Calibration of Highway Safety Manual Work Zone Crash Modification Factors

Final Report
June 2014

SWZDI
Smart Work Zone Deployment Initiative

Mizzou
University of Missouri - Columbia

Sponsored by
Smart Work Zone Deployment Initiative
(TPF-5(081))
Federal Highway Administration
(InTrans Project 06-277)

About SWZDI

Iowa, Kansas, Missouri, and Nebraska created the Midwest States Smart Work Zone Deployment Initiative (SWZDI) in 1999 and Wisconsin joined in 2001. Through this pooled-fund study, researchers investigate better ways of controlling traffic through work zones. Their goal is to improve the safety and efficiency of traffic operations and highway work.

About InTrans

The mission of the Institute for Transportation (InTrans) at Iowa State University is to develop and implement innovative methods, materials, and technologies for improving transportation efficiency, safety, reliability, and sustainability while improving the learning environment of students, faculty, and staff in transportation-related fields.

Disclaimer Notice

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The opinions, findings and conclusions expressed in this publication are those of the authors and not necessarily those of the sponsors.

The sponsors assume no liability for the contents or use of the information contained in this document. This report does not constitute a standard, specification, or regulation.

The sponsors do not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.

Non-Discrimination Statement

Iowa State University does not discriminate on the basis of race, color, age, religion, national origin, sexual orientation, gender identity, genetic information, sex, marital status, disability, or status as a U.S. veteran. Inquiries can be directed to the Director of Equal Opportunity and Compliance, 3280 Beardshear Hall, (515) 294-7612.

Iowa Department of Transportation Statements

Federal and state laws prohibit employment and/or public accommodation discrimination on the basis of age, color, creed, disability, gender identity, national origin, pregnancy, race, religion, sex, sexual orientation or veteran's status. If you believe you have been discriminated against, please contact the Iowa Civil Rights Commission at 800-457-4416 or the Iowa Department of Transportation affirmative action officer. If you need accommodations because of a disability to access the Iowa Department of Transportation's services, contact the agency's affirmative action officer at 800-262-0003.

The preparation of this report was financed in part through funds provided by the Iowa Department of Transportation through its "Second Revised Agreement for the Management of Research Conducted by Iowa State University for the Iowa Department of Transportation" and its amendments.

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Iowa Department of Transportation or the U.S. Department of Transportation Federal Highway Administration.

Technical Report Documentation Page

1. Report No. InTrans Project 06-277	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Calibration of Highway Safety Manual Work Zone Crash Modification Factors		**5. Report Date** June 2014
		6. Performing Organization Code
7. Author(s) Sun, C., Edara, P., Brown, C., Zhu, Z. and Rahmani, R.		**8. Performing Organization Report No.** InTrans Project 06-277
9. Performing Organization Name and Address University of Missouri-Columbia Department of Civil & Environmental Engineering E 2509 Lafferre Hall Columbia, MO 65211		**10. Work Unit No. (TRAIS)**
		11. Contract or Grant No. 081MO
12. Sponsoring Organization Name and Address Smart Work Zone Deployment Initiative Iowa Department of Transportation 800 Lincoln Way Ames, Iowa 50010	Federal Highway Administration U.S. Department of Transportation 1200 New Jersey Avenue SE Washington, DC 20590	**13. Type of Report and Period Covered** Final Report
		14. Sponsoring Agency Code TPF-5(081)

15. Supplementary Notes

Visit www.intrans.iastate.edu/smartwz for color PDFs of this and other SWZDI research reports.

16. Abstract

The Highway Safety Manual is the national safety manual that provides quantitative methods for analyzing highway safety. The HSM presents crash modification factors related to work zone characteristics such as work zone duration and length. These crash modification factors were based on high-impact work zones in California. Therefore there was a need to use work zone and safety data from the Midwest to calibrate these crash modification factors for use in the Midwest. Almost 11,000 Missouri freeway work zones were analyzed to derive a representative and stratified sample of 162 work zones. The 162 work zones was more than four times the number of work zones used in the HSM. This dataset was used for modeling and testing crash modification factors applicable to the Midwest. The dataset contained work zones ranging from 0.76 mile to 9.24 miles and with durations from 16 days to 590 days. A combined fatal/injury/non-injury model produced a R2 fit of 0.9079 and a prediction slope of 0.963. The resulting crash modification factors of 1.01 for duration and 0.58 for length were smaller than the values in the HSM. Two practical application examples illustrate the use of the crash modification factors for comparing alternate work zone setups.

