

ESTIMATION OF SAFETY EFFECTIVENESS OF COMPOSITE SHOULDERS
ON RURAL TWO-LANE HIGHWAYS

By

Huanghui Zeng

Submitted to the graduate degree program in Civil, Environmental and Architectural Engineering
and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements
for the degree of Master of Science.

Chairperson Dr. Steven D. Schrock

Dr. Yong Bai

Dr. Thomas E. Mulinazzi

Date Defended: February 21, 2012

The Thesis Committee for Huanghui Zeng

certifies that this is the approved version of the following thesis:

ESTIMATION OF SAFETY EFFECTIVENESS OF COMPOSITE SHOULDERS
ON RURAL TWO-LANE HIGHWAYS

Chairperson Dr. Steven D. Schrock

Date approved: February 21, 2012

ABSTRACT

ESTIMATION OF SAFETY EFFECTIVENESS OF COMPOSITE SHOULDERS ON RURAL TWO-LANE HIGHWAYS

Huanghui Zeng

The University of Kansas

Advisor: Dr. Steven Schrock

February 2012

A paved shoulder has been regarded as an effective safety improvement to reduce crashes. There is belief that there is a diminishing safety benefit for each additional increment of paved shoulder width. Thus there may be opportunities for greater system-wide safety benefits from paving longer roadway segments with a composite shoulder than paving shorter roadway segments with a full-width paved shoulder. The objective of this study was to determine the safety benefits of composite shoulders - such as a small paved shoulder combined with turf outside of that. This approach was part of the Kansas Department of Transportation's effort to find "Practical Improvements" to maximize benefits relative to the input costs required.

Among the 8,300 miles of rural two-lane highways in Kansas, approximately 25 percent of them are equipped with composite shoulders consisting of three ft of pavement with the remainder turf. Their safety effectiveness was studied using the Empirical Bayes (EB) approach and the cross-sectional approach. Three developed Safety Performance Functions (SPFs) were used to create Kansas-specific Crash Modification Factors (CMFs) for composite shoulders compared with segments with no or unpaved shoulders. It was found that upgrading narrow unpaved shoulders to composite shoulders can reduce shoulder related crashes by up to 61 percent and fatal and injury crashes by 31 percent. Based on these results, 20-year projections were developed projecting the safety benefit that can be achieved through implementing these safety improvements.

ACKNOWLEDGEMENT

First of all, I would like to give my best appreciation to my parents. I would like to dedicate this thesis to you for being my best friends and supporters in my life. Mom and Dad, you always believe in me in my every decision, enjoy the happiness with me for my every accomplishment, and encourage me whenever I am down. Thank you and I love you.

My appreciation goes to my advisor, Dr. Steven Schrock. He brought me to KU, worked closely with me as a great boss and a patient advisor all the time. I especially appreciate that he allowed me to explore my own interests both in courses and research while also provided the necessary guidance to complete my master's study. This thesis cannot be completed without his time, effort, and resources.

The members of my committee, Dr. Thomas Mulinazzi and Dr. Yong Bai, deserve thanks from me. Thanks for being patient and helpful, and giving me valuable advice and comments on my work.

People in the Transportation Research group have worked closely with me. They are also my best friends and supporters. Special thanks go to Robert Rescot for helping me settle down at Lawrence and always being ready to help and to Eric Fitzsimmons for being a great reviewer and giving me useful comments in writing this thesis.

The research was funded by the Kansas Department of Transportation. I wish to thank Jonathan Marburger, Howard Lubliner, Ruby Bradley, Tina Cramer, Matthew Soper, and Kyle Gonterwitz of KDOT for providing crash and roadway data for this study. Especially, I appreciate Howard and Jonathan's comments on this research.

Table of Contents

ABSTRACT.....	III
ACKNOWLEDGEMENT	IV
CHAPTER 1 INTRODUCTION	1
1.1 Background.....	1
1.2 Problem Statement.....	2
1.3 Research Objectives.....	6
1.4 Contribution to Highway Safety.....	6
1.5 Organization of Thesis.....	7
CHAPTER 2 LITERATURE REVIEW	9
2.1 Shoulder Studies	9
2.1.1 Research Before 2000.....	9
2.1.2 Recent Research.....	12
2.1.3 Findings in the Highway Safety Manual	13
2.1.4 Detrimental Tendencies of Shoulders	15
2.2 Methodologies to Develop CMFs.....	15
2.2.1 Before-After with Comparison Group.....	16
2.2.2 The Empirical Bayes Approach.....	17
2.2.3 Full Bayes Method.....	19
2.2.4 The Cross-Sectional Approach	19
2.2.5 Case-Control Studies	20
2.3 Development of Crash Statistic Models	21
2.3.1 The Beginning of Crash Statistic Models	21

2.3.2 Poisson Regression and Negative Binominal Regression Models	22
CHAPTER 3 METHODOLOGY	27
3.1 The Empirical Bayes Approach	27
3.2 The Cross-Sectional Approach	30
CHAPTER 4 DATA DESCRIPTION	34
4.1 General Consideration	34
4.2 Data for the Empirical Bayes Approach	37
4.3 Data for the Cross-Sectional Approach	40
CHAPTER 5 DATA ANALYSIS	43
5.1 Safety Performance Functions	43
5.1.1 Variable Selection	43
5.1.2 Test Statistics	44
5.1.3 Final Models	45
5.2 The Empirical Bayes Approach Results	49
5.3 The Cross-Sectional Approach Results	52
5.4 Summary	57
CHAPTER 6 COMPOSITE SHOULDER SAFETY ECONOMIC BENEFIT ANALYSIS.....	60
6.1 Crash Avoided Prediction.....	60
6.2 CMFs Used to Calculate Crash Cost Avoided.....	61
CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS	63
7.1 Discussion and Conclusions	63
7.2 Contributions and Recommendations	64
7.3 Limitations and Future Studies	65

REFERENCES	66
APPENDIX A.....	69
APPENDIX B.....	70
APPENDIX C.....	71

List of Tables

TABLE 1 Effect of Shoulder Widening for Related Crash Types on Rural Two-Lane Highways	10
TABLE 2 Effect of Lane and Shoulder Widening for Related Crashes on Rural Two-Lane Highways	11
TABLE 3 CMFs for Average Shoulder Width (ft).....	13
TABLE 4 CMF for Related Crashes Based on Shoulder Width (CMF_{wra})	14
TABLE 5 CMF for Related Crashes Based on Shoulder Types (CMF_{tra})	14
TABLE 6 Tabulation for Simple Case-Control Analysis	20
TABLE 7 Vehicle Crash Database Variables and Description	34
TABLE 8 Winter Crashes Distribution for Total Crashes.....	36
TABLE 9 Winter Crashes Distribution for Related Crashes	36
TABLE 10 Descriptive Statistics for Segments in the Reference and Treatment Groups	38
TABLE 11 Crash Information on Class D Roads by Year	40
TABLE 12 Descriptive Statistics for Segments in the Cross-Sectional Approach	41
TABLE 13 Safety Performance Functions for Composite Shoulder.....	45
TABLE 14 CMFs for Composite Shoulder from Initial EB Methods and Naïve Method	50
TABLE 15 CMFs for Composite Shoulder from Combined EB Method	51
TABLE 16 The Cross-Sectional Approach Models Summary.....	53
TABLE 17 CMFs (Confidence Interval) for Composite Shoulders and Wide Unpaved Shoulders from the Cross-Sectional Study	54
TABLE 18 Regressed Results for θ	55
TABLE 19 Summary of Estimated Crash Cost Avoided from Composite	61

List of Figures

FIGURE 1 Wide paved shoulder on US Highway 59, Maxwell, Kansas.....	3
FIGURE 2 Unpaved shoulder on Co Rd 1029, Lecompton, Kansas.....	4
FIGURE 3 Composite shoulders on US Highway 24, Williamstown, Kansas.	5
FIGURE 4 Diagram of thesis research performed.....	8
FIGURE 5 Illustration of regression-to-the-mean and the EB estimation.....	18
FIGURE 6 Scatter plot of crashes and residuals.....	48
FIGURE 7 Annual crashes vs. AADT for total crashes.	56
FIGURE 8 Annual crashes vs. AADT for FI crashes.....	56
FIGURE 9 Annual crashes vs. AADT for related crashes.....	57
FIGURE 10 Cumulative crashes avoided from upgrading 10 miles of shoulder per year for 10 years.....	60

STAY HUNGRY. STAY FOOLISH.

-- *Steve Jobs*

CHAPTER 1 INTRODUCTION

1.1 Background

Federal highway agencies have many options for how to spend safety funds each year. Safety is one pillar in many states' comprehensive highway funding. This pillar is executed through a multitude of maintenance activities and strategic investments in updated roadway design. Given the fiscal constraints and a desire to continually strive to make the roadways safer, it would, therefore, be desirable for highway agencies to seek to maximize safety benefits relative to the input costs required. Quantitative safety analysis allows transportation engineers, designers, and planners to better understand the trade-offs of safety versus cost.

To address the safety effectiveness quantifiably, research over the last few decades have led to develop Crash Prediction Models (CPMs). CPMs can estimate, and ideally predict the expected safety performance of a highway based on its traffic, geometric, and traffic control features. With an increase in computer processing power and efforts at the federal level, the safety-based decision making transportation industry has gained momentum. The largest step toward that goal was the publication of the Highway Safety Manual (HSM) in 2010, by the American Association of State Highway and Transportation Officials (AASHTO). The primary goal of the HSM is to provide a science-based technical approach to quantitative safety analysis.

The safety effectiveness of geometric and traffic control features can be indicated by Crash Modification Factors (CMFs). A CMF is a multiplicative factor used to compute the expected number of crashes after implementing a given highway safety countermeasure at a specific location. A CMF greater than 1.0 indicates an expected increase in crashes, while a CMF less than 1.0 indicates an expected reduction in crashes. The HSM CPM equation has taken

different CMFs into account to predict the crash frequency for the specific conditions of the modeled highway as follows (1):

$$N_{predicted} = N_{spfx} \times (CMF_{1x} \times CMF_{2x} \times \dots \times CMF_{yx}) \times C_x \quad \text{Equation 1}$$

Where:

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

N_{spfx} = predicted crashes from the Safety Performance Function (SPF);

CMF_{yx} = Crash Modification Factors; and

C_x = calibration factor to adjust for local conditions.

It's speculated that CMFs will differ from each other among individual highway safety countermeasures and various jurisdictions. As transportation engineers, planners, and designers become increasingly interested in quantitative safety evaluations, CMFs are becoming an essential safety indicator for highway safety countermeasures. Although CMF research has progressed in recent years, there are safety strategies of interest to practitioners that have not been investigated thoroughly by previous research.

1.2 Problem Statement

One of the elements which remain to be investigated is to determine the safety benefits relative to incremental increases in paved shoulder widths. While a full-width paved shoulder might be the most desirable, there is belief that there is a diminishing safety benefit for each additional increment of paved shoulder width. Thus, there may be opportunities for greater system-wide safety benefits from paving longer roadway segments with a narrower shoulder than paving shorter roadway segments with a wider shoulder. Determining the benefits of various shoulder

improvement approaches fits within Kansas Department of Transportation’s (KDOT) “Practical Improvements” approach to maximize benefits relative to the input costs required.

Paved shoulder options come in many varieties. While the ideal shoulder type is pavement (asphalt or concrete), other options also exist. These include turf, gravel, and various combinations of the aforementioned options. Each shoulder surface type has its own unique possibilities and limitations and would possibly have a unique incremental shoulder width safety benefit compared to the other options. Figures 1 to 3 show various shoulder types that are common on rural two-lane highways in Kansas.

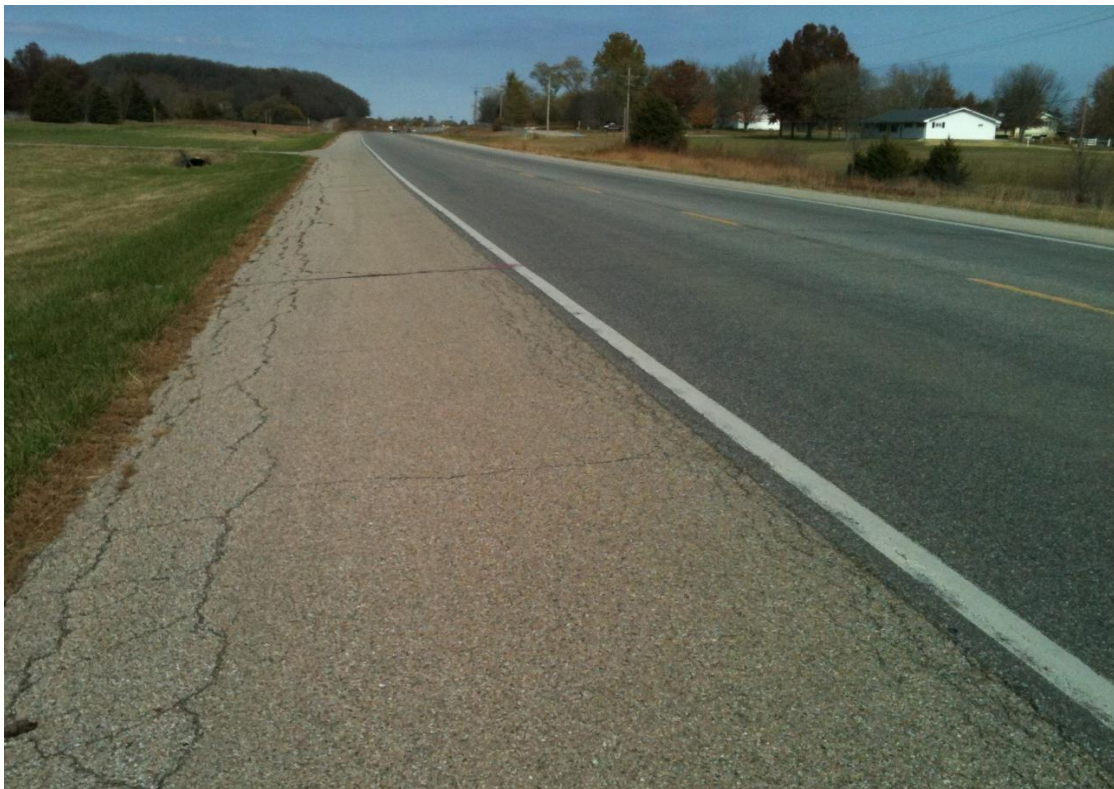


FIGURE 1 Wide paved shoulder on US Highway 59, Maxwell, Kansas.

A wide paved shoulder (six to 12 ft) can be paved with asphalt, concrete, or aggregated materials. In Kansas, this type of shoulders is usually installed on highways with high traffic volume, which account for approximately 16 percent of rural two-lane highways.



FIGURE 2 Unpaved shoulder on Co Rd 1029, Lecompton, Kansas.

The most common type of unpaved shoulders is turf shoulders on rural two-lane highways in Kansas. More than 80 percent of turf shoulders are narrow shoulders less than five ft. Turf shoulders are installed on low volume highways, which account for nearly half of rural two-lane highways.



FIGURE 3 Composite shoulders on US Highway 24, Williamstown, Kansas.

Composite shoulders are shoulders with a small paved shoulder combined with an unpaved section outside. The paved section could be paved by asphalt, concrete, or other stabilized materials, while the unpaved section could be turf, gravel, or other unstabilized materials.

