Report No. K-TRAN: KU-10-1 ▪ FINAL REPORT ▪ November 2013

Evaluation of Interactive Highway Safety Design Model Crash Prediction Tools for Two-Lane Rural Roads on Kansas Department of Transportation Projects

Steven Schrock, Ph.D., P.E.
Ming-Heng Wang, Ph.D.

The University of Kansas

A cooperative transportation research program between
Kansas Department of Transportation,
Kansas State University Transportation Center, and
The University of Kansas

Kansas Department of Transportation

This page intentionally left blank.

1 **Report No.** K-TRAN: KU-10-1	2 **Government Accession No.**	3 **Recipient Catalog No.**
4 **Title and Subtitle** Evaluation of Interactive Highway Safety Design Model Crash Prediction Tools for Two-Lane Rural Roads on Kansas Department of Transportation Projects		5 **Report Date** November 2013 6 **Performing Organization Code**
7 **Author(s)** Steven Schrock, Ph.D., P.E.; Ming-Heng Wang, Ph.D.		8 **Performing Organization Report No.**
9 **Performing Organization Name and Address** The University of Kansas Civil, Environmental & Architectural Engineering Department 1530 West 15th Street Lawrence, Kansas 66045-7609		10 **Work Unit No. (TRAIS)** 11 **Contract or Grant No.** C1815
12 **Sponsoring Agency Name and Address** Kansas Department of Transportation Bureau of Research 2300 SW Van Buren Street Topeka, Kansas 66611-1195		13 **Type of Report and Period Covered** Final Report May 2009–May 2013 14 **Sponsoring Agency Code** RE-0530-01
15 **Supplementary Notes** For more information, write to address in block 9..		

16 **Abstract**

Historically, project-level decisions for the selection of highway features to promote safety were based on either engineering judgment or adherence to accepted national guidance. These tools have allowed highway designers to produce facilities that have demonstrated an improving safety record in recent decades. However, these tools do not allow for comparison of the safety performance of dissimilar facilities or roadway attributes. To address this gap, researchers have been working for decades to develop Crash Prediction Models (CPMs) that can estimate, and ideally predict the expected safety performance of a highway based on its geometric and traffic control features. The main focus of this research was to evaluate the use of CPMs for rural two-lane highways in Kansas. Both CPMs provided in the Highway Safety Manual (HSM) and ones developed specifically based on Kansas data were considered.

Many useful insights and tools were developed through this research study that focused on non-intersection related crashes. The primary conclusions were that single statewide calibration factors were calculated and recommended for rural two-lane highway segments and 3- and 4-leg stopped controlled intersections. A calibration function was also developed for highway segments that can be used to better account for animal crashes, which account for a significant number of rural two-lane highway crashes.

17 **Key Words** Safety, Rural, Crash Prediction Models, Two-Lane Highway		18 **Distribution Statement** No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161	
19 **Security Classification (of this report)** Unclassified	20 **Security Classification (of this page)** Unclassified	21 **No. of pages** 50	22 **Price**

Form DOT F 1700.7 (8-72)

Evaluation of Interactive Highway Safety Design Model Crash Prediction Tools for Two-Lane Rural Roads on Kansas Department of Transportation Projects

Final Report

Prepared by

Steven Schrock, Ph.D., P.E.
Ming-Heng Wang, Ph.D.

The University of Kansas

A Report on Research Sponsored by

THE KANSAS DEPARTMENT OF TRANSPORTATION
TOPEKA, KANSAS

and

THE UNIVERSITY OF KANSAS
LAWRENCE, KANSAS

November 2013

PREFACE

The Kansas Department of Transportation's (KDOT) Kansas Transportation Research and New-Developments (K-TRAN) Research Program funded this research project. It is an ongoing, cooperative and comprehensive research program addressing transportation needs of the state of Kansas utilizing academic and research resources from KDOT, Kansas State University and the University of Kansas. Transportation professionals in KDOT and the universities jointly develop the projects included in the research program.

NOTICE

The authors and the state of Kansas do not endorse products or manufacturers. Trade and manufacturers names appear herein solely because they are considered essential to the object of this report.

This information is available in alternative accessible formats. To obtain an alternative format, contact the Office of Transportation Information, Kansas Department of Transportation, 700 SW Harrison, Topeka, Kansas 66603-3754 or phone (785) 296-3585 (Voice) (TDD).

DISCLAIMER

The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the views or the policies of the state of Kansas. This report does not constitute a standard, specification or regulation.

Abstract

Historically, project-level decisions for the selection of highway features to promote safety were based on either engineering judgment or adherence to accepted national guidance. These tools have allowed highway designers to produce facilities that have demonstrated an improving safety record in recent decades. However, these tools do not allow for comparison of the safety performance of dissimilar facilities or roadway attributes. To address this gap, researchers have been working for decades to develop Crash Prediction Models (CPMs) that can estimate, and ideally predict the expected safety performance of a highway based on its geometric and traffic control features. The main focus of this research was to evaluate the use of CPMs for rural two-lane highways in Kansas. Both CPMs provided in the Highway Safety Manual (HSM) and ones developed specifically based on Kansas data were considered.

Many useful insights and tools were developed through this research study that focused on non-intersection related crashes. The primary conclusions were that single statewide calibration factors were calculated and recommended for rural two-lane highway segments and 3- and 4-leg stopped controlled intersections. A calibration function was also developed for highway segments that can be used to better account for animal crashes, which account for a significant number of rural two-lane highway crashes.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Historically, project-level decisions on the selection of highway features to promote safety were based on either engineering judgment or adherence to accepted national guidance, such as A Policy on the Geometric Design of Highways and Streets, also known as The Green Book (AASHTO, 2011). These tools have allowed highway designers to produce facilities that have demonstrated an improving safety record in recent decades. However, these tools do not allow for the comparison of the safety performance of dissimilar facilities or roadway attributes. For example, the Green Book details the recommended minimum shoulder width for a freeway facility carrying 20,000 vehicles per day. However, it provides no quantifiable safety benefits of using this shoulder width, nor the costs and benefits of using a narrower or wider shoulder.

To address this gap, researchers have been working for decades to develop Crash Prediction Models (CPMs) that can estimate, and ideally predict an expected safety performance of a highway based on its geometric and traffic control features. With increases in computer processing technology and efforts at the national level, a method for safety-based decision making in the field of transportation engineering has gained momentum as a procedure for decision-making at the programmatic and project level. The largest step toward that goal was the adoption of the Highway Safety Manual (HSM) in 2010, published by the American Association of State Highway and Transportation Officials (AASHTO, 2010). The primary goal of the HSM is to provide a science-based technical approach to quantitative safety analysis.

1.1 Problem Statement and Methodology

The Kansas Department of Transportation (KDOT) has over 8,600 centerline miles of rural two-lane highways that it is in charge of maintaining. Having more advanced, predictive tools will allow road designers the ability to make better informed decisions that will allow for efficient decision making related to highway safety.

1.2 Research Objectives

The main focus of this research study was to evaluate the use of CPMs for rural two-lane highways in Kansas which included two major efforts:

- Calibration and evaluation of the CPMs provided in the HSM; and
- Development and evaluation of a Kansas-specific CPMs for roadway segments.

1.2.1 Development of HSM CPM for Kansas

The first objective of the study was to calibrate and validate the HSM CPM for rural two-lane two-way roadway segments using the Kansas highway system. Equation 1.1 (The HSM CPM equation), has a calibration factor intended to adjust the model for jurisdiction-specific conditions.

$$N_{predicted} = N_{spfx} \times (CMF_{1x} \times CMF_{2x} \times ... \times CMF_{yx}) \times C_x$$ **Equation 1.1**

Where:

$N_{predicted}$ = predicted average crash frequency for a specific year;

N_{spfx} = Safety Performance Function;

CMF_{yx} = Crash Modification Factors; and

C_x = calibration factor to adjust for local conditions.

As shown in Equation 1.1, in addition to the calibration factor, C_x, there are two other elements of the equation, the SPF and crash modification factors (CMFs). These elements are included to first predict a base number of crashes for a given traffic volume and then adjust the prediction to the specific conditions of the modeled roadway.