17. Key Words crash modification factors—highway safety—work zones		**18. Distribution Statement** No restrictions.	
19. Security Classification (of this report) Unclassified.	**20. Security Classification (of this page)** Unclassified.	**21. No. of Pages** 40	**22. Price** NA

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

CALIBRATION OF HIGHWAY SAFETY MANUAL WORK ZONE CRASH MODIFICATION FACTORS

Final Report
June 2014

Carlos Sun, Ph.D., P.E., J.D., Associate Professor
Praveen Edara, Ph.D., P.E., Associate Professor
Henry Brown, P.E., Research Engineer
Zhongyuan (Eric) Zhu, Graduate Research Assistant
Roozbeh Rahmani, Graduate Research Assistant

University of Missouri-Columbia
Department of Civil & Environmental Engineering
E 2509 Lafferre Hall,
csun@missouri.edu (573) 884-6330

Sponsored by the Midwest Smart Work Zone Deployment Initiative and
the Federal Highway Administration (FHWA) Pooled Fund Study TPF-5(081):
Iowa (lead state), Iowa, Kansas, Missouri, Nebraska, Wisconsin

Preparation of this report was financed in part through funds provided by the Iowa Department of Transportation through its Research Management Agreement with the Institute for Transportation (InTrans Project 06-277)

TABLE OF CONTENTS

ACKNOWLEDGMENTS vii

EXECUTIVE SUMMARY ix

1. INTRODUCTION 1

    1.1 Crash Modification Function and Factor Studies 2
    1.2 HSM Work Zone CMF 4

2. LITERATURE REVIEW 5

3. DATA 7

    3.1 Databases 7
    3.2 Sampling 7
    3.3 Data Fusion 10
    3.4 Data Descriptive Statistics 13

4. METHODOLOGY 15

5. MODEL ESTIMATION 17

    5.1 Modeling All Crashes (Model 1) 17
    5.2 Modeling Fatal/Injury Crashes (Model 2) 19
    5.3 Modeling Non-Injury Crashes (Model 3) 21

6. SAMPLE APPLICATIONS 23

    6.1 Work Zone Length Example 23
    6.2 Work Zone Duration Example 25

7. CONCLUSION 27

REFERENCES 29

LIST OF FIGURES

Figure 1.1 Assessing traffic and safety impacts for planning work zones......................................1
Figure 3.3.1 MoDOT TMS Accident Browser example...11
Figure 3.3.2 MoDOT TMS ARAN Viewer example 11
Figure 5.1.1 Model 1 versus observed number of crashes..18
Figure 5.1.2 HSM model versus observed number of crashes 19
Figure 5.2.1 Model 2 versus observed number frequencies 21
Figure 5.3.1 Model 3 versus observed number of crashes..22

LIST OF TABLES

Table ES1 Summary of Work Zone Data Characteristics ix
Table ES2 Comparison of SWZDI and HSM CMF x
Table 3.2.1 Work Zone Data Stratification...8
Table 3.3.2 Stratified Sampling Results 9
Table 3.2.3 Stratified Modeling Data 9
Table 3.2.4 Stratified Testing Data 10
Table 3.3.1 Examples of Work Zone Characteristics 12
Table 3.4.1 Descriptive Statistics of the Work Zone Sample (n=162) 13
Table 3.4.2 Crashes during Work Zone and Before Work Zone 14
Table 5.1.1 Combined Model for Injury and Non-Injury Crashes (Model 1) 17
Table 5.1.2 Comparison of SWZDI and HSM CMF..18
Table 5.2.1 Model for Fatal/Injury Crashes (Model 2)...20
Table 5.3.1 Model for Non-Injury Crashes (Model 3)..21
Table 6.1 Societal Costs of Crashes...25

ACKNOWLEDGMENTS

This research was conducted under the Smart Work Zone Deployment Initiative (SWZDI) and Federal Highway Administration (FHWA) Pooled Fund Study TPF-5(081), involving the following state departments of transportation:

- Iowa (lead state)
- Kansas
- Missouri
- Nebraska
- Wisconsin

The authors would like to thank the FHWA, the Iowa Department of Transportation (DOT), and the other pooled fund state partners for their financial support and technical assistance.

The authors gratefully acknowledge the contributions made by the following individuals:

- Calvin Fales, undergraduate research assistant, University of Missouri, for compiling and processing work zone related data
- Myrna Tucker, Missouri DOT (MoDOT), Transportation Management Systems, for assisting with the querying of work zone, traffic, and crash data
- Mike Curtit and John Miller, MoDOT, Traffic and Highway Safety, for guidance on MoDOT highway safety policies and procedures

EXECUTIVE SUMMARY

The Highway Safety Manual (HSM) is the national manual on highway safety just as the Highway Capacity Manual, the AASHTO Green Book, and the Manual of Uniform Traffic Control Devices are national manuals on capacity/level-of-service, geometric design, and traffic control devices. With the advent of the HSM, transportation engineers now have a tool for quantifying safety along with other impacts such as capacity and delay. However, the first edition of HSM has many gaps in terms of facility and geographic coverage. The tools presented for work zones are brief and based on limited research. This report describes the research conducted to improve the HSM work zone Crash Modification Factors (CMFs) by using data from the Midwest. Specifically, HSM models were calibrated using data from the Midwest, a larger sample size than the HSM was used, and separate models were designed for all, fatal/injury, and non-injury crashes.