Kansas has 8,300 miles of rural two-lane highway. It was found that more than one-third include composite shoulders, and approximately 70 percent of the composite shoulders consist of the first three ft of pavement with the remainder turf.

Supporters of composite shoulders often state that compared to fully-paved shoulders, besides costing less, composite shoulders are easier to upgrade if it is needed. Since limited

research exists that focuses specifically on the safety effectiveness of composite shoulders, it is necessary to develop CMFs for composite shoulders in Kansas.

1.3 Research Objectives

The overall goal of this research is to quantify the safety effectiveness of composite shoulders to reduce vehicle crashes. Part of this effort is to apply the HSM findings and recommended methodologies to the local jurisdictions. Quantitative safety analysis for composite shoulders aims at helping transportation engineers, designers, and planners to better understand the trade-offs of safety versus cost.

The specific research objectives of this research were:

- Develop a detailed procedure of how to apply the Empirical Bayes (EB) approach to estimate the safety effectiveness of a specific safety treatment.
- Create the Kansas-specific CMFs and Safety Performance Functions (SPFs) for upgrading non- or unpaved shoulders to composite shoulders.
- Conduct safety economic analysis using CMFs to help engineers and planners understand the trade-offs of safety and cost.

This thesis only researched composite shoulders that consist of the first three ft paved and the remainder unpaved. Unless being specifically stated, the composite shoulder concept in next sections should be referred to this typical type of composite shoulders.

1.4 Contribution to Highway Safety

This research will include data driven insight into the safety benefits (and limitations) of composite shoulders on rural two-lane highways. It is expected that the result will be valuable as KDOT considers various options and their benefit-cost ratios for the investment of state

maintenance funds. It is expected that this project will aid decision makers and engineers in continuing to build safe and efficient highway systems in Kansas. In addition, this research will use methodologies recommended in the HSM. It will support the study of the HSM at the local level.

1.5 Organization of Thesis

This thesis is organized into seven chapters. Figure 4 demonstrates the execution of the research, as well as the thesis's organization.

Chapter 1 is an introduction into the background of CMFs, research needed on composite shoulders in Kansas, and proposed objectives of this thesis. Chapter 2 provides a literature review to identify the primary research related with shoulders, to introduce methodologies for the development of CMFs, and to describe the development of an important crash model. Chapter 3 is a description of data used in this thesis. It introduces data collection and a modification process, as well as a summary of collected data. Two methodologies were described in-depth in Chapter 4. Chapter 5 covers the data analysis that led to the creation of CMFs for composite shoulders from the EB approach and the cross-sectional approach. Chapter 6 shows an example on how CMFs can be used in a safety economic benefit analysis. The last chapter summarizes this thesis. This chapter concludes the findings from this thesis and addresses the limitations and recommendations for future research.

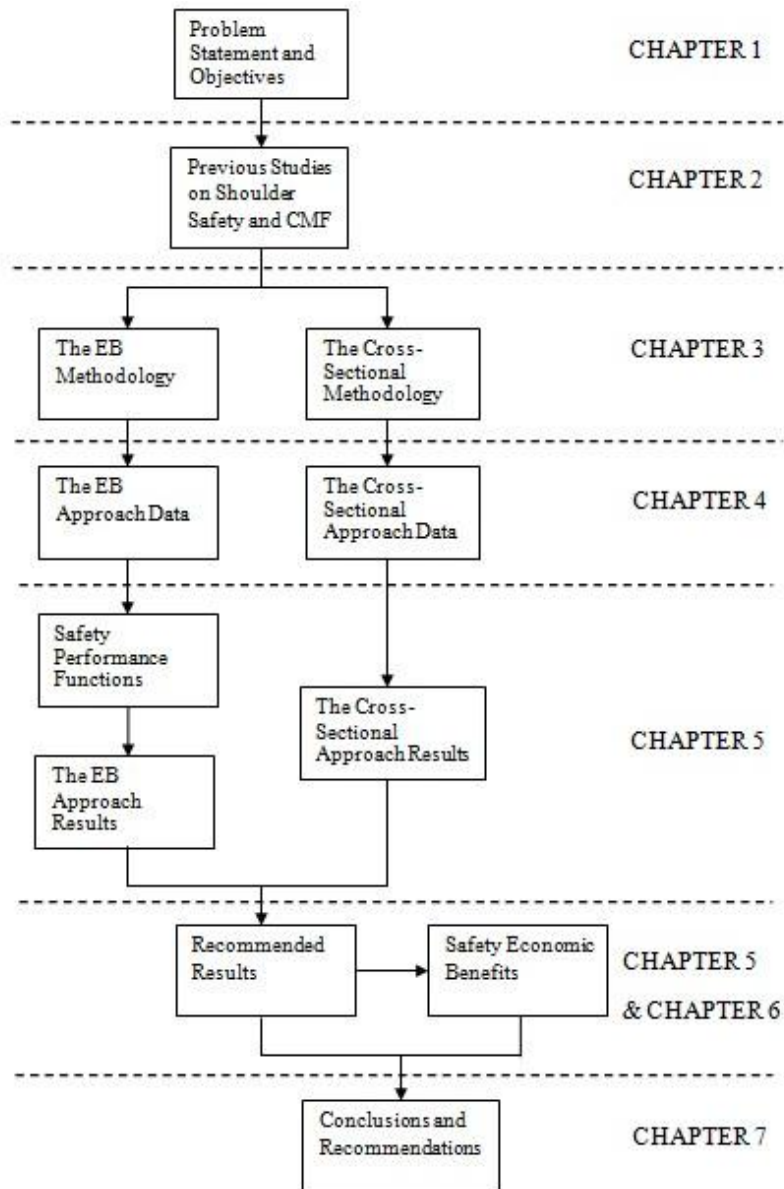


FIGURE 4 Diagram of thesis research performed.

CHAPTER 2 LITERATURE REVIEW

2.1 Shoulder Studies

The safety benefits gained from widening roadway shoulders have been studied for decades. According AASHTO's A Policy on Geometric Design of Highway and Street, shoulders located adjacent to travel lanes accomplish several functions including emergency stop and pull off, a recovery area for driver error, and pavement edge support (2). This section is to review literature on safety effectiveness studies for shoulders.

2.11 Research Before 2000

Zegeer, et al. conducted a comparative study on the effect of lane and shoulder widths on rural two-lane roads using data from Ohio and Kentucky (3). The authors identified that roadway shoulders could benefit crash types as run-off-road (ROR), head on, and opposite-direction sideswipe. Geometric features, crashes, and traffic volume data were obtained for more than 15,000 miles of roads. It was found that a 21 percent reduction in related crashes would be expected on roadways with 9-ft wide shoulders when compared with roadways absent of shoulders. A diminishing safety benefit for each additional increment of paved shoulder width was found when shoulders were wider than three ft. Based on Zegeer's results, for roadways with lane widths greater than 10 feet, it was found to be economically beneficial to widen the shoulders if there were at least five ROR and/or opposite direction crashes in one year. For roads without shoulders, the optimal shoulder width to install was found to be five feet.

Another study by Zegeer, et al. applied statistical models to investigate the relationship between crashes and roadway geometry features. Run-off-road, head-on, and sideswipe (same direction and opposite direction) crashes were considered as shoulder related (4). The research data were collected on 4,950 miles of two-lane roadways in seven states, including detailed

traffic, crash, roadway, and roadside data. This study applied statistical testing along with a crash prediction model to determine the expected crash reductions related to geometric improvements such as lane width and shoulder width. The authors also conducted a before-and-after study on control sites for comparison. The effects of shoulder widening on related crashes was determined for paved and unpaved shoulders and were shown in Tables 1. Table 2 shows the combined effects of lane widening and shoulder widening.

TABLE 1 Effect of Shoulder Widening for Related Crash Types on Rural Two-Lane Highways (4)

Amount of Shoulder Widening (ft.) per Side	Percent Reduction in Related Crash Types	
	Paved	Unpaved
2	16	13
4	29	25
6	40	35
8	49	43

TABLE 2 Effect of Lane and Shoulder Widening for Related Crashes on Rural Two-Lane Highways (4)

Amount of Lane Widening (ft)	Existing Shoulder Condition (Before Period) Shoulder Width Surface Type		Percent Related Crashes Reduced							
			Shoulder Condition in After Period							
			2 ft.		4 ft.		6 ft.		8 ft.	
		Paved	Unpaved	Paved	Unpaved	Paved	Unpaved	Paved	Unpaved	
3	0	N/A	43	41	52	49	59	56	65	62
	2	Paved	32	--	43	--	52	--	59	--
	2	Unpaved	34	32	44	41	53	49	60	56
	4	Paved	--	--	32	--	43	--	52	--
	4	Unpaved	--	--	36	32	46	41	54	49
	6	Paved	--	--	--	--	32	--	43	--
	6	Unpaved	--	--	--	--	37	32	47	41
	8	Paved	--	--	--	--	--	--	32	--
	8	Unpaved	--	--	--	--	--	--	39	32
2	0	N/A	35	33	45	42	53	50	61	56
	2	Paved	23	--	35	--	45	--	53	--
	2	Unpaved	25	23	37	33	46	42	55	50
	4	Paved	--	--	23	--	35	--	45	--
	4	Unpaved	--	--	27	23	38	33	48	42
	6	Paved	--	--	--	--	23	--	35	--
	6	Unpaved	--	--	--	--	29	23	40	33
	8	Paved	--	--	--	--	--	--	23	--
	8	Unpaved	--	--	--	--	--	--	31	23
1	0	N/A	26	24	37	34	47	43	55	50
	2	Paved	12	--	26	--	37	--	47	--
	2	Unpaved	14	12	28	24	39	34	48	43
	4	Paved	--	--	12	--	26	--	37	--
	4	Unpaved	--	--	17	12	30	24	41	34
	6	Paved	--	--	--	--	12	--	26	--
	6	Unpaved	--	--	--	--	19	12	31	24
	8	Paved	--	--	--	--	--	--	12	--
	8	Unpaved	--	--	--	--	--	--	21	12

Compared with the first study, the second study obtained a higher reduction rate in related crashes (up to 50 percent). These two studies were review by the expert panel for the HSM, and were adopted for use in the HSM since they provided valid results with comprehensive study.

2.1.2 Recent Research

Örnek and Drakopoulos investigated the effectiveness of shoulders on reducing run-off-road crashes using the PRÉCIS database for the Wisconsin state trunk highway system (5). Data were collected on rural two-lane highways between 1998 and 2002. The authors found that providing additional unpaved shoulders on rural two-lane highways with three ft of paved shoulders would reduce run-off-road crash rates. However, the shoulder effectiveness would decrease for additional unpaved shoulder widths in excess of seven ft. It was also found that additional shoulder width in excess of 10 ft provided no additional safety benefits.

Another study by Gross and Jovanis used case control and cohort methods to estimate the safety benefits of shoulder widening (6). A total of 26,000 rural two-lane undivided highway segments in Pennsylvania between 1997 and 2001 were investigated. The results of both methodologies indicated that overall crashes decreased as shoulder width increased. In the case control approach it was found that widening shoulders from 2 ft to 8 ft provided a CMF of 0.80. The cohort approach provided a CMF of 0.86 for the same improvement. A confidence interval (95%) was created for each case to determine significance.

Gross, et al. evaluated various lane-shoulder width configurations for fixed paved widths (sum of paved lane and shoulder width) as a countermeasure for roadway departure crashes (7). This study applied a matched case-control analysis to five years of geometric, traffic volume, and crash data in Pennsylvania and Washington. Crash reductions were estimated for wider paved widths, wider lanes, and wider shoulders. Specifically, a 12 ft lane provided the optimal safety benefit for 26 to 32 ft total paved widths. While an 11 ft lane provided the optimal safety benefit for a 34 ft total paved width. Both provided the optimal safety benefits for a 36 ft total paved width.

Stamatiadis, et al. conducted a study to quantify the safety effectiveness of shoulder width and median width (8). The research team reviewed the past literature, recommended CMF values for the HSM, and CMFs from NCHRP Project 15-27: Safety Impacts of Design Element Trade-offs. The authors adjusted results from previous research and developed a list of CMFs for all crashes and the average shoulder width as shown in Table 3.

TABLE 3 CMFs for Average Shoulder Width (ft)^A (8)

Category	Average Shoulder Width (ft) ^B						
	0	3	4	5	6	7	8
Undivided	1.22	1.00	0.94	0.87	0.82	0.76	0.71
Divided	1.17	1.00	0.95	0.90	0.85	0.81	0.77

^A The CMFs are for all crashes and all severities.

^B The average shoulder width for undivided is the average of the right shoulders; for divided, it is the average of left and right shoulder in the same direction.

Gross and Donnell studied case-control and cross-sectional methods for estimating CMFs for fixed roadway lighting and the configurations of lane and shoulder widths (9). The authors hoped to provide an alternative to a before-after methodology, when it is determined to be impractical due to data restrictions. Data from previous study in Pennsylvania were used to conduct this shoulder study. Both shoulder width and additional shoulder width (the difference between the total shoulder width and paved shoulder width) were treated as test variables. It was found that providing at least four ft of unpaved shoulder beyond existing paved shoulders produced a significant incremental safety effect at the 0.05 level of confidence.

2.1.3 Findings in the Highway Safety Manual

The HSM provided CMFs for various widths and types of shoulders. The CMF for a specific shoulder is comprised of the separate CMF values for shoulder width (CMF_{wra}) and shoulder type (CMF_{tra}) using Equation 2. Various equations were used to calculate CMF_{wra} based on

different traffic volumes. These equations were based on several previous researches (3, 4, 15).

The equations necessary to calculate these CMFs are displayed in Tables 4 and 5.

TABLE 4 CMFs for Related Crashes Based on Shoulder Width (CMF_{wra}) (1)

Shoulder Width (ft)	AADT (veh/day)		
	<400	400 to 2000	>2000
0	1.10	$1.10 + 2.5 \times 10^{-4}(AADT - 400)$	1.50
2	1.07	$1.07 + 1.43 \times 10^{-4}(AADT - 400)$	1.30
4	1.02	$1.02 + 8.125 \times 10^{-5}(AADT - 400)$	1.15
6	1.00	1.00	1.00
8 or more	0.98	$0.98 + 6.875 \times 10^{-5}(AADT - 400)$	0.87

The CMF_{wra} in Table 4 is calculated by dividing the CMF for the after-improvement condition by the CMF for the before condition.

TABLE 5 CMFs for Related Crashes Based on Shoulder Types (CMF_{tra}) (1)

Shoulder Type	Shoulder Width (ft)						
	0	1	2	3	4	6	8
Paved	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Gravel	1.00	1.00	1.01	1.01	1.01	1.02	1.02
Composite ^A	1.00	1.01	1.02	1.02	1.03	1.04	1.06
Turf	1.00	1.01	1.03	1.04	1.05	1.08	1.11

^A The values for composite shoulders represent a 50/50 paved/turf shoulder width

The CMF_{tra} in Table 5 is calculated by dividing the CMF for the after-improvement condition by the CMF for the before condition.

$$CMF_{2r} = (CMF_{wra} \times CMF_{tra} - 1.0) \times p_{ra} + 1.0 \quad \text{Equation 2}$$

Where:

CMF_{2r} = Crash Modification Factor for the effect of shoulder width and type on total crashes;

CMF_{wra} = Crash Modification Factor for related crashes based on shoulder width;

CMF_{tra} = Crash Modification Factor for related crashes based on shoulder type; and

p_{ra} = proportion of total crashes constituted by related crashes.