1.2.2 Development of a Kansas-Specific CPM

The second objective of this study was to create a CPM developed from Kansas data. The HSM recommends this step as a way to possibly improve the model accuracy since it’s based on data specific to a given jurisdiction. Several methods are available to develop such models and are listed in the HSM.

1.3 Contribution to the State of the Art

Based on the results of this research study, Kansas is considered a leader in the use of CPMs for project-level transportation decisions. The findings of this research study have been used nationally to shape future research in the area of CPM applications.

1.4 Report Structure and Other Related Resources

The research effort undertaken with this KTRAN study was a large effort including multiple reports, dissertations and conference proceedings. Those resources provide a more thorough and complete documentation of all of the efforts performed and alternatives considered as part of this study. This report has been formatted to summarize the major findings of the study that are most applicable to practitioners.

Chapter 2: Literature Review Background

The literature review for this project was originally performed in 2011 to guide the development of the project research plan. The review of literature was extensive and can be found in the other publications related to this research. Since the science of crash prediction modeling has become promising, there has been a vast amount of research performed and published on the subject since that time. For this reason and because the information is available elsewhere, it was determined there would be of little value to this report to republish the entire literature review. Selected Kansas specific literature was utilized during this study and since that work is especially relevant and that body of knowledge is relatively unchanged it is included herein.

2.1 Kansas Crash Prediction Research

The safety of the highway system and drivers is a paramount issue to The Kansas Department of Transportation (KDOT). Continuously improving the safety of its highway system, KDOT has commissioned numerous studies to address to specifically address safety. Three of the most recent contemporary studies addressed crash prediction on rural two-lane highway segments.

Similar to many other transportation organizations, KDOT has relied on research for more efficient ways to screen its already robust roadway system inventories and crash data for identifying relationships between highway features and safety. Najjar and Mandavilli (2009) used Artificial Neural Networks (ANN) in an effort to identify these relationships for Kansas highways. Their research specifically investigated the six major types of roadway network in Kansas: rural Kansas Turnpike Authority (KTA), rural two-lane, rural expressway, rural freeway, urban freeway, and urban expressway. The models developed evaluated not only the total crash rate, but also the fatal, injury, and severe injury crash rates. For rural two-lane highways, Najjar and Mandavilli identified eight significant and differing variables that were shown to impact crashes: section length, surface width, route class, shoulder width (outside), shoulder type (outside), average annual daily traffic (ADT), average percent of heavy trucks, and average speed limit.

The ANN models produced by Najjar and Mandavilli (2009) were measured against training, testing, and validation data sets. The rural two-lane model produced a coefficient of determination factor (R^2) of 0.4655. The total crash rate model would be the most similar to the HSM model being investigated with this research. The R^2 value for the total crash rate ANN model was 0.1728.

The research developed by Najjar and Mandavilli (2009) reported to be the “first in the nation to utilize the ANN mining approach to extract new and reliable traffic-crash correlations from historical databases.” This methodology was found to potentially provide a valid framework for future applications. However, specific results for rural two-lane highways in Kansas appeared to be inconsistent with engineering judgment, other research, and current practices. For example, one such result was the safety performance of similar width shoulders with differing pavement types. Due to identified practical limitations, the ANN model has not been implemented into practice by KDOT.

The only identified research to investigate animal crashes on Kansas highways was performed by Meyer (2006) as part of a research program sponsored by KDOT. This study, Assessing the Effectiveness of Deer Warning Signs, used a multiple layer regression, logistic regression, and principal component analysis to model the safety effectiveness of roadside deer warning signs based on a before-and-after data analysis where signs had been installed. While this analysis failed to produce a viable statistical model to aid in predicting the safety benefit of installing deer signs, or being able to prioritize segments for installation of signs, there were several important statistical findings (Meyer, 2006):

- The absence of the variable “presence of deer warning sign” suggested that there is limited or no relationship between deer warning signs and crash rate.
- The most significant variable found was the amount of surrounding area that was wooded. Most likely, the amount of wooded area was acting in these data as a surrogate for deer population.
- The sole direct measure of deer population (harvest density) was only available at an extremely coarse geographical resolution for this application.

- Other than percent wooded area, the other variable that was found to have a significant influence on crash rate were traffic volume, speed, sight distance (indirectly implied by the curvature ratio and side slope), and clear zone width.

With current guidance on how to perform statically accurate before-and-after study, it is possible that a developed model could be constructed to better quantify and qualify factors impacting deer crashes. However, the findings of this research are still valid and can aid in informing future consideration on the nature of animal crashes in Kansas.

The lack of measurable statistical benefits from the use of deer crossing signs was supported by a study conducted by Knapp (2005). The study synthesized all available research at the time on the safety benefits of deer crash roadway countermeasures. This research found that using exclusionary fencing and wildlife crossings indicated a positive safety benefit for reducing deer-vehicle crashes.

A study conducted by Rhys et al. (2010) evaluated the benefits of adding centerline rumble strips to two different rural two-lane highways in Kansas using a before-and-after analysis. Utilizing the Empirical Bayes (EB) method, the study found an 85 percent reduction in the targeted crash types: head-on and opposite sideswipe. They also found a 33 percent reduction in total crashes. It is worth noting that this study defined “total crashes” as excluding animal crashes. The findings of this study stated that “it can be assumed that overall results found in Kansas are comparable to results found by other states.” This is somewhat difficult to compare results of Rhys et al. to the HSM due to the fact that the CMF for centerline rumble strips also applies to one-half of run-off-the-road crashes.

However, the value given for reduction of target crashes for the centerline CMF is 0.79 (a 21 percent reduction). Therefore, the study conducted by Rhys et al. (2010) demonstrated a larger safety benefit for centerline rumble strips than what is currently shown in the HSM.

An additional finding of the Rhys et al. (2010) study was the creation of Safety Performance Functions (SPFs) for roads similar to the two test sections analyzed. This was developed to specifically isolate the safety benefits of the rumble strips. The equation developed by the research team for similar rural two-way highways is as follows:

$$ACC = e^{\beta_0} \times e^{(AADT_{before} \times \beta_1)}$$ **Equation 2.1**

Where:

ACC = expected number of crashes (per mile per year) in a section with the same characteristics to the section of interest;

$AADT_{before}$ = average AADT for the before period;

β_0 = -1.4019 (section A), -1.2229 (section B); and

β_1 = 0.0004 (section A), 0.0007 (section B).

An overdispersion factor was also calculated for the Equation 1.1. It equaled -0.0793 for section A and -0.1475 for section B. The two sections cited in this report *A* and *B*, refer to the two different sections that were investigated for crash reduction due to the addition of a segment of centerline rumble strips. Highways with similar traffic volumes, road geometry, and crash history were used to develop an SPF for each roadway type.

Chapter 3: Data Collection

Existing publications including the HSM have identified various roadway elements that have been shown previously to be statically significant and impact the likelihood of vehicle crashes. The major component to this KTRAN study was to collect data for as many of these roadway elements as practical in order to develop the most complete understanding of rural two-lane highway crashes in Kansas.

3.1 Data Sources

KDOT maintains roadway and crash databases along with existing as-built plans which were the main data sources for much of the study. This was due to the fact that the data were more easily accessible, and would be convenient for KDOT practitioners to access these elements in the future. When the databases failed to contain critical and necessary roadway elements, other data sources were used to supplement missing elements.

3.1.1 CANSYS Database

The CANSYS database contained most roadway features for KDOT using combined sources and coded at different intervals, which also introduced inconsistencies as to where specific changes did occur. The database also gave information to rule out: urban or multi-lane facilities in the Kansas road system, locations of crashes, and intersections. The data used from the CANSYS database included: shoulder width, lane width, and 2007 average annual daily traffic (AADT).