There is a tremendous amount of effort required for performing work zone safety studies, because different data sources need to be cleaned and then fused together. Three types of data are required: work zone characteristics such as length, duration, location, and type; work zone traffic characteristics; and work zone crash characteristics. Thus the work zone CMFs presented in the HSM were developed using 36 work zones. Of the 10,973 Missouri freeway work zones that were analyzed in this report, only 536 were suitable for use in safety modeling. Stratified sampling was performed to obtain a sample of 162 work zones from the 536. Table ES1 presents the characteristics of the work zones such as length, duration, AADT and number of crashes. The table shows that the work zone data represented a wide variety of work zones.

Table ES1 Summary of Work Zone Data Characteristics

Work Zone Characteristic	Minimum	Maximum	Average
Length (miles)	0.76	187.7	9.24
Duration (days)	16	590	105
AADT (veh./day)	1,990	88,017	21,789
Pre WZ Crashes	0	293	16.74
During WZ Crashes	0	281	16.76

In modeling work zone Crash Modification Functions, a before-after study was conducted using data from both before and during the work zone deployment. Thus the same site was investigated with and without a work zone. As is common in safety modeling, a negative binomial model form was used for modeling work zone crash frequency. The models included variables such as AADT, duration, length, and injury. The resulting models performed well in terms of fit and prediction. The combined fatal/injury/non-injury model resulted in a R^2 fit of 0.9079 and a prediction slope of 0.963. The fatal/injury model resulted in a R^2 fit of 0.8814 and a prediction slope of 1.13. And the non-injury model resulted in a R^2 fit of 0.9062 and a prediction slope of 0.9371. Table ES2 shows how the new Smart Work Zone Deployment Initiative (SWZDI) CMFs are smaller than the HSM values. This is not surprising since the HSM model was based on high-impact work zones in California.

Table ES2 Comparison of SWZDI and HSM CMF

CMF	SWZDI	HSM
Duration	1.01	1.11
Length	0.58	0.67

Two illustrative examples are presented to show how this report can be used for assessing the safety of different work zone plans. One example shows how to compare the safety of two different work zone lengths. A second example shows how to compare different work zone durations.

1. INTRODUCTION

A safe and efficient road network plays an important role in the quality of life of individuals living in any society. To ensure that safety and efficiency goals of roadways are met, transportation agencies regularly carry out construction, rehabilitation, and maintenance activities. According to FHWA (2009), there are more than 3000 work zones during the peak construction season in the US and an estimated 12 billion vehicle miles traveled a year through active work zones. These activities often necessitate closure of travel lanes and/or shoulders or reduction of lane widths, resulting in a temporary reduction in roadway capacity. The changes to the geometrics, presence of workers, work equipment, and other factors contribute to the risk of crashes in work zones. There are over 40,000 injuries per year occurring at work zones with someone being injured in a work zone every 13 minutes (FHWA, 2009).

Transportation agencies in charge of planning and scheduling roadwork activities must take into account the impact of lane closures resulting from work zones. As shown in Figure 1.1, road user costs are computed by estimating the traffic and safety impacts of a work zone.

