial	4.76	0.094	-725.02

The results indicated a consistent fit of the data across all model functional forms; however, the Hoerl model provided the best overall validation statistics. As shown in Figure 23, the CURE plot for the Hoerl model indicates a relatively good fit of the data; the final nine percent of observations exceed the 95 percent confidence intervals. The area where the observations exceed the confidence intervals is above 110,000 AADT, for which there were relatively few sites.

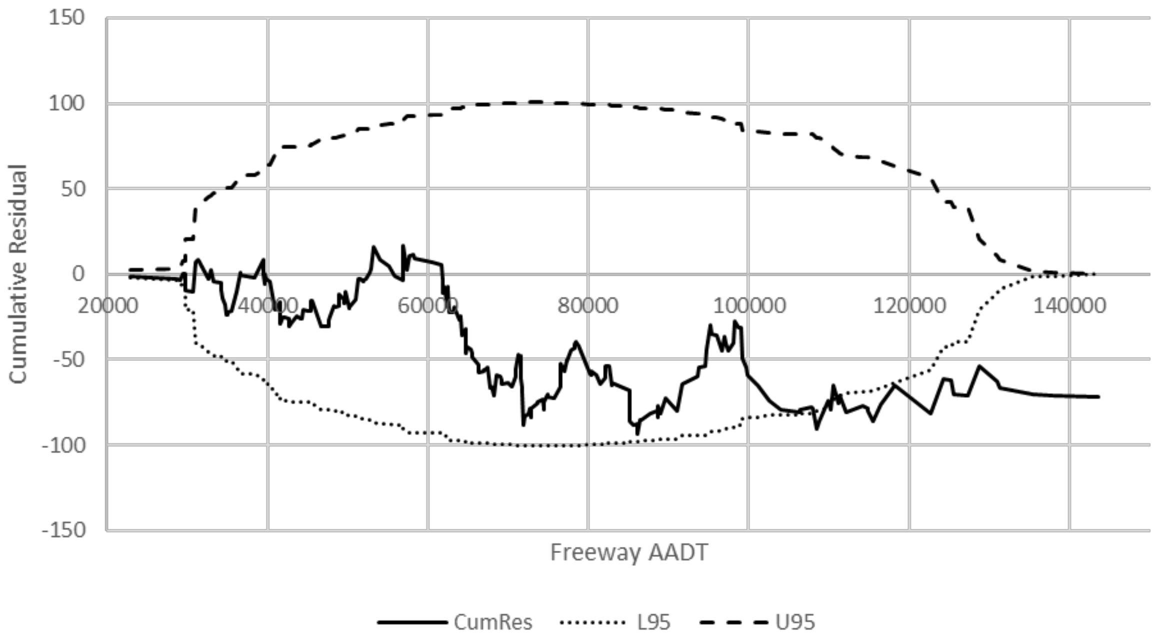

Figure 23. CURE Plot for Urban Six Lane Freeway Segment FI Crashes SPF.

Urban Eight or More Lane Freeway Segment Total Crashes

Table 29 provides the summary statistics for each model functional form for urban eight or more lane freeway segment total crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 29. Urban Eight or More Lane Freeway Segment Total Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	23.65	0.126	-330.52
Power	28.29	0.091	-343.54
Hoerl	19.53	0.066	-353.22
Polynomial	19.38	0.126	-330.40

The results indicated a consistent fit of the data for Hoerl and polynomial functional forms; however, in this case, the polynomial model provided the best overall validation statistics. As indicated with rural four-lane freeways, the Hoerl model was selected based on foundational theory and a lack of substantial improvement for the polynomial model. As shown in Figure 24, the CURE plot for the Hoerl model indicates a good fit of the data; only the final two observations exceed the 95 percent confidence intervals. The area where the observations exceed the confidence intervals is above 160,000 AADT, for which there were relatively few sites.

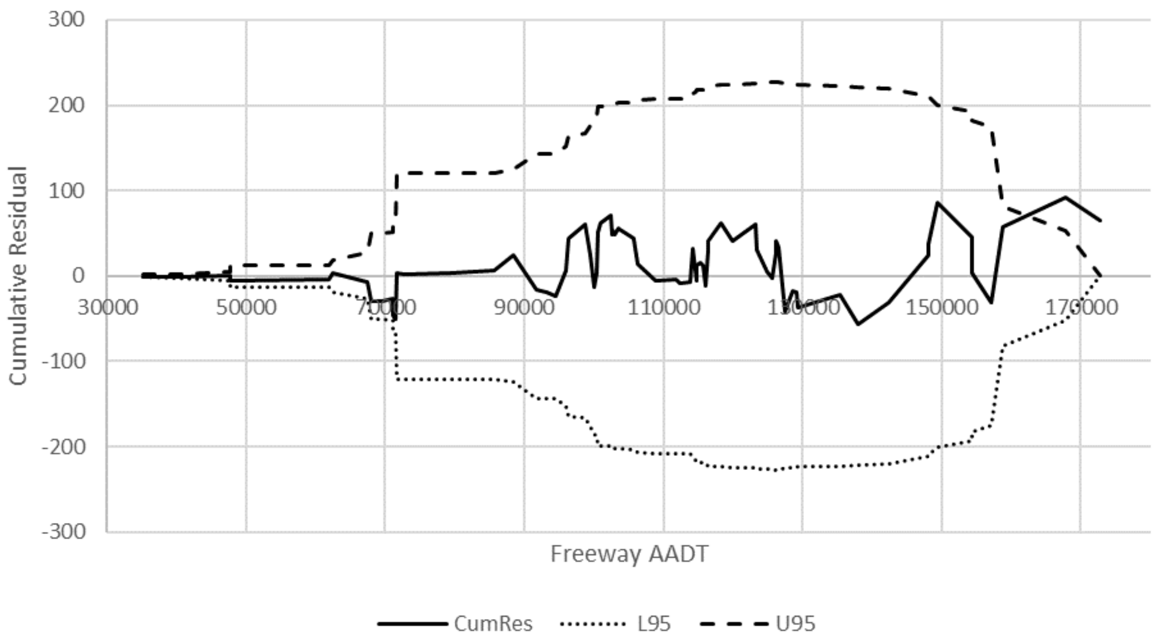

Figure 24. CURE Plot for Urban Eight or More Lane Freeway Segment Total Crashes SPF.

Urban Eight or More Lane Freeway Segment FI Crashes

Table 30 provides the summary statistics for each model functional form for urban eight or more lane freeway segment FI crashes. Note that the MAD provides an indication of the average error per segment over a five-year period. The MAD, divided by five, represents the annual error in crash prediction.

Table 30. Urban Eight or More Lane Freeway Segment Total Crashes GOF Measures.

Functional Form	MAD	Modified R^2	Log Likelihood
Exponential	6.72	0.118	-249.07
Power	7.59	0.091	-256.70
Hoerl	6.10	0.135	-244.48
Polynomial	5.97	0.136	-244.02

The results indicated a consistent fit of the data for Hoerl and polynomial functional forms; however, in this case, the polynomial model provided the best overall validation statistics. As indicated with rural four-lane freeways, the Hoerl model was selected based on foundational theory and a lack of substantial improvement for the polynomial model. As shown in Figure 25, the CURE plot for the Hoerl model indicates a good fit of the data; only the final observation exceeds the 95 percent confidence intervals.

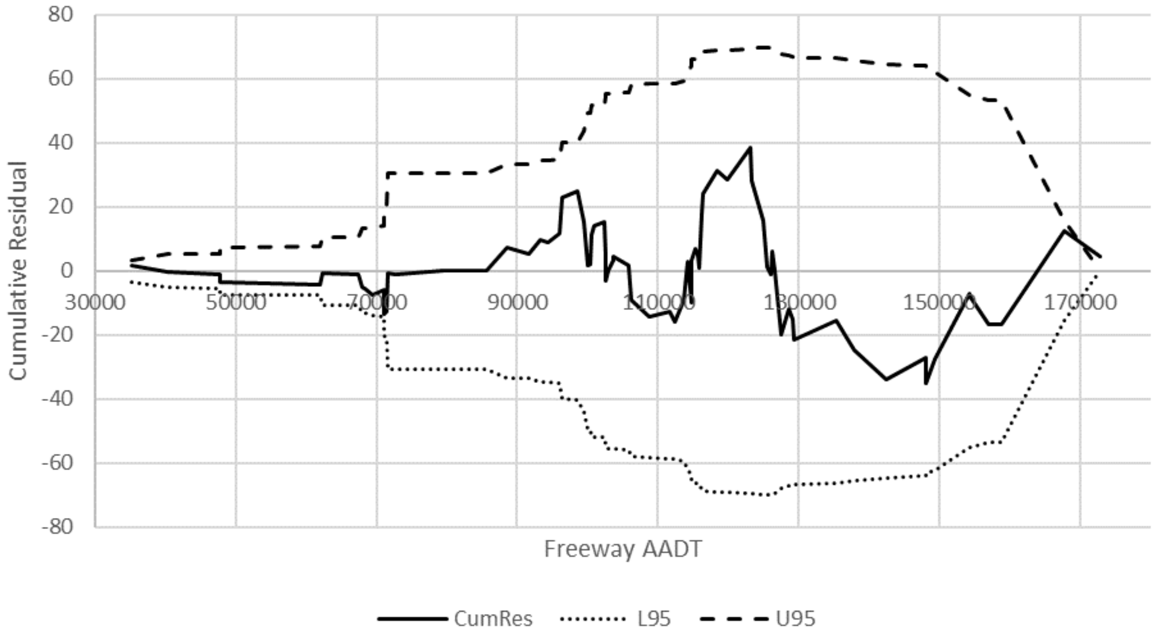

Figure 25. CURE Plot for Urban Eight or More Lane Freeway Segment FI Crashes SPF.

Rural Four Lane Freeway Interchange Segment Total Crashes

Figure 26 provides the CURE plot for the final rural four lane freeway interchange segment total crash model. The CURE plot for the power model indicates a good fit of the data; all observations fall within the 95 percent confidence interval. There is a slight systematic bias from 30,000 to 37,000 AADT, but the cumulative residuals fall well within the confidence interval through this section.

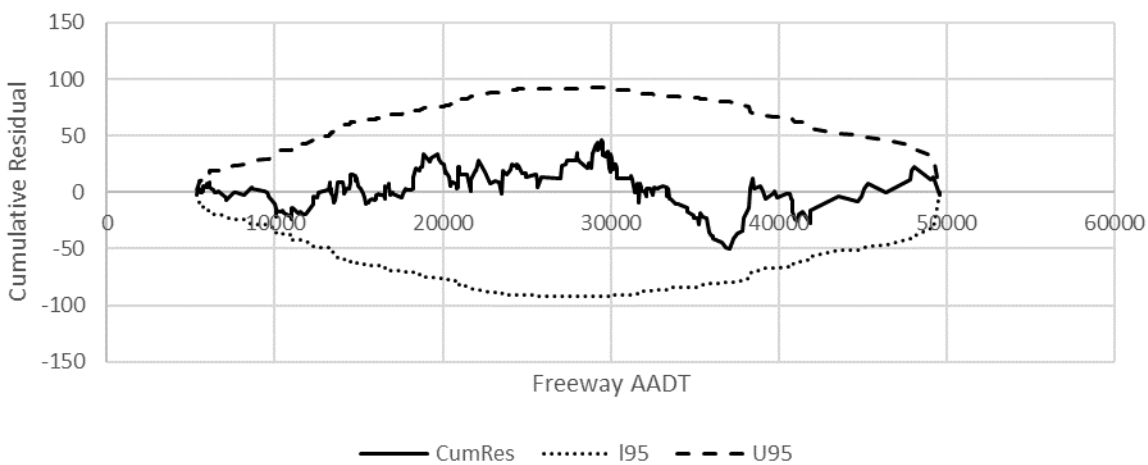

Figure 26. CURE Plot for Rural Four Lane Freeway Interchange Segment Total Crashes SPF.

Rural Four Lane Freeway Interchange Segment FI Crashes

Figure 27 provides the CURE plot for the final rural four lane freeway interchange segment FI crash model. The CURE plot for the power model indicates a good fit of the data; all but three observations fall within the 95 percent confidence interval. The cumulative residuals oscillate about the zero line, which indicates little systematic bias.

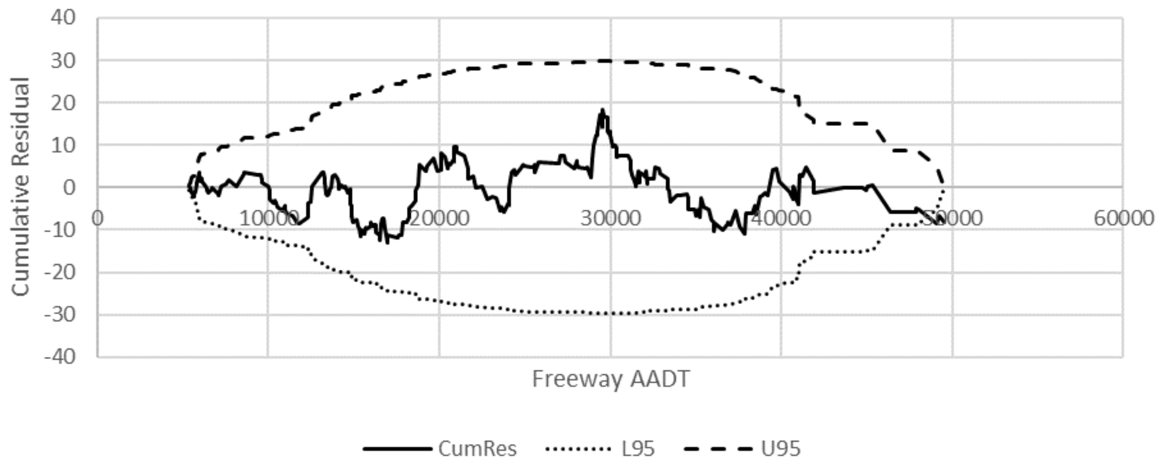

Figure 27. CURE Plot for Rural Four Lane Freeway Interchange Segment FI Crashes SPF.