3.1.2 KDOT Crash Database

At the time of this study, KDOT ran a separate crash database that provided information on each crash incident including: the location, the type, and the severity of the crash among many other details. All crash reports from 2005 through 2007 were collected and reduced. The crash database also allowed the research team to sort crashes by type and frequency. A significant effort was needed to merge the crash and roadway feature databases to form the final dataset.

Crashes listed in the crash report as occurring at an intersection, or as being intersection related were associated with the intersection CPMs. Any crash not shown as occurring at one of those two locations was associated with the segment CPM.

3.1.3 Highway Design Plans

As stated previously, not all roadway geometric features were coded into the CANSYS database. Therefore, some of the roadway features were extracted from actual as-built plans. Data mining from these plan sets were found to be a time consuming process; however, it provided critical data elements for horizontal curvature, vertical grades, some of the information needed to determine the roadside hazard rating (RHR), and allowed the research team to verify the data that were coded in the CANSYS database.

3.1.4 Aerial Imagery

Aerial Imagery taken from Google Earth and Google Maps were aided the research team in determining a RHR by giving an estimated width of the clear zone area. These online applications were also useful to determine the density of private driveways along a selected segment.

3.1.5 KDOT Maps

KDOT currently had a database of various maps that was also useful in obtaining the remaining missing data elements needed. Statewide vehicular traffic count maps were able to provide the AADT for years other than 2007. A map reporting the posted speed limits for rural highways in the state of Kansas was also used for obtaining the speed limit on each highway.

3.2 Random Segment Generation

One critical element of the data collection process identified by the research team was that the HSM and standard research protocol recommend: if the entire jurisdiction’s data for that facility type is not available, that the use of randomly selected locations for data collection be utilized.

As part of a previous KTRAN project, Review and Analysis of the Kansas Department of Transportation Maintenance Quality Assurance Program (Schrock et al. 2009), the University of Kansas developed a random segment generator to aid with the maintenance quality assurance (MQA) program. For this study, a modified version of that generator was developed. The generator for this study was populated by the same data used for the MQA program. The primary difference was that the generator allowed the user to vary the length of the random roadway segment. While any method can be used to randomly select roadway segments for performing the model calibration, this generator investigated the entire Kansas highway system and adjusted for proper highway termini. Two negatives of the generator found by the research team are that it required manual screening of two-lane rural sections and provided the data in state mileposts. Since other KDOT data sources generated data in county milepost, the data therefore had to be converted. This conversion was accomplished by manually reviewing a state milepost to county milepost conversion chart and changing the values.

3.3 Minor Road Traffic Volumes

The research team found that one of the most difficult pieces of data to collect was the side road volume for minor roads for the intersection CPM. Exposure to crashes at a location is one of the elements with the highest correlation to crash expectancy. For that reason, it was critical that traffic volumes were collected for all side roads of intersections identified in the calibration sections. Published traffic volume data were available for highways and rural secondary (RS) routes. However, KDOT currently only develops estimated traffic volumes on local roads when requested for a specific funded project, and then only the roads projected to have over 200 vehicles per day are analyzed. No data are typically provided for the remaining low volume side roads. The process for developing traffic estimates for low volume side roads was fairly time intensive, however it is believed by the research team to be closer to the accepted trip generation model.

Based on interviews with KDOT planning staff, traffic volumes on local roads were estimated by developing a ‘travel shed’ for each local road as it intersects a highway. The process of developing a travel shed is similar to developing a watershed. The planner investigates the local road in relation to the rest of the road network and determines the area in which people are likely to drive from their destination, along the side road, and through the intersection being analyzed. Next, the planner will count the number of traffic generators within the given travel shed, including homes and businesses. The Traffic Generation Manual can then be consulted to estimate the daily volume produced by each traffic generator within the travel shed. Finally, the volumes generated for each site within the travel shed are totaled and an estimated traffic volume is developed for the local road.

A travel shed for each local road intersection was developed to produce a representative sample of the minor leg volumes to be expected for 2-lane rural roads in Kansas. Travel shed volumes were also developed at RS routes for comparison to the published volumes but were not used in any other analysis.

3.4 Summary

- All of the data both required and desirable were collected for the CPMs provided in the HSM. Since no default values were utilized, this study examined the full capacity of the HSM CPM for rural two-lane highways.
- Once established processes were developed most data elements were relatively easily collected through the available data sources listed in previous sections. The exception was traffic volumes for minor, low volume roads which proved especially resource consuming.

Chapter 4: Methodology

Once the data collection methodology and sources were gathered, the next step in the research study was to screen the data and reduce it into appropriate groups to perform the calibration, model development, and validation. Important distinctions and methods were developed through this project and were not cited in any previous research.

4.1 Definition of "Rural"

The HSM utilizes the Federal Highway Administration (FHWA) definition of "rural" which is any highway located outside a city (or incorporated area) with a population of 5,000 or more. During the data mining process an inconsistency was discovered in the application of the FHWA definition of "rural" for Kansas highways as it applies to the HSM model. Some of the random highway segments that were generated for the analysis contained portions that traveled through cities with populations under 5,000 people. The typical sections for the highways in these cities were two-lane, or short four-lane, so they would otherwise qualify for analysis using the HSM model. However, other features of the highway were not consistent with the two-lane rural model. Some sections included: curb and gutter, storm sewer, on-street parking, sidewalks, and downtown-style development. These sections, which qualified under the HSM model definition, could not accurately be modeled using the rural two-lane model. For this reason, the definition of "rural" for applications on Kansas highways was modified to exclude segments traveling through cities of any population. This was noted as a significant finding because at the time of this research, Kansas contained roughly 587 cities with a population fewer than 5,000 and nearly all of them were served directly by a highway. All data for this study were modified to exclude any section that passed through a city of any size.

4.2 Segments

The HSM two-lane rural CPM has 18 variables that are used to calculate expected crashes. Additionally, existing research recommended several additional variables associated with horizontal and vertical alignments that should be considered when developing a state specific CPM. To satisfy the different needs through the evolution of this research, all of these data elements were collected for three distinct segment data groups.

4.2.1 Data Group 1

The HSM recommends a minimum number of crashes per year to provide an appropriately sized calibration data set. Data Group 1 was developed to produce a data set that met this size requirement and minimized bias. Data for the sections in this group were for 2005 through 2007.

The use of randomly selected highway segments provided the least biased data for the purpose of calibration and model development. Ten-mile long sections were selected to minimize the likelihood that a crash occurred within the study section and was inappropriately assigned outside of the section. Additionally, longer sections made data collection more efficient by reducing the total number of existing plans that would need to be utilized.

Fifty random ten-mile sections were generated using the modified version of the program developed to choose random highway segments for KDOT's MQA program. Nine of the sections were removed from future consideration because they had elements that violated the HSM two-lane rural model parameters. These violations included sections that were in urban areas and some four-lane sections. The combined CANSYS and crash database information was then referenced to determine how many segment crashes occurred within each ten mile segment.

It was determined that going through the list of random sections until the minimum number of crashes was reached would bias the data set to sections with high crash frequency. To address this potential bias, a statistical analysis of crash frequency on KDOT highway segments was performed from the remaining 41 sections. The mean number of crashes for the 41 sections was 18 and the standard deviation was 15. These values are for the full three-year period (2005 – 2007) that crash data were collected.

It was then decided to use a conservative value for the number of sections that would be evaluated to develop the calibration value. Therefore, the calculation to determine the necessary number of sections was based on two standard deviations from the mean to produce the HSM minimum recommendation of 100 segment crashes per year. Assuming a normal distribution of crashes per ten mile section, it was estimated that 19 ten mile sections were appropriate for Data Group 1.

The list of 41 ten-mile sections was again used to select the 19 sections that would be Data Group 1. Some bias was intentionally added to the section selection to assure a geographic distribution throughout the state of Kansas. To accomplish this geographic distribution, a minimum of three sections were selected from each of KDOT's six geographic districts. Sections were then chosen from the top of the randomly generated list until each district had at least three sections.