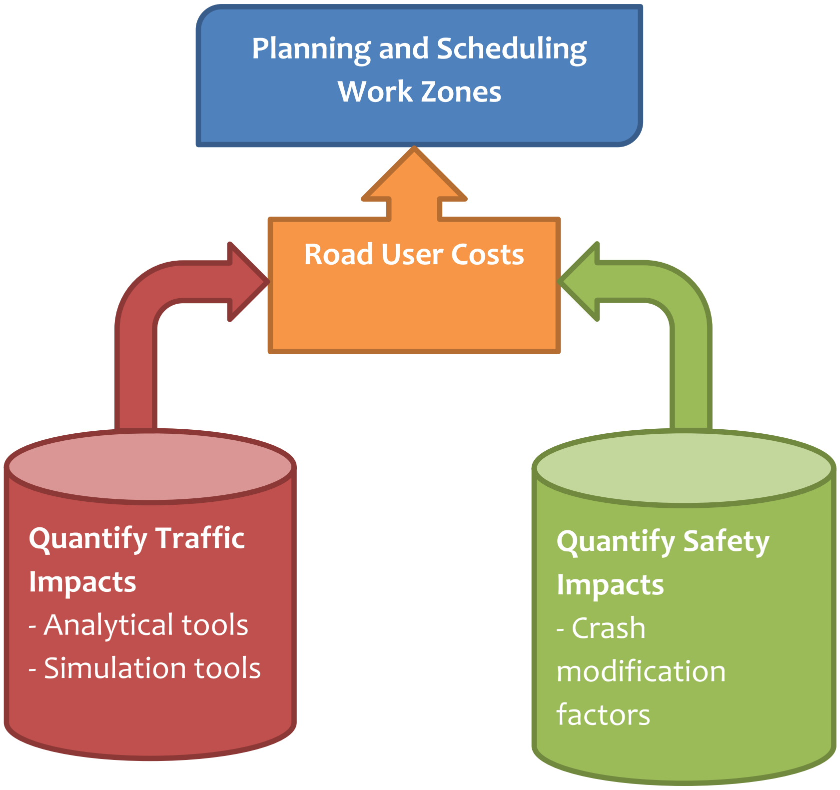


Figure 1.1 Assessing traffic and safety impacts for planning work zones

The traffic impact assessment has been studied in great detail in previous research. As a result, several tools are now available to practitioners to quantify traffic impacts. Analytical tools based on the Highway Capacity Manual (2010) procedures are recommended for sketch planning, especially for simple work zones. Simulation tools are recommended for complex work zones and to derive a wide range of performance measures. A discussion of the traffic impact analysis tools can be found in Edara (2009) and Edara et al. (2013).

Unlike traffic impact analysis, assessing safety impacts of work zones has not been well researched in the past. Very few studies exist on the topic of quantifying safety impacts of work zones. The Highway Safety Manual (HSM) synthesized previous research to provide practitioners a way to quantitatively evaluate highway safety, including work zone safety. Although the safety prediction models take into account a wide range of geometric and operational conditions, they do not cover all geographical differences. The crash data used to develop the models typically comes from a handful of states. For work zones, the HSM crash modification factors (CMFs) were developed using data from California only. CMFs will be defined formally and explained in detail in the next section, Section 1.1.

The key research question posed in this project is, are the HSM recommended work zone CMFs accurate enough for use in the Midwest? This question was answered by developing CMFs from crash data of work zones in Missouri and comparing them with the HSM CMF values. Although the crash data used was exclusively from Missouri, it is recommended that the other Smart Work Zone Deployment Initiative (SWZDI) states in the Midwest use the work zone CMFs developed in this study rather than the HSM values due to the similarities in the geographical and driving population characteristics within the Midwest. The developed CMFs help produce reliable crash frequency models and are an effective tool to assist traffic engineers, designers, and contractors in planning and scheduling work zones.

Several tasks were executed to accomplish this research project. This report presents the details of each of these tasks. The first task was to conduct a thorough review of previous literature on the safety assessment of work zones. Specifically, the review focused on studies that developed quantitative assessment methods. The second task was to extract the data needed to develop statistical models for predicting work zone crashes. As with any data-driven study, the second task involved very time intensive data sampling, fusion, and extraction. The processed data was then used to develop models using appropriate safety statistical models. Models were developed for three severity classifications: 1) all crashes, 2) injury only crashes, and 3) non-injury only crashes. The models for each classification provided the respective crash modification factors for work zones.

1.1 Crash Modification Function and Factor Studies

In evaluating the safety of work zones two related measures are typically used: crash modification factor (CMF) and crash modification function (CMFcn), also known as accident modification function. The former is a single point estimate, and is a multiplicative factor used to estimate the crash frequency after implementing changes to a site, such as implementing a work zone (FHWA 2010). The latter is a continuous function that varies the crash modification factor across a range of variables or combinations of variables (Elvik 2009). Thus CMFcn is a generalization of a single CMF across different characteristics. For example, CMFcn could involve variables such as the duration and length of work zone so that the crash frequency increases with longer durations or lengths.

The studies of CMF and CMFcn are divided into the two categories of experimental and observational (Carter et al. 2012). Experimental studies are planned studies, while observational studies use data that is collected retrospectively. Experimental studies are seldom, if ever, conducted for investigating work zone safety, since work zones are necessitated by actual road work. Also, there are ethical concerns with road safety experimentation involving actual drivers who do not give their consent.

One popular type of observational study is the before-after study. In a before-after study, the safety performance of a site before a treatment is applied is compared with the performance of the same site after the application of a treatment (FHWA 2010). The most straightforward before-after study is the naïve study which does not account for some changes unrelated to a treatment. This was the method used by Khattak and Council (2002) which produced the results documented in the HSM. Some changes that occur from year to year, such as traffic volume (AADT), were addressed by the method as AADT was a variable used in modeling. This is also the method presented in this report. More complicated before-after studies can correct for more confounding factors but also involve significantly more resources in terms of data requirements and labor. Confounding factors are variables that completely or partially account for the relationship between an outcome and a predictor variable.

One variation of the before-after study involves the use of a comparison group. The use of an untreated comparison group of sites similar to the treated ones helps to account for changes unrelated to the treatment such as changes in traffic patterns, land-use, and driver behavior. Another variation is the Empirical Bayes before-after study which uses additional data from reference sites with similar traffic and physical characteristics as the treated sites. Thus Empirical Bayes computes the number of expected crashes using both the observed before period and data from reference sites. This method has the advantage of correcting for regression-to-the-mean. Unfortunately, in work zone safety modeling, the data requirements are already so extensive for the naïve study, that the use of Empirical Bayes is very costly. The full Bayes approach is a more comprehensive approach to the Empirical Bayes and uses a distribution of likely crash frequency values instead of just a point estimate. As such, it is even more costly to implement than Empirical Bayes.