This study provided CMFs for composite shoulders. These CMFs only considered a situation in which half of the shoulder width was paved and the remainder was turf. However, most composite shoulders in Kansas have wider unpaved sections than paved sections. In addition, the values of CMFs were determined simply by averaging those CMFs for paved shoulders and turf shoulders, which may be not true in reality.

2.1.4 Detrimental Tendencies of Shoulders

Research studies were also found to disagree in the safety benefits of wider shoulders. Agent and Pigman collected Kentucky crash data between 1985 through 1987 to study the problem of crashes involving vehicles on shoulders of limited access highways (10). The authors found that 11 percent of fatal freeway crashes were related to drivers that stopped on the shoulder and abandoned their vehicles.

Hauer had conducted a literature review investigating CMFs and SPFs on shoulder widths and reanalyzed some data sets (11). Hauer found that a wider shoulder allows for the safe recovery of stray vehicles. However, negative safety effects may include: inviting some voluntary shoulder stops, higher travel speeds, the possibility of steeper roadside slopes, and shoulders used for travel.

2.2 Methodologies to Develop CMFs

Gross et al. developed a federal guidance on the development of CMFs (12). The authors discussed several specific study designs and addressed their associated strengths and weakness. Before-after methods, including comparison group studies and the EB approach, were usually

preferred. However, there were also situations that called for an alternative approach because before-after methods are not practical. These alternative methods include: full bayes, cross-sectional, and case-control methods. The following sections will provide an overview for these methods discussed by Gross et al. More detail descriptions will be given in subsequent chapters of the cross-sectional and the EB approaches. All of the information in the following is summarized from the Gross et al.'s study.

2.2.1 Before-After with Comparison Group

This method can be applied when the treatment is sufficiently similar among treatment sites, before and after data are available for both treated and untreated sites, and untreated sites can be used to account for non-treatment related crash trends.

An untreated comparison group of sites similar to the treated sites are used in a before-after with comparison group study to account for changes in crashes unrelated to the treatment such as time and traffic volume trends. A ratio is calculated by dividing the observed crash frequency of the comparison group in the after period by the crash frequency in the before period. In order to estimate the expected number of crashes in the treatment group had a treatment been not applied, the observed crash frequency in the before period of a treatment group is multiplied by the determined comparison ratio. This number of crashes is then compared to the observed number of crashes of the treatment group in the after period. Treatment effectiveness is determined by the comparison.

A key step for a comparison group study is to select a suitable comparison group. Hauer developed a time series test and used a sequence of sample odds ratios to quantitatively determine the suitability of a candidate comparison group (13).

Compared to other methods, the before-after with comparison group study is relatively simple. It accounts for non-treatment related time trends and changes in traffic volume. However, it is difficult to account for possible regression-to-the-mean influence. Regression-to-the-mean is the natural tendency of observed crashes to regress (return) to the mean in the year following an unusually high or low crash count. As a result, if a treatment had been installed at those locations with randomly high short-term crash counts, one would tend to over-estimate the treatment's safety effectiveness if the regression-to-the-mean influence is not properly addressed in the analysis.

2.2.2 The Empirical Bayes Approach

The EB approach is considered the most accurate before-after method to estimate the safety effectiveness of a treatment. The EB approach precisely predicts the number of crashes that would have occurred at an individual treated site in the after period if a treatment was not implemented (E_A). Similar to the before-after with comparison group method described in section 2.2.1, safety effectiveness is estimated by comparing the total crash prediction for all treated sites with the observed number of crashes in the after period.

The EB approach differs from the before-after with comparison group study in that the EB approach includes more complex steps to predict E_A . E_A is based on the number of crashes expected in the before period without the treatment (E_B). E_B is a weighted average of information from two sources:

- The observed number of crashes in the before period at the treated sites (O_B), and
- The crashes predicted at the treated sites based on reference sites with similar traffic and geometric features (N_B).

N_B can be calculated using Safety Performance Functions (SPFs) that indicate crashes' relationships with traffic volume and geometric features. SPFs regressed from information of an untreated "reference" group. Sites in the reference group have similar features as treated sites in the before period. Figure 5 shows how EB estimation works for a single site.

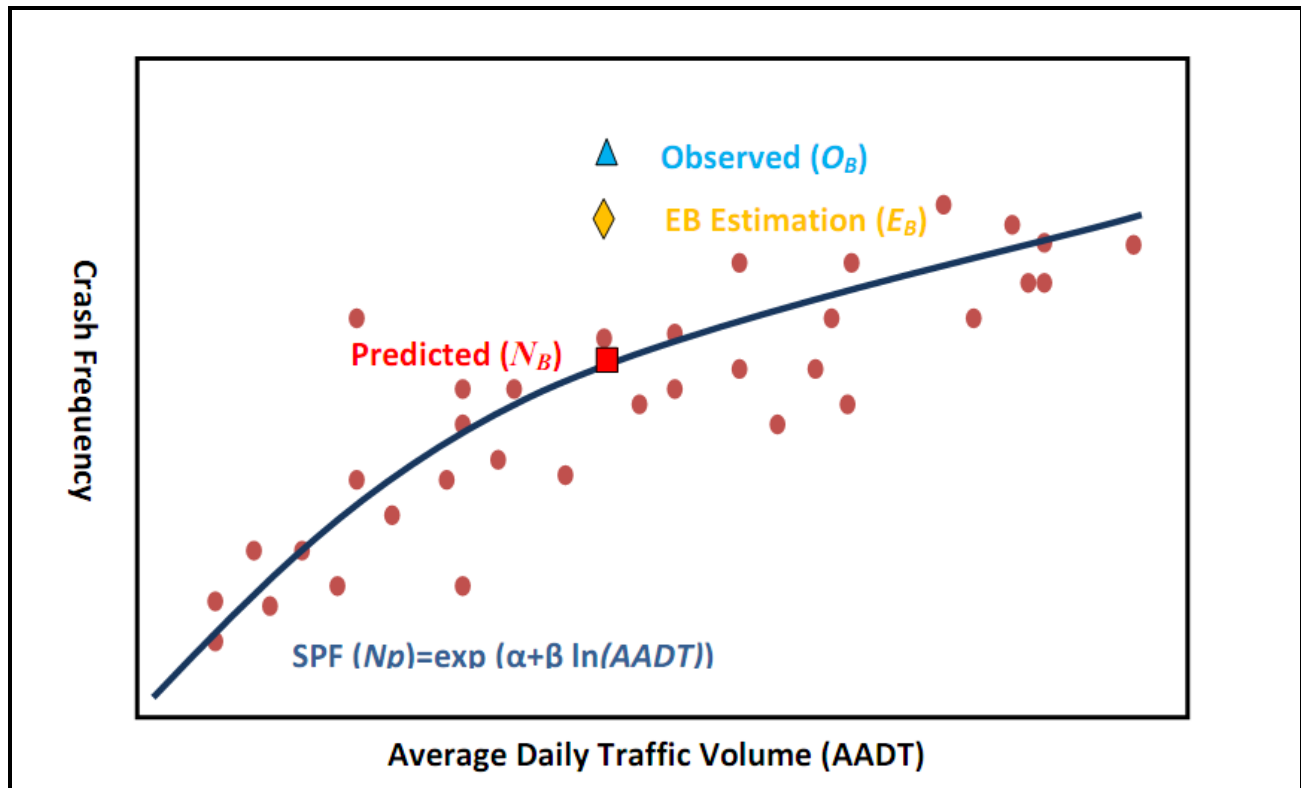


FIGURE 5 Illustration of regression-to-the-mean and the EB estimation.

As shown in Figure 8, E_B falls somewhere between the values of O_B and N_B . Regression-to-the-mean effect is the difference between observed crashes and EB estimation.

A ratio is calculated by dividing the predicted crash frequency of the treated sites (N_A) in the after period by that in the before period. In order to estimate the expected crashes in the treatment group had no treatment been applied, the EB estimated crash frequency in the before period in a treatment group is multiplied by the ratio.

2.2.3 Full Bayes Method

The Full Bayes method is a modeling approach which has a similar function as SPFs in an EB study, or the generalized linear models in a cross-sectional study. A Full Bayes study is useful for before-after or cross-section studies when: complex model forms are required, there is a need to consider spatial correlation among sites, and previous model estimates or CMF estimates are to be introduced into the modeling process.

In a Full Bayes method applied to a before-after study, a reference population is used to generate a distribution of likely values instead of a point estimate. The long-term expected crash frequency is then estimated by combining the distribution of likely values and the observed number of crashes. This methodology provides reliable results with small sample sizes. It can include prior knowledge, spatial correlation, and complex model forms in the evaluation process.

2.2.4 The Cross-Sectional Approach

A cross-sectional study examines the crash experience of locations with or without some feature and investigates the safety difference. In its most basic application, the CMF is estimated as the ratio of the average crash frequency for sites with and without the feature. Cross-sectional studies are useful when before-after data are limited, but there are a sufficient number of sites that are similar except for the treatment of interest.

In practice, it is difficult to find enough locations with similar features that may affect crash risk. This results in analyses that are often accomplished by multiple variable regression models. These models attempt to account for all variables that affect safety. Variables might include traffic volume and roadway geometric characteristics. CMFs are derived from the model parameters by calculating the change in the dependent variable (crashes) that results from a unit

change in a specific variable. Negative binomial regression models are recommended since they are able to account for overdispersion effects that commonly exist in crash data.

An issue that arises with cross-sectional studies is the difficulty in properly accounting for unknown, or known but unmeasured factors. The results derived from the models would be skewed. Additionally, inaccurate CMFs may be derived from inappropriate model forms, omitted variable bias, or correlation among variables.

2.2.5 Case-Control Studies

Case-control studies are based on cross-sectional data. Unlike cross-sectional studies, case-control studies group observations based on the outcome status such as whether crashes exist at a specific site. Outcome groups are called “cases” or “controls.” In each outcome group, the status of prior treatment is determined. An advanced case-control study can match cases with controls that are identical in crash risk factors. This method can control for potential confounding variables.

The odds ratio (*OR*) is a direct estimate of the CMF. Table 6 and Equation 3 indicate a simple example of developing CMF using case-control studies.

TABLE 6 Tabulation for Simple Case-Control Analysis

Treatment	Number of Cases	Number of Controls
With	A	B
Without	C	D

$$OR = CMF = \frac{A/B}{C/D} = \frac{AD}{BC} \quad \text{Equation 3}$$

As shown, in a simple case-control analysis, CMF can be determined by the related ratios of number of cases and number of controls.

2.3 Development of Crash Statistical Models

Crash statistic models indicate a relationship between crashes and roadway features. These models play significant roles in developing CMFs. They have been researched for many years, and have had several development stages. This portion of the literature review is to highlight several documents that demonstrate the evolution of crash statistical models from their early inception to their current form.

2.3.1 The Beginning of Crash Statistical Models

Zeeger et al. developed the first quantitative model to predict crashes using data from previous studies in Ohio and Kentucky (3) that studied relationships between lane and shoulder widening as well as the presence of obstructions along the roadway (14). The following model was developed using a weighted, least-squares fit method with normal distribution assumed:

$$AR = 4.1501(0.8907)^L (0.9562)^S (1.0026)^{LS} (0.9403)^P (1.0040)^{LP} \quad \text{Equation 4}$$

Where:

AR = number of run-off-road and opposite-direction crashes per million vehicle miles;

L = lane width (ft);

S = shoulder width including stabilized and unstabilized components (ft); and

P = stabilized component of the shoulder (et).

This equation was only a starting point for predictive models, and it was not able to account for crash risk factors other than lane width, shoulder width, and shoulder type.

Zeeger et al. conducted another study of statistical models, with a more comprehensive study of roadway geometry and its effects on crashes (4). Data were collected from seven states, which provided more variety in geographic characteristics, like terrain type. The model analyzed

various combinations of thirty-four variables, including the number of railroad crossings, number of intersections, and type of development adjacent to the roadway. After checking the interactions and correlations of the variables, a best-fit equation was found to be as follows:

$$A = 0.0019(ADT)^{0.8824} (0.8786)^W (0.9192)^{PA} (0.9316)^{UP} (1.2365)^H (0.8822)^{TER_1} (1.3221)^{TER_2}$$

Equation 5

Where:

A = number of crashes per mile per year;

ADT = average daily traffic;

W = lane width (ft);

PA = width of paved shoulder (ft);

UP = width of unpaved shoulder (ft);

H = average roadside hazard rating;

TER_1 = 1 for flat terrain, 0 otherwise; and

TER_2 = 1 for mountainous terrain, 0 otherwise.

This model assumed a normal distribution, and its goodness of fit was indicated by the regular R^2 statistic (0.456).

2.3.2 Poisson Regression and Negative Binominal Regression Models

As relationships between road improvements and the reduction in crashes became clearer and preliminary equations were developed to predict the number of crashes on a roadway, researchers began to explore these models to more accurately predict crashes.

Miaou and Lum compared four different types of models to find the model of best fit to estimate the number of truck crashes along a highway segment (15). These models included additive and multiplicative linear regression models, and multiplicative Poisson regression with an exponential rate function and a nonexponential rate function. The Poisson regression models were found to predict better as crashes are rare and random events, and the crash counts were nonnegative numbers. The Poisson regression model was closer to a probability model as compared to the multiple linear regression models. The best fit model is as follows:

$$P(y_i) = \frac{(\lambda_i v_i)^{y_i} e^{-\lambda_i v_i}}{y_i!} \quad \text{Equation 6}$$

Where:

y_i = number of trucks involved in crashes on the highway segment;

$P(y_i)$ = probability that y_i trucks will be involved in crashes;

λ_i = mean crash rate (number of trucks per million truck-miles) on the segment; and

v_i = truck exposure (millions of truck-miles).

λ_i is predicted using the following equation:

$$\lambda_i = 0.0818 + 0.1022x_{1i} + 0.0949x_{2i} + 0.0426x_{3i} + 0.0341x_{4i} - 0.0263x_{5i} \quad \text{Equation 7}$$

Where on the i^{th} section:

x_{1i} = average daily traffic (ADT) per lane (in thousands of vehicles);

x_{2i} = horizontal curvature (in degrees per hundred feet);

x_{3i} = x_{2i} multiplied by horizontal curve length;

x_{4i} = deviation of stabilized outside shoulder width in each direction; and

x_{5i} = percent trucks.

A limitation of using a Poisson regression model is that the conditional variance of data is restrained to be equal to the conditional mean. However, the crash data were found to display extra variations or overdispersions relative to a Poisson model. The negative binomial regression model was recommended to account for overdispersion as it allows for additional variance.

A followed-up study by Miaou compared a Poisson regression, zero-inflated Poisson (ZIP), and the negative binomial regression statistical methods in predicting truck crashes (16). No model was proved to be better than the other models. The author concluded that a Poisson regression can be used to establish the relationship between highway geometry and crashes. Additionally, ZIP or negative binomial regressions were preferred if the Poisson regression is found to have overdispersion.

The EB approach has become the forerunner in crash prediction study. The EB approach combines statistic models with crash history to make accurate estimates. Negative binominal models were found to work better than other form of models in the EB procedure (17, 18).