Rural Six Lane Freeway Interchange Segment Total Crashes

Figure 28 provides the CURE plot for the final rural six lane freeway interchange segment total crash model. The CURE plot for the power model indicates a good fit of the data; all but three observations fall within the 95 percent confidence interval. The cumulative residuals oscillate about the zero line, which indicates little systematic bias.

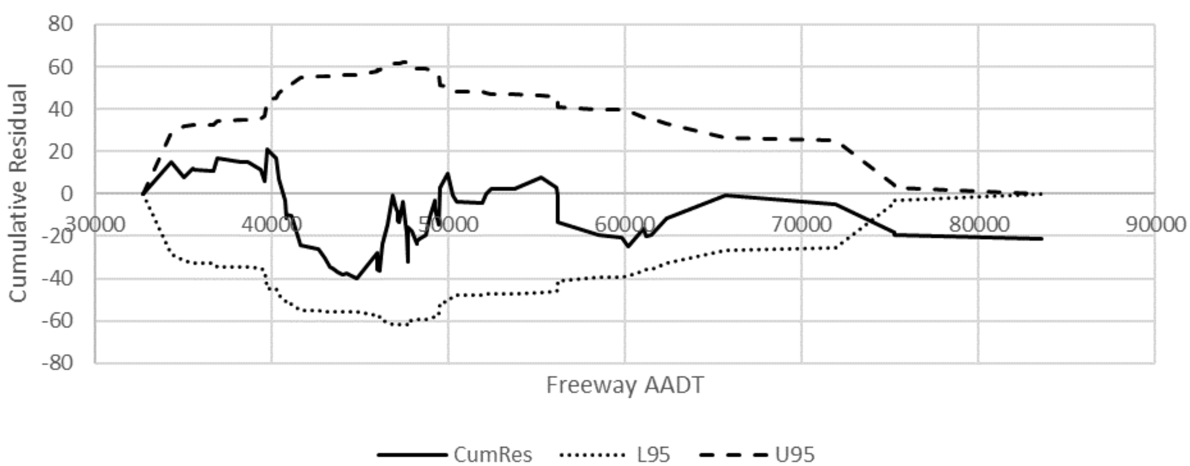

Figure 28. CURE Plot for Rural Six Lane Freeway Interchange Segment Total Crashes SPF.

Rural Six Lane Freeway Interchange Segment FI Crashes

Figure 29 provides the CURE plot for the final rural six lane freeway interchange segment FI crash model. The CURE plot for the power model indicates a good fit of the data; all but three observations fall within the 95 percent confidence interval. There appears to be some systematic bias from 47,000 to 50,000 AADT, but the trend line falls well within the 95 percent confidence interval through this section.

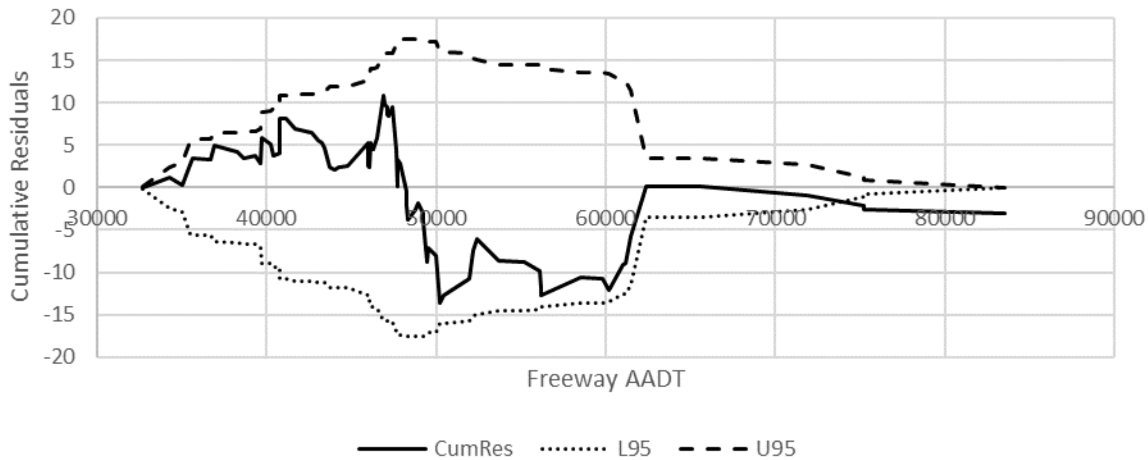

Figure 29. CURE Plot for Rural Six Lane Freeway Interchange Segment FI crashes SPF.

Urban Four Lane Freeway Interchange Segment Total Crashes

Figure 30 provides the CURE plot for the final urban four lane freeway interchange segment total crash model. The CURE plot for the power model indicates a modest fit for the data. The cumulative residual exceeds the 95th percentile confidence interval for approximately 10 percent of sites. This is due to a systematic bias from 25,000 AADT to 45,000 AADT. In this section, the predictions tend to exceed the observed values.

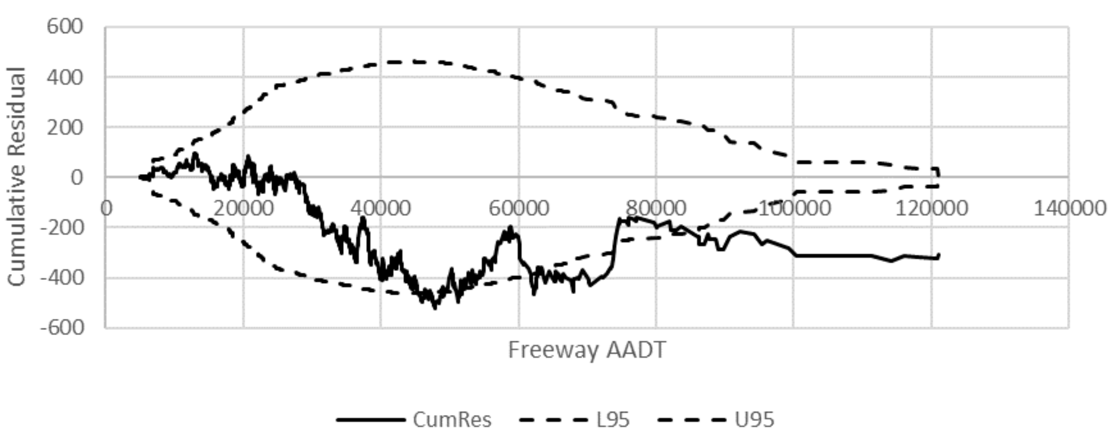

Figure 30. CURE Plot for Urban Four Lane Freeway Interchange Segment Total Crashes SPF.

Urban Four Lane Freeway Interchange Segment FI Crashes

Figure 31 provides the CURE plot for the final urban four lane freeway interchange segment FI crash model. The CURE plot for the power model indicates a modest fit for the data. The cumulative residual exceeds the 95th percentile confidence interval for approximately one percent of sites. This is due to a systematic slight overprediction from 20,000 to 40,000 AADT and slight systematic underprediction from 40,000 to 75,000 AADT. However, the vast majority of these sites fall within the 95 percent confidence interval.

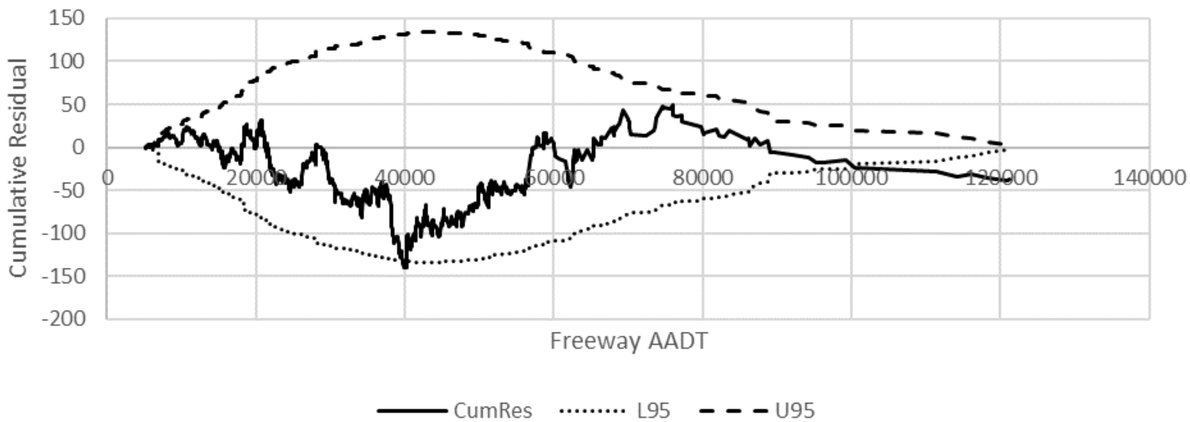

Figure 31. CURE Plot for Urban Four Lane Freeway Interchange Segment FI Crashes SPF.

Urban Five/Six Lane Freeway Interchange Segment Total Crashes

Figure 32 provides the CURE plot for the final urban five and six lane freeway interchange segment total crash model. The CURE plot for the power model indicates a modest fit for the data. The cumulative residual exceeds the 95th percentile confidence interval for approximately 11 percent of sites. This is due to a systematic bias from 60,000 AADT to 75,000 AADT. In this section, the predictions tend to exceed the observed values.

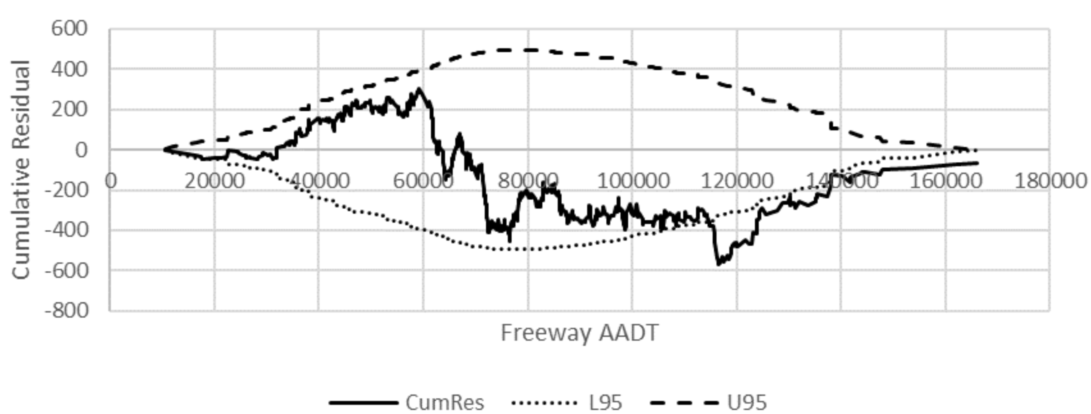

Figure 32. CURE Plot for Urban Five/Six Lane Freeway Interchange Segment Total Crashes SPF.

Urban Five/Six Lane Freeway Interchange Segment FI Crashes

Figure 33 provides the CURE plot for the final urban five and six lane freeway segment FI crash model. The CURE plot for the power model indicates a modest fit for the data. The cumulative residual exceeds the 95th percentile confidence interval for approximately 11 percent of sites. This is due to a systematic bias from 60,000 AADT to 75,000 AADT. In this section, the predictions tend to exceed the observed values.

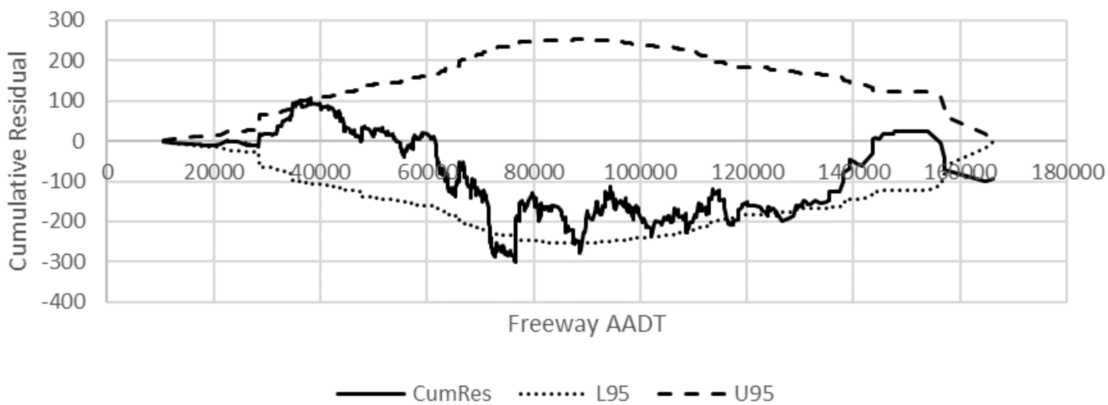

Figure 33. CURE Plot for Urban Five/Six Lane Freeway Interchange Segment FI Crashes SPF.

Urban Seven or More Lane Freeway Interchange Segment Total Crashes

Figure 34 provides the CURE plot for the final urban seven or more lane freeway interchange segment total crash model. The CURE plot for the power model indicates a good fit for the data. The cumulative residual exceeds the 95th percentile confidence interval for approximately four percent of sites. These sites are all at the extreme upper end of the AADT range, where there are relatively few observations.