Another class of observational studies is various types of cross-section studies. Instead of observing the same site with and without treatment, cross-section studies rely on observing untreated sites that are similar to the treatment sites. The basic cross-sectional study uses data from a single point in time in contrast to before-after studies. Ensuring that all factors affecting crash risk are similar between the treated and untreated sites is a practical challenge in cross-sections studies. A related study type is the case-control study. Case-control studies also use cross-sectional data but differ in that sample sites are selected based on the outcome, i.e. whether a crash occurred or not. A cohort study assigns sites into specific cohorts based on the current treatment status, and then the cohorts are tracked over time. Generally, cross-section and related studies are not used as often as before-after studies in safety research (FHWA 2010).

1.2 HSM Work Zone CMF

Chapter 16 of HSM presents CMFcns related to work zones. Specifically, Section 16.4.2 presents CMFcns involving work zone design treatments such as work zone duration and length. These CMFcns were based on research performed by Khattak and Council (2002). The crash modification function involving work zone duration is

$$CMF_{d,all} = 1.0 + \frac{(\%\ \text{increase in duration}\ x\ 1.11)}{100}$$

where $CMF_{d,all}$ is the CMFcn for all crash severities as a function of duration (HSM, 2010). The crash modification function involving work zone length is

$$CMF_{l,all} = 1.0 + \frac{(\%\ \text{increase in duration}\ x\ 0.67)}{100}$$

where $CMF_{l,all}$ is the CMFcn for all crash severities as a function of length (HSM, 2010). The multipliers of 1.11 and 0.67 from the two CMFcns above were taken directly from Khattak and Council's (2002) negative binomial Model 1. Each CMFcn specifies a linear relationship between the duration or length and the CMF value. Subsequent sections of this report describe how these HSM CMFcns were re-derived using data from the Midwest. Chapter 16A.4 describes additional work zone traffic control and operational elements such as signs and signals, delineation, changeable speed warning signs, temporary speed limit signs and speed zones, innovative flagging procedures, changeable message signs, radar drones, and police speed enforcement. However, there are no CMFs presented for any of these elements.

2. LITERATURE REVIEW

While existing research is somewhat inconsistent on the impact of work zone presence on crash severity, most studies show that work zone presence has a negative impact on crash frequency. According to a recent review from Yang et.al (2014), 48% of previous studies on work zone crash severity indicate no clear evidence that there is an increase in crash severity during work zone conditions. On the other hand, the majority of previous studies regarding work zone crash frequency show an obvious increase in crash frequencies during work zone operations. Crash frequency is usually used as a safety evaluation measure for work zones and is expressed in the total number of crashes in a given time period.

Although there are many studies on crash frequency modeling, only a few of them focus on work zone presence. Pal and Sinha (1996) conducted a study on Indiana highway work zones and found that crash rates in work zones were significantly higher than non-work zone conditions. They developed two normal regression models to compare the predicting crash rate of different types of lane closure. Although normal regression model seemed to have better prediction power over the negative binomial and Poisson models, it produced negative crash rates in several cases. To ensure non-negative predicting results, researchers started using and fine tuning negative binomial models and Poisson models. Venugopal and Tarko (2000) developed two negative binomial models with duration of work, type of work, AADT and work zone length as main variables. The two models were calibrated for approaches to work zones and inside work zones separately. They also added cost of work to the model as an indicator of the intensity of work and showed AADT, work zone length and duration to be major safety related factors. Khattak and Council (2002) developed a negative binomial model using before-and-after data with a coefficient of 0.65 crashes per million vehicle kilometers without work zones and 0.79 crashes per million vehicle kilometers with work zones. Thus the models they developed showed higher crash tendency for work zones. Their findings were consistent with previous studies which suggested that a higher AADT, and a longer work zone duration and work zone length led to a higher crash rate. The current HSM CMF for work zone condition is derived from the aforementioned model by Khattak and Council (2002). To account for zero-crash work zones, researchers have suggested using zero-inflated negative binomial models. Although there were studies comparing zero-inflated negative binomial models with negative binomial models for crash frequency predicting modeling (Lord et. al 2005; Lord and Mannering 2010), no one has tested and compared zero-inflated negative binomial models with other models using work zone data. Qi et al.(2005) built a zero-inflated negative binomial model but did not compared it to the truncated negative binomial model in their study. Srinivasan et al. (2011) developed negative binomial safety performance functions for all crashes, injury crashes, and PDOs, and then used the empirical Bayes method to estimate different CMFs for daytime and nighttime work zones. Recently, Ozturk et. al (2013) developed a negative binomial-based model with further temporal adjusted daytime and nighttime traffic volumes and found that “work zone duration,” “length of work zone” and “traffic volumes” had the most impact on work zone safety. Chen and Tarko (2014) proposed a new fixed-parameter negative binomial model with random effects as an alternative to random parameters model, and obtained similar crash frequency prediction accuracies.