The base SPFs in the HSM were based on a study by Vogt and Bared (19). They collected roadway geometric features for rural two-lane highways in Washington and Minnesota. This study investigated the Poisson regression model, negative binomial regression, and an extended negative binomial regression, which broke segments into homogeneous subsegments. The extended negative binomial regression technique was recommended as it accounted for overdispersion and worked well with the EB method when past crash data were available. The following equation was created by the extended negative binomial regression:

$$N_{br} = EXPO \times \exp(0.6409 + 0.1388STATE - 0.0846LW - 0.0591SW + 0.0668RHR + 0.0084DD)$$

$$(\sum WH_i \exp(0.0450DEG_i))(\sum WV_j \exp(0.4652V_j))(\sum WG_k \exp(0.1048GR_k))$$

Equation 8

Where:

N_{br} = predicted number of crashes along a highway segment;

$EXPO$ = exposure in million vehicle-miles of travel per year = $(ADT)(365)(L)(10^{-6})$;

ADT = average daily traffic volume (veh/day) on highway segment;

L = length of roadway segment (mi);

$STATE$ = which state the segment is in (0 = Minnesota, 1 = Washington);

LW = lane width (ft); average if different in each direction;

SW = shoulder width (ft); average if different in each direction;

RHR = roadside hazard rating; takes values from 1 to 7 and represents how hazardous the roadside can be;

DD = driveway density (driveways per miles) on highway segment;

WH_i = weight factor for the i^{th} horizontal curve in the highway segment; proportion of total highway segment length represented by the portion of the i^{th} horizontal curve that lies in the segment (the weights, WH_i , must sum to 1.0);

DEG_j = degree of curvature for the i^{th} horizontal curve in the highway segment (degrees per 100 ft);

WV_j = weight factor for the j^{th} crest vertical curve in the roadway segment; proportion of total highway segment length represented by the portion of the j^{th} vertical curve that lies in the segment (the weights, WV_j , must sum to 1.0);

V_j = crest vertical curve grade rate for the j^{th} crest vertical curve that lies within the segment in percent change in grade per 100 ft = $|g_{j2}-g_{j1}|/l_j$;

g_{j1}, g_{j2} = highway grades at the beginning and end of the j^{th} vertical curve (percent);

l_j = length of j^{th} vertical curve (in hundreds of feet);

WG_k = weight factor for the k^{th} straight grade segment in the roadway segment; proportion of total highway segment length represented by the portion of the k^{th} straight grade segment that lies in the segment (the weights, WG_k , must sum to 1.0); and

GR_k = absolute value of grade for the k^{th} straight grade on the segment (percent).

Estimates of safety based on statistical models, like that used by Vogt and Bared (19), can be a very accurate method for predicting expected crashes. However, statistical models can also show inverse or disproportionate weighting of variables that are not consistent with engineering principles. This can often be caused by variables serving as surrogates for other factors. In addition, the statistical models do not necessarily show a cause and effect relationship, only a correlation. In order to more accurately account for the impact of various highway elements on safety, additional scrutiny of the model was needed.

In summary, Chapter 2 reviews previous research on shoulder safety, methodologies to develop CMFs, and statistical models to predict crashes. According to the results from literature review, as well as data availability, the EB approach and the cross-sectional approach were applied to study composite shoulders' safety effectiveness. Detail discussions on the EB approach and the cross-sectional approach are given in Chapter 3.

CHAPTER 3 METHODOLOGY

For decades, observational before-after studies have been considered the industry standard for evaluation of highway safety treatments which resulted in the development of CMFs. Harwood et al. documented that there are three common ways to carry out a before-after study: naïve before-after evaluations, comparison group evaluations, and the EB approach (20). Of these three evaluations listed, the EB approach was widely recommended in the HSM (1). According to Hauer's study, the EB method is able to account for regression-to-the-mean effects, as well as traffic volume and other roadway characteristic changes, by combining SPF estimates with the observed number of crashes (13). This allows the EB approach to overcome limitations identified by other two evaluations and provide more accurate estimates. Moreover, KDOT conducted a significant number of shoulder upgrade projects between 2003 and 2007. This allowed for an adequate sample of study sites. All these left the EB approach as the most suitable and desirable method for this study.

The cross-sectional study was used as a supporting method in estimating the safety effectiveness of composite shoulders. It can be an alternative to the EB approach in case there are not adequate sample sites for fatal and injury (FI) crash and related crash analysis. Additionally, its results can be compared to the EB results to validate each other.

3.1 The Empirical Bayes Approach

The general EB procedure has been studied or described by many researchers (e.g. Persaud et al.(21), Abdel-Aty et al. (22), Srinivasan et al. (23), Harwood et al. (24), and Hauer (13)). One key step for the EB approach is to develop or select a SPF. A well-developed SPF will properly account for traffic volume and geometric feature changes. In addition, developing specific SPFs

for various types of crashes is necessary since most treatments affect various crash impacts and severity types differently. A reference group was generated by containing rural two-lane highway segments with no or unpaved shoulders. Geometric, traffic flow, and crash history data were obtained from KDOT. Three types of SPFs, including a SPF for total crashes, a SPF for fatal and injury crashes, and a SPF for related crashes, were created using negative binomial regression. The related crashes included run-off-road, head-on, and sideswipe (same direction and opposite direction) crashes.

SPFs were applied to initially predict annual crash frequency of segments in the treatment group. The treatment group included segments where shoulders were upgraded from unpaved or non-existent to composite shoulder with a three ft paved section and the remainder unpaved. For every individual segment, the next step was to combine the sum of initial predictions (N_B) with the sum of observed count of crashes (O_B). This process was completed through the use of an overdispersion parameter (k), and resulted in a good estimation of crash frequency (E_B) for the expected number of crashes in the before period (13). The related variance ($Var(E_B)$) was also estimated. This estimate of E_B was:

$$E_B = wN_B + (1-w)O_B, \quad w = \frac{1}{1+kN_B} \quad \text{Equation 9}$$

$$Var(E_B) = (1-w)E_B \quad \text{Equation 10}$$

According to Harwood et al., the EB procedure works best if the roadway segments contained at least a specified minimum number of predicted crashes (13). The minimum crash frequency is generally $1/k$, where k is the overdispersion parameter of the relevant base model. Segments were combined for EB analysis if they did not meet the minimum crash frequency

requirement. The analysis should account for both the assumption that both are perfectly correlated (H_0) and the alternative assumption that the different entities are statistically independent (H_1). It was implemented by the following equations (24):

$$E_0 = w_0 \sum_i N_i + (1 - w_0) \sum_i O_i, w_0 = \frac{1}{1 + \sqrt{k}} \quad \text{Equation 11}$$

$$E_1 = w_1 \sum_i N_i + (1 - w_1) \sum_i O_i, w_1 = \frac{1}{1 + k \sum_i N_i} \quad \text{Equation 12}$$

$$E_B = \frac{E_0 + E_1}{2} \quad \text{Equation 13}$$

Where:

w_0 = weight placed on predicted crash frequency when crash frequencies for different roadway segments are perfectly correlated (H_0);

w_1 = weight placed on predicted crash frequency when crash frequencies for different roadway segments are statistically independent (H_1);

E_0 = expected crash frequency based on H_0 ;

E_1 = expected crash frequency based on H_1 ;

E = expected crash frequency for combined segments;

k = overdispersion parameter;

N_i = predicted crash number for segment i ; and

O_i = observed crash number for segment i .

With the above results and the prediction of number of crashes (N_A) from the SPF for the same segment, the expected number of crashes in the after period without upgrading the shoulder could be estimated by the following equation:

$$E_A = E_B \frac{N_A}{N_B} \quad \text{Equation 14}$$

To estimate the index of safety effectiveness, or CMF, one needs to sum E_A over all road segments in the treatment group (E_{Asum}) and then compare with the total observed crash number (O_{Asum}) during the after period in the same group (Equation 15) (13). The standard deviation of CMF is determined by Equation 16 (13).

$$CMF = \frac{O_{Asum} / E_{Asum}}{1 + \text{Var}(E_{Asum}) / E_{Asum}^2} \quad \text{Equation 15}$$

$$\sigma = \sqrt{\frac{CMF^2 (\text{Var}(E_{Asum}) / E_{Asum}^2 + \text{Var}(O_{Asum}) / E_{Asum}^2)}{(1 + \text{Var}(E_{Asum}) / E_{Asum}^2)^2}} \quad \text{Equation 16}$$

3.2 The Cross-Sectional Approach

A cross-sectional study can be used as a supplemental method when before-after data are limited. As discussed previously, its analyses are often accomplished through multiple variable regression models. Many model forms have been recorded by researchers (Miaou (16), Vogt and Bared (19)) and the negative binomial specification has become the forerunner in crash count modeling. For this specific research, negative binomial regression models were applied both in the cross-sectional approach and to develop SPFs in the EB approach. While the variance in the number of crashes at a site is equal to the mean in the Poisson distribution, it is greater than the

mean under the negative binomial distribution. This phenomenon is known as overdispersion. The negative binomial model takes the form (19):

$$P(y_i) = \frac{\Gamma(y_i + \frac{1}{K})}{y_i! \Gamma(\frac{1}{K})} \left(\frac{K\mu_i}{1 + K\mu_i}\right)^{y_i} \left(\frac{1}{1 + K\mu_i}\right)^{\frac{1}{K}} \quad \text{Equation 17}$$

Where:

$P(y_i)$ = the probability of y_i crashes observed at site number i ;

μ_i =the mean number of crashes to be expected at site number i ; and

K = the overdispersion parameter.

In order to represent overdispersion, a quadratic term is added to the variance as shown in Equation 18. If K equals 0, the negative binomial reduces to the Poisson model. The greater the value of K , the more variability there is in the data over and above that associated with the mean.

$$Var = \mu_i + K(\mu_i)^2 \quad \text{Equation 18}$$

The vector of coefficients β and K are estimated by maximizing the log-likelihood function for the negative binomial distribution as shown in Equation 19 (19).

$$L(\beta, K) = \sum_i \left[\left(\sum_{j=0}^{y_i} \log(1 + Kj) \right) - \log(1 + Ky_i) + y_i \log \mu_i - \left(y_i + \frac{1}{K} \right) \log(1 + K\mu_i) - \log(y_i!) \right]$$

$$\text{Equation 19}$$

In practice, negative binomial regression models are estimated by statistical software such as SPSS, SAS, and STATA using information about traffic volume, crash record, and roadway features (SPSS code is included in Appendix A). A common model form for a roadway segment is indicated as follows:

$$n = \alpha(\text{SegmentLength})^{\beta_L} (\text{AADT})^{\beta_{AADT}} e^{\sum \beta_i x_i} e^{\sum \beta_j x_j} \quad \text{Equation 20}$$

Where:

n = expected crashes for a site;

x_i = crash risk factors that are treated as continuous variables;

x_j = crash risk factors that are treated as categorical variables;

$\beta_L, \beta_{AADT}, \beta_i, \beta_j$ = coefficients.

Using the estimated coefficients from the model, CMFs can be inferred. CMFs represent the changes in expected crashes when the value of a variable is changed. For continuous variables, CMF function can be developed by the Equation 21.

$$CMF = \exp(\beta_i \Delta x_i) \quad \text{Equation 21}$$

Where:

Δx_i = the changed value of the variable.

For categorical variables, one variable that is most likely to be treated is set to be the reference group, and its coefficient is defaulted as 0. CMFs for other variables are equal to values of $\exp(\beta_j)$. For example, three types of shoulders were included in the model in this research: narrow unpaved shoulders (≤ 5 ft), wide unpaved shoulders (>5 ft), and composite

shoulders. If segments with narrow unpaved shoulders were treated as the reference group, CMFs for widening the shoulders and upgrading to composite shoulders can be indicated by their individual coefficients.

CHAPTER 4 DATA DESCRIPTION

4.1 General Consideration

A considerable number of shoulder upgrading projects were conducted in the state of Kansas between 2003 and 2007. The study period for this thesis was set from 2000 to 2009. This allowed for adequate crash data to perform a before-after study. Data were extracted by the Geometric and Accident Data Unit of KDOT. Two separated databases were used to obtain roadway characteristics and crash history information. Rural two-lane highways in Kansas were broken into 5,682 segments. Individual segments contained information such as county name, route name and number, beginning and ending county milepost, segment length, AADT, shoulder type/width, lane width, and record year. Vehicle crash database included crashes that occurred between 2000 and 2009. Individual crash record contains the variables listed in Table 7.

TABLE 7 Vehicle Crash Database Variables and Description

Variables	Description
ACCIDENT_KEY	crash ID in KDOT crash database, it can link to the original crash report
DATE_OF_ACCIDENT	it indicates the date of the crash
ROUTE	it shows in which route the crash occurred
ACC_CMP	accident county milepost, it indicates the crash's location on the route
NBR_OF_FATALITIES	number of fatalities involved in the crash
NBR_OF_INJURIES	number of injuries involved in the crash
AccidentLocation	it indicates whether the crash is intersection related crash, non-intersection crash, parking lot crash, etc.
CWOV_FHE	first harmful event, it indicates the type of crash such as head on, rear end, sideswipe, etc.

During the modification of data, ACC_CMP was used to match the crash to the related segment; NBR_OF_FATALITIES and NBR_OF_INJURIES were used to determine whether the crash was a FI crash; AccidentLocation was used to remove non-segment crashes for the purpose of this research; and CWOV_FHE was used to calculate the number of shoulder related crashes.

The first step of data collection was to select study sites. The EB approach and cross-sectional approach have their own criteria to pick study segments. The shoulder type and width were tracked every year to see whether and when the segment's shoulder had been upgraded or not. Using this information, an initial list of segments was generated. Staff at KDOT evaluated whether any major construction or alignment changes on the selected study sites occurred during the study period. Segments that had major construction or alignment changes were removed and a final list of study segments was created.

After the study sites were selected, crash data were matched to the related segments based on county milepost using Matlab. A code was written to determine the numbers of total crashes, fatal and injury (FI) crashes, and shoulder related crashes in each year for individual segments. This code is included in Appendix B. This automated procedures reduced the time needed to match the data, and increased the accuracy of the processing. The final datasets included three lists of study segments with roadway features and crash history: one contained the reference group data; one contained the treatment group data for the EB approach; and a third one contained data for the cross-sectional approach. Appendix C shows the first ten records in the dataset for the cross-sectional approach.

It was assumed that study site shoulders had a consistent impact on crashes during the entire year, and the winter crashes were considered. Table 8 and Table 9 show the distribution of winter crashes.

TABLE 8 Winter Crashes Distribution for Total Crashes

Year	Winter	Non-winter	Crashes in a Year	Percentage
2000	125	141	266	47%
2001	133	163	296	45%
2002	114	162	276	41%
2003	122	161	283	43%
2004	122	150	272	45%
2005	132	119	251	53%
2006	126	173	299	42%
2007	130	166	296	44%
2008	162	158	320	51%
2009	99	140	239	41%
Average	127	153	280	45%

TABLE 9 Winter Crashes Distribution for Related Crashes

Year	Winter	Non-winter	Crashes in a Year	Percentage
2000	14	15	29	48%
2001	18	18	36	50%
2002	13	18	31	42%
2003	13	31	44	30%
2004	15	25	40	38%
2005	18	17	35	51%
2006	21	26	47	45%
2007	20	24	44	45%
2008	27	38	65	42%
2009	18	29	47	38%
Average	18	24	42	43%

In Kansas, winter weather lasts an average of five months from November to March. It is speculated that the percentage of winter crashes should be approximately 42 percent. It was found that only a slightly increase of crashes between November and March occurred. Additionally, similar trends were found between total crashes and related crashes. For example, both of total crashes and related crashes had the highest winter crash percentage in 2005. This resulted in not removing winter crashes because it was speculated to not skew the results of this study.