4.2.2 Data Group 2

The primary function of Data Group 2 was to develop a data set that most closely mimicked how the HSM CPM would be utilized by KDOT. For that reason, sections in Data Group 2 were selected that corresponded to a highway reconstruction project that was performed between 1999 and 2003. This timeframe allowed sufficient data after the project was constructed to compare the predicted versus observed crash performance. Selection of segments that experienced a geometric improvement project would also properly assess the model's ability to use existing crash data on the unimproved system to predict safety performance on the future improved section. This is more consistent with KDOT practice than analyzing segments that are static over time.

Ten projects were selected from a list of "Modernization – Safety & Shoulder Improvements" which included projects greater than 2.5 miles long in the order they were provided from the database query. To provide a mixed geographical representation, bias was added to this selection to ensure that at least one project was selected from each of KDOT's six districts. Some final modifications were performed within the limits of the ten selected projects to remove any sections that passed through a city. Table 4.1 contains a list of the validation projects/sections that were selected for analysis.

TABLE 4.1
Selected KDOT Projects for Validation

Section	Project Number	Route	County	District	County Milepost	
					Begin	End
1	K-5393-01	K-383	Norton	3	0	13.618
2	K-5384-01	US-50	Chase	2	20.671	28.486
3	K-5745-01	US-56	Marion	2	32.051	39.815
4	K-5767-01	US-77	Butler	5	0	12.713
5	K-5391-01	US-283	Ness	6	13.944	30.202
6	K-5761-01	US-73	Atchison	1	0	4.142
7	K-5757-01	K-47	Wilson	4	5.573	7.747
8	K-5741-01	US-36	Rawlins	3	28.472	36.393
9	K-5749-01	K-150	Barton	5	18.61	35.81
10	K-5743-01	US-50	Hamilton	6	17.217	28.498

To be consistent with anticipated future practices, crash data for Data Group 2 were requested for the three years prior to the project construction and then for all of the years from the project completion through 2009. If construction was completed in the middle of a year, that full year was dropped to avoid biasing the data with seasonal impacts on crash frequency.

4.2.3 Data Group 3

To account for the more robust data needs of state-specific CPM construction a third data set was developed. A process identical to that used to develop Data Group 2 (an explained in previous sections) was used to select nine additional projects to produce Data Group 3 as shown in Table 4.2. Four of six KDOT districts were represented in Data Group 3's project list due to the limited number of projects that fit the time period and type of projects required.

TABLE 4.2
Selected KDOT CPM Validation Projects

Section	Project Number	Route	County	District	County Milepost	
					Begin	End
1	K-6777-01	K-150	Marion	2	0	8.008
2	K-5754-01	US-36	Rawlins	3	20.294	28.662
3	K-5769-01	K-150	Chase	2	23.135	29.998
4	K-6372-01	US-24	Osborne	3	4.693	11.393
5	K-5768-01	US-77	Marion	2	16.992	27.959
6	K-5358-01	US-50	Marion	2	16.766	20.995
7	K-5752-01	US-283	Norton	3	32.049	32.049
8	K-5740-01	K-27	Sherman	3	20.44	30.684
9	K-5738-01	K-27	Sherman	3	16.299	20.44

4.3 Intersections

At the time of this study, KDOT did not have an intersection database to provide such details as the number of intersections on two-lane rural highways to be able to calculate the number of intersections that would reasonably satisfy the minimum data requirements for calibration as recommended in the HSM. Due to data scarcity, a data set was selected based on a reasonable data collection effort. Originally, 30 random ten mile sections were generated for this data set. Four of those sections were screened out because they had elements that violated the HSM two-lane rural model parameters. The remaining 26 ten-mile sections were then carried forward to develop the intersection calibration.

The 26 sections in the calibration data set yielded a total of 278 intersections which experienced 37 crashes over the three year study period (2005-2007). This is compared to the 300 minimum crashes per intersection type recommended by the HSM to calibrate the model for that same period. It was also found that none of the intersections in the calibration data set were signalized. Consequently, it was determined that there were not enough signalized intersections on rural two-lane highways in Kansas to justify calibration of these models. Additionally, the crash frequency for stop-controlled intersection along these routes were low enough that a single calibration function was developed for the four-leg and three-leg minor stop control intersection models.

4.4 Summary

- Application of the HSM for Kansas rural highways should only account for sections of highways that do not travel through a city of any size. This is a level of screening that was not previously considered in any other studies. It’s more limiting than the HSM definition, which follows the FHWA definition of segments outside a city of a population 5,000 or greater.
- Data Groups 2 and 3 were developed in a manner that was most consistent with how the HSM CPMs would be utilized in practice. This is unique to this particular study as compared to any previous research.
- Crashes attributed to intersections along rural two-lane highways in Kansas are so infrequent that, given the level of effort needed to collect additional data, minimum thresholds recommended in the HSM for calibration are not achievable by KDOT until a full intersection inventory of the system can be completed.

Chapter 5: Model Calibration/Development

As stated previously, the primary goal of this research study was to develop a calibrated crash prediction model for rural two-lane highways in Kansas. Many established and non-traditional methods of developing new models or calibrating existing models were considered during this study and can be found in other documented sources by the research team. In an effort to provide a succinct document that is practitioner friendly, only most promising results are presented herein.

5.1 Crash Distribution

The first step recommended by the HSM is to calibrate the CPMs and replace default crash type and severity distribution tables with ones based on data from the local jurisdiction under investigation. Since this process also provided an insight into the nature of crashes on Kansas two-lane rural highways and how that experience might differ from the HSM models, this was the first step in this research methodology. All data from the combined CANSYS/Crash Report databases attributes to rural two-lane highways were utilized to develop these distributions and replacement HSM tables as shown in the following tables. The HSM recommends replacement of only certain default values for two-lane rural highways as shown in Table 5.1.

TABLE 5.1
Default Crash Distributions Used in Part C Predictive Models Which May Be Calibrated by Users to Local Conditions

Table of Equation Number	Data element or Distribution That May Be Calibrated to Local Conditions
Table: 10-3	Crash severity by facility type for roadway segments
Table: 10-4	Collision type by facility type for roadway segments
Table: 10-5	Crash severity by facility type for intersections
Table: 10-6	Collision type by facility type for intersections
Equation: 10-18	Driveway-related crash as a proportion of total crashes (p_{dwy})
Table: 10-12	Nighttime crashes as a proportion of total crashes by severity level
Table: 10-15	Nighttime crashes as a proportion of total crashes by severity level and by intersection type

(Source: Table A-3, HSM)

Prior to 2007, the Kansas Highway Patrol (KHP) motor vehicle crash reports did not include the type of intersection at which a crash had occurred. The HSM requires separate tables for all of the different intersection types, but for the purposes of this study only one distribution was developed for both three-leg and four-leg stop controlled intersection types. While this did not provide the accuracy called for in the HSM, it provided a more accurate distribution than if the default tables provided by the HSM were used. Since Kansas highways did not have nearly enough rural signalized intersection crashes to develop its own distribution, the default distribution was used for analysis of four-leg signalized intersections.

Standard KHP motor vehicle crash reports list crash severity, collision type, whether the crash is intersection related or not, what type of traffic control were present, and light conditions. The crash reports also have a driveway-related crash location called, “Access to Parking Lot/Driveway.” This was used to develop a KDOT specific value that was inserted into HSM Equation 10-18 as shown in Table 5.1. Therefore, KDOT specific values were able to be calculated for the all of the recommended segment tables and equations with minor modifications needed to the basic report data provided.

Examples of interpretation of the standard KHP fields were needed to categorize the collision types into similar categories as were provided by the HSM. Shown in Table 5.2 is the distribution for crash severity level on rural two-lane, two-way roadway segments developed for KDOT based on HSM Table 10-3. This distribution was developed by analyzing all crashes in the data set that were considered not intersection or intersection-related. Each crash was counted only once and was attributed to the highest severity level. For example, if a crash had both incapacitating injuries and non-incapacitating injuries, it was only counted as incapacitating.