Since previous studies have shown reliable results on using the negative binomial model in work zone crash frequency modeling, this study also developed negative binomial models. Missouri data was used for model development.

3. DATA

3.1 Databases

Work zone crash data and segment characteristics were collected from 3 MoDOT databases: the work zone database, the crash database, and the segment database. Data fusion was used to merge the information contained in these three databases. Because of the complexities involved in this data fusion process, large sample sizes are not typically encountered in work zone crash modeling. For example, Khattak and Council (2002) used a sample size of 36 work zones in building the model used in the HSM.

The work zone database includes a unique work zone ID, a roadway segment ID, start and end date, time of the work, and start and end location. The crash database contained archived highway patrol reports. Even though there is a column in the crash reports indicating work zone presence, it was not relied upon because it is biased on a police officer's judgment. There are crashes that occurred in the work zones that are not reported as work zone-related crashes in crash reports. To account for advance warning areas, all crashes recorded within 0.5 mile before the beginning and 0.5 mile after the end of the work zone were classified as work zone crashes. The 0.5 mile threshold was the same threshold used by Khattak (2002) and used in the HSM.

As previous studies concluded, AADT values play an important role in crash frequency modeling. Most previous studies assumed the same AADT before and after work zone presence (Yang et.al 2014); however, this study uses the corresponding year's AADT from the segment database. The AADT for the entire work zone area is calculated as a weighted total of all the segments within the work zone.

3.2 Sampling

There were 10,973 freeway work zones in Missouri from January of 2010 to the end of 2012. Short work zones with a short duration usually have zero crashes and are less useful. So work zones shorter than 0.5 mile and with duration of less than 15 days were omitted. This minimum threshold is similar to the 0.51 mile and 16 days used by Khattak and Council (2002). Since crashes recorded within 0.5 mile before the beginning and 0.5 mile after the end of the work zone were classified as work zone crashes, work zones that had beginning log-miles of less than 0.5 were not used due to lack of non-Missouri crash data. After eliminating work zones according to the aforementioned criteria, the remaining work zones were 536.

Stratified random sampling refers to the grouping of population data into similar strata (Salkind 2006). Because the data points within a stratum are similar, the variance is less than the population as a whole. The variance of the sample estimator is the weighted sum of the within-stratum variances. Since the weighted sum variance is less than the variance with no stratification, this sampling technique is more precise than simple random sampling. Stratified random sampling is also more representative of the overall population, since the number of work zone samples is biased towards the shorter duration and length work zones. To perform stratified random sampling, the 536 work zones were categorized into strata with different lengths and durations. The number of work zones in each stratum is shown in Table 3.2.1. This sampling technique requires the strata to be mutually exclusive and collectively exhaustive.

Table 3.2.1 Work Zone Data Stratification

	Work zone duration (days)					
Work zone length (mile)	**<30**	**30-119**	**120-209**	**210-300**	**>300**	**Total**
<2	98	59	14	9	8	188
2-4	35	57	12	7	0	111
4-6	24	31	6	3	0	64
6-8	21	19	2	2	0	44
8-10	12	14	2	1	0	29
10-12	6	8	4	2	0	20
12-14	7	12	0	0	0	19
14-16	4	9	2	1	0	16
>16	16	27	0	2	0	45
Total	**223**	**236**	**42**	**27**	**8**	**536**

Some of the strata in Table 3.2.1 are empty which means there is no data for the stratum. As the models are supposed to represent all existing lengths and durations, a maximum number of 6 work zones were selected randomly from each stratum. Table 3.2.2 shows the number of randomly selected work zones in each stratum. Among the 162 work zones, 138 were used for making the model (Table 3.2.3) and the rest, 24, for testing the model (Table 3.2.4).

Table 3.3.2 Stratified Sampling Results

	Work zone duration (days)					
Work zone length (mile)	**<30**	**30-119**	**120-209**	**210-300**	**>300**	**Total**
<2	6	6	6	6	5	29
2-4	6	6	6	6	0	24
4-6	6	6	6	3	0	21
6-8	6	6	2	2	0	16
8-10	6	6	2	1	0	15
10-12	6	6	4	2	0	18
12-14	6	6	0	0	0	12
14-16	4	6	2	1	0	13
>16	6	6	0	2	0	14
Total	**52**	**54**	**28**	**23**	**5**	**162**

There were 162 freeway work zones that were randomly sampled from the work zone database. Of the samples, 138 were used for building models and the rest were used for model testing. It is common to set aside approximately 25% of the data for testing (Hastie et al. 2001).