4.2 Data for the Empirical Bayes Approach

Two datasets were created: one for the reference group and the other one for the treatment group. Information from the reference group were used to develop the SPFs, while information from the treatment group was used to conduct before-after studies. The selection criteria for the reference group segments are as follows:

- i. Segments with unpaved shoulders or without shoulders on rural two-lane roads;
- ii. No major construction occurred that changed the segment characteristics significantly;
- iii. Information concerning AADT, crashes, and geometry were available; and
- iv. A length of the segment range from 0.25 mile to 5 miles.

The selection criteria for the treatment group are stated as follows:

- i. Rural two-lane road segments with the shoulder upgraded from unpaved or non-existent to paved by three ft and the remainder unpaved;
- ii. No major construction occurred that change the segment characteristics significantly;
- iii. Information concerning AADT, crashes, and geometry were available for the before period of at least three years and the after period of at least one year; and
- iv. A length of the segment range from 0.25 mile to 5 miles.

Based on the listed criteria, 61 rural two-lane roads which included 204 segments were selected into the reference group. Twenty nine segments were defined as the treatment group, with nine of them having shoulder upgrades between 2003 and 2004, 12 between 2004 and 2005, and eight between 2005 and 2006. Segments in the treatment group were also included into the reference group since their properties in the before period also met the reference group criteria. In all, segments covered 395.6 miles of rural two-lane highways in Kansas. 1,758 observations

were in the reference group, 115 observations were in the before periods of the treatment group, and 117 observations were in the after periods of the treatment group. One observation was a year's record for one segment, including crash numbers, traffic volume, and geometry data.

Table 10 offers a summary of the databases described in the previous sections.

TABLE 10 Descriptive Statistics for Segments in the Reference and Treatment Groups

Groups	Variable	Minimum	Maximum	Mean	Std. Deviation	Exposure ^A
Reference Group	<i>Total crash</i>	0	17	1.04	1.64	0.57
	<i>FI crash</i>	0	7	0.19	0.55	0.10
	<i>Related crash</i>	0	13	0.17	0.67	0.09
	<i>Lane width (ft)</i>	8.5	14.0	11.80	0.49	
	<i>AADT (veh/day)</i>	65	4580	940	789	
	<i>Length (miles)</i>	0.25	5.00	1.82	1.23	
Treatment Group-Before	<i>Total crash</i>	0	6	0.97	1.21	0.47
	<i>FI crash</i>	0	2	0.15	0.38	0.07
	<i>Related crash</i>	0	2	0.11	0.37	0.06
	<i>Lane width (ft)</i>	11.5	14.0	12.01	0.35	
	<i>AADT(veh/day)</i>	380	2340	978	500	
	<i>Length (miles)</i>	0.25	4.98	2.06	1.21	
Treatment Group-After	<i>Total crash</i>	0	5	1.21	1.31	0.56
	<i>FI crash</i>	0	2	0.16	0.41	0.07
	<i>Related crash</i>	0	2	0.10	0.36	0.05
	<i>Lane width (ft)</i>	11.5	14.0	12.02	0.45	
	<i>AADT(veh/day)</i>	540	2150	983	376	
	<i>Length (miles)</i>	0.25	4.98	2.11	1.26	

^A Average number of crashes per mile per year.

As shown in Table 10, the average number of total crashes per mile per year in the treatment group-after was larger than that in the treatment group-before. However, the safety effectiveness cannot be indicated simply by these average numbers since they were not able to account for traffic volume and other changes. The EB approach was able to account for these changes, as well as different durations between the before and the after period. The average AADT in the treatment group is slightly higher than that in the reference group. It was also found that approximately half of segments in the treatment group had narrow unpaved shoulders, and all converted to wide composite shoulders in the after period.

Three types of SPFs needed to be created. As a result, non-intersection crash data were segregated into three types: total number of crashes, number of FI crashes, and number of related crashes (run-off-road, head-on, and sideswipe).

Two dummy variables were defined including: *shoulder size* and *treatment group*. A value of 0 for *shoulder size* was used if a segment's entire shoulder width was equal to or less than five ft. A value of 1 was used if the width was greater than five ft. in the reference group, approximately 47 percent of the segments had wide unpaved shoulders (> 5 ft). In the treatment group, approximately 50 percent of the segments had narrow shoulders (≤ 5 ft) in the before period, and they were widened to greater than five ft in the after period.

A second dummy variable was created to test and correct a common EB approach issue concerning how well the reference group represented the treatment group. Persaud and Lyon stated that a reference group must be similar to the treatment group in terms of geometric design, traffic volumes, vehicle fleet, and so on in order to represent the treatment group well (25). Although the two groups used in the study were similar, it was still possible that the relationships between the number of crashes and independent variables were different between the two groups. If the relationships were different, the study sample would suffer "the reference group issue", which indicates that the reference group can not represent well the treatment group. As a result, there were two purposes for the variable *treatment group*: (i) test how well the reference group represented the treated entities; and (ii) provide a possible solution if the treated entities were not represented well by the reference group. A value of 1 was used if the segment was included into both the reference and treatment groups, otherwise it was 0.

Besides the above information, the total annual crashes were collected in the same class (Class D roads) as the roadways in the reference group. In Kansas, “Class D” routes are less traveled, most of which are rural two-lane roads, and include approximately 3,270 miles. It was used to create a control variable as a direct measure of exposure to risk. This method was used by Noland et al. (26), and Quddus (27) to account for the time series effect in crash prediction models. Table 11 summarizes crash information for Class D roads by year.

TABLE 11 Crash Information on Class D Roads by Year

Year	Total Crashes	FI Crashes	Related Crashes	VMT^A
2000	5,917	1,480	1,516	^B
2001	5,657	1,374	1,514	2,206.04
2002	5,678	1,385	1,583	2,183.44
2003	5,419	1,283	1,465	2,209.42
2004	5,447	1,258	1,416	2,188.36
2005	5,157	1,175	1,448	2,148.83
2006	4,876	1,168	1,427	2,127.44
2007	5,073	1,146	1,504	2,113.09
2008	4,751	1,093	1,404	2,138.00
2009	4,208	891	925	1,992.41

^A. Annual vehicle miles traveled in million.

^B. Data not available in KDOT’s database, its value is assumed to equate to the average value of 2001 and 2002.

As shown in the table, crashes had a decreasing trend between 2000 and 2009, with 2009 experiencing the most significant decrease in crashes.

4.3 Data for the Cross-Sectional Approach

For this research, a cross-sectional approach was used as a supplemental method to compared the safety effectiveness of composite shoulders with wide (>5 ft) unpaved shoulders and narrow (≤5 ft) unpaved shoulders. The reference group data for the EB approach were still used to provide information for wide shoulders and narrow shoulders. Segments in the treatment group contained composite shoulder information in their after period. However, additional composite shoulder

data were required to be fully represented. Seventy more segments with composite shoulders between 2000 and 2009 were randomly selected using similar criteria as the reference group for the EB approach. A total of 284 segments were included in the dataset. There were 941 observations for narrow unpaved shoulders, 828 observations for wide unpaved shoulders, and 817 observations for composite shoulders. Table 12 summarizes the descriptive statistic for the cross-sectional study data.

TABLE 12 Descriptive Statistics for Segments in the Cross-Sectional Approach

Shoulder Types	Variable	Minimum	Maximum	Mean	Std. Deviation	Exposure ^A
Narrow Unpaved Shoulders (≤ 5 ft)	<i>Total crash</i>	0	17	1.19	1.79	0.58
	<i>FI crash</i>	0	7	0.26	0.65	0.13
	<i>Related crash</i>	0	13	0.26	0.87	0.12
	<i>Lane width (ft)</i>	8.5	14.0	11.70	0.62	
	<i>AADT (veh/day)</i>	65	3320	667	534	
Wide Unpaved Shoulders (>5 ft)	<i>Total crash</i>	0	11	0.86	1.43	0.56
	<i>FI crash</i>	0	4	0.11	0.38	0.07
	<i>Related crash</i>	0	2	0.07	0.27	0.04
	<i>Lane width (ft)</i>	11.0	12.0	11.91	0.25	
	<i>AADT (veh/day)</i>	380	2340	978	500	
Composite Shoulders	<i>Total crash</i>	0	17	1.63	2.47	0.75
	<i>FI crash</i>	0	6	0.24	0.58	0.11
	<i>Related crash</i>	0	5	0.16	0.49	0.07
	<i>Lane width (ft)</i>	11.5	14.0	12.00	0.17	
	<i>AADT (veh/day)</i>	495	4950	1562	937	
	<i>Length (miles)</i>	0.34	4.98	2.16	1.42	

^A Average number of crashes per mile per year.

As shown in Table 12, segments with composite shoulders had the highest AADT, while segments with narrow shoulders had the least AADT. Additionally, segments with composite shoulders experienced more crashes per mile per year than segments with the other two types of shoulders. However, they showed less average FI crashes and related crashes than segments with narrow shoulders. Wide unpaved shoulders had the least exposures in all three types of crashes.

During the development of regression models, shoulder types were treated as categorical variables, with *narrow shoulders* as the reference variable.

The above information in Chapter 4 introduces the data collection procedure and summaries for data used in the EB approach and the cross-sectional approach. Data analysis and results are shown in Chapter 5.

CHAPTER 5 DATA ANALYSIS

As discussed before, the EB approach was the main analysis method to study the safety effectiveness of composite shoulders. A cross-sectional approach was used as a supplemental methodology. Chapter 5 describes the two methods, as well as a comparison of their results.

5.1 Safety Performance Functions

5.1.1 Variable Selection

Previous research studies have investigated various SPF forms. The most widely used form is a negative binomial regression model as shown in Equations 22 and 23:

$$n = \alpha(\text{SegmentLength})^{\beta_1} (\text{AADT})^{\beta_2} e^{(\beta_3 x_1 + \beta_4 x_2 + \dots)} \quad \text{Equation 22}$$

or $\ln(n) = \alpha + \beta_1 \ln(\text{SegmentLength}) + \beta_2 \ln(\text{AADT}) + \beta_3 x_1 + \beta_4 x_2 + \dots \quad \text{Equation 23}$

Where n is the predicted annual crash number, α and β_i are coefficients, and x_i are other explanatory variables than segment length and AADT. This form was determined to offer the best performance for this research.

For every individual SPF, including the natural logarithms of AADT, segment length, and the relevant number of crashes on Class D roads, there were ten variables that could be included in the model. This resulted in 521 possible subsets of explanatory variables. Akaike's Information Criterion (AIC), a small-is-better criterion, was used to decide which subset of explanatory variables should be included in the model. Miaou (28) and Abdel-Aty and Radwan (29) have reported the use of AIC in crash modeling. Various models with different subsets of explanatory variables were examined, and a best model was selected based on the smallest AIC value. The total crash SPF included four variables: $\ln(\text{AADT})$, $\ln(\text{Length})$, shoulder size , and

lane width; the FI crash SPF included four variables: *ln (AADT)*, *ln (Length)*, *shoulder size*, and *treatment group*; and the related crash SPF included six variables: *ln (AADT)*, *ln (Length)*, *shoulder size*, *lane width*, *treatment group*, and *ln (number of related crashes on Class D roads)*. It was found that dropped variables were not significant at the 0.1 level of confidence.

5.1.2 Test Statistics

Compared to ordinary least squares (OLS) models, negative binominal regression models have different test statistics for both parameters and goodness-of-fit.

The Wald Chi-Square test is a defaulted parametric statistical test for negative binominal regressions in many statistical programs. Under the Wald test, the maximum likelihood estimate of the coefficient (β) is compared with the proposed value (β_0), with the assumption that the difference between the two will be approximately normal. The Wald Chi-Square can be determined by Equation 24, which is compared against a Chi-square distribution.

$$WaldChiSquare = \frac{(\beta - \beta_0)^2}{Var(\beta)} \quad \text{Equation 24}$$

The ordinary R-squared is not a preferred goodness-of-fit measure for negative binominal regressions. It cannot account for non-normal distribution, as well as overdispersion effects. This research used the Pseudo R-squared test, advocated by Miaou (28). The test is based explicitly on the overdispersion parameter as shown in Equation 25.

$$PseudoR^2 = 1 - \frac{K}{K_{max}} \quad \text{Equation 25}$$

Where K is the overdispersion parameter estimated in the model, and K_{max} is the overdispersion parameter estimated in a intercept-only model.

Besides the above test statistics, the log-likelihood ratio Chi-squared was applied to compare the fitted model against the intercept-only model.

5.1.3 Final Models

SPSS was applied to develop SPF models using data of the reference group. A code was included in appendix A. Table 13 displays the result of the developed SPFs, as well as their goodness-of-fit information.

TABLE 13 Safety Performance Functions for Composite Shoulder

Variable	SPF for Total Crashes		SPF for FI Crashes		SPF for Related Crashes	
	Coefficient (Std. Error)	Wald Chi-Square	Coefficient (Std. Error)	Wald Chi-Square	Coefficient (Std. Error)	Wald Chi-Square
<i>Intercept</i>	-3.18 (0.87)	13.33*** ^A	-9.99 (0.76)	173.11***	8.25 (8.60)	0.92
<i>Ln AADT</i>	0.71 (0.05)	231.83***	1.19 (0.11)	121.09***	1.50 (0.13)	124.97***
<i>Ln Length</i>	1.00 ^B		1.39 (0.12)	128.05***	1.53 (0.15)	101.39***
<i>Shoulder size</i>	-0.35 (0.07)	26.82***	-1.17 (0.15)	62.51***	-1.64 (0.20)	89.43***
<i>Lane width</i>	-0.16 (0.07)	5.58*			-0.37 (0.16)	5.06**
<i>Treatment group</i>		^C	-0.94 (0.27)	12.02***	-0.93 (0.34)	7.29***
<i>Ln Crashes on Class D Road</i>					-2.33 (1.20)	3.80**
k	0.46 (0.05)		0.71 (0.20)		1.40 (0.30)	
Log-likelihood ratio Chi-Square	228.76***		265.53***		300.91***	
AIC	4354.71		1560.39		1323.03	
Pseudo R ²	0.65		0.73		0.74	

^A * indicates statistically significant at the 0.1 level;

** indicates statistically significant at the 0.05 level; and

*** indicates statistically significant at the 0.01 level;

^B This variable was used as an offset in the model;

^C Cells were left blank because the variables were not included into the model.

The three SPFs have the expected positive or negative coefficients. It was found that the variable *treatment group* was significant in both the FI crash model and the related crash model at the 0.01 level of confidence, while it was not significant in the total crash model at the 0.1 level. This means that the relationship between number of crashes and the independent variables in the treated segments differs from that in the untreated segments with respect to FI crashes and related crashes. In this case, the reference group may not represent the treated group well in both

FI crashes and related crashes. For a later analysis, including the *treatment group* variable helped to improve but not remove the reference group issue. The coefficients of shoulder size in all three SPFs were negative and significant at the 0.01 level. As a result, the models showed that wider shoulders (> 5 ft) were associated with fewer crashes than narrower shoulders (≤ 5 ft).