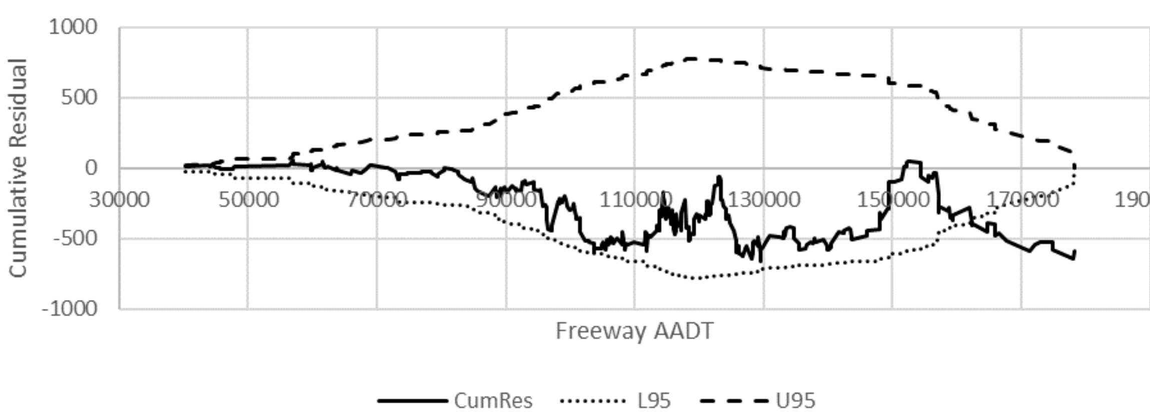

Figure 34. CURE Plot for Urban Seven or More Lane Freeway Interchange Segment Total Crashes SPF.

Urban Seven or More Lane Freeway Interchange Segment FI Crashes

Figure 35 provides the CURE plot for the final urban seven or more lane freeway interchange segment fatal and injury crash model. The CURE plot for the power model indicates a good fit for the data. The cumulative residual exceeds the 95th percentile confidence interval for approximately one percent of sites. These sites are all at the extreme upper end of the AADT range, where there are relatively few observations.

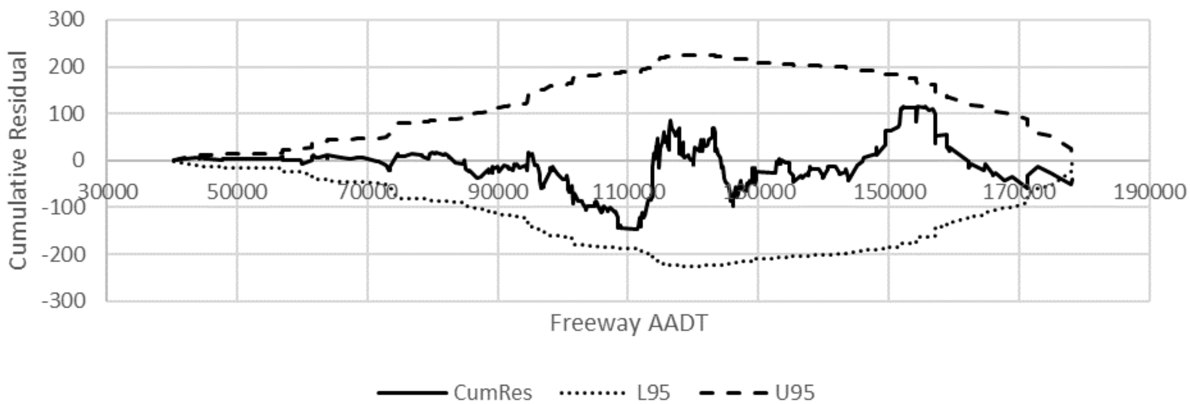

Figure 35. CURE Plot for Urban Seven or More Lane Freeway Interchange Segment FI Crashes SPF.

Appendix D: Project Design-Level SPF Development

Introduction

This appendix presents the process for developing SPFs for project design-level analysis. Project design-level SPFs include both bi-directional and one-directional analysis approaches. In each case, the project team developed separate analysis databases for the following SPF categories:

1. Network screening databases:
    a. Base segments.
    b. Interchange segments.
2. Bi-directional project-level database:
    a. Base segments.
3. Directional project-level databases:
    a. Base segments.
    b. Entry speed change segments.
    c. Exit speed change segments.

Note that consistent with the current methods provided in Chapter 18 of the HSM, entry and exit speed change lanes are directional in nature and crash frequency is only predicted for directional crashes.

Project Design-Level Analysis SPF Methodology

Project design-level SPFs are used as part of an overall CPM in Part C of the HSM. A CPM is used to estimate the predicted average crash frequency of a specific type of site (e.g., segment) with specific geometric design elements and traffic control features. Each CPM in the HSM has the following general form (Equation 14):

$$N_p = C \times N_{SPF} \times (AF_1 \times \ldots \times AF_n) \qquad \text{Equation 14}$$

where:

N_p = predicted average crash frequency, crashes/yr.

C = local calibration factor.

N_{SPF} = predicted crash frequency for site with base conditions.

AF_i = Adjustment factor for geometric element, or traffic control feature i (i = 1 to n).

n = total number of AFs.

Each CPM includes an SPF, one or more AFs, and a local calibration factor (C). The SPF is used to predict the crash frequency N_{SPF} for a site having characteristics that match a specified set of “base conditions.” These conditions describe the typical site’s design elements and control features (e.g., 12-foot lane width). The set of AFs are used to adjust N_{SPF} such that the CPM can provide reliable estimates of the predicted crash frequency N_p for sites that do not match all base conditions.

The calibration factor simply adjusts the base SPF estimate up or down and may not truly improve the fit of the predictive CPM. Developing an Ohio-specific SPF can improve the fit of the CPM, but the new SPF combined with national AFs may not provide sufficiently reliable estimates.

The project team developed SPFs that can be combined with the existing AFs in Chapter 18 of the HSM; however, the primary purpose of the project design-level SPF development was to include geometric and operational characteristics in predictive models for developing inferred AFs for a new CPM specific to ODOT. The project team developed separate SPFs by the following characteristics:

- Segment type (base segment, entry speed-change lane, exit speed-change lane).
- Crash type (multivehicle versus single vehicle crashes).
- Crash severity (fatal and injury versus property damage only crashes).

As with the current models in the HSM, the project team separately included effects for area type and number of lanes but pooled data to bolster sample size (particularly for speed-change lane sites). The project team developed project design-level SPFs consistent with segments types in Chapter 18 of the HSM for crash and severity types.

The project team developed separate bi-directional and one-directional CPMs for freeway segments, entry speed-change lanes, and exit speed-change lanes. Equation 15 serves as the foundation for the freeway segments CPM.

$$N_{tot,fs} = L \times \exp(b_0 \ln[AADT_{fr}]) \times (AF_1 \times \ldots \times AF_n) \quad \text{Equation 15}$$

where:

$N_{tot,fs}$ = predicted average total crashes for freeway segment site, crashes.

L = segment length.

b_0 =regression coefficient freeway AADT.

$AADTfr$ = freeway AADT.

AF_i = adjustment factor for freeway geometric design element, or traffic control feature i.

For the full model, AFs are inferred from the multiple-variable regression model. While the functional form of most AFs are consistent with those in Chapter 18, some AFs differ from those found in the HSM because of differences in ODOT’s data. For example, the AF for lane changing in this research includes the proportion of the segment that is within 0.5 miles of the entrance or exit gore (accounting for the proportion of the segment actually impacted by lane changing). Additionally, emphasis was placed on developing functions, where the CMF includes continuous variables rather than indicators, and interactions were considered when practical.

While the CPM predicts crash frequency (by combined severity categories and crash types), the project team developed SDFs to predict the proportion of K, A, B, C, and O crash severity categories. The probability of each severity category is predicted as a function of traffic volume, geometry, and other roadway characteristics. The proportion is multiplied by the predicted crash frequency to obtain an estimate of the crash frequency for the corresponding severity category.

The generalized database was used to estimate the new bidirectional and directional CPMs. The new CPMs were coded as multiple-variable regression models and regression analysis was how the regression coefficients are computed. Supplemental regression analyses were undertaken to assess model validity and to produce an overdispersion parameter that is suitable for site-specific evaluation using the EB method described in the HSM.

For entry speed-change lanes and exit speed-change lanes, the foundational equation was modified to Equation 16:

$$N_{tot,sc} = L \times \exp(b_0 \ln\left[AADT_{fr}\right] + b_1 \ln[AADT_r]) \times (AF_1 \times \ldots \times AF_n) \qquad \text{Equation 16}$$

where:

$N_{tot,sc}$ = predicted average total crashes for speed change lane site, crashes.

L = speed change lane segment length.

b_0 =regression coefficient freeway AADT.

$AADT_{fr}$ = freeway AADT.

b_1 =regression coefficient ramp AADT.

$AADT_r$ ramp AADT.

AF_i = adjustment factor for freeway geometric design element, or traffic control feature i.

One version of this equation was developed for entry speed change lanes and a second version was developed for exit ramp speed change lanes. This equation includes the entry or exit ramp AADT and its associated regression parameter. Additionally, the ramp entrance length was included as an additional AF in each predictive model. Both the entry speed change lane CPM and the exit speed change lane CPM were estimated only for directional crashes (consistent with the existing HSM methodology). As with freeway segments, the CMFs developed for the predictive method were inferred from a multiple-variable regression model such that they can be used together in addition to substituting base values for the existing HSM Chapter 18 model. Note that this model predicts total crashes near the speed-change lane. Separate SDFs were developed to predict the proportions of crash severities.

Crash Prediction Model Development Methodology

For crash frequency modeling, count models are traditionally used to quantify the relationship between crash frequency and traffic volumes, design elements, and traffic control features. Negative binomial regression has most commonly been applied to account for the overdispersion inherently found in crash data. The overdispersion parameter estimated from the modeling process is used in the development of the weight factor in the EB analysis method.

Relatively recently, researchers have been applying more sophisticated versions of count models to account for temporal and spatial correlations. Depending on the assumption of the correlation between unobserved effects and right-hand side variables, fixed- and random-effects models have been applied to account for temporal and spatial correlations. As models become more sophisticated, bias and inconsistency are generally lessened, improving the transferability of the model. However, more sophisticated models can be more time consuming to estimate, limiting the number of models that can be estimated during the modeling process. Additionally, more sophisticated models can prove to be more difficult to reach convergence with a larger number of predictor variables, limiting the number of geometric and operational features that can be included in the model specification.

The project team considered fixed effects, random effects, and mixed effects models to account for unobserved correlations inherent in the data. The project team applied negative binomial count models as is the current state-of-the-practice.

The project team used a multinomial logit model to estimate SDFs for each of the CPMs. The multinomial logit model allows for some variables to be constrained to have the same effect on each severity level while allowing other variables to have a variable effect among levels. For SDFs, the database was restructured such that the observed unit was the crash instead of the road segment. The total crashes CPM can be combined with the SDF to estimate the number of crashes of different severity levels.

The following section describes the results of the project design-level SPF development.

Project Level SPF Development

The VHB project team developed one-directional and bi-directional project design-level SPFs for rural and urban freeways based on a sample of sites for which the project team incorporated additional data elements. Figure 36 provides an overview of the freeway segments included in the project design-level SPF development dataset, including the following freeways:

- Interstate 70.
- Interstate 71.
- Interstate 75.
- Interstate 77.
- Interstate 80.
- Interstate 90.
- State Route 2.
- State Route 11

Figure 36. Sample Corridors for Project-Level SPFs.

ODOT provided supplemental crash data identifying if crashes occurred in the cardinal or non-cardinal direction of the freeway. The project team used this information to classify the freeway direction for developing directional project design-level SPFs. Additionally, ODOT provided horizontal curve data for select curves on the freeway network and a barrier inventory for supplementing median type (which included the presence of median barrier) and for inclusion of outside barrier. However, this dataset did not provide the offset to the barrier from the edge of pavement.

The project team further collected data identifying gore points and end of tapers for entrance and exit speed change lanes. The project team classified whether gore points were for speed change lanes, were for a lane add or lane drop, and for whether the ramp is on the inside or outside. Figure 37 provides an example of how the project team marked gore and taper points in both directions of a freeway segment. The project team used the mileposts for the gore and taper points and subtracted those segments out of the roadway inventory file to identify speed change lanes from freeway segments. The project team stored the speed change lane segments separately from freeway segments for separate analyses. The project team additionally used the gore point information to identify the distance from study segments to upstream entrance ramps and to downstream exit ramps. For bidirectional project-level SPFs, the project team used the extents of concurrent freeway segments (without speed change lanes) to identify bidirectional freeway segments. The project team only estimated speed change lane SPFs for directional segments (consistent with the approach currently included in the HSM).

Figure 37. Example Ramp Points Collected at Interchange for Both Directions.

Consistent with the first edition of the HSM, the project team evaluated four crash outcomes for bidirectional and directional freeway segment models:

- Fatal and injury multi-vehicle crashes (fi_mv). These crashes involve more than one vehicle and have at least one injury in the crash.
- Fatal and injury single vehicle crashes (fi_sv). These crashes involve no more than one vehicle and have at least one injury in the crash.
- Property damage only multi-vehicle crashes (pdo_mv). These crashes involve more than one vehicle and have no reported injuries.
- Property damage only single vehicle crashes (pdo_sv). These crashes involve no more than one vehicle and have no reported injuries.

Table 31 provides an overview of the data elements the project team collected and considered in project-level SPFs. The individual sections for crash types by project-level SPF type provide the summary statistics for each variable, including the minimum and maximum values.

The following sections provide the freeway segment crash prediction models (broken into SPFs and AFs) for directional project design-level SPFs, speed-change lane SPFs, and bi-directional project design-level SPFs. The speed-change lane SPFs, like those currently found in the HSM, are one-directional and may be used with either one-directional or bi-directional project design-level SPFs. Each set of SPFs is provided individually by crash type. The final sections provide an overview of severity distribution functions.