Each work zone was paired with two “pre-work zone” periods from the previous two years at the same location. To account for seasonal variation, the “pre-work zone” period was chosen as the same months as when the work zone occurred. For example, if a work zone occurred from August 1st to August 7th, 2012, then the crash and traffic data were also collected from August 1st to August 7th, in 2010 and 2011 respectively, to create two “pre-work zone” samples. As a result, 414 samples were used to build the models (3 x 138) versus the 72 samples used in HSM.

Table 3.2.3 Stratified Modeling Data

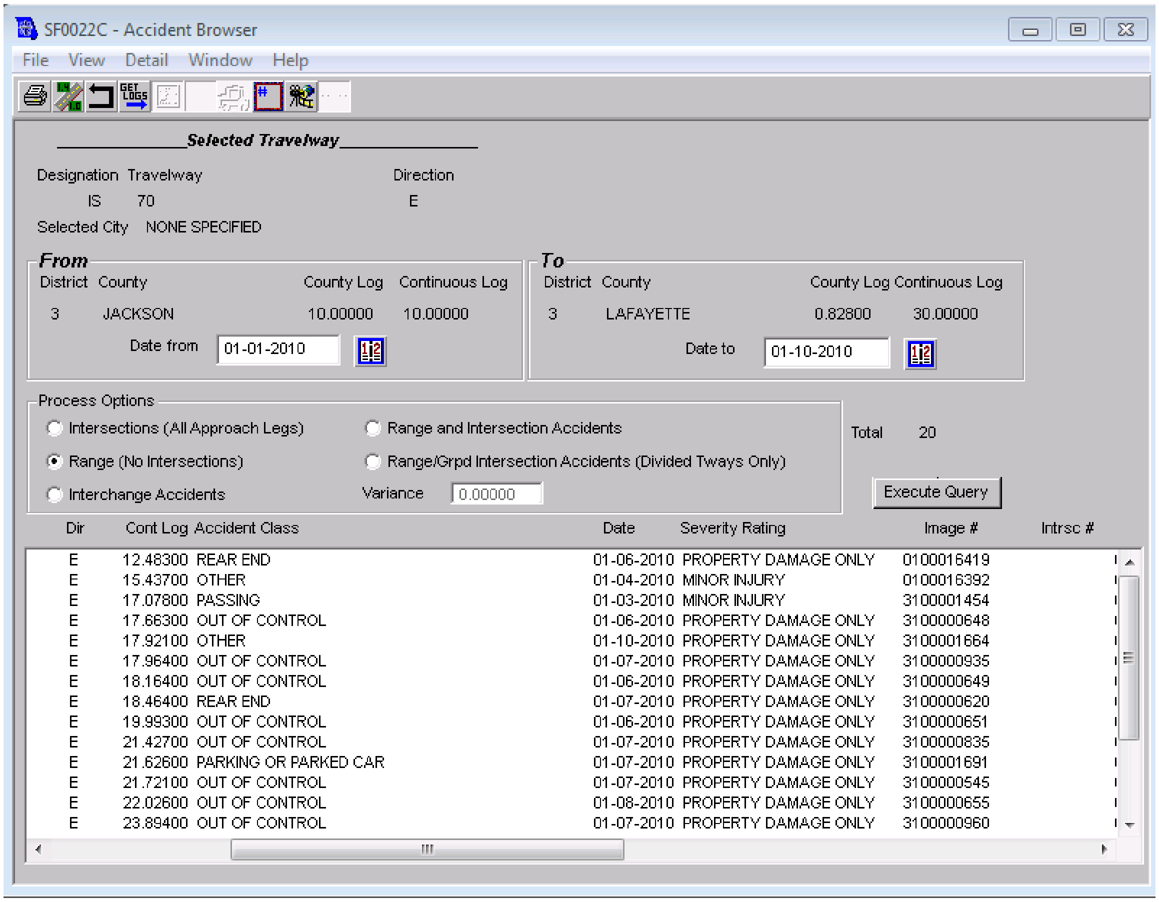
	Work zone duration (days)					
Work zone length (mile)	**<30**	**30-119**	**120-209**	**210-300**	**>300**	**Total**
<2	5	5	5	5	5	25
2-4	5	5	5	5	0	20
4-6	5	5	5	3	0	18
6-8	5	5	2	2	0	14
8-10	5	5	2	1	0	13
10-12	5	5	3	2	0	15
12-14	5	5	0	0	0	10
14-16	3	5	2	1	0	11
>16	5	5	0	2	0	12
Total	**43**	**45**	**24**	**21**	**5**	**138**

Table 3.2.4 Stratified Testing Data

	Work zone duration (days)				
Work zone length (mile)	**<30**	**30-119**	**120-209**	**210-300**	**Total**
<2	1	1	1	1	4
2-4	1	1	1	1	4
4-6	1	1	1	0	3
6-8	1	1	0	0	2
8-10	1	1	0	0	2
10-12	1	1	1	0	3
12-14	1	1	0	0	2
14-16	1	1	0	0	2
>16	1	1	0	0	2
Total	**9**	**9**	**4**	**2**	**24**