TABLE 5.2
Default Distribution for Crash Severity Level on Rural Two-Lane, Two-Way Roadway Segments

Crash Severity Level	KDOT		HSM
	Count	Percent (%)	Percent (%)
Fatal	270	1.5%	1.3%
Incapacitating (Disabled) Injuries	495	2.7%	5.4%
Non-Incapacitating Injuries	1,574	8.7%	10.9%
Possible Injury	966	5.3%	14.5%
Total Fatal and Injury	3,305	18.3%	32.1%
Property Damage Only	14,791	81.7%	67.9%
Total	18,096	100.0%	100.0%

Shown in Table 5.3 is the default distribution by collision type for specific crash severity levels on two-lane, two-way roadway segments developed for KDOT. As shown, the same crashes used in Table 5.2 were used, but then broken down further by collision type. Once the crashes were distributed into Property Damage Only (PDO) and Total Fatal and Injury, the crashes were assigned using the collision types available in the standard KHP motor vehicle crash reports.

TABLE 5.3
Crashes by District and Severity

Collision Type	Fatal & Injury		PDO		Total Crashes	
	Count	(%)	Count	(%)	Count	(%)
Collision with Animal	345	10.4%	10,320	69.8%	10,665	58.9%
Collision with Pedestrian	22	0.7%	0	0.0%	22	0.1%
Collision with Cyclist	13	0.4%	0	0.0%	13	0.1%
Overturned	893	27.0%	559	3.8%	1,452	8.0%
Ran Off Road	481	14.5%	754	5.1%	1,235	6.8%
Collision with Legally Parked Vehicle	13	0.4%	89	0.6%	102	0.6%
Collision with Railway Train	5	0.2%	0	0.0%	5	0.0%
Collision with Fixed Object	644	19.5%	1,312	8.9%	1,953	10.8%
Collision with Other Object	13	0.4%	138	0.9%	151	0.8%
Other Non-Collision	64	1.9%	300	2.0%	364	2.0%
Total Single Vehicle Crashes	**2,493**	**75.4%**	**13,472**	**91.1%**	**15,965**	**88.2%**
Angle Collision	192	5.8%	221	1.5%	413	230.0%
Head-On Collision	167	5.0%	27	0.2%	194	1.1%
Read-End Collision	266	8.0%	471	3.2%	737	4.1%
Sideswipe: Opposing Direction	135	4.1%	187	1.3%	322	1.8%
Sideswipe: Same Direction	36	1.1%	203	1.4%	239	1.3%
Backed Into	6	0.2%	92	0.6%	98	0.5%
Other	11	0.3%	113	0.8%	124	0.7%
Unknown	2	0.1%	2	0.0%	4	0.0%
Total Multiple Vehicle Crashes	**815**	**24.6%**	**1,316**	**8.9%**	**2,131**	**11.8%**

Since the collision types available in the standard KHP motor vehicle crash report failed to match those provided in the HSM, additional sorting was required in order to compare values between sources. For the variables single vehicle crashes collisions with legally parked vehicles, fixed objects, and other objects were assigned to “Ran Off Road.” Due to all of these elements existing outside of the normal roadway, it was assumed a vehicle departed the roadway and collided with them. “Collision with Railway Train” was combined with “Other Non-Collision” under the heading “Other Single Vehicle Accident.” Similarly, in the Multiple-Vehicle crashes category, the “Backed Into” and “Unknown” collision types were assigned to the “Other” category. After performing this sorting method, the following collision type distribution was developed for KDOT data to replace Table 10-4 in the HSM as shown in Table 5.4.

TABLE 5.4
Collision Type Distribution for KDOT Data as Compared to HSM Distribution

Collision Type	KDOT			HSM		
	Fatal & Injury	**PDO**	**Total**	**Fatal & Injury**	**PDO**	**Total**
Collision with Animal	10.4%	69.8%	58.9%	3.8%	18.4%	12.1%
Collision with Pedestrian	40.0%	0.0%	0.1%	0.7%	0.1%	0.3%
Collision with Cyclist	0.4%	0.0%	0.1%	0.7%	0.1%	0.3%
Overturned	27.0%	3.8%	8.0%	3.7%	1.5%	2.5%
Ran Off Road	34.8%	15.5%	19.0%	54.5%	50.5%	52.1%
Other Single Vehicle	2.1%	2.0%	2.1%	0.7%	2.9%	2.1%
Total Single Vehicle Crashes	**75.4%**	**91.1%**	**88.2%**	**63.8%**	**73.5%**	**69.3%**
Angle Collision	5.8%	1.5%	2.3%	10.1%	7.2%	8.5%
Head-On Collision	5.0%	0.2%	1.1%	3.4%	0.3%	1.6%
Read-End Collision	8.0%	3.2%	4.1%	16.5%	12.2%	14.2%
Sideswipe Collision	5.2%	2.7%	3.1%	3.8%	3.8%	3.7%
Other Multiple Vehicle	0.6%	1.3%	1.2%	2.6%	3.0%	2.7%
Total Multiple Vehicle Crashes	**24.6%**	**8.9%**	**11.8%**	**36.2%**	**26.3%**	**30.7%**

Table 5.5 was developed for KDOT specific jurisdiction based on the HSM Table 10-5, default distribution for crash severity level on rural two-lane, two-way intersections. This distribution was developed by analyzing all crashes in the data set that were labeled as intersection or intersection-related. Similar to the segment crashes, each crash was counted only once and was attributed to the highest severity level.

TABLE 5.5
Distribution for Crash Severity Level on Rural Two-Lane, Two-Way Intersections for KDOT Data as Compared to HSM Distribution

Crash Severity Level	KDOT		HSM		
	Crash Count	**(%)**	**Three-Leg Stop Controlled**	**Four-Leg Stop Controlled**	**Four-Leg Signalized**
Fatal	62	2.5%	1.7%	1.8%	0.9%
Incapacitating (Disabled) Injury	135	5.4%	4.0%	4.3%	2.1%
Non-Incapacitating Injury	418	16.6%	16.6%	16.2%	10.5%
Possible Injury	281	11.2%	19.2%	20.8%	20.5%
Total Fatal and Injury	896	35.6%	41.5%	43.1%	64.0%
PDO	1,618	64.4%	58.5%	56.9%	66.0%
Total	2,514	100.0%	100.0%	100.0%	100.0%

Table 5.6 was developed for KDOT specific jurisdiction based on the HSM Table 10-6, Default Distribution by Collision Type and Manner of Collision at Rural Two-Way Intersections. As shown in Table 5.6, the same crashes used for the HSM Table 10-5 were utilized, but were further refined by collision type. Once the crashes were distributed into PDO and total fatal and injury, the crashes were assigned using the collision types available in the standard KHP motor vehicle crash reports. The results of the distribution are shown in Table 5.6.

TABLE 5.6
Distribution by Collision Type and Manner of Collision at Rural Two-Way Intersections for KDOT Data as Compared to HSM Distribution

Collision Type	Fatal and Injury		PDO		Total Crashes	
	Count	(%)	Count	(%)	Count	(%)
Collision with Animal	5	0.6%	233	14.4%	238	9.5%
Collision with Pedestrian	4	0.4%	2	0.1%	6	0.2%
Collision with Cyclist	1	0.1%	1	0.1%	2	0.1%
Overturned	53	5.9%	41	2.5%	94	3.7%
Ran Off Road	0	0.0%	0	0.0%	0	0.0%
Collision with Legally Parked Vehicle	1	0.1%	4	0.2%	5	0.2%
Collision with Railway Train	0	0.0%	0	0.0%	0	0.0%
Collision with Fixed Object	97	10.9%	192	11.8%	289	11.5%
Collision with Other Object	0	0.0%	13	0.8%	13	0.5%
Other Non-Collision	12	1.3%	35	2.2%	47	1.9%
Total Single Vehicle Crashes	**173**	**19.4%**	**521**	**32.1%**	**694**	**27.6%**
Angle Collision	388	43.4%	474	29.2%	862	34.3%
Head-On Collision	31	3.5%	19	1.2%	50	2.0%
Read-End Collision	250	28.0%	388	23.9%	638	25.4%
Sideswipe: Opposing Direction	12	1.3%	39	2.4%	51	2.0%
Sideswipe: Same Direction	37	4.1%	154	9.5%	191	7.6%
Backed Into	1	0.1%	21	1.3%	22	0.9%
Other	1	0.1%	4	0.2%	5	0.2%
Unknown	0	0.0%	1	0.1%	1	0.0%
Total Multiple Vehicle Crashes	**720**	**80.6%**	**1,100**	**67.9%**	**1,820**	**72.4%**

A similar sorting method described previously for segments of the crashes was necessary for the intersections. After performing this sorting, the following collision type distribution was developed for KDOT data to replace HSM Table 10-6 as shown in Table 5.7. It should be noted that the HSM default values also provided for contrast.