Table 31. Data Elements Collected for Project-Level SPF Development.

Variable	Definition
Routename	Provides the combined route type and route number for the segment
County	Provides the county in which the segment is located
District	Provides the ODOT district in which the segment is located
Areatype	Indicates whether the study segment is urban or rural
Thru_lanes	Indicates the number of thru lanes on study segment
Lane_width	Average lane width in feet
Median_type	Indicates the type of median (1: rigid barrier, 2: semi-rigid barrier, 3: flexible barrier, 4: raised median with curb, 5: depressed median, 6: flush paved median, 7: HOV lanes, 8: railroad or other rapid transit, 9: other divided)
Median_width	Median width in feet
Shoulder_out	Outside shoulder width in feet
Shoulder_in	Inside shoulder width in feet
Posted_speed	Posted speed limit in mph
Lane_add	Indicator for a lane add within the study segment
Lane_add_aadt	Ramp AADT for lane added by ramp
Lane_drop	Indicator for a lane drop within the study segment
Lane_drop_aadt	Ramp AADT for lane dropped by ramp
Weave_type	Indicates the type of weaving section (0: no weave, 1: Type A weave, 2: Type B weave, 3: Type C weave)
Freeway_AADT	Directional or bidirectional AADT in vehicles per day
Down_ex_length	Length to downstream exit in miles
Up_en_length	Length to upstream entrance in miles
Down_ex_AADT	Downstream exit AADT in vehicles per day
Up_en_AADT	Upstream entrance AADT in vehicles per day
Seg_length	Study segment length in miles
Length_ex	Segment length within 0.5 miles of downstream exit
Length_en	Segment length within 0.5 miles of upstream entrance
Barrier_in	Length of inside barrier in miles
Barrier_out	Length of outside barrier in miles
Lighting	Indicator for presence of highway or interchange lighting
SCL_type	Speed change lane type (1: entrance, 2: exit)
Ramp_AADT	Exit or entrance ramp AADT
FI	Number of fatal and injury crashes
PDO	Number of property damage only crashes
Fi_mv	Number of fatal and injury multivehicle crashes
Fi_sv	Number of fatal and injury single vehicle crashes
PDO_mv	Number of property damage only multivehicle crashes
PDO_sv	Number of property damage only single vehicle crashes

One-Direction Freeway Segment Project-Level SPF for FI MV Crashes

Equations 17 and 18 provide the base SPFs for fatal and injury multiple-vehicle crashes for rural and urban base condition segments. Table 32 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 32 also provides the base condition for each variable. The SPF is based on 6,175 segments, or 2,242 miles, of data.

The SPF is based on five years of data, including 10,013 fatal and injury multivehicle crashes.

$$N_{fi\_mv\_r} = L \times f\_aadt^{1.565} \times e^{(-16.173)} \times AF_i \quad \text{Equation 17}$$

$$N_{fi\_mv\_u} = L \times f\_aadt^{1.565} \times e^{(-15.889)} \times AF_i \quad \text{Equation 18}$$

Table 32. Summary Data for One-Direction Freeway Segment FI MV Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.003	8.139	0.363	N/A
Directional freeway AADT	2,750	78,400	27,900	N/A
Urban	0	1	0.72	N/A
Inside shoulder width in feet	0	12	6.888	6
Outside shoulder width in feet	0	12	10.021	10
Depressed median width in feet	0	260	34.472	60
Proportion of segment within half mile of downstream exit (unitless)	0	1	0.279	0
Length to downstream exit in miles	0	2	1.200	0
Downstream exit ramp AADT	0	88,433	2,640	0
Lane add AADT	0	119,950	2,970	0
Lane drop AADT	0	58,400	2,550	0
Presence of Type A weaving section	0	1	0.020	0
Average horizontal curve degree	0	8.185	0.154	0
Fatal and injury multivehicle crashes	0	25	1.622	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for fatal and injury multiple-vehicle crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Directional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{1.565}$.
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.030(ISW - 6)}$.
- Shoulder_width_out_max = Segment outside shoulder width in feet. If the outside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.026(OSW - 10)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00190(MW - 60) \times DM}$, where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.

- Prop_downstream_lc = Proportion of lane change based on downstream exit. The proportion is calculated as Equation 19: Equation 19

$$Proportion\ Downstream\ LC = \left(\frac{Segment\ Length\ within\ 0.5\ Miles\ of\ Downstream\ Exit}{Total\ Segment\ Length}\right) \times \frac{\frac{Downstream\ Exit\ AADT}{1{,}000}}{Downstream\ Exit\ Length\ in\ miles}$$

- The AF is calculated as $e^{0.000740(Proportion\ Downstream\ LC\ -\ 0)}$.
- La_aadt_thous = Interaction for presence of a lane add within the segment and the entrance ramp AADT in thousands of vehicles. The AF is calculated as $e^{0.00478(R_{AADT}\ /\ 1{,}000\ -\ 0) \times LA}$ where LA is an indicator for a lane addition (1 if yes; 0 otherwise).
- Ld_aadt_thous = Interaction for presence of a lane drop within the segment and the exit ramp AADT in thousands of vehicles. The AF is calculated as $e^{0.00734(R_{AADT}\ /\ 1{,}000\ -\ 0) \times LD}$ where LD is an indicator for a lane drop (1 if yes; 0 otherwise).
- Weavea = Indicator for the presence of a Type A weaving section. The AF is calculated as $e^{(0.159) \times Type\ A}$ (where Type A is 1 if a Type A weaving section is present; otherwise 0).
- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.082 \times (Ave\_Degree)}$.

One-Direction Freeway Segment Project Design-Level SPF for PDO MV Crashes

Equations 20 through 23 provide the base SPFs for property damage only multiple-vehicle crashes for rural and urban base condition segments. The equations are also provided separately for two directional lanes versus 3 or more directional lanes. Table 33 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 33 also provides the base condition for each variable. The SPF is based on 5,886 segments, or 2,148 miles, of data. The SPF is based on five years of data, including 24,905 property damage only multivehicle crashes.

$N_{pdo\_mv\_r\_2l} = L \times f\_aadt^{1.656} \times e^{(-16.050)} \times AF_i$ Equation 20

$N_{pdo\_mv\_r\_3+l} = L \times f\_aadt^{1.656} \times e^{(-16.188)} \times AF_i$ Equation 21

$N_{pdo\_mv\_u\_2l} = L \times f\_aadt^{1.656} \times e^{(-15.755)} \times AF_i$ Equation 22

$N_{pdo\_mv\_u\_3+l} = L \times f\_aadt^{1.656} \times e^{(-15.892)} \times AF_i$ Equation 23

Table 33. Summary Data for One-Direction Freeway Segment PDO MV Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.003	8.139	0.365	N/A
Directional freeway AADT	2,750	78,400	26,950	N/A
Urban	0	1	0.709	N/A
Inside shoulder width in feet	0	12	6.822	6
Depressed median width in feet	0	260	35.721	60
Proportion of segment within half mile of downstream exit (unitless)	0	1	0.278	0
Length to downstream exit in miles	0	2	1.210	0
Downstream exit ramp AADT	0	88,433	2,500	0
Proportion of segment within half mile of upstream entrance (unitless)	0	1	0.258	0
Length to upstream entrance in miles	0	2	1.251	0
Upstream entrance ramp AADT	0	88,433	2,550	0
Lane add AADT	0	119,950	3,050	0
Lane drop AADT	0	58,400	2,350	0
Presence of Type A or B weaving section	0	1	0.023	0
Average horizontal curve degree	0	8.185	0.140	0
Median barrier	0	1	0.332	0
Property damage only multivehicle crashes	0	93	4.231	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for property damage only multiple-vehicle crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Directional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{1.656}$.
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.015(ISW - 6)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00143(MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Prop_downstream_lc = Proportion of lane change based on downstream exit. The proportion is calculated as Equation 24: Equation 24

$$Proportion\ Downstream\ LC = \left(\frac{Segment\ Length\ within\ 0.5\ Miles\ of\ Downstream\ Exit}{Total\ Segment\ Length}\right) \times \frac{\frac{Downstream\ Exit\ AADT}{1{,}000}}{Downstream\ Exit\ Length\ in\ miles}$$

- The AF is calculated as $e^{0.00155(Proportion\ Downstream\ LC - 0)}$.

- Prop_upstream_lc = Proportion of lane change based on upstream entrance. The proportion is calculated as Equation 25: Equation 25

$$Proportion\ Upstream\ LC = \left(\frac{Segment\ Length\ within\ 0.5\ Miles\ of\ Upstream\ Entrance}{Total\ Segment\ Length}\right) \times \frac{\frac{Upstream\ Entrance\ AADT}{1{,}000}}{Upstream\ Entrance\ Length\ in\ miles}$$

- The CMF is calculated as $e^{0.000738(Proportion\ Upstream\ LC\ -\ 0)}$.
- La_aadt_thous = Interaction for presence of a lane add within the segment and the entrance ramp AADT in thousands of vehicles. The AF is calculated as $e^{0.00536(R_{AADT}\ /\ 1{,}000\ -\ 0)\ \times\ LA}$ where LA is an indicator for a lane addition (1 if yes; 0 otherwise).
- Ld_aadt_thous = Interaction for presence of a lane drop within the segment and the exit ramp AADT in thousands of vehicles. The AF is calculated as $e^{0.0113\ (R_{AADT}\ /\ 1{,}000\ -\ 0)\ \times\ LD}$ where LD is an indicator for a lane drop (1 if yes; 0 otherwise).
- Weaveab = Indicator for the presence of a Type A or Type B weaving section. The AF is calculated as $e^{(0.230)\ \times\ Type\ AB}$ (where Type AB is 1 if a Type A or Type B weaving section is present; otherwise 0).
- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.139\ \times\ (Ave\_Degree)}$.
- Median_barrier = Indicator for presence of median barrier based on ODOT median type (1, 2, or 3). The AF is calculated as $e^{0.046\ \times\ MB}$ (where MB is 1 if median barrier is present; otherwise 0).

One-Direction Freeway Segment Project Design-Level SPF for FI SV Crashes

Equations 26 through 29 provide the base SPFs for fatal and injury single vehicle crashes for rural and urban base condition segments. The equations are also provided separately for two directional lanes versus 3 or more directional lanes. Table 34 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 34 also provides the base condition for each variable. The SPF is based on 6,393 segments, or 2,340 miles, of data. The SPF is based on five years of data, including 6,448 fatal and injury single vehicle crashes.

$$N_{fi\_sv\_r\_2l} = L \times f\_aadt^{0.710} \times e^{(-7.622)} \times AF_i$$ Equation 26

$$N_{fi\_sv\_r\_3+l} = L \times f\_aadt^{0.710} \times e^{(-7.522)} \times AF_i$$ Equation 27

$$N_{fi\_sv\_u\_2l} = L \times f\_aadt^{0.710} \times e^{(-7.416)} \times AF_i$$ Equation 28

$$N_{fi\_sv\_u\_3+l} = L \times f\_aadt^{0.710} \times e^{(-7.316)} \times AF_i$$ Equation 29

Table 34. Summary Data for One-Direction Freeway Segment FI SV Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.003	8.139	0.366	N/A
Directional freeway AADT	2,750	78,400	28,950	N/A
Urban	0	1	0.732	N/A
Inside shoulder width in feet	0	12	6.931	6
Depressed median width in feet	0	260	43.269	60
Presence of median barrier	0	1	0.341	0
Average horizontal curve degree	0	8.185	0.153	0
Presence of outside barrier	0	1	0.503	0
Fatal and injury single vehicle crashes	0	20	1.009	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for fatal and injury single vehicle crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Directional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{0.710}$.
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.035(ISW - 6)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00335(MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Median_barrier = Indicator for presence of median barrier based on ODOT median type (1, 2, or 3). The AF is calculated as $e^{-0.129 \times MB}$ (where MB is 1 if median barrier is present; otherwise 0).
- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.120 \times (Ave\_Degree)}$.
- Barrier_outside = Indicator for presence of outside barrier based on ODOT barrier file. The AF is calculated as $e^{-0.027 \times OB}$ (where OB is 1 if outside barrier is present; otherwise 0).

One-Direction Freeway Segment Project Design-Level SPF for PDO SV Crashes

Equation 30 through 33 provide the base SPFs for property damage only single vehicle crashes for rural and urban base condition segments. The equations are also provided separately for two directional lanes versus 3 or more directional lanes. Table 35 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 35 also provides the base condition for each variable. The SPF is based on 6,524 segments, or 2,584 miles, of data. The SPF is based on five years of data, including 6,448 property damage only single vehicle crashes.

$$N_{pdo\_sv\_r\_2l} = L \times f\_aadt^{0.347} \times e^{(-2.666)} \times AF_i \quad \text{Equation 30}$$

$$N_{pdo\_sv\_r\_3l} = L \times f\_aadt^{0.347} \times e^{(-2.550)} \times AF_i \quad \text{Equation 31}$$

$$N_{pdo\_sv\_u\_2l} = L \times f\_aadt^{0.347} \times e^{(-2.592)} \times AF_i \quad \text{Equation 32}$$

$$N_{pdo\_sv\_u\_3+l} = L \times f\_aadt^{0.347} \times e^{(-2.476)} \times AF_i \quad \text{Equation 33}$$

Table 35. Summary Data for One-Direction Freeway Segment PDO SV Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.003	8.139	0.396	N/A
Directional freeway AADT	2,750	78,400	29,100	N/A
Urban	0	1	0.731	N/A
Depressed median width in feet	0	260	33.458	60
Presence of median barrier	0	1	0.343	0
Average horizontal curve degree	0	8.185	0.153	0
Posted speed limit in mph	40	70	65.78	70 rural; 65 urban
Property damage only single vehicle crashes	0	123	31,326	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for property damage only single vehicle crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Directional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{0.347}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00112(MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Median_barrier = Indicator for presence of median barrier based on ODOT median type (1, 2, or 3). The AF is calculated as $e^{0.043 \times MB}$ (where MB is 1 if median barrier is present; otherwise 0).

- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.162 \times (Ave\_Degree)}$.
- Posted_speed = Posted speed limit on freeway mainline segment. The base condition is 70 mph for rural freeways and 65 mph for urban freeways. The AF is calculated as $e^{0.012 \times (PSL - BL)}$ (where PSL is posted speed limit in mph and BL is base line for rural or urban segment).

Entrance Speed-Change Lane Project Design-Level SPF for FI Crashes

Equations 34 and 35 provide the base SPFs for fatal and injury crashes on base condition entrance speed-change lane segments. The equations are provided separately for two directional lanes versus 3 or more directional lanes. Table 36 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 36 also provides the base condition for each variable. The SPF is based on 743 segments, or 111 miles, of data. The SPF is based on five years of data, including 1,400 fatal and injury crashes.

$$N_{ent\_fi\_2l} = L \times f\_aadt^{1.452} \times r\_aadt^{0.176} \times e^{(-14.372)} \times CMF_i \quad \text{Equation 34}$$

$$N_{ent\_fi\_3+l} = L \times f\_aadt^{1.452} \times r\_aadt^{0.176} \times e^{(-14.524)} \times CMF_i \quad \text{Equation 35}$$

Table 36. Summary Data for Entrance Speed-Change Lane FI Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.023	0.293	0.149	N/A
Directional freeway AADT	2,750	73,650	29,300	N/A
Ramp AADT	100	20,600	4,100	N/A
Left side ramp indicator	0	1	0.004	0
Inside shoulder width	0	12	6.786	6
Depressed median width in feet	0	224	32.584	60
Presence of median barrier	0	1	0.361	0
Proportion of segment that is curve	0	1	0.111	0
Fatal and injury crashes	0	32	1.884	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for fatal and injury crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Directional AADT on freeway mainline. The effect of AADT is calculated as *crashes* = $AADT^{1.452}$.
- r_aadt = AADT on entrance ramp in thousands of vehicles. The effect of AADT is calculated as *crashes* = $R\_AADT^{0.176}$ / 1,000.
- left_ramp = Indicator for left-side entrance ramp. The AF is calculated as $e^{0.605 \times Left}$ (where Left is 1 if entrance ramp is on the left side; otherwise 0).
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.032(ISW - 6)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00474(MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Median_barrier = Indicator for presence of median barrier based on ODOT median type (1, 2, or 3). The AF is calculated as $e^{-0.116 \times MB}$ (where MB is 1 if median barrier is present; otherwise 0).
- Prop_curve = Proportion of segment that is horizontal curve. If multiple curves are present within the segment, their combined length is used in the calculation. The AF is calculated as $e^{0.166 \times (Prop\_curve)}$.

Exit Speed Change Lane Project Design-Level SPF for FI Crashes

Equation 36 provides the base SPF for fatal and injury crashes on base condition exit speed-change lane segments. Table 37 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 37 also provides the base condition for each variable. The SPF is based on 722 segments, or 84 miles, of data. The SPF is based on five years of data, including 1,168 fatal and injury crashes.

$$N_{exit\_fi} = L \times f\_aadt^{1.359} \times r\_aadt^{0.041} \times e^{(-13.240)} \times CMF_i \qquad \text{Equation 36}$$

Table 37. Summary Data for Exit Speed-Change Lane FI Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.013	0.286	0.116	N/A
Directional freeway AADT	2,750	76,200	29,050	N/A
Ramp AADT	100	136,900	4,300	N/A
Left side ramp indicator	0	1	0.021	0
Inside shoulder width	0	12	6.682	6
Depressed median width in feet	0	224	33.351	60
Proportion of segment that is curve	0	1	0.107	0
Fatal and injury crashes	0	19	1.618	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for fatal and injury crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Directional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{1.359}$.
- r_aadt = AADT on entrance ramp in thousands of vehicles. The effect of AADT is calculated as $crashes = R\_AADT^{0.041}$ / 1,000.
- left_ramp = Indicator for left-side entrance ramp. The AF is calculated as $e^{0.822 \times Left}$ (where Left is 1 if entrance ramp is on the left side; otherwise 0).
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.032(ISW - 6)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00374(MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Prop_curve = Proportion of segment that is horizontal curve. If multiple curves are present within the segment, their combined length is used in the calculation. The AF is calculated as $e^{0.133 \times (Prop\_curve)}$.

Entrance Speed-Change Lane Project Design-Level SPF for PDO Crashes

Equations 37 and 38 provide the base SPFs for property damage only crashes on base condition entrance speed-change lane segments. The equations are provided separately for rural and urban entrance speed-change lane segments. Table 38 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 38 also provides the base condition for each variable. The SPF is based on 743 segments, or 111 miles, of data. The SPF is based on five years of data, including 4,669 property damage only crashes.

$$N_{ent\_pdo\_r} = L \times f\_aadt^{0.983} \times r\_aadt^{0.040} \times e^{(-8.397)} \times CMF_i \quad \text{Equation 37}$$

$$N_{ent\_pdo\_u} = L \times f\_aadt^{0.983} \times r\_aadt^{0.040} \times e^{(-8.139)} \times CMF_i \quad \text{Equation 38}$$

Table 38. Summary Data for Entrance Speed Change Lane PDO Project-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.023	0.293	0.149	N/A
Directional freeway AADT	2,750	73,650	29,300	N/A
Ramp AADT	100	20,600	4,100	N/A
Left side ramp indicator	0	1	0.004	0
Inside shoulder width	0	12	6.786	6
Depressed median width in feet	0	224	32.584	60
Proportion of barrier on outside	0	1	0.138	0
Segment average degree of curve	0	3.351	0.151	0
Property damage only crashes	0	96	6.284	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for property damage only crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Directional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{0.983}$.
- r_aadt = AADT on entrance ramp in thousands of vehicles. The effect of AADT is calculated as $crashes = R\_AADT^{0.040}$ / 1,000.
- left_ramp = Indicator for left-side entrance ramp. The AF is calculated as $e^{0.996 \times Left}$ (where Left is 1 if entrance ramp is on the left side; otherwise 0).
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.015(ISW - 6)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00210 \times (MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Prop_barrier_out = Proportion of segment with outside barrier of any type. The AF is calculated as $e^{0.308 \times POB}$.
- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.149 \times (Ave\_Degree)}$.

Exit Speed-Change Lane Project Design-Level SPF for PDO Crashes

Equation 39 provides the base SPF for property damage only crashes on base condition exit speed-change lane segments. Table 39 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 39 also provides the base condition for each variable. The SPF is based on 722 segments, or 84 miles, of data. The SPF is based on five years of data, including 4,016 property damage only crashes.

$$N_{exit\_pdo} = L \times f\_aadt^{1.137} \times r\_aadt^{0.026} \times e^{(-9.688)} \times CMF_i \qquad \text{Equation 39}$$

Table 39. Summary Data for Exit Speed-Change Lane PDO Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.013	0.286	0.116	N/A
Directional freeway AADT	2,750	76,200	29,050	N/A
Ramp AADT	100	136,900	4,300	N/A
Left side ramp indicator	0	1	0.021	0
Inside shoulder width in feet	0	12	6.682	6
Outside shoulder width in feet	0	12	10.066	10
Depressed median width in feet	0	224	33.351	60
Segment average degree of curve	0	5.730	0.147	0
Property damage only crashes	0	90	5.563	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for property damage only crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Directional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{1.137}$.
- r_aadt = AADT on entrance ramp in thousands of vehicles. The effect of AADT is calculated as $crashes = R\_AADT^{0.026}$ / 1,000.
- left_ramp = Indicator for left-side entrance ramp. The AF is calculated as $e^{0.732 \times Left}$ (where Left is 1 if entrance ramp is on the left side; otherwise 0).
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.029(ISW - 4)}$.
- Shoulder_width_out_max = Segment outside shoulder width in feet. If the outside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.029(OSW - 6)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00377 \times (MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median

width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.

- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.234 \times (Ave\_Degree)}$.

Bi-Directional Freeway Segment Project Design-Level SPF for FI MV Crashes

Equations 40 and 41 provide the base bi-directional SPFs for fatal and injury multiple-vehicle crashes on base condition freeway segments. The equations are provided separately for rural and urban facilities. Table 40 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 40 also provides the base condition for each variable. The SPF is based on 2,060 segments, or 911 centerline miles, of data. The SPF is based on five years of data, including 8,791 fatal and injury multivehicle crashes.

$$N_{fi\_mv\_r} = L \times f\_aadt^{1.567} \times e^{(-16.516)} \times CMF_i \quad \text{Equation 40}$$

$$N_{fi\_mv\_u} = L \times f\_aadt^{1.567} \times e^{(-16.071)} \times CMF_i \quad \text{Equation 41}$$

Table 40. Summary Data for Freeway Segment FI MV Bi-Directional Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.002	8.138	0.442	N/A
Bi-directional freeway AADT	5,500	156,800	56,100	N/A
Urban	0	1	0.568	N/A
Inside shoulder width in feet	0	12	6.928	6
Presence of median barrier	0	1	0.334	0
Depressed median width in feet	0	260	34.046	60
Proportion of segment within half mile of downstream exit (unitless)	0	1	0.279	0
Length to downstream exit in miles	0	2	1.234	0
Downstream exit ramp AADT	0	119,947	5,700	0
Lane add AADT	0	88,433	842	0
Lane drop AADT	0	88,433	725	0
Presence of Type A weaving section	0	1	0.037	0
Average horizontal curve degree	0	8.185	0.174	0
Fatal and injury multivehicle crashes	0	55	4.267	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for fatal and injury multiple-vehicle crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Bidirectional AADT on freeway mainline. The effect of AADT is calculated as *crashes* = $AADT^{1.567}$.
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.033(ISW - 6)}$.
- Median_barrier = Indicator for presence of median barrier based on ODOT median type (1, 2, or 3). The AF is calculated as $e^{-0.075 \times MB}$ (where MB is 1 if median barrier is present; otherwise 0).
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00203\,(MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Prop_downstream_lc = Proportion of lane change based on downstream exit. The proportion is calculated using Equation 42: Equation 42

$$Proportion\ Downstream\ LC = \left(\frac{Segment\ Length\ within\ 0.5\ Miles\ of\ Downstream\ Exit}{Total\ Segment\ Length}\right) \times \frac{\frac{Downstream\ Exit\ AADT}{1{,}000}}{Downstream\ Exit\ Length\ in\ miles}$$

- The AF is calculated as $e^{0.00129\,(Proportion\ Downstream\ LC\ -\ 0)}$.
- La_aadt_thous = Interaction for presence of a lane add within the segment and the entrance ramp AADT in thousands of vehicles. The AF is calculated as $e^{-0.00771\,(R_{AADT} / 1{,}000\ -\ 0) \times LA}$ where LA is an indicator for a lane addition (1 if yes; 0 otherwise).
- Ld_aadt_thous = Interaction for presence of a lane drop within the segment and the exit ramp AADT in thousands of vehicles. The AF is calculated as $e^{0.00827\,(R_{AADT} / 1{,}000\ -\ 0) \times LD}$ where LD is an indicator for a lane drop (1 if yes; 0 otherwise).
- Weavea = Indicator for the presence of a Type A weaving section. The AF is calculated as $e^{(0.170) \times Type\ A}$ (where Type A is 1 if a Type A weaving section is present; otherwise 0).
- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that

is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.109 \times (Ave\_Degree)}$.

Bi-Directional Freeway Segment Project Design-Level SPF for PDO MV Crashes

Equation 43 to 44 provide the base bi-directional SPFs for property damage only multiple-vehicle crashes on base condition freeway segments. The equations are provided separately for rural and urban facilities as well as four bi-directional lanes versus six or more bi-directional lanes. Table 41 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 41 also provides the base condition for each variable. The SPF is based on 1,927 segments, or 838 centerline miles, of data. The SPF is based on five years of data, including 20,806 property damage only multivehicle crashes.

$$N_{pdo\_mv\_r\_4l} = L \times f\_aadt^{1.661} \times e^{(-16.540)} \times CMF_i \quad \text{Equation 43}$$

$$N_{pdo\_mv\_r\_6+l} = L \times f\_aadt^{1.661} \times e^{(-16.676)} \times CMF_i \quad \text{Equation 44}$$

$$N_{pdo\_mv\_u\_4l} = L \times f\_aadt^{1.661} \times e^{(-16.045)} \times CMF_i \quad \text{Equation 45}$$

$$N_{pdo\_mv\_u\_6+l} = L \times f\_aadt^{1.661} \times e^{(-16.181)} \times CMF_i \quad \text{Equation 46}$$

Table 41. Summary Data for Freeway Segment PDO MV Bi-Directional Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.002	8.138	0.435	N/A
Directional freeway AADT	5,500	156,800	53,350	N/A
Urban	0	1	0.542	N/A
Inside shoulder width in feet	0	12	6.877	6
Outside shoulder width in feet	0	12	10.011	10
Depressed median width in feet	0	260	35.093	60
Proportion of segment within half mile of downstream exit (unitless)	0	1	0.278	0
Length to downstream exit in miles	0	2	1.252	0
Downstream exit ramp AADT	0	119,950	2,500	0
Lane add AADT	0	86,004	654	0
Lane drop AADT	0	58,400	502	0
Presence of Type A or B weaving section	0	1	0.044	0
Average horizontal curve degree	0	8.185	0.157	0
Proportion of outside barrier	0	1	0.235	0
Property damage only multivehicle crashes	0	266	10.797	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for property damage only multiple-vehicle crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Bidirectional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{1.661}$.
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.017(ISW - 6)}$.
- Shoulder_width_out_max = Segment outside shoulder width in feet. If the outside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.011(OSW - 10)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00172\,(MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Prop_downstream_lc = Proportion of lane change based on downstream exit. The proportion is calculated as Equation 47: Equation 47

$$Proportion\ Downstream\ LC = \left(\frac{Segment\ Length\ within\ 0.5\ Miles\ of\ Downstream\ Exit}{Total\ Segment\ Length}\right) \times \frac{\frac{Downstream\ Exit\ AADT}{1{,}000}}{Downstream\ Exit\ Length\ in\ miles}$$

- The AF is calculated as $e^{0.00236\,(Proportion\ Downstream\ LC - 0)}$.
- La_aadt_thous = Interaction for presence of a lane add within the segment and the entrance ramp AADT in thousands of vehicles. The AF is calculated as $e^{0.00487\,(R_{AADT} / 1{,}000 - 0) \times LA}$ where LA is an indicator for a lane addition (1 if yes; 0 otherwise).
- Ld_aadt_thous = Interaction for presence of a lane drop within the segment and the exit ramp AADT in thousands of vehicles. The AF is calculated as $e^{0.0201\,(R_{AADT} / 1{,}000 - 0) \times LD}$ where LD is an indicator for a lane drop (1 if yes; 0 otherwise).
- Weaveab = Indicator for the presence of a Type A or Type B weaving section. The AF is calculated as $e^{(0.221) \times Type\ AB}$ (where Type AB is 1 if a Type A or Type B weaving section is present; otherwise 0).
- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that

is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.255 \times (Ave\_Degree)}$.