For the exposure to risk, the natural logarithms of the relevant number of crashes on Class D roads were not statistically significant at 0.1 level of confidence on total crash number and FI crash number. However, a significant relationship was found between the related crashes and the natural logarithm of the related crashes on Class D roads.

The related crash SPF did not have a significant intercept at the 0.1 level. In this case, it is unlikely that the line of related crash SPF would pass through the origin where $\ln(n)=0$, and $x_i = 0$. The estimation of coefficients will be biased if excluding the intercept (31). As a result, the intercept was not removed in order to avoid biasing the model. In fact, models with insignificant intercepts were also found in previous research (19, 24).

It is necessary to discuss the transferability of these SPFs, especially for FI crash SPF and related SPF. They can work best on projects regarding rural two-lane roads in Kansas, since these roads share similar characteristics with Class D roads and segments in the reference group. People can set the default value for the *treatment group* variable as 0 under the assumption that the new treatment group can be represented well by the SPFs. Additionally, calibration procedures recommended by HSM can be conducted to gain a more accurate result.

Figure 6 displays the scatter plots of the dependent variables in these three SPFs and their relevant residuals. The X axes of the scatter plot indicate the observed number of crashes for a sample site in a certain year, while the Y axes indicate the difference between the observed

number and the SPF prediction, which was mentioned as SPF residual here. With these plots, readers are better able to notice the general performance of the SPF.

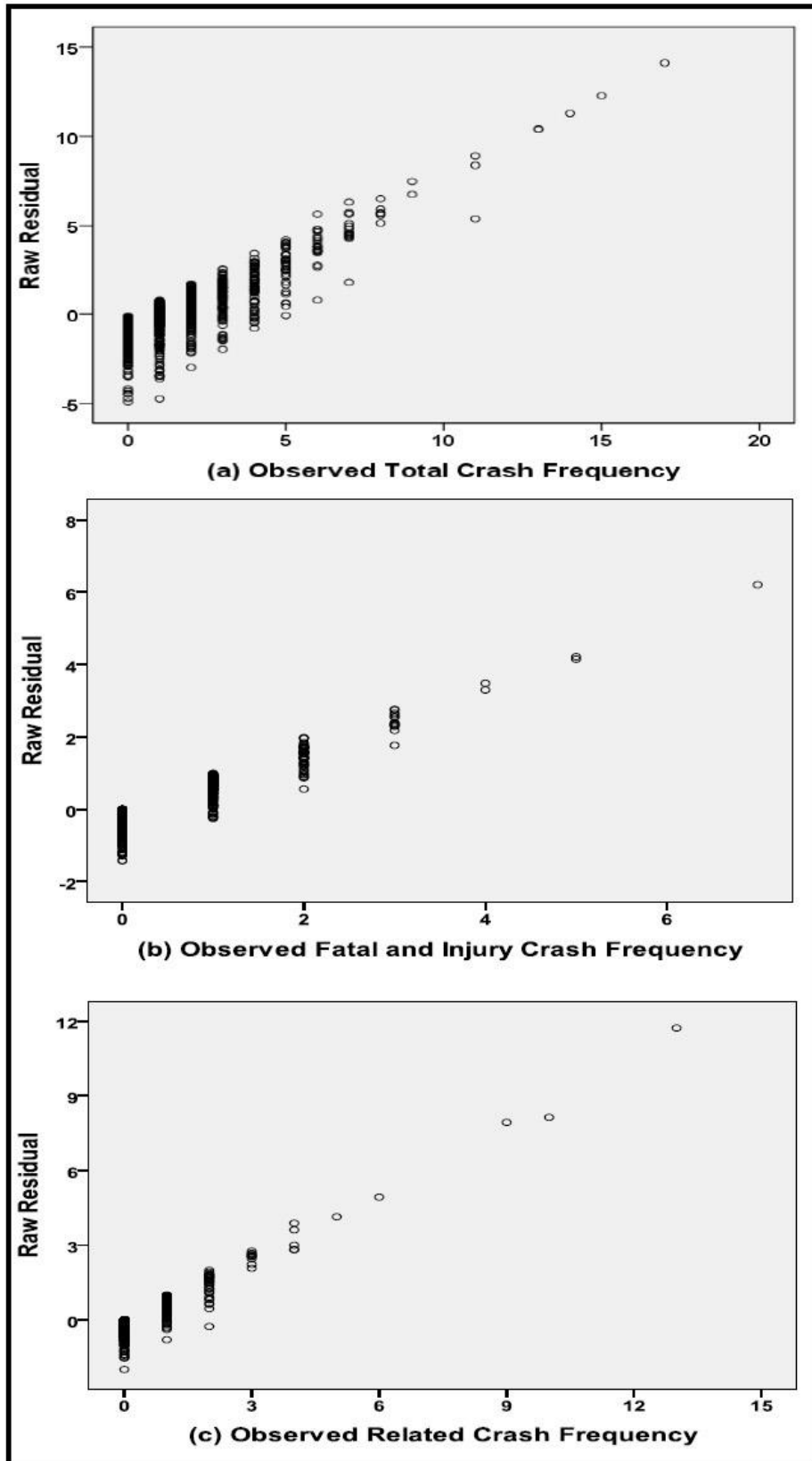


FIGURE 6 Scatter plot of crashes and residuals.

As shown in the Figure 6, total crash SPF predicts the number of crashes well with small residuals when the observed total crash frequency is between zero and four, while the other two SPFs perform better when the relevant crash number is zero or one. Total crash SPF's better-performing area covers 96.8 percent of the total observations, and FI SPF and related crash SPF's better-performing areas cover 97.4 percent and 97.9 percent, respectively.

5.2 The Empirical Bayes Approach Results

Two runs were conducted for the EB before-after approach. One applied a relevant SPF for each of the three crash types. Considering that the treatment group was not represented well by the reference group with respect to FI crashes and related crashes, a second EB before-after study was conducted using only the total SPF. Hauer illustrated a tutorial on how to approach this method (31). It initially predicted the numbers of FI crashes and related crashes in the before period by multiplying the initial total crash predictions by the relevant proportion to total crashes. A similar procedure was then applied to estimate the expected crash numbers in the after period. In this case, only the SPF for total crash was used.

An aggregated analysis was performed for the entire treatment group, which came out with aggregated CMFs for the three types of crashes. Additionally, separated analyses were conducted on upgrading narrow unpaved shoulders (≤ 5 ft) and upgrading wide unpaved shoulders (> 5 ft) to composite shoulders. A limitation identified with the dataset was that most of the studied segments with wide unpaved shoulders were successive sections of US Highway 24. This small and concentrated sample data could skew the resulted CMFs for upgrading wide unpaved shoulders. As a result, these CMFs were not recommended for upgrading wide unpaved shoulders to composite shoulders, and the cross-sectional approach results, which came from more segment data, were used to help to investigate its safety effectiveness in Section 5.3.

For better comparison, a naïve before-after study on those treated segments was also conducted. The estimated CMFs are shown in Table 14.

TABLE 14 CMFs for Composite Shoulder from Initial EB Methods and Naïve Method

Methods		Total Crash (Standard Deviation)	FI Crash (Standard Deviation)	Related Crash (Standard Deviation)
Aggregated analysis	EB with separated SPFs	1.114 (0.129)	0.930 (0.265)	0.639(0.237)
	EB with a same SPF	-- ^A	0.958 (0.251)	0.708 (0.224)
	Naïve B-A study	1.211 (0.174)	0.988 (0.355)	0.738 (0.308)
Shoulder widths ≤5 ft (1.5 m) in the before periods	EB with separated SPFs	0.861(0.145)	0.651(0.230)	0.335(0.159)
	EB with a same SPF	--	0.733(0.250)	0.442(0.205)
	Naïve B-A study	1.019(0.210)	0.643(0.274)	0.393(0.204)
Shoulder widths >5 ft (1.5 m) in the before periods	EB with separated SPFs	1.420(0.224)	-- ^B	-- ^B
	EB with a same SPF	--	--	--
	Naïve B-A study	1.381(0.269)	--	--

^A the total crash CMFs did not change since total crash SPFs were still used.

^B the number of crashes is not large enough to provide a valid CMF.

Only 3 FI crashes and 2 related crashes were observed among all segments with shoulders wider than 5 ft in the before period. This resulted in CMFs with large standard errors and these CMFs were considered to be invalid. The results were consistent across these three estimation methods; the EB approach's results had less estimation errors compared to the naïve before-after study; it also provided a more favorable result for the upgrade of shoulders from no or unpaved shoulders to composite shoulders. Upgrading narrow shoulders had the smallest CMFs, while upgrading wide shoulders to composite shoulders had CMFs greater than 1.00.

Differences were found between the two EB methods. While the EB approach with separated SPFs suffered from the reference group issue, with which the reference group can not

represent well the treatment group, the EB approach with the same SPF had difficulty in properly accounting for the treatment’s different effects on various crash impacts and severity types. For example, shoulder projects may impact the run-off-the-road crashes and the rear-end crashes differently. Thus, only one SPF was not able to fully indicate situations in all types of crashes. The author recommended averaging those two results to make an even better estimation which can take both considerations into account. The following two equations show how to combine these two CMFs, as well as their standard deviations (σ). The combine results are shown in Table 15.

$$CMF_c = \frac{CMF_1 + CMF_2}{2} \quad \text{Equation 26}$$

$$\sigma_c = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{2} \quad \text{Equation 27}$$

TABLE 15 CMFs for Composite Shoulder from Combined EB Method

	Total Crash (Standard Deviation)	FI Crash (Standard Deviation)	Related Crash (Standard Deviation)
Aggregated analysis	1.11 (0.129)	0.94 (0.183)	0.67 (0.163)
Narrow unpaved shoulders (≤ 5 ft)	0.86 (0.145)	0.69 (0.170)	0.39 (0.130)
Wide unpaved shoulders (> 5 ft)	1.42 (0.224) ^A	-- ^B	-- ^B

^A Base on a limited dataset, see discussion below.

^B The number of crashes is not large enough to provide a valid CMF.

According to the results of the EB approach, upgrading narrow shoulders to composite shoulders has significant safety effect on shoulder related crashes, with a CMF of 0.39. It also has considerable effects in FI crashes, with a CMF of 0.69. Overall, the total crash CMF for this treatment was estimated to be 0.86.

Determining a definitive CMF for shoulder widths greater than five ft was not possible with the given dataset. While the results indicated a CMF of 1.42, limited data were available and most of the studied segments with wide unpaved shoulders were successive sections of US Highway 24. This small and concentrated sample data could have skewed the results. More robust results can be provided by either additional data from neighboring states, or applying alternative methodologies that do not require before-after crash data such as cross sectional studies and case control studies.

5.3 The Cross-Sectional Approach Results

The development of regression models in the cross-sectional approach was similar to the development of SPFs in the EB approach. However, they had different list of variables. The cross-sectional models did not have *treatment group* and *shoulder size* variables. New categorical variables were created to indicate the shoulder type of segments. They included *narrow shoulders*, *wide shoulders*, and *composite shoulders*. AIC was also used to determine which subsets of variables provided the best fit models. It was found that all three final regression models shared the same independent variables including: *Ln (AADT)*, *Ln (Length)*, *Lane width*, and *Shoulder types (narrow, wide, and composite shoulders)*. Table 16 shows resulting models for the three types of crashes.

TABLE 16 The Cross-Sectional Approach Models Summary

Variable	Total Crashes		FI Crashes		Related Crashes	
	Coefficient (Std. Error)	Wald Chi-Square	Coefficient (Std. Error)	Wald Chi-Square	Coefficient (Std. Error)	Wald Chi-Square
<i>Intercept</i>	-2.84 (0.83)	11.58*** ^A	-6.14 (1.56)	15.43***	-5.91 (1.86)	10.14***
<i>Ln AADT</i>	0.73 (0.04)	331.87***	1.03 (0.08)	161.47***	1.23 (0.10)	143.28***
<i>Ln Length</i>	1.07 (0.04)	723.27***	1.24 (0.09)	196.10***	1.26 (0.11)	126.60***
<i>Lane width</i>	-0.21 (0.07)	9.62**	-0.24 (0.13)	3.60*	-0.38 (0.15)	6.71**
<i>CS</i> ^B	-0.39 (0.07)	29.79***	-1.14 (0.14)	65.81***	-1.56 (0.18)	77.88***
<i>WS</i> ^B	-0.33 (0.07)	23.82***	-1.03(0.14)	52.03***	-1.52 (0.19)	67.00***
<i>NS</i> ^B	0 ^C		0		0	
k	0.50 (0.04)		0.56(0.14)		1.45 (0.25)	
Log-likelihood ratio Chi-Square	1003.19***		265.53***		298.25***	
AIC	6818.773		1560.39		2011.53	
Pseudo R Square	0.65		0.75		0.73	

^A * indicates statistically significant at the 0.1 level;

** indicates statistically significant at the 0.05 level; and

*** indicates statistically significant at the 0.01 level;

^B CS = Composite Shoulder, NS = Narrow Unpaved Shoulder, WS = Wide Unpaved Shoulder.

^C Set to zero because this variable is the reference variable.

Table 16 shows the estimated overdispersion parameters and test statistics. With this information, $\exp(\beta_j)$ was applied to calculate CMFs for upgrading narrow shoulders to wide shoulders or composite shoulders. For example, the coefficient of wide shoulder is -0.33. As a result, total crash CMF for widening narrow shoulders to wide shoulders can be calculated as follows:

$$\text{CMF} = \exp(\beta_{WS}) = \exp(-0.33) = 0.71$$

The 95 percent confidence interval can be determined with the coefficient and standard error (σ):

$$(\exp(\beta_{WS}-1.96\sigma), \exp(\beta_{WS}+1.96\sigma)) = (\exp(-0.47), \exp(-0.19)) = (0.63, 0.82)$$

Table 17 shows CMFs, as well as their 95 percent intervals, from the cross-sectional study.

TABLE 17 CMFs (95 Percent Confidence Interval) for Composite Shoulders and Wide Unpaved Shoulders from the Cross-Sectional Study

	Total Crashes	FI Crashes	Related Crashes
Wide Unpaved Shoulders	0.71 (0.63, 0.82)	0.35 (0.27, 0.47)	0.22 (0.15, 0.32)
Composite Shoulders	0.68 (0.59, 0.78)	0.32 (0.24, 0.42)	0.21 (0.15, 0.30)

Compared with narrow unpaved shoulders, composite shoulders can expect a 32 percent reduction in total crashes, with 68 percent and 79 percent in FI crashes and related crashes, respectively. Wide unpaved shoulders have slightly smaller safety effectiveness than composite shoulders.

The next step is to test whether there were significantly different safety effectiveness between composite shoulders and wide unpaved shoulders, compared with narrow unpaved shoulders. A new coefficient Θ was generated to conduct the test, and it indicated the difference between the coefficient of *composite shoulder* (CS) and the coefficient of *wide unpaved shoulder* (WS). The null hypothesis was stated as follows:

$$H_0: \Theta = \beta_{CS} - \beta_{WS} = 0$$

The regular cross-sectional model is shown in Equation 28.

$$\ln(n) = \alpha + \beta_1 \ln(AADT) + \beta_2 \ln(Length) + \beta_3 LaneWidth + \beta_{CS} CS + \beta_{WS} WS \quad \text{Equation 28}$$

Θ was introduced into the model by letting $\beta_{CS} = \Theta + \beta_{WS}$, resulting in a converted model format as Equation 29.

$$\ln(n) = \alpha + \beta_1 \ln(AADT) + \beta_2 \ln(Length) + \beta_3 LaneWidth + \Theta CS + \beta_{WS} (WS + CS) \quad \text{Equation 29}$$

The converted model was regressed for each type of crashes. Table 18 shows θ 's information in the new regression models.