25

TABLE 5.7
Collision Type Distribution for KDOT Data as Compared to HSM Distribution

Collision Type	KDOT			HSM (3ST)			HSM (4ST)			HSM (4SG)		
	F&I[A]	PDO	Total	F&I	PDO	Total	F&I	PDO	Total	F&I	PDO	Total
SINGLE VEHICLE CRASHES												
Collision with Animal	0.6%	14.4%	9.5%	0.8%	2.6%	1.9%	0.6%	1.4%	1.0%	0.0%	0.3%	0.2%
Collision with Pedestrian	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Collision with Cyclist	0.4%	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	10.0%	0.1%	0.1%
Overturned	5.9%	2.5%	3.7%	2.2%	0.7%	1.3%	0.6%	0.4%	0.5%	0.3%	0.3%	0.3%
Ran Off Road	11.0%	12.9%	12.2%	24.0%	24.7%	24.4%	9.4%	14.4%	12.2%	3.2%	8.1%	6.4%
Other Single Vehicle	1.3%	2.2%	1.9%	1.9%	2.0%	1.6%	0.4%	1.0%	0.8%	0.3%	1.8%	0.5%
Total Single Vehicle Crashes	19.4%	32.1%	27.6%	28.3%	30.2%	29.4%	11.2%	17.4%	14.7%	4.0%	10.7%	7.6%
MULTIPLE VEHICLE CRASHES												
Angle Collision	43.4%	29.2%	34.3%	27.5%	21.0%	23.7%	53.2%	35.4%	43.1%	33.6%	24.2%	27.4%
Head-On Collision	3.5%	1.2%	2.0%	8.1%	3.2%	5.2%	6.0%	2.5%	4.0%	8.0%	4.0%	5.4%
Read-End Collision	28.0%	23.9%	25.4%	26.0%	29.2%	27.8%	21.0%	26.6%	24.2%	40.3%	43.8%	42.6%
Sideswipe Collision	5.5%	11.9%	9.6%	5.1%	13.1%	9.7%	4.4%	14.4%	10.1%	5.1%	15.3%	11.8%
Other Multiple Vehicle	0.2%	1.6%	1.1%	5.0%	3.3%	4.2%	4.2%	3.7%	3.9%	9.0%	2.0%	5.2%
Total Multiple Vehicle Crashes	80.6%	67.9%	72.4%	71.7%	69.8%	70.6%	88.8%	82.6%	85.3%	96.0%	89.3%	92.4%

[A]Fatal and Injury

The HSM Equation 10-18 allows for replacement of a jurisdiction's specific value for the percentage of driveway-related crashes as a portion of total number of crashes. Based on the data extracted from the KDOT crash database, a total of 18,096 segment crashes were found. According to the crash data, 284 of them were driveway or parking lot related. Therefore, this yields a proportion of p_{dwy} equal to 0.016.

Another table described by the HSM for roadway segments Table 10-12, Nighttime Crash Proportions for Unlighted Roadway Segments. The KHP motor vehicle crash reports have five listed values for roadway lighting conditions:

- Daylight,
- Dawn,
- Dusk,
- Dark: Street Lights On,
- Dark: No Street Lights, and
- Unknown.

For the purpose of determining the proportions necessary for the HSM Table 10-12, the crashes labeled either "Dark: Street Lights On" or "Unknown" were discarded in the total count of crashes. Crashes for dawn and dusk were summed and were then half assigned to each light and dark variable. Shown in Table 5.8 are the numbers of segment crashes in each category.

TABLE 5.8
Number of Kansas Segment Crashes by Lighting Condition and Severity

Lighting Condition	Fatal and Injury	PDO	Total
Light	2,660	6,390	9,050
Dark: Street Lights On	231	1,147	1,378
Dark: Street Lights Off	1,304	8,835	10,139
Unknown	7	36	13
Total	**3,964**	**15,225**	**19,189**

From this data shown in Table 5.8, Table 5.9 was created to replacement values that are presented in the HSM Table 10-12.

TABLE 5.9
Values for HSM Table 10-12 based on Kansas Crash Data and Table 5.8

Data Source	Proportion of Total Crashes by Severity Type		Proportion of Crashes That Occur at Night
	Fatal and Injury (p_{inr})	PDO (p_{pnr})	(p_{pnr})
KDOT	0.129	0.871	0.53
HSM	0.382	0.618	0.37

The final table found in the HSM for roadway segments in Table 10-15, Nighttime Crash Proportions for Unlighted Intersections. Similar to previously created tables for roadway segments, a similar methodology was used to sort intersections by lighting condition. Summarized in Table 5.10 are the total numbers of intersection crashes in Kansas for various lighting conditions.

TABLE 5.10
Kansas Intersection Crashes for Intersections

Lighting Condition	Total
Light	1,850.5
Dark: Street Lights On	172.0
Dark: Street Lights Off	487.5
Unknown	6.0
Total	2,338.0

Using the total number of crashes for each category in Table 5.10, replacement values for the HSM Table 10-15 was developed and shown in Table 5.11.

TABLE 5.11
Values for HSM Table 10-15 Based on Kansas Intersection Crash Data and Table 5.10

Data Source	Proportion of Crashes that Occur at Night
	(pnr)
KDOT	0.209
HSM (3-Way Stop Controlled)	0.260
HSM (4-Way Stop Controlled)	0.244
HSM (4-Way Signalized)	0.286

5.2 Analysis

Analysis of the second distribution regarding collision types is the most telling regarding how the nature of crashes on KDOT highways could impact how those crashes are modeled. Fifty-eight-point-nine percent of segment crashes on KDOT highways were collisions with animals. This is compared to only 12.1 percent of crashes in the default distribution. This is significant first because the KDOT value is almost five times higher than the default value. It is also significant because animal collision crashes account for a majority of crashes on KDOT highways. Animal collisions are typically rare and random crashes and difficult to attribute them to the geometric design of a highway. The ability to model animal collisions has a significant impact on crash prediction on KDOT highways. Therefore, this issue will be examined in further depth in the following sections.

5.3 Segment Analysis

Crashes attributed to roadway segments as opposed to intersections along rural two-lane highways in Kansas account for 93 percent of the crashes along these facilities. For this reason, a significant amount of effort in model development and calibration completed in this study was dedicated to roadway segments. Though many different models and calibration techniques were considered, the three most promising are presented in this report. These are: development of a single statewide calibration factor for the HSM CMP, a calibration by county based on a state-specific calibration function for the HSM CMP, and a state specific CPM.

5.3.1. Statewide Calibration Factor

To account for jurisdictional differences in climate, driver populations, animal populations, crash reporting thresholds, and crash reporting system procedures a calibration factor (‘C’) can be developed. The factors are based on a ratio of the number of observed crashes for a particular site versus the number of predicted crashes for the same site. The HSM suggests developing different calibration factors within a given jurisdiction if there are significant variations in climate or topography.

Data Group 1 was analyzed for such variations relative to large geographic regions in the state of Kansas. Since no trends could be found for these large regions, a single statewide calibration factor for rural two lane highway segments was developed using Equation 5.1.

$$C_r \left(\text{or } C_i\right) = \frac{\sum_{\text{all sites}} \text{observed crashes}}{\sum_{\text{all sites}} \text{predicted crashes}}$$

Equation 5.1

Shown in Table 5.12 are the results of the analysis including the final statewide calibration factor of 1.48.