- Barrier_out_prop = Proportion of outside barrier on segment. The AF is calculated as $e^{0.177 \times POB}$.

Bi-Directional Freeway Segment Project Design-Level SPF for FI SV Crashes

Equations 48 and 49 provide the base bi-directional SPFs for fatal and injury single vehicle crashes on base condition freeway segments. The equations are provided separately for rural and urban facilities. Table 42 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 42 also provides the base condition for each variable. The SPF is based on 1,997 segments, or 603 centerline miles, of data. The SPF is based on five years of data, including 4,261 fatal and injury single vehicle crashes.

$$N_{fi\_sv\_r} = L \times f\_aadt^{0.748} \times e^{(-7.905)} \times CMF_i \quad \text{Equation 48}$$

$$N_{fi\_sv\_u} = L \times f\_aadt^{0.748} \times e^{(-7.618)} \times CMF_i \quad \text{Equation 49}$$

Table 42. Summary Data for Freeway Segment FI SV Bi-Directional Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.002	5.572	0.302	N/A
Directional freeway AADT	5,500	156,800	57,700	N/A
Urban	0	1	0.592	N/A
Inside shoulder width in feet	0	12	6.906	6
Depressed median width in feet	0	260	33.782	60
Presence of median barrier	0	1	0.334	0
Average horizontal curve degree	0	8.185	0.173	0
Presence of Type A weaving section	0	1	0.042	0
Fatal and injury single vehicle crashes	0	30	2.134	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for fatal and injury single vehicle crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Bidirectional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{0.748}$.
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{-0.019(ISW - 6)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.00172\,(MW - 60) \times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median

width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.

- Median_barrier = Indicator for presence of median barrier based on ODOT median type (1, 2, or 3). The AF is calculated as $e^{-0.079 \times MB}$ (where MB is 1 if median barrier is present; otherwise 0).
- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.280 \times (Ave\_Degree)}$.
- Weavea = Indicator for the presence of a Type A weaving section. The AF is calculated as $e^{(0.145) \times Type\ A}$ (where Type A is 1 if a Type A weaving section is present; otherwise 0).

Bi-Directional Freeway Segment Project Design-Level SPF for PDO SV Crashes

Equations 50 and 51 provide the base bi-directional SPFs for fatal and injury single vehicle crashes on base condition freeway segments. The equations are provided separately for rural and urban facilities. Table 43 provides the summary statistics, including the minimum, maximum, and average values. The minimum and maximum values provide an indication of the conditions for which the AF is applicable. Table 43 also provides the base condition for each variable. The SPF is based on 1,452 segments, or 190 centerline miles, of data. The SPF is based on five years of data, including 6,089 property damage only single vehicle crashes.

$$N_{pdo\_sv\_r} = L \times f\_aadt^{0.310} \times e^{(-1.664)} \times CMF_i \qquad \text{Equation 50}$$

$$N_{pdo\_sv\_u} = L \times f\_aadt^{0.310} \times e^{(-1.444)} \times CMF_i \qquad \text{Equation 51}$$

Table 43. Summary Data for Freeway Segment PDO SV Bi-Directional Project Design-Level SPFs.

Input Variable	Min	Max	Average	Base Condition
Segment length	0.002	0.752	0.131	N/A
Directional freeway AADT	5,500	156,800	56,500	N/A
Urban	0	1	0.602	N/A
Inside shoulder width in feet	0	12	6.681	6
Depressed median width in feet	0	260	34.316	60
Average horizontal curve degree	0	8.185	0.166	0
Property damage only single vehicle crashes	0	39	4.194	N/A

The following provides a set of specific definitions for input characteristics and AFs applicable to the base models for property damage only single vehicle crashes:

- L = segment length in miles. This is a unit offset as input by the analyst.
- f_aadt = Bidirectional AADT on freeway mainline. The effect of AADT is calculated as $crashes = AADT^{0.310}$.
- Shoulder_width_in_max = Segment inside shoulder width in feet. If the inside shoulder width is greater than 12 feet wide, then 12 feet should be used. The AF is calculated as $e^{0.022(ISW-6)}$.
- Depressed_med_wid = Depressed median width in feet. This includes an indicator to determine if the median type is a depressed median, and if so, the width in feet. The AF is calculated as $e^{-0.000663\,(MW-60)\times DM}$ where DM is an indicator for depressed median (1 if yes; 0 otherwise). The minimum median width for depressed medians is 12 feet. If a narrower median is present, then the indicator should be set to 0.
- Ave_degree = Average degree of curvature for segment. This is calculated by taking the proportion of segment that is curved and multiplying by the horizontal curve degree. A value of 0 is used for the proportion of segment that is tangent. If multiple curves are present within the segment, the length-weighted average is used. The AF is calculated as $e^{0.210\times(Ave\_Degree)}$.

Severity Distribution Functions

The project team evaluated the severity of crash outcomes through severity distribution functions (SDFs) rather than estimating separate SPFs by crash severity. The severity of outcomes is defined through probability of the outcome's occurrence. The probability of outcome M is defined by Equation 52.

$$P_M = \frac{e^{V_M}}{\sum_{i=1}^{M} e^{V_M}} \qquad \text{Equation 52}$$

Where:

P_M = The probability of the outcome M.

V_M = The deterministic component of outcome M.

M = The total number of possible outcomes modeled.

In this case, the project team developed SDFs for injury crash outcomes (K, A, B, and C), with level C considered to be the base scenario. The project team rearranged the dataset to use the crash outcome as the observational unit, rather than the segment. The segment attributes were retained for each crash outcome to estimate the effects of geometric variables on crash severity. Initial efforts indicated that crash sample sizes for fatalities (i.e., K) were low; therefore, K and A severity crashes were combined for fatal and severe injury crashes.

The probability for each outcome is shown as Equations 53 to 55:

$$P_{KA} = \frac{e^{V_{KA}}}{1 + e^{V_{KA}} + e^{V_B}} \quad \text{Equation 53}$$

$$P_B = \frac{e^{V_B}}{1+e^{V_{KA}}+e^{V_B}} \quad \text{Equation 54}$$

$$P_C = 1.0 - (P_{KA} + P_B) \quad \text{Equation 55}$$

The resulting proportions can be combined with fatal and injury crash frequency models to estimate the predicted crash frequency by severity level. Analysts can use an Ohio-specific proportion factor to separate K and A severities as needed.

Severity Distribution Functions for Freeway Segment Injury Crashes

Table 44 provides the SDFs for freeway segment injury crashes, where C-injury crashes serve as the baseline. The SDFs indicate the following characteristics are associated with crash severity on freeway segments:

- Posted speed limit. An increase in posted speed limit (in mph) is associated with an increase in the likelihood of KA and B severity levels. The magnitude of effect is higher (as expected) for KA crashes than for B crashes.
- Outside shoulder width. An increase in outside shoulder width (in feet) is associated with an increase in the likelihood of KA and B severity levels. The magnitude is the same for both KA and B crashes.
- Inside shoulder width. An increase in inside shoulder width (in feet) is associated with a decrease in the likelihood of KA severity.
- Median width. An increase in median width (in feet) is associated with an increase in the likelihood of KA and B severity levels. The magnitude is greater for KA crashes than for B crashes.
- Proportion of outside barrier. An increase in the proportion of segment with outside barrier is associated with a decrease in the likelihood of KA severity.
- Average degree of curve. An increase in the average degree of curvature is associated with an increase in the likelihood of B severity.
- Urban area type. The segment having urban area type characteristics is associated with a decrease in the likelihood of KA and B severity. The magnitude is greater for B crashes than for KA crashes.
- Presence of median barrier. The presence of median barrier is associated with an increase in the likelihood of KA and B severity. The magnitude is greater for KA crashes than for B crashes.
- Presence of highway lighting. The presence of highway lighting is associated with a reduction in the likelihood of KA and B severity. The magnitude is greater for KA crashes than for B crashes.

Equations 56 and 57 provide the results of the models for calculating the deterministic component of each crash severity level for freeway segments. These equations can be combined as shown in the introduction to this section to estimate the probability of each severity level. Equation 56

$$V_{KA} = -4.2662 + 0.0478 \times PSL + 0.0297 \times OSW - 0.0570 \times ISW + 0.0006 \times MW - 0.1795 \times POB - 0.1478 \times URB + 0.1356 \times PMB - 0.2141 \times PHL$$

Equation 57

$$V_B = -2.7407 + 0.0391 \times PSL + 0.0297 \times OSW + 0.0003 \times MW + 0.0206 \times DOC - 0.2368 \times URB + 0.0780 \times PMB - 0.0877 \times PHL$$

For separating the probability of KA crashes, the data indicated that 12.62 percent of KA crashes are K and 87.38 percent are A.

Table 44. Parameter Estimates for Freeway Segment SDFs.

Variable	KA		B	
	Value	t-stat	Value	t-stat
Alternative specific constant (ASC)	-4.2662	0.5290	-2.7407	0.3216
Posted speed limit in mph (PSL)	0.0478	0.0074	0.0391	0.0043
Outside shoulder width in feet (OSW)	0.0297	0.0124	0.0297	0.0124
Inside shoulder width in feet (ISW)	-0.0570	0.0078	N/A	N/A
Median width in feet (MW)	0.0006	0.0003	0.0003	0.0002
Proportion of outside barrier (POB)	-0.1795	0.0697	N/A	N/A
Average degree of curve (DOC)	N/A	N/A	0.0431	0.0206
Urban area type (URB)	-0.1478	0.0748	-0.2368	0.0471
Presence of median barrier (PMB)	0.1356	0.0589	0.0780	0.0355
Presence of highway lighting (PHL)	-0.2141	0.0726	-0.0877	0.0438

Severity Distribution Functions for Speed Change Lanes

Table 45 provides the SDFs for speed change lane injury crashes, where C-injury crashes serve as the baseline. The SDFs indicate the following characteristics are associated with crash severity on speed change lanes:

- Posted speed limit. An increase in posted speed limit (in mph) is associated with an increase in the likelihood of KA and B severity levels. The magnitude of effect is the same for KA and B crashes.
- Average degree of curve. An increase in the average degree of curvature is associated with an increase in the likelihood of KA severity.
- Outside shoulder width. An increase in outside shoulder width (in feet) is associated with a decrease in the likelihood of KA severity.
- Median width. An increase in median width (in feet) is associated with a decrease in the likelihood of B severity.
- Ramp AADT. An increase in adjacent ramp AADT is associated with a decrease in the likelihood of KA and B severity. The magnitude is greater for KA crashes than for B crashes.

- Urban area type. The segment having urban area type characteristics is associated with a decrease in the likelihood of KA and B severity. The magnitude is greater for KA crashes than for B crashes.
- Proportion of median barrier. An increase in the proportion of segment with median barrier is associated with a decrease in the likelihood of B severity.

Table 45. Parameter Estimates for Speed Change Lane SDFs.

Variable	KA		B	
	Value	t-stat	Value	t-stat
Alternative specific constant (ASC)	-2.9208	0.7529	-1.8458	0.6342
Posted speed limit in mph (PSL)	0.0345	0.0091	0.0345	0.0091
Average degree of curve (DOC)	0.1427	0.0991	N/A	N/A
Outside shoulder width in feet (OSW)	-0.0394	0.0399	N/A	N/A
Median width in feet (MW)	N/A	N/A	-0.0032	0.0014
Ramp AADT in thousands (RAADT)	-0.0175	0.0134	-0.0102	0.0059
Urban area type (URB)	-0.4659	0.1914	-0.2828	0.1369
Proportion of median barrier (PMB)	N/A	N/A	-0.2320	0.0936

Equations 58 and 59 provide the results of the models for calculating the deterministic component of each crash severity level for speed change lanes. These equations can be combined as shown in the introduction to this section to estimate the probability of each severity level.