TABLE 18 Regressed Results for θ

	Total Crashes	FI Crashes	Related Crashes
θ Value	-0.06	-0.11	-0.04
Standard Error	0.06	0.15	0.19
P Value	0.35	0.47	0.83

It was found that there was not enough significance to reject the hypothesis at the 0.05 level in all three types of crashes. No significantly different safety effectiveness can be expected between composite shoulders and wide shoulders in total crashes, FI crashes, and related crashes.

Consider a five-mile long segment with 12-ft lanes, the crash frequency per year can be expected based on its traffic volume using the regression models indicated in Table 16. Figures 7 to 9 indicate the relationship between traffic volume and the expected annual crashes on the segment.

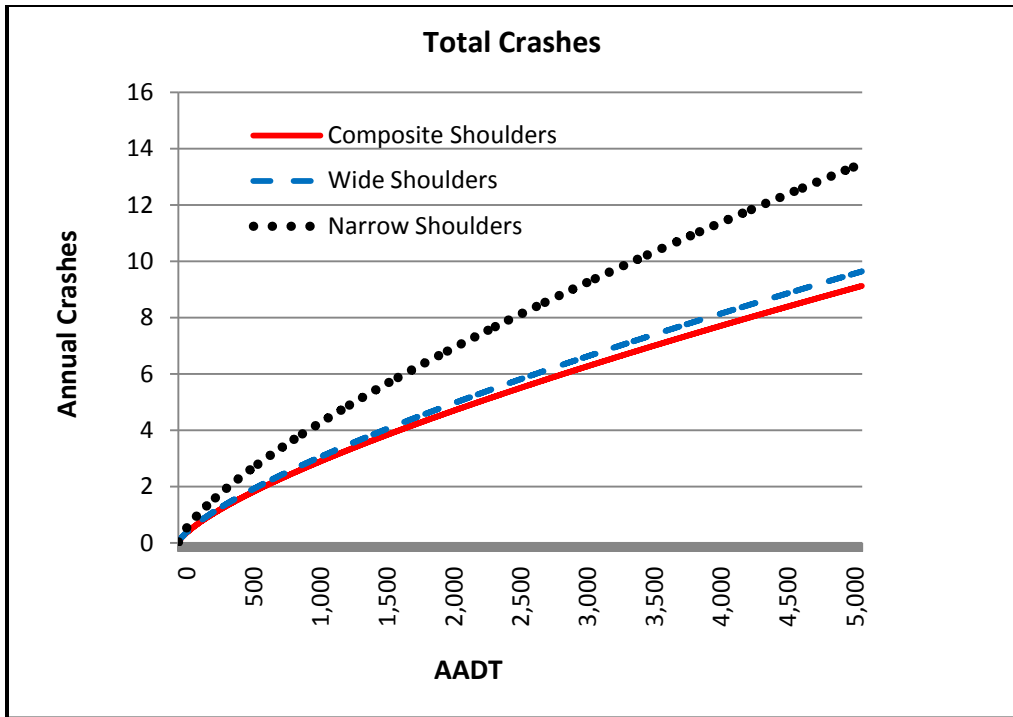


FIGURE 7 Annual crashes vs. AADT for total crashes.

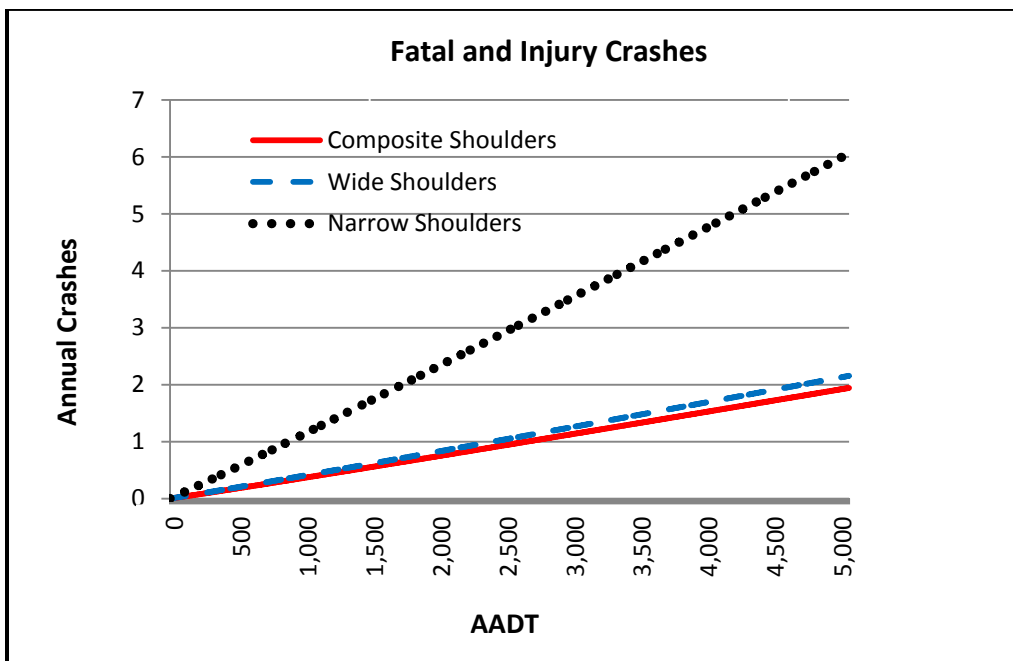


FIGURE 8 Annual crashes vs. AADT for FI crashes.

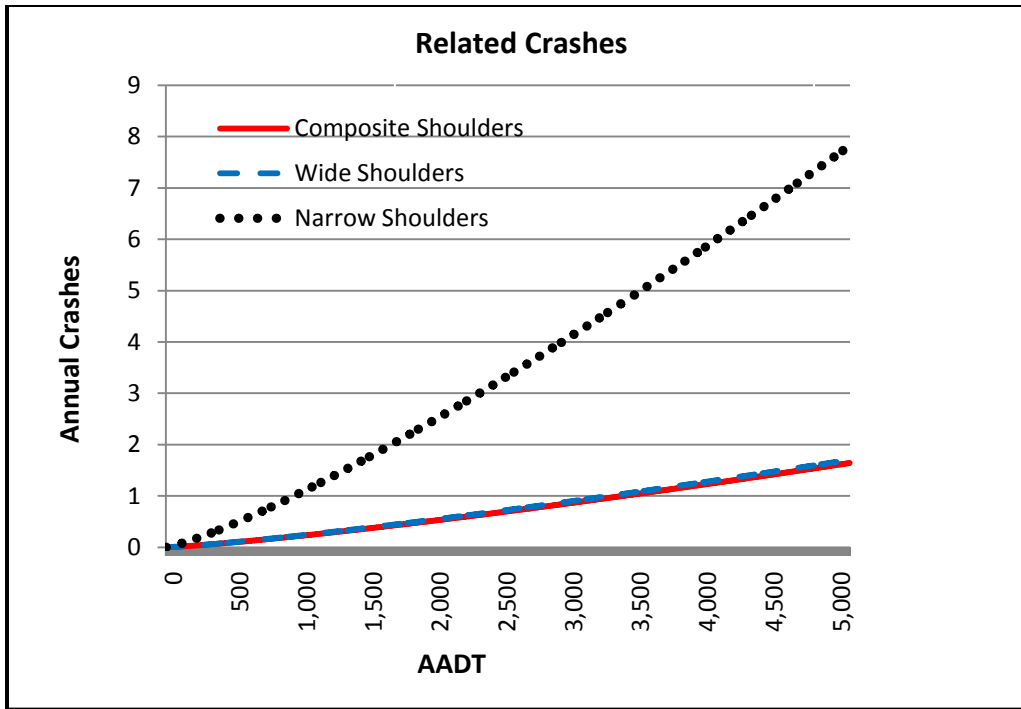


FIGURE 9 Annual crashes vs. AADT for related crashes.

As can be seen from Figures 7, 8, and 9, composite shoulders were expected to offer the most safety effectiveness, especially in respect to related crashes. Wide shoulders had similar results.

5.4 Summary

Two methods were used in the analysis section: the EB approach and the cross-sectional approach. Kansas-specific SPFs and CMFs for composite shoulders were generated through the EB approach.

According to results of the EB approach, upgrading narrow unpaved shoulders to composite shoulders has a significant safety effect on shoulder related crashes, with a CMF of 0.39. It also has considerable effects in FI crashes, with a CMF of 0.69. Overall, the total crash CMF for this treatment was estimated to be 0.86.

However, upgrading wide unpaved shoulders (> 5 ft) to composite shoulders may increase the overall number of expected crashes by 42 percent based on this research's data. Limited data were available and most of the studied segments with wide unpaved shoulders were successive sections of US Highway 24. This small and concentrated sample data could have skewed the results.

The cross-sectional approach also created a list of CMFs for composite shoulders. Compared with narrow unpaved shoulders, composite shoulders' CMFs were found to be 0.68, 0.32, and 0.21 for total crashes, FI crashes, and related crashes respectively. When compared with wide unpaved shoulders, no significantly different safety effectiveness was found in composite shoulders for all three types of crashes.

Regarding upgrading narrow unpaved shoulders to composite shoulders, findings from the EB approach and the cross-sectional approach were consistent among all three types of crashes. CMFs from the cross-sectional approach were found to have lower values. First, both predicted reductions in total crashes, FI crashes, and related crashes. In addition, they both indicated similar trends among different types of crashes: related crashes will experience the most reduction by percentage, and total crashes will have the least reduction by percentage.

The two results differed from each other regarding upgrading wide shoulders to composite shoulders. The cross-sectional approach results showed advantages for the reason of adequate sample size, though before-after data were not included.

In summary, the EB approach results were recommended to indicate the safety effectiveness of upgrading narrow unpaved shoulders to composite shoulders because the EB approach has been proved to be more accurate than the cross-sectional approach when there are

adequate data. However, the cross-sectional approach was also a useful alternative method when there were only limited data for conducting a proper EB analysis to investigate the safety effectiveness of converting wide unpaved shoulders to composite shoulders. It also showed its value in comparing the safety effectiveness between composite shoulders and wide unpaved shoulders.

CHAPTER 6 COMPOSITE SHOULDER SAFETY ECONOMIC BENEFIT ANALYSIS

6.1 Crash Avoided Prediction

About 0.6 crash per mile were found to occur annually in the reference group segments between 2007 and 2009. FI crashes and related crashes contribute about 18.3 percent and 16.1 percent of the total crashes, respectively. Combined with the CMF results, this information can be used to predict crashes.

As an example, if one assumes that 10 miles of turf shoulders (≤ 5 ft) will be upgraded to composite shoulders each year for the next 10 years, 20-year projections were developed estimating the benefits in reductions that could be achieved through implementing these safety improvements. By the end of the tenth year, it was estimated that 8.4 total crashes, 3.4 FI crashes, and 5.9 related crashes could be avoided per year.

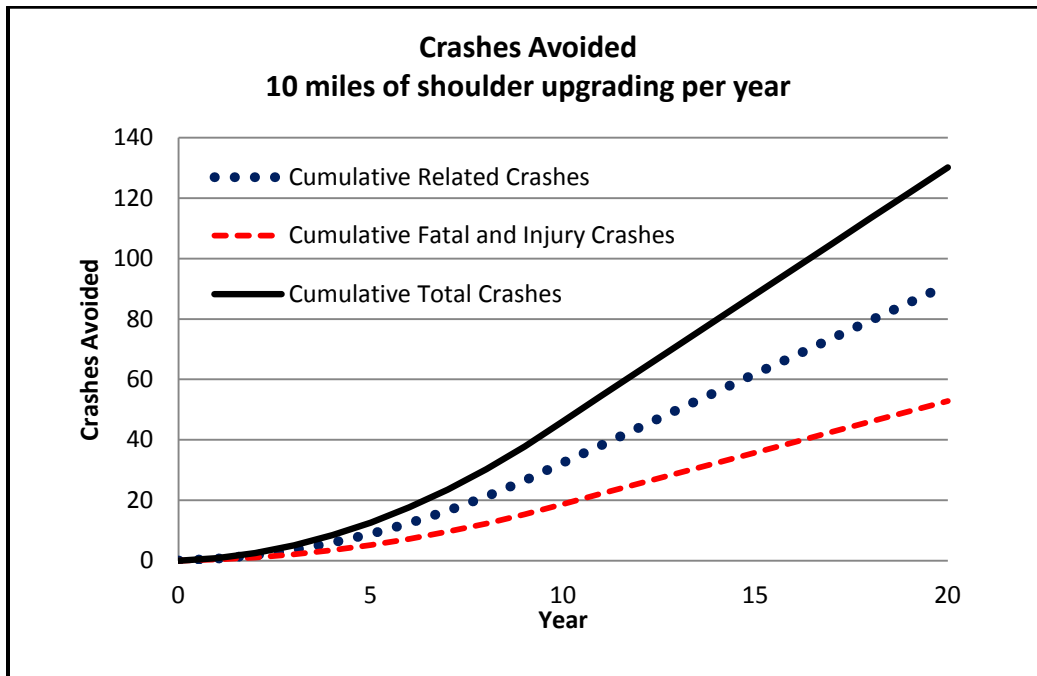


FIGURE 10 Cumulative crashes avoided from upgrading 10 miles of shoulder per year for 10 years.

As shown in Figure 10, by the end of the twentieth year, the cumulative total crashes, FI crashes, and related crashes avoided was estimated to reach 130, 53 and 91, respectively.

6.2 CMFs Used to Calculate Crash Cost Avoided

As state budgets are tightened, there is pressure to show that proposed safety improvements are effective in achieving safety goals for a minimum cost. Using the estimated crashes avoided or increased from this research, it is possible to estimate the crash cost that can be avoided. In order to complete such an analysis, it required estimating the cost of an individual fatal, injury, and property damage only (PDO) crashes (32). In the case of this research, upgrading narrow unpaved shoulders to composite shoulders was applied as an example. Table 18 summarizes the benefit-cost analysis from composite shoulder project.

TABLE 19 Summary of Estimated Crash Cost Avoided from Composite Shoulder Project.

Crash Type	Crashes Avoided in 20 Years	Estimated Crash Cost Avoided	Total
Fatal	4.3 ^A	\$20,556,600 ^{B,C}	
Incapacitating Injury	8.0	\$ 2,447,500	
Nondisabled Injury	25.2	\$ 1,981,100	\$25,748,000
Possible Injury	15.5	\$ 578,000	
PDO	77	\$ 184,800	

^A Distribution of FI crashes on rural two-lane roads in Kansas (33):

Fatality	8.2%	Incapacitating injury	15.0%
Nondisabled injury	47.6%	Possible injury	29.2%

^B Average comprehensive crash cost by injury severity (32):

Fatality	\$4,300,000 per person
Incapacitating injury	\$ 216,800 per person
Nonincapacitating injury	\$ 55,300 per person
Possible injury	\$ 26,300 per person
PDO	\$ 2,400 per crash

^C Average persons involved in computation of Kansas costs per crash type (34):

Fatality	1.1 persons involved
Injury	1.42 persons involved
PDO	N/A

It was found that approximately 26 million dollars crash cost can be avoided by this project. This avoided cost can be identified as highway shoulders' incremental benefit besides their benefits on highway maintenance and operation. With these estimates, state highway agencies are better able to articulate to legislators or other budgetary authorities the estimated benefits for constructing such improvements.