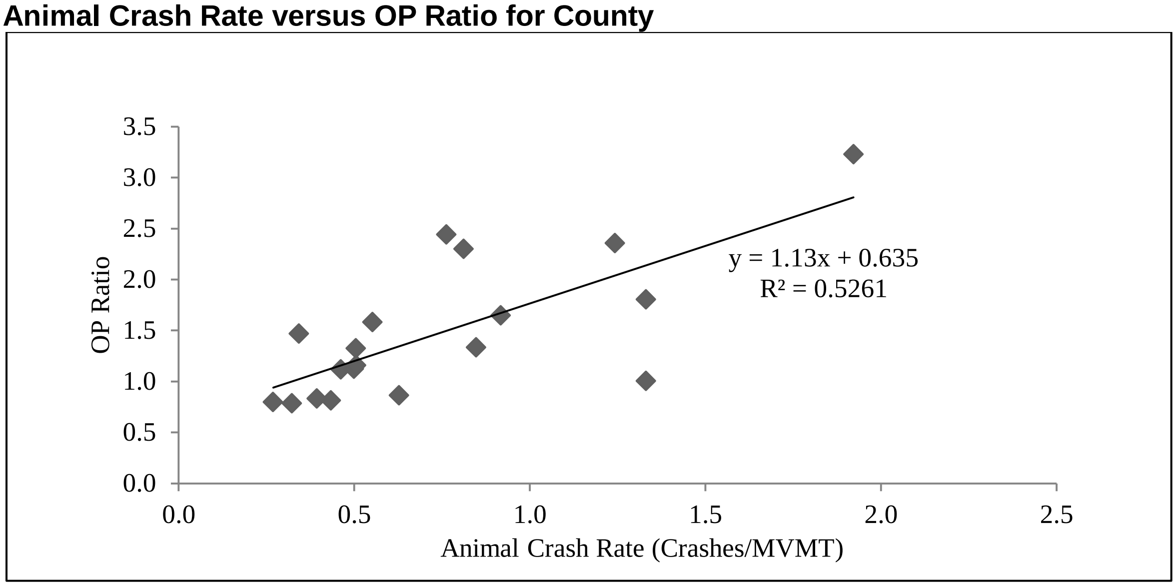
TABLE 5.12
Crash Prediction Results for Data Group 1

Section	District	AADT	Predicted	Observed	OP Ratio
1	6	1457	12.26	18	1.47
2	5	1389	30.12	26	0.86
3	6	159	3.76	3	0.80
4	2	1388	9.86	8	0.81
5	6	497	3.83	3	0.78
6	2	778	6.80	9	1.32
7	6	1000	8.05	9	1.12
8	4	3498	16.54	42	1.58
9	4	3921	26.98	36	1.33
10	3	406	2.99	3	1.00
11	5	2140	17.01	28	1.65
12	5	1925	14.86	35	2.36
13	3	1941	14.46	12	0.83
14	3	1704	13.30	24	1.80
15	1	3038	25.24	58	2.30
16	4	4365	32.05	36	1.12
17	2	1337	10.53	34	3.23
18	1	4030	30.24	35	1.16
19	1	795	7.38	18	2.44
Total			**296.26**	**437**	**1.48**

5.3.2 KDOT Calibration Function

Through the analysis of Data Group 1, the predominant crash type for rural two-lane highways in Kansas was animal collisions. Not only did this crash type represent such a substantial hazard for Kansas drivers on this type of facility, the Kansas experience deviated greatly from the states whose data were used to develop the original HSM models. To further investigate this, the research team investigated multiple ways to model the impact of animal crashes on the total crash predictions. The most promising method determined for addressing this proved to be a calibration function where a different calibration value was calculated for each county depending on the animal crash rate in the specific county. The higher the rate of animal crashes in a given county, the higher the calibration value would be, as defined by the OP ratio. The OP ratio is simply the ratio of the observed crashes to the predicted crashes. Figure 5.1 shows this specific relationship.

FIGURE 5.1
Animal Crash Rate versus OP Ratio for County



From this relationship the following Kansas specific equation was developed from Data Group 1:

$$C_{county} = 1.13 \times ACR_{county} + 0.635$$ **Equation 5.2**

Where:

CCounty = Calibration factor for a county; and

ACRcounty = Deer crash rate for a county.

5.3.3 KDOT Specific Crash Prediction Models

There are several statistical methods which have shown to be applicable in creating a CPM. These include Poisson regression, ZIP regression, and negative binomial regression models. In a review of literature, Miaou (1994) found that different regression models produced similar equations and there was no one superior model. Therefore, a negative binomial regression was selected as the statistical method due to the HSM's preference to replace it as a base SPF because it accounts for overdispersion.

To create a CPM, data was first collected. Data Groups 1 and 2 were run through SPSS software to determine the coefficients of the CPMs. The coefficients for the variables were kept if they were found to be significant at the 95 percent level of confidence. Several equations were created including forms similar to the HSM's safety performance function (SPF) having exponents on both the AADT and Segment Length, and separating out prediction of animal and non-animal crashes. The accuracy of the developed equations were determined by considering several statistical tests as there was no one test that showed which model was best. Statistical tests included Pearson's *R* to consider the correlation, the t-test which indicated significance, Mean Prediction Bias (MPB) which considered the overdispersion, and the Mean Absolute Deviation (MAD) which gave the extent of variability. Considering all of these tests together gives the best picture of which models will perform optimally. Shown in Equation 5.3 is the equation that removed the animal crashes.

$$N_{pred-no-an} = AADT^{1.01} L^{0.85} e^{(-10.07+0.58 \times RHR)}$$ **Equation 5.3**

Where:

AADT = Average Annual Daily Traffic

L = Length (miles)

RHR = Roadside Hazard Rating

The animal crash variable was removed due to the difficulty of determining where these crashes would occur. However, to better predict animal or total crashes, either county-specific CPMs that take into account the number of animal crashes reported or a county-specific calibration factor would still need to be applied that would be similar to that made in Kansas calibration process.

5.4 Intersection Analysis

As detailed in the data collection section, a single statewide calibration factor was calculated that combined 3-leg and 4-leg stop-controlled intersection. The method for calculating the intersection calibration factor was identical to the method for calculating the segment calibration factor. Since animal crashes represented a small percentage of crashes reported as intersection or intersection-related, no additional studies were performed beyond development of this single statewide calibration factor. Shown in Table 5.13 is the overall breakdown of intersection related crashes and the resulting calibration factor of 0.21.

TABLE 5.13
Overall Breakdown of Intersection Related Crashes in Kansas

Intersection Type	Predicted	Actual	Calibration Factor	Number of Intersections
All	176.5	37	0.21	278
3-Leg	43.0	12	0.28	99
4-Leg	133.6	25	0.19	179

Since intersection crashes represented a small percentage of the crashes for rural two-lane roads, it was determined by the research team that a large effort was needed to improve data collection at the time of this study and would improve the model accuracy enough to be warranted.

5.5 Summary

Research efforts were focused on roadway segment crashes because they accounted for the vast majority of crashes for this facility type.

- While many different calibration methods and model forms were developed for roadway segments for this research study, only three were found to be promising enough and carried forward for validation:
    - Single statewide calibration factor;
    - Calibration function addressing animal crashes; and
    - Kansas specific CPM looking at non-animal crashes.
- A calibration factor of 0.21 was calculated for both 3-leg and 4-leg stop controlled intersections and no other intersection types were available or considered for calibration for this study.

Chapter 6: Validation

To evaluate the accuracy of the preferred calibration methods and Kansas specific CPM, a validation step was performed. The goal of this validation step was to not only compare the different methodologies, but also to evaluate the overall accuracy of CPMs for use on rural two-lane highways in Kansas. As noted previously, the validation sets developed for this study were most consistent with the intended model application and similar methods were not found in any previous studies.