Equation 58

$$V_{KA} = -2.9208 + 0.0345 \times PSL + 0.1427 \times DOC - 0.0394 \times OSW - 0.0175 \times RAADT - 0.4659 \times URB$$

Equation 59

$$V_{B} = -1.8458 + 0.0345 \times PSL - 0.0032 \times MW - 0.0102 \times RAADT - 0.2828 \times URB - 0.2320 \times PMB$$

For separating the probability of KA crashes, the data indicated that 13.24 percent of KA crashes are K and 86.76 percent are A.

Appendix E: Project Design-Level SPF Validation

Introduction

This appendix presents the process for validating the performance of the project design-level SPFs developed in this research and alternative forms of predictive analysis for freeway segments, including the previously calibrated version of the HSM. The approaches include bi-directional and one-direction SPFs along with a shortened set of AFs developed from this research versus the full set of AFs available from the HSM predictive method. The project team evaluated the performance of the following predictive model forms, using validation sites and sample project applications:

- Uncalibrated HSM bi-directional models (UHSM).
- Calibrated HSM bi-directional models (CHSM). The project team used Ohio-specific calibration factors available in ODOT's ECAT based on previous efforts.
- Ohio-specific bi-directional SPF with HSM AFs (OHSM).
- Ohio-specific bi-directional predictive method (OBPM).
- Ohio-specific directional predictive method (ODPM).

The project team further evaluated the predictive performance of each model for severity-specific crash frequency from the sample project applications to determine which method performed the best when considering crash severity (evaluated as K, A, B, C, or O on the KABCO scale).

Validation Sites Overview

The project team collected the full datasets required for both HSM analyses and the new predictive models for 52 freeway segments, 50 entrance speed change lanes, and 49 exit speed change lanes. Table 46 provides details on the segments for used for validation of basic freeway segment models. Table 46 includes information on the facility ID, beginning and ending milepost, area type, and number of lanes in each direction. Table 47 provides details for entrance speed-change lanes and Table 48 provides details for exit speed-change lanes. Consistent with the predictive model development, the project team used 2014 to 2018 data for validation. The validation dataset consisted of sites on Interstate 74, Interstate 675, US 24, and US 35. The validation dataset also consisted of sites with two directional lanes and three directional lanes as well as both urban and rural sites.

Table 46. Summary of Basic Freeway Segment Validation Sites.

Site	Facility	Begin MP	End MP	NLFID	Area Type	Lanes Right	Lanes Left
1	I-74	0.00	1.14	SHAMIR00074**C	Urban	2	2
2	I-74	2.05	3.13	SHAMIR00074**C	Urban	2	2
3	I-74	4.06	5.24	SHAMIR00074**C	Urban	2	2
4	I-4	9.72	10.95	SHAMIR00074**C	Urban	2	2
5	I-74	11.73	14.15	SHAMIR00074**C	Urban	2	2
6	US 35	1.13	1.67	SGREUS00035**C	Urban	2	2
7	US 35	1.78	2.35	SGREUS00035**C	Urban	2	2
8	US 35	8.06	9.03	SGREUS00035**C	Urban	2	2
9	US 35	9.57	9.65	SGREUS00035**C	Urban	2	2
10	US 35	10.54	11.67	SGREUS00035**C	Urban	2	2
11	US 35	12.44	14.08	SGREUS00035**C	Urban	2	2
12	US 35	14.87	20.46	SGREUS00035**C	Rural	2	2
13	US 35	21.40	22.20	SGREUS00035**C	Rural	2	2
14	US 35	23.13	27.23	SGREUS00035**C	Rural	2	2
15	US 35	1.87	4.70	SFAYUS00035**C	Rural	2	2
16	US 35	4.83	11.49	SFAYUS00035**C	Rural	2	2
17	US 35	12.22	13.73	SFAYUS00035**C	Rural	2	2
18	US 35	14.26	14.99	SFAYUS00035**C	Urban	2	2
19	US 35	15.77	16.88	SFAYUS00035**C	Urban	2	2
20	US 35	17.58	24.02	SFAYUS00035**C	Rural	2	2
21	US 35	5.17	7.34	SROSUS00035**C	Rural	2	2
22	US 35	8.10	14.24	SROSUS00035**C	Rural	2	2
23	US 35	15.18	16.92	SROSUS00035**C	Urban	2	2
24	US 35	17.25	18.13	SROSUS00035**C	Urban	2	2
25	US 35	18.91	19.44	SROSUS00035**C	Urban	2	2
26	US 35	25.54	33.66	SROSUS00035**C	Rural	2	2
27	I-74	15.12	17.67	SHAMIR00074**C	Urban	2	3
28	US 24	0.00	3.22	SPAUUS00024**C	Rural	2	2
29	US 24	4.21	12.14	SPAUUS00024**C	Rural	2	2
30	US 24	13.18	19.21	SPAUUS00024**C	Rural	2	2
31	I-675	0.91	1.62	SMOTIR00675**C	Urban	2	2
32	I-675	2.80	3.98	SMOTIR00675**C	Urban	3	3
33	I-675	5.30	6.94	SMOTIR00675**C	Urban	3	3
34	I-675	0.46	2.66	SGREIR00675**C	Urban	3	3
35	I-675	3.58	5.33	SGREIR00675**C	Urban	3	3
36	I-675	8.55	9.22	SGREIR00675**C	Urban	3	3
37	I-675	10.78	12.48	SGREIR00675**C	Urban	2	2
38	I-675	13.76	14.93	SGREIR00675**C	Urban	2	2
39	I-675	15.92	17.07	SGREIR00675**C	Urban	2	2
40	US 30	0.00	3.46	SRICUS00030**C	Urban	2	2
41	US 30	4.44	7.06	SRICUS00030**C	Urban	2	2
42	US 30	8.03	9.33	SRICUS00030**C	Urban	2	2
43	US 30	10.17	10.48	SRICUS00030**C	Urban	2	2
44	US 30	10.99	11.54	SRICUS00030**C	Urban	2	2
45	US 30	12.52	12.61	SRICUS00030**C	Urban	2	2
46	US 30	13.20	13.76	SRICUS00030**C	Urban	2	2
47	US 30	14.44	15.06	SRICUS00030**C	Urban	2	2
48	US 30	15.71	16.14	SRICUS00030**C	Urban	2	2
49	US 30	3.59	4.14	SRICUS00030**C	Urban	2	2
50	US 35	0.00	1.72	SFAYUS00035**C	Rural	2	2
51	US 35	0.00	4.39	SROSUS00035**C	Rural	2	2

Table 47. Summary of Entrance Speed-Change Lane Segment Validation Sites.

Site	Facility	Begin MP	End MP	NLFID	Area Type	Lanes Right	Lanes Left
1	I-74	1.32	1.14	SHAMIR00074**N	Urban	2	2
2	I-74	1.922	2.037	SHAMIR00074**C	Urban	2	2
3	I-74	3.162	3.292	SHAMIR0074**N_	Urban	2	2
4	I-74	3.814	3.944	SHAMIR00074**C	Urban	2	2
5	US-35	14.63	4.764	SROSUS00035**N	Urban	2	2
6	US-35	14.97	5.124	SROSUS00035**C	Urban	2	2
7	US-35	16.95	7.121	SROSUS00035**C	Urban	2	2
8	US-35	18.13	8.259	SROSUS00035**N	Urban	2	2
9	US-35	18.76	8.898	SROSUS00035**C	Urban	2	2
10	US-35	19.41	9.532	SROSUS00035**N	Urban	2	2
11	US-35	19.93	0.053	SROSUS00035**C	Urban	2	2
12	I-74	17.75	7.938	SHAMIR00074**N	Urban	3	3
13	I-74	15.03	5.195	SHAMIR00074**C	Urban	2	3
14	I-74	14.14	4.323	SHAMIR00074**N	Urban	2	2
15	US-35	25.36	5.539	SROSUS00035**C	Rural	2	2
16	US-35	33.66	3.776	SROSUS00035**N	Rural	2	2
17	US-35	34.21	4.352	SROSUS00035**C	Rural	2	2
18	I-74	10.96	1.113	SHAMIR00074**N	Urban	2	2
19	I-74	11.60	1.735	SHAMIR00074**C	Urban	2	2
20	US-35	0.607	0.718	SGREUS00035**C	Urban	2	2
21	US-35	0.925	1.057	SGREUS00035**C	Urban	2	2
22	US-35	2.355	2.447	SGREUS00035**N	Urban	2	2
23	US-35	2.988	3.092	SGREUS00035**C	Urban	2	2
24	US-35	7.583	7.753	SGREUS00035**C	Urban	2	2
25	US-35	9.208	9.303	SGREUS00035**N	Urban	2	2
26	US-35	9.418	9.571	SGREUS00035**C	Urban	2	2
27	US-35	9.833	9.937	SGREUS00035**N	Urban	2	2
28	US-35	10.42	10.54	SGREUS00035**C	Urban	2	2
29	US-35	11.84	11.97	SGREUS00035**N	Urban	2	2
30	US-35	12.2	12.33	SGREUS00035**C	Urban	2	2
31	US-35	14.24	14.34	SGREUS00035**N	Rural	2	2
32	US-35	14.74	14.87	SGREUS00035**C	Rural	2	2
33	US-35	20.63	20.74	SGREUS00035**N	Rural	2	2
34	US-35	21.23	21.41	SGREUS00035**C	Rural	2	2
35	US-35	22.37	22.54	SGREUS00035**N	Rural	2	2
36	US-35	22.96	23.13	SGREUS00035**C	Rural	2	2
37	US-35	1.72	1.87	SFAYUS00035**N	Rural	2	2
38	US-35	4.69	4.83	SFAYUS00035**C	Rural	2	2
39	US-35	11.49	11.61	SFAYUS00035**N	Rural	2	2
40	US-35	12.12	12.22	SFAYUS00035**C	Rural	2	2
41	US-35	13.74	13.85	SFAYUS00035**N	Urban	2	2
42	US-35	14.13	14.26	SFAYUS00035**C	Urban	2	2
43	US-35	14.99	15.10	SFAYUS00035**N	Urban	2	2
44	US-35	15.65	15.77	SFAYUS00035**C	Urban	2	2
45	US-35	16.87	17.00	SFAYUS00035**N	Urban	2	2
46	US-35	17.47	17.58	SFAYUS00035**C	Urban	2	2
47	US-35	4.39	4.51	SROSUS00035**N	Rural	2	2
48	US-35	5.06	5.17	SROSUS00035**C	Rural	2	2
49	US-35	7.31	7.47	SROSUS00035**N	Rural	2	2
50	US-35	8	8.10	SROSUS00035**C	Rural	2	2

Table 48. Summary of Exit Speed-Change Lane Segment Validation Sites.

Site	Facility	Begin MP	End MP	NLFID	Area Type	Lanes Right	Lanes Left
1	I-74	1.176	1.372	SHAMIR00074**C	Urban	2	2
2	I-74	1.853	2.043	SHAMIR00074**N	Urban	2	2
3	I-74	3.244	3.424	SHAMIR00074**C	Urban	2	2
4	I-74	3.816	3.947	SHAMIR00074**N	Urban	2	2
5	I-74	5.381	5.475	SHAMIR00074**C	Urban	2	2
6	I-74	11.42	11.53	SHAMIR00074**N	Urban	2	2
7	I-74	14.28	14.39	SHAMIR00074**C	Urban	2	2
8	US-35	1.04	1.13	SGREUS00035**N	Urban	2	2
9	US-35	1.67	1.78	SGREUS00035**C	Urban	2	2
10	US-35	2.5	2.68	SGREUS00035**C	Urban	2	2
11	US-35	2.92	3.03	SGREUS00035**N	Urban	2	2
12	US-35	8.14	8.25	SGREUS00035**N	Urban	2	2
13	US-35	9.04	9.18	SGREUS00035**C	Urban	2	2
14	US-35	9.59	9.72	SGREUS00035**N	Urban	2	2
15	US-35	9.81	9.95	SGREUS00035**C	Urban	2	2
16	US-35	10	10.13	SGREUS00035**C	Urban	2	2
17	US-35	10.29	10.4	SGREUS00035**N	Urban	2	2
18	US-35	10.47	10.59	SGREUS00035**N	Urban	2	2
19	US-35	11.69	11.81	SGREUS00035**C	Urban	2	2
20	US-35	12.51	12.6	SGREUS00035**N	Urban	2	2
21	US-35	14.08	14.18	SGREUS00035**C	Rural	2	2
22	US-35	14.85	14.96	SGREUS00035**N	Rural	2	2
23	US-35	20.54	20.64	SGREUS00035**C	Rural	2	2
24	US-35	21.35	21.45	SGREUS00035**N	Rural	2	2
25	US-35	22.32	22.42	SGREUS00035**C	Rural	2	2
26	US-35	23.09	23.19	SGREUS00035**N	Rural	2	2
27	US-35	4.74	4.84	SFAYUS00035**N	Rural	2	2
28	US-35	11.49	11.6	SFAYUS00035**C	Rural	2	2
29	US-35	12.1	12.2	SFAYUS00035**N	Rural	2	2
30	US-35	13.8	13.88	SFAYUS00035**C	Urban	2	2
31	US-35	14.13	14.2	SFAYUS00035**N	Urban	2	2
32	US-35	15.01	15.12	SFAYUS00035**C	Urban	2	2
33	US-35	15.59	15.69	SFAYUS00035**N	Urban	2	2
34	US-35	16.9	17.01	SFAYUS00035**C	Urban	2	2
35	US-35	17.43	17.55	SFAYUS00035**N	Urban	2	2
36	US-35	4.42	4.54	SROSUS00035**C	Rural	2	2
37	US-35	4.95	5.06	SROSUS00035**N	Rural	2	2
38	US-35	7.44	7.56	SROSUS00035**C	Rural	2	2
39	US-35	7.93	8.04	SROSUS00035**N	Rural	2	2
40	US-35	14.24	14.35	SROSUS00035**C	Urban	2	2
41	US-35	15.07	15.18	SROSUS00035**N	Urban	2	2
42	US-35	17.14	17.25	SROSUS00035**N	Urban	2	2
43	US-35	18.25	18.38	SROSUS00035**C	Urban	2	2
44	US-35	18.68	18.81	SROSUS00035**N	Urban	2	2
45	US-35	19.45	19.58	SROSUS00035**C	Urban	2	2
46	US-35	20	20.11	SROSUS00035**N	Urban	2	2
47	US-35	25.32	25.41	SROSUS00035**N	Rural	2	2
48	US-35	33.76	33.88	SROSUS00035**C	Rural	2	2
49	US-35	34.18	34.3	SROSUS00035**N	Rural	2	2

The project team also used sample project applications where ECAT was formerly applied to evaluate the performance of the new predictive models. Table 49 provides an overview of the project application sites. In total, there are 26 sample project applications, including both rural and urban contexts with two or three directional lanes in rural applications and two to five lanes in urban applications. The project team used 2015 to 2019 crash data when available; however fewer years were used in some cases if the year was deemed not to be representative (due to construction or other ongoing activities).