3.3 Data Fusion

The work zone length, duration, number of closed lane, and road type (urban or rural) were collected from MoDOT’s work zones database. After making the sample of 162 work zones, the data for each work zone, including AADT and number of crashes, were queried from MoDOT using an Open Database Connectivity. The AADT for a work zone location was computed by weighted averaging throughout the length of a freeway segment. The weighted average of AADT is more realistic than choosing the AADT from a single point on a segment. Injury and non-injury crashes with and without the presence of a work zone were collected from the MoDOT Transportation Management System (TMS) Accident Browser as shown in Figure 3.3.1. The number of lanes was recorded from the MoDOT TMS Automated Road Analyzer (ARAN) viewer as seen in Figure 3.3.2. The number of lanes and number of closed lanes were used in initial models, but the results showed that they did not improve the model’s prediction power. Table 3.3.1 shows the work zone characteristics for some of the samples used in modeling. Some of these examples have few or no crashes, since they have short lengths and durations.



Selected Travelway

Designation IS
Travelway 70
Direction E
Selected City NONE SPECIFIED

From

District	County	County Log	Continuous Log
3	JACKSON	10.00000	10.00000

Date from 01-01-2010

To

District	County	County Log	Continuous Log
3	LAFAYETTE	0.82800	30.00000

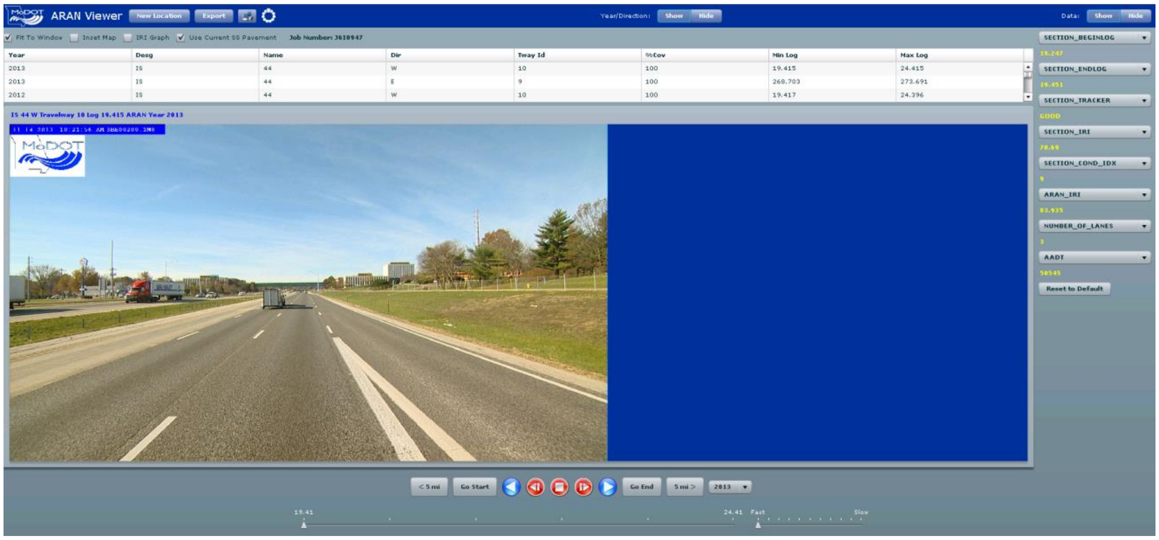
Date to 01-10-2010

Process Options
○ Intersections (All Approach Legs)
○ Range and Intersection Accidents
● Range (No Intersections)
○ Range/Grpd Intersection Accidents (Divided Tways Only)
○ Interchange Accidents
Variance 0.00000

Total 20

Dir	Cont Log	Accident Class	Date	Severity Rating	Image #	Intrsc #
E	12.48300	REAR END	01-06-2010	PROPERTY DAMAGE ONLY	0100016419	
E	15.43700	OTHER	01-04-2010	MINOR INJURY	0100016392	
E	17.07800	PASSING	01-03-2010	MINOR INJURY	3100001454	
E	17.66300	OUT OF CONTROL	01-06-2010	PROPERTY DAMAGE ONLY	3100000648	
E	17.92100	OTHER	01-10-2010	PROPERTY DAMAGE ONLY	3100001664	
E	17.96400	OUT OF CONTROL	01-07-2010	PROPERTY DAMAGE ONLY	3100000935	
E	18.16400	OUT OF CONTROL	01-06-2010	PROPERTY DAMAGE ONLY	3100000649	
E	18.46400	REAR END	01-07-2010	PROPERTY DAMAGE ONLY	3100000620	
E	19.99300	OUT OF CONTROL	01-06-2010	PROPERTY DAMAGE ONLY	3100000651	
E	21.42700	OUT OF CONTROL	