6.1 Empirical Bayes Method

The HSM promotes the use of Empirical Bayes (EB) method to improve the accuracy of crash prediction by combining the results of the predictive model with observed crash data. This method can help to address the random nature of crashes and the negative effects of crash spikes on prediction. This phenomenon is called regression-to-the-mean in statistics. The EB method can be used to predict the crashes on a highway that is not being improved. If the highway is being improved, the scope of the improvements needs to be considered. The EB method should not be used on projects where new alignments are being considered, the number of through lanes are changing, or that have intersections planned for major reconfiguration. If a project varies in scope, it is acceptable to only apply the EB method to relatively unaffected segments.

Since this method is universally recommended to improve model accuracy, all validation analysis was performed utilizing the EB method. Specifically, the project level EB procedure provided in the HSM was utilized in the validation process. To allow for this, Data Groups 2 and 3 involved additional screening to remove any sections where a major realignment took place that would no longer make it an appropriate application of the EB method. Additional information can be found on the improvement gained by implementing the EB method in the other publications developed as a result of this research study.

6.2 Segment Validation

Validation of the statewide calibration factor and the calibration function based on animal crashes were performed on Data Group 2. Since Data Group 2 was used in the development of the Kansas specific CPM, Data Group 3 was then used for validation of that model. Unfortunately this did not provide a pure contrast of the different methodologies, but the results are provided below for general comparison. The evaluation metrics described above were used to evaluate the three recommended methodologies shown in Table 6.1.

TABLE 6.1
Evaluation Methodologies

Method	Data Group	r	MPB	MAD	p-value
Statewide Calibration Factor	2	0.973	-0.165	1.194	0.476
Calibration Function	2	0.980	-0.631	1.110	0.872
Kansas CPM	3	0.886	-0.450	0.730	0.447

The results shown in Table 6.1 indicate that the Kansas-specific CPM does not provide a greater level of accuracy in the final crash prediction model as compared to the calibration of the HSM models. Since the Kansas specific CPM only considers the RHR as an input and does not provide greater accuracy, it's not recommended for implementation and was not analyzed further in this study.

Of the two remaining calibration methods, the statewide calibration factor performed better on aggregate measures that allow high and low predictions to cancel out. Ultimately, the total percent difference between predicted and actual crashes of the validation data using the statewide calibration factor was only -1.0 percent. This compares to -6.6 percent using the calibration function. However, in metrics that investigate consistent model prediction and do not allow high and low predictions to cancel out the calibration function performs better. When investigating the average absolute percent difference between the predicted and actual crashes for the validation data the calibration function averaged 16.6 percent accuracy. This compared to 20.4 percent using the statewide calibration factor.

6.3 Intersection Validation

To perform the validation of intersections, all of the 3-leg and 4-leg stop controlled intersections found within the limits of Data Group 2 were analyzed. The statewide calibration factor and EB method were used in the analysis and the results are shown in Table 6.2.

TABLE 6.2
Results of Validation for 3 and 4-Leg Intersections

Project Number	Observed	Predicted
1	4	5.57
2	12	8.65
3	3	4.74
4	14	7.53
5	3	5.78
6	0	1.35
8	1	0.88
9	9	12.58
10	2	1.97
Total	**48**	**49.05**

Since only the one method was being analyzed and the crash data volumes were low, advanced statistic measures were not needed to evaluate these data. However, similar to the segment results the single statewide calibration factor did well at predicting the aggregate number of crashes. Additionally, the accuracy per project holds up fairly well given the difficulty in predicting something as random as crash occurrence.

6.4 Summary

- The statewide calibration factor and calibration function based on animal crashes were recommended for implementation on prediction of crashes on rural two-lane highway segments.
- The statewide calibration factor for combined 3-leg and 4-leg stop controlled intersections is also recommended for implementation.

Chapter 7: Conclusion

The use of CPMs for rural two-lane highways in Kansas were analyzed and validated with this study. Ultimately, it is recommended that these models were able generate enough accuracy in their crash predictions to be a useful tool in aiding project-level decisions. The HSM CPM for segments should be utilized with the Kansas-specific calibration factor of 1.48 or the calibration function shown in Equation 7.1 to account for animal crashes.

$$C_{county} = 1.13 \times ACR_{county} + 0.635$$ **Equation 7.1**

Kansas-specific CPMs were also developed and evaluated in this study. Their use is not recommended due the limited improvement in prediction accuracy and limitation in the number of elements that can be analyzed. This was likely caused by the relative homogeneity of data of for Kansas rural two-lane highways.

The statewide calibration factor of 0.21 should be used for both 3- and 4-leg stop-controlled intersections. No calibration factor was developed for signalized rural intersections due to their limited application on Kansas highways.

7.1 Future Research Avenues Uncovered from the CPM Research

In the process of filling some gaps in the existing research of the HSM CPM for rural two-lane roadways, this research study also exposed some potentially new areas that could be addressed by future research studies.

7.1.1 National Research

The most significant finding of this research relative to national application of the HSM CPM, is the fundamental definition of what sections qualify as rural. Those looking to apply the HSM CPM in the future could benefit from determination of the impact of this finding on previous studies and/or from confirmation of this discrepancy in other jurisdictions.

Similarly, future research could also benefit from identifying how highways through small communities should be modeled. Specifically, it should be determined if modifications can be made to the rural two-lane model so these road can be analyzed, or do these roads perform in a way that is more consistent with the urban/suburban arterial model. It is also unknown if the higher crash rates along these relatively short sections of highway can skew analysis that groups them with rural sections that have no portion through a community.

Since the alternative method for calibrating the HSM CPM improved the accuracy of the CPM for Kansas, it should be considered for use by other jurisdictions. This method could prove especially helpful for jurisdictions that have a significant cause of crashes that is not considered by the HSM CPM and is not related to the roadway geometry or traffic control.

7.1.2 Kansas Research

To assist KDOT with future research in crash prediction on rural two-lane highways, the organization should consider adding a field to the CANSYS database to determine if a section of highway goes through a community of any size.

The calibration values developed in this research study are only good for three years after the last year of data analyzed which was 2007. Therefore, a new calibration value should be developed when the 2008 through 2010 crash data are available for Kansas. Since the IHSDM input files were prepared, the recalibration should be much simpler.

Due to the promising results of the prediction results from the HSM rural two-lane models it is recommended that the other CPMs available in the HSM be calibrated and validated for Kansas highways.

References

American Association of State Highway and Transportation Officials. 2010. *Highway Safety Manual*. AASHTO, Washington, D.C.

American Association of State Highway and Transportation Officials. 2011. *A Policy on Geometric Design of Highway and Streets*. AASHTO, Washington, D.C.

Knapp, K.K. 2005. "Crash Reduction Factors for Deer-Vehicle Crash Countermeasures." *Journal of the Transportation Research Board*, No. 1908, Transportation Research Board of the National Academies, Washington, D.C., pp.172-179.

Meyer, E., *Assessing the Effectiveness of Deer Warning Signs*. Report No. K-TRAN: KU-03-6. Kansas Department of Transportation, Topeka, KS, 2006.

Miaou, S-P. The Relationship between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial Regressions. In *Accident Analysis and Prevention,* Vol. 26, No.4, 1994, pp. 417-482.

Najjar, Y. and S. Mandavilli, *Data Mining the Kansas Traffic-Crash Database*. Report No. K-TRAN: KSU-05-6. Kansas Department of Transportation, Topeka, KS, 2009.

Rhys, M.J., D.E. Karkle, A. Vijayakumar, R. Makarla, and E. Russell, *Promoting Centerline Rumble Strips to Increase Rural, Two-Lane Highway Safety*. Report No. K-TRAN: KSU-08-3. Kansas Department of Transporation, Topeka, KS, 2010.

Schrock, S., C.B. Young, and D. Chellamani, *Review and Analysis of the Kansas Department of Transportation Maintenance Quality Assurance Program*. Report No. K-TRAN: KS-09-4. Kansas Department of Transportation, Topeka, KS, 2009.

K-TRAN

KANSAS TRANSPORTATION RESEARCH AND NEW-DEVELOPMENT PROGRAM

AD ASTRA PER ASPERA

Kansas

Department of Transportation