Table 49. Overview of Project Application Sites.

Site Number	Description	Area Type	Directional Lanes
1	I-70 between I-675 ramp and SR 235 EB	Rural	3
2	I-71 between I-270 and Polaris/Gemini NB	Urban	5
3	I-71 between I-270 and Polaris/Gemini SB	Urban	4
4	I-71 between SR 48 and SR 123 NB	Rural	2
5	I-71 between US 30 and SR 39 SB	Rural	3
6	I-71 between US 36 exits NB	Rural	3
7	I-75 between I-275 and Union Centre NB	Urban	4
8	I-75 between SR 4 and Stanley NB	Urban	3
9	I-75 between US 35 and 2nd NB	Urban	5
10	I-75 between Western Hills Viaduct and Hopple Street NB	Urban	5
11	I-76 between SR 44 and SR 14 WB	Rural	2
12	I-76 between SR 534 and Bailey Road EB	Rural	2
13	I-77 between Archwood and I-76 NB	Urban	3
14	I-77 between High Street and US 39 NB	Urban	2
15	I-77 between Wilbeth and Archwood NB	Urban	3
16	I-90 between Bassett Road and SR 252 WB	Urban	3
17	I-90 between Kirtland Road and SR 615 EB	Rural	3
18	I-270 between I-670 and Easton NB	Urban	5
19	I-475 between Corey Road and Tallmadge Road EB	Urban	2
20	I-670 between Airport and just west of I-270 EB	Urban	4
21	I-670 between Leonard and 5th EB	Urban	3
22	SR 2 between SR 44 and Richmond Street EB	Urban	2
23	SR 8 between Perkins and Market SB	Urban	4
24	SR 129 between SR 747 and SR 4 WB	Urban	2
25	US 30 between Apple Creek and Carr Road EB	Rural	2
26	US 422 between SR 44 and SR 306 WB	Rural	2

Prediction Model Evaluation Methodology

The project team compared several performance measures across predictive model types, calculated for each site and then aggregated for comparison, including the following:

- MAD. This measures the average magnitude of variability of prediction. Smaller values are preferred to larger values in comparing two or more predictions to observed crash history.
- Root-mean-square error (RMSE). This measures the differences between model predictions and observed crash frequency as the square root of the second sample moment. The advantage of this measure is that the value is proportional to the size of the squared error; meaning larger errors have a larger overall effect on the RMSE. Therefore, it is sensitive to outliers. Smaller values are preferred to larger values in comparing two or more predictions to observed crash history.
- MAPE. This measure provides an indication of the relative error for model predictions. By using the absolute value, it is not sensitive to the direction of the error; however, there is some bias toward models that underpredict crashes (i.e., the underprediction cannot be represented by more than 100 percent, but there is no limit to the extent of overprediction). Additionally, observations with no crash history cannot be considered as this would result in a division by zero error. Moreover, the values are more susceptible to values with fewer crashes (i.e., smaller denominator). Smaller values are preferred to larger values in comparing two or more predictions to observed crash history.
- CURE plots: CURE plots provide a graphical representation of cumulative residuals (which are observed crashes minus predicted crashes for each segment) against a variable of interest sorted in ascending order (e.g., major road traffic volume). CURE plots help to identify the following concerns:
    - Long trends: trends in the CURE plot (increasing or decreasing) indicate regions of bias.
    - Percent exceeding the confidence limits: cumulative residuals outside the confidence limits indicate a poor fit over that range of the variable. Cumulative residuals frequently outside the confidence limits indicate a notable bias in the SPF. A reasonable upper threshold for the percent of the CURE plot exceeding the 95 percent limits is 5 percent.
    - Vertical changes: large vertical changes in the CURE plot are potential indicators of outliers, which require further examination.

Project Level SPF Comparison Results

Table 50 provides an overview of the comparison results for the project design-level SPF validation sites. The comparison does not clearly indicate that one predictive model approach is superior to the others. However, the comparison does indicate that both the calibrated HSM models and the one-direction approach with a reduced set of AFs outperform the other approaches. The calibration factors indicate that the calibrated HSM models generally underpredict crash frequency on freeway segments (except for single vehicle fatal and injury crashes) and entrance speed-change lanes, while the directional predictive method tends to slightly over-predict crash frequency for the validation sites. In total, the general measures tend to slightly favor the calibrated method, but there are exceptions where one measure (such as RMSE) favors the directional predictive method.

The final column in Table 50 provides an indication of the percentage of sites for which the individual predictions were closer between the calibrated HSM model and the one-direction predictive method. The one-direction predictive method was closer more often in 3 out of the 4 crash types for freeway segments and the models were split for entrance and exit speed-change lanes.

The project team also evaluated the prediction comparison measures by number of lanes and by area type and found no difference in results from those reported for aggregated measures.

The project team further assessed CURE plots (provided separately) for calibrated HSM predictions and the one-direction predictive method by crash type and severity. The CURE plots clearly indicate the one-directional predictive model outperforms the calibrated HSM model for multiple-vehicle fatal and injury (MV FI) crashes, multiple-vehicle property damage only (MV PDO) crashes, and single vehicle fatal and injury (SV FI) crashes. The calibrated HSM model performs better for single vehicle property damage only (SV PDO) crashes. The two approaches perform similarly for FI crashes for entrance speed-change segments, but the calibrated HSM model does outperform the one-direction predictive method for PDO crashes for entrance speed-change lanes. The two approaches perform similarly for both FI and PDO crashes for exit speed-change lanes.

While the one-direction models do provide for better CURE plots for freeway segments, it is not surprising there was no improvement for speed-change lane segments as compared to the calibrated HSM predictive method. The HSM predictive method already treats speed-change lanes as one-direction segments and includes a broader dataset than that which was captured specifically for Ohio. However, there does seem to be more gain for basic freeway segments than there is loss for speed-change lane segments in terms of cumulative residuals (particularly when accounting for the magnitude of crash prediction for speed-change lane segments versus basic freeway segments.

Table 50. Comparison Measure Summary for Validation Sites.

Segment and Crash Type	Model	Calibration Factor	MAD	RMSE	MAPE*	Percent Closer Prediction**
Freeway Segment MVFI	UHSM	1.28	3.26	5.41	45.48	N/A
	CHSM	1.18	3.41	5.55	47.31	55
	OHSM	0.73	5.04	8.61	75.12	N/A
	OBPM	0.74	4.49	7.27	75.58	N/A
	ODPM	0.85	3.91	6.55	57.67	45
Freeway Segment MVPDO	UHSM	1.95	10.05	17.62	50.14	N/A
	CHSM	1.51	8.82	15.59	44.39	45
	OHSM	0.75	11.48	20.11	71.40	N/A
	OBPM	0.66	13.93	26.49	85.28	N/A
	ODPM	0.86	10.37	18.02	58.84	55
Freeway Segment SVFI	UHSM	0.70	3.69	4.72	100.41	N/A
	CHSM	0.76	3.29	4.34	84.78	29
	OHSM	0.78	2.77	3.62	82.96	N/A
	OBPM	0.66	3.95	5.07	92.04	N/A
	ODPM	0.88	2.48	3.24	56.11	71
Freeway Segment SVPDO	UHSM	1.60	13.39	19.47	44.62	N/A
	CHSM	1.10	11.43	17.15	52.15	43
	OHSM	0.57	25.42	31.95	134.93	N/A
	OBPM	0.58	23.20	29.49	120.22	N/A
	ODPM	0.79	12.00	16.87	67.02	57
Entrance Speed Change Lane FI	UHSM	1.49	0.67	1.01	61.09	N/A
	CHSM	1.55	0.66	1.03	61.24	46
	OHSM	1.23	0.66	0.95	61.63	N/A
	ODPM	1.59	0.65	1.03	64.22	54
Entrance Speed Change Lane PDO	UHSM	1.94	1.64	3.28	51.65	N/A
	CHSM	1.34	1.63	2.99	53.76	64
	OHSM	0.86	1.91	3.01	75.53	N/A
	ODPM	0.81	2.04	3.18	85.55	36
Exit Speed Change Lane FI	UHSM	0.91	0.62	0.85	59.32	N/A
	CHSM	0.96	0.61	0.85	61.25	37
	OHSM	0.87	0.62	0.88	61.33	N/A
	ODPM	0.95	0.60	0.88	61.75	63
Exit Speed Change Lane PDO	UHSM	1.21	1.18	1.74	55.21	N/A
	CHSM	0.82	1.38	1.78	72.56	59
	OHSM	0.74	1.51	1.95	88.64	N/A
	ODPM	0.76	1.51	1.98	86.74	41

*Only applies to sites with nonzero observed crashes during comparison period – consistent across model types.

**Applies only to calibrated HSM predictive method compared to one-direction predictive method.

Table 51 provides an overview of the aggregate validation measures for sample project application sites. In the first evaluation case, model predictions include aggregated predictions across crash type and severity as well as for freeway segments and speed change lanes. The measures compare project totals. As with the results in Table 50, the results in Table 51 for aggregate evaluation sites indicate the calibrated HSM model and one-direction predictive methods clearly outperform the other two models. Similarly, the performance measures are similar between the two approaches. Further analysis indicated that calibrated HSM models provided better predictions when there are two directional lanes and three directional lanes. The one-direction predictive method outperformed the calibrated HSM models when there were 4 or 5 directional lanes. Additionally, Table 51 shows the prediction measures comparison by severity level. The results indicate the one-direction predictive method performs better for all levels of injury crashes, but the calibrated HSM model performs better for property damage only crashes.

Table 51. Comparison of Validation Measures for Project Application Sites.

Evaluation	Model	MAD	RMSE	MAPE*	Percent Closer Prediction**
Aggregate Evaluation Sites	UHSM	7.92	11.28	40.29	N/A
	CHSM	5.80	7.79	39.21	58
	OHSM	7.24	10.00	54.67	N/A
	ODPM	6.08	7.92	46.56	42
2 Directional Lanes	CHSM	3.61	5.82	21.12	78
	ODPM	5.25	6.33	42.63	22
3 Directional Lanes	CHSM	7.24	8.89	51.55	56
	ODPM	7.77	10.41	65.80	44
4 Directional Lanes	CHSM	9.62	10.35	64.47	25
	ODPM	7.94	8.53	32.00	75
5 Directional Lanes	CHSM	9.77	9.86	26.88	50
	ODPM	8.03	7.37	28.94	50
KA	CHSM	0.26	0.31	78.91	42
	ODPM	0.17	0.21	52.46	58
B	CHSM	0.90	1.26	67.23	46
	ODPM	0.89	1.23	72.90	54
C	CHSM	0.93	1.22	63.68	38
	ODPM	0.88	1.18	68.93	62
PDO	CHSM	4.70	6.30	43.83	54
	ODPM	4.90	6.42	55.39	46

Conclusions and Recommendations

Overall, the calibrated HSM predictive method and one-direction predictive method developed from this research clearly outperform the other model application approaches. In general, the two methods provide very similar measures of error, while the predictions themselves differ between the two approaches. The calibrated HSM predictive method tends to underpredict crashes while the one-direction approach tends to over-predict crashes. It should be noted that many of the performance measures tend to favor underprediction to overprediction when conducting comparisons. The CURE plots generally indicate the new, one-direction predictive method provides better fits to the observed crash data at validation sites for basic freeway segments, and similar fits (but not quite as good) for speed-change lane segments.

The calibrated HSM predictive method applies more variables and is therefore more sensitive in application to geometric factors being considered. The one-direction predictive method was developed from a dataset with fewer predictor variables; therefore, it is easier to implement, but is less sensitive to geometric considerations. The hybrid approach, of developing an Ohio-specific SPF and reducing to a base SPF for use with HSM AFs is shown to not be a good approach in this case. Due to the number of missing variables in model development, the SPF clearly suffers from omitted-variable bias and this approach should not be used. The one-direction predictive method also (nearly) universally outperformed the bi-directional predictive method.

Both methods can provide reliable results depending on the objective of the analysis. The one-direction predictive method can be implemented more easily, provides better overall fits for injury severity crash frequency (by severity level) and provides a better fit for basic freeway segments. However, the one-direction predictive method can be further improved by collecting the full set of elements included in the HSM predictive method. Moreover, it is clear that freeway crash frequency is sensitive to horizontal curve data, for which only limited data were available for this research. A richer curve dataset will provide clearer findings on the relationship between curve radius and crash frequency on Ohio freeways. Furthermore, better information on roadside clear zone, barrier presence, and offset will improve the fits of the models. The project team noted, that while barrier presence was included, there seemed to be inaccuracies in the dataset provided. Offset to barrier will also improve the estimation of the relationship with crash frequency and severity outcomes.