Alternatively, if provided with a limited budget where not all desired safety improvements can be built, it is possible to rank alternative improvements and select on the basis of the more economical benefit. Incorporating direct economic benefits or congestion reduction benefits into an analysis could also be used to determine a more accurate benefit-cost analysis of any proposed programs, but such a concept falls beyond the scope of this research.

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Discussion and Conclusions

As transportation engineers, planners, and designers become increasingly interested in quantitative safety evaluations, CMFs have become essential safety indicators for highway safety countermeasures. The research of CMF allows transportation practitioners to better understand the trade-offs of safety versus cost.

Kansas has 8,300 miles of rural two-lane roads. It was found that one-third include composite shoulders. The most typical types of composite shoulders consist of the first three-ft section being paved and the remaining section unpaved. This research employed the EB approach and the cross-sectional approach to estimate the safety effectiveness of the most typical types of composite shoulders.

Both methods demonstrate that upgrading narrow unpaved shoulders (≤ 5 ft) to composite shoulders is an effective countermeasure to potentially reduce crashes on rural two-lane highways in Kansas. This treatment allows the most reduction by percentage in shoulder related crashes, as well as significant reduction in fatal and injury crashes. The estimated CMFs from the EB approach were 0.86, 0.69, and 0.39 for total crashes, FI crashes, and related crashes, respectively. Results were consistent with other studies related to shoulder widening or paving projects (1, 4, 8). It can be concluded that the upgraded shoulders are safer both due to being wider and to being paved. However, it is hard to indicate individual contributions respectively only based on this study.

In Kansas, the current practice of upgrading wide unpaved shoulders (> 5 ft) usually does not change the shoulder size, and only paves the first three ft of the shoulders. According to the

cross-sectional approach results, no significant safety improvements were provided by this treatment. However, composite shoulders still have advantages in highway maintenance and operation over unpaved wide shoulders.

It is necessary to examine the reference group issue stated previously when developing SPFs in the EB approach. In this thesis, the treatment group was found to not be represented well by the reference group with respect to FI crashes and related crashes. Results from the EB analysis with separated SPFs and the EB analysis with only total crash SPFs were combined to make the final estimates: one used separate SPFs for each type of studied crash, and the other one used the same SPF for all crash types, which performed similarly in both the reference and treatment groups. The combined EB approach resulted in better estimates with lower standard errors.

7.2 Contributions and Recommendations

This research has shown the value in determining local SPFs and CMFs for use in estimating the safety benefits of proposed geometric improvement programs using the EB approach. The methodology improved upon national SPFs and CMFs and provided more realistic values for local agencies. It is recommended that the local agencies apply these SPFs and CMFs to the Crash Prediction Models to provide more accurate safety estimations on rural two-lane roads in Kansas.

This research also shows examples of how CMFs can be used to conduct a benefit-cost analysis. This procedure can produce data driven insight into the safety benefits and limitations of composite shoulders on rural two-lane highways. It is expected that the result will be valuable as KDOT and other state highway agencies considers various options and their benefit-cost ratios for the investment of state maintenance funds. It is expected that this research will also aid

engineers and decision makers in continuing to build a safe and efficient highway systems in Kansas.

7.3 Limitations and Future Studies

An identified limitation with this research is that limited data were available on unpaved shoulders upgraded to composite shoulders. Approximately 2,000 miles of roadway have the studied types of composite shoulders prior to 2000, while only about 100 miles were upgraded during the studied period. As a result, only a small size of sample data was eligible for the EB study, especially for estimating the safety effects for upgrading wide unpaved shoulders. This is also the reason why the widths of shoulders were separated by narrow and wide, not by every incremental foot. To improve this research, one suggestion is to extend the study period, from 1990 to 2009. Using an extended period, there would be more potential roadways with this treatment. However, more variance will be introduced with respect to crash features, driving behaviors, and other factors. Another suggestion is to apply additional data from neighboring states. Many previous safety studies were based on data from more than one state, which can be seen in Section 2.1. A possible obstacle is with data collection, as different states may have varying database systems.

Another limitation is the variables in the development of SPFs and other regression models. This research only included variables that were believed to be most essential to safety. However, many other factors were believed to have impacts in crash risk. These factors include the existence of edge line, whether the shoulder has the safety edge treatment, lighting, crash time, heavy vehicle percentage, the existence of rumble strip, roadside hazard rating, etc. In practice, it is impossible to include all crash risk factors into the model. However, it will improve the research if more possible variables are included in future research.

REFERENCES

1. American Association of State Highway and Transportation Officials. *Highway Safety Manual*. AASHTO, Washington, D.C., 2010.
2. American Association of State Highway and Transportation Officials. *A Policy on Geometric Design of Highway/Street*. AASHTO, Washington, D.C., 2004.
3. Zegeer, C.V., R.C. Dean and J.G. Mayes. Effect of Lane and Shoulder Widths on Accident Reduction on Rural, Two-Lane Roads. In *Transportation Research Record: Journal of the Transportation Research Board, No. 806*, Transportation Research Board of the National Academies, Washington, D.C., 1981, pp. 33–43.
4. Zegeer, C.V., J. Hummer, D. Reinfurt, L. Herf, and W. Hunter. Safety Effects of Cross-Section Design for Two-Lane Roads. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1195*, Transportation Research Board of the National Academies, Washington, D.C., 1988, pp. 20–32.
5. Örnek, E., and A. Drakopoulos. Analysis of Run-Off-Road Crashes in Relation to Roadway Features and Driver Behavior. Proceedings of the 2007 Mid-Continent Transportation Research Symposium, Ames, Iowa, 2007.
6. Gross, F. and P.P. Jovanis. Estimation of Safety Effectiveness of Changes in Shoulder Width with Case Control and Cohort Methods. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2019*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 237–245.
7. Gross, F., P.P. Jovanis, K. Eccles, and K. Chen. *Safety Evaluation of Lane and Shoulder Width Combinations on Rural, Two-Lane, Undivided Roads*. Publication FHWA-HRT-09-031. FHWA, U.S. Department of Transportation, 2009.
8. Stamatiadis, N., J. Sacksteder, W. Ruff, and D. Lord. *Impact of Shoulder Width and Median Width on Safety*. NCHRP Report 633. Transportation Research Board of the National Academies, Washington, D.C., 2009
9. Gross, F. and E.T. Donnell. Case-Control and Cross-Sectional Methods for Estimating Crash Modification Factors: Comparisons from Roadway Lighting and Lane and Shoulder Width Safety Effect Studies. *Journal of Safety Research*, Vol. 42, 2011, pp. 117-129.
10. Agent, K.R., and J.G. Pigman. *Accidents Involving Vehicles Parked on Shoulders of Limited Access Highways*. Report KTC-89-36. Kentucky Transportation Center, Lexington, KY, 1989.
11. *Highway Safety Manual Knowledge Base*. NCHRP 17-27, Chapter 3, pp. 26-34. http://www.cmfclearinghouse.org/collateral/HSM_knowledge_document.pdf. Accessed July 20, 2011.

12. Gross, F., B. Persaud, and C. Lyon. *A Guide to Developing Quality Crash Modification Factors*. Publication FHWA-SA-10-032. FHWA, U.S. Department of Transportation, 2010.
13. Hauer, E. *Observational Before-After Studies in Road Safety*. Pergamon Press, Elsevier Science Ltd., Oxford, England, 1997.
14. Zeeger, C.V. and J. Deacon. *Effect of Lane Width, Shoulder Width, and Shoulder Type on Highway Safety: A Synthesis of Prior Literature*, in *Relationship Between safety and Key Highway Features: A Synthesis of Prior Research*, Transportation Research Board of the National Academics: Washington, DC., 1986.
15. Miaou, S.-P. and H. Lum, Modeling Vehicle Accidents and Highway Geometric Design Relationships. in *Accident Analysis and Prevention*, Vol. 25, No. 6, 1993.
16. Miaou, S.-P., The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions. *Accident Analysis and Prevention*, Vol. 26, No. 4, 1994.
17. Mountain, L., B. Fawaz, and D. Jarrett, Accident Prediction Models for Roads with minor Junctions. in *Accident Analysis and Prevention*, Vol. 28, No. 6, 1996.
18. Persaud, B.N., Accident Prediction Models for Rural Roads. in *Canadian Journal of Civil Engineering*, Volume 21, 1993.
19. Vogt, J. and J.G. Bared, *Accident Models for Two-Lane Rural Roads: Segment and Intersections*. Report No. FHWA-RD-98-133. FHWA, U.S. Department of Transportation, 1998.
20. Harwood, D., K. Bauer, I. Potts, D. Tobric, K. Richard, E. Kohlman Rabbani, E. Hauer, and L. Elefteriadou. Safety Effectiveness of Intersection Left- and Right- Turn Lanes. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1840, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 131-139.
21. Persaud, B.N., R.A. Retting, and C.A. Lyon. Crash Reduction Following Installation of Centerline Rumble Strips on Rural Two-Lane Roads. *Accident Analysis and Prevention*, Vol. 36, No. 6, 2004, pp. 1073-1079.
22. Abdel-Aty, M., P.C. Devarasetty, and A. Pande. Safety Evaluation of Multilane Arterials in Florida. *Accident Analysis and Prevention*, Vol. 41, No. 4, 2009, pp. 777-788.
23. Srinivasan, R., J. Baek, D. Carter, B. Persaud, C. Lyon, C., Eccles, K., Gross, F., and N. Lefler. *Safety Evaluation of Improved Curve Delineation*. Publication FHWA-HRT-09-045. FHWA, U.S. Department of Transportation, 2009.
24. Harwood, D.W., F.M. Council, E. Hauer, W.E. Hughes, and A. Vogt. *Prediction of the Expected Safety Performance of Rural Two-Lane Highways*. Publication FHWA-RD-99-207. FHWA, U.S. Department of Transportation, 2000.

25. Persaud, B., and C. A. Lyon. Empirical Bayes Before-After Safety Studies: Lessons Learned from Two Decades of Experience and Future Directions. *Accident Analysis and Prevention*, Vol. 39, No. 3, 2007, pp. 546-555.
26. Noland, R.B., M. A. Quddus, and W. Y. Ochieng. The Effect of the Congestion Charge on Traffic Casualties in London: an Intervention Analysis. Presented at 85th Annual meeting of the Transportation Research Board, Washington, D.C., 2006.
27. Quddus, M.A. Time Series Count Data Models: An Empirical Application to Traffic Accidents. *Accident Analysis and Prevention*, Vol. 40, No. 5, 2008, pp. 1732-1741.
28. Miaou, S.P. *Measuring the Goodness-of-Fit of Accident Prediction Models*. Publication FHWA-RD-96-040. FHWA, U.S. Department of Transportation, 1996.
29. Abdel-Aty, M.A., and A. E. Radwan. Modeling Traffic Accident Occurrence and Involvement. *Accident Analysis and Prevention*, Vol. 32, No. 5, 2000, pp. 633-642.
30. Fridstrom, L., J. Ifver, S. Ingebrigtsen, R. Kulmala, and L. K. Thomsen. Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident Counts. *Accident Analysis and Prevention*, Vol. 27, No. 1, 1995, pp. 1-20.
31. Wooldridge, J. M. *Introductory Econometrics: A Modern Approach, Fourth Edition*, South-Western Cengage Learning, Inc., Mason, OH, 2009
32. Hauer, E., D. W. Harwood, F. M. Council, and M. S. Griffith. Estimating Safety by the Empirical Bayes Method: A Tutorial. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1784*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 126-131.
33. National Safety Council.
http://www.nsc.org/NEWS_RESOURCES/INJURY_AND_DEATH_STATISTICS/Pages/EstimatingtheCostsofUnintentionalInjuries.aspx, Accessed July 20, 2011.
34. Lubliner, H. *Evaluation of the Highway Safety Manual Crash Prediction Model for Rural Two-lane Highway segments in Kansas*. The University of Kansas, Lawrence, Kansas, Dissertation, 2011.
35. 2008 Kansas Traffic Safety Facts Book.
<http://www.ksdot.org/burtransplan/prodinfo/acstat/2008factsbook.pdf>. Accessed July 20, 2011.

APPENDIX A – SPSS Code for Developing Negative Binominal Regression Models

```
GENLIN Total WITH ReG LogADT LaneWidth FSShou LogTC AADT
/MODEL LogADT FSShou LaneWidth INTERCEPT=YES OFFSET =LogL
DISTRIBUTION=NEGBIN(MLE) LINK=LOG
/CRITERIA METHOD=NEWTON SCALE=1 COVB=MODEL MAXITERATIONS=100
MAXSTEPHALVING=30 PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012
ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FULL
/MISSING CLASSMISSING=EXCLUDE
/PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION CORB
/SAVE RESID (R).
```

APPENDIX B - Matlab Code for Matching Data

```
[pickeddata, pickedtext, pickedall] =  
xlsread('\people.soecs.ku.edu\ehhuizeng\Home\Desktop\2011 Fall\Matlab road segments.xlsx');  
  
[pooldata, pooltext, poolall] =  
xlsread('\people.soecs.ku.edu\ehhuizeng\Home\Desktop\Composite Shoulder\RTLRCrashes2.xlsx','2000');  
  
n=numel(poolall)/10-1;  
crash=zeros(70,3);  
  
for e=1:70;  
    t=pickedtext(e+1,2);  
    i=pickeddata(e,3);  
    j=pickeddata(e,4);  
    for ep=1:n;  
        TF=strcmp(t,pooltext(ep+1,3));  
        if TF==1,  
            mile=pooldata(ep, 1);  
            if mile>=i && mile<=j,  
                crash(e,1)=crash(e,1)+1;  
                if pooldata(ep,4)~=0 || pooldata(ep,5)~=0,  
                    crash(e,2)=crash(e,2)+1;  
                end  
            end  
        end  
    end  
end;  
  
xlswrite ('crash data', crash).
```


APPENDIX C- Sample Observations

CANSYS_ROUTE	BEG_CMP	END_CMP	LGTH	Shoulder Size	Lane Width	Year	Total	FI	<u>Rel</u>	AADT
001K0020200-EB	0.000	0.698	0.698	0.000	8.500	2003	0.000	0.000	0.000	195
001K0020200-EB	0.000	0.698	0.698	0.000	8.500	2004	1.000	1.000	0.000	170
001K0020200-EB	0.000	0.698	0.698	0.000	8.500	2005	0.000	0.000	0.000	165
001K0020200-EB	0.000	0.698	0.698	0.000	8.500	2006	0.000	0.000	0.000	230
001K0020200-EB	0.000	0.698	0.698	0.000	8.500	2007	0.000	0.000	0.000	230
001K0020200-EB	0.000	0.698	0.698	0.000	8.500	2008	0.000	0.000	0.000	170
001K0020200-EB	0.000	0.698	0.698	0.000	8.500	2009	0.000	0.000	0.000	170
001K0022400-EB	0.000	1.039	1.039	1.000	12.000	2000	1.000	0.000	0.000	975
001K0022400-EB	0.000	1.039	1.039	1.000	12.000	2001	1.000	0.000	0.000	955
001K0022400-EB	0.000	1.039	1.039	1.000	12.000	2002	3.000	0.000	0.000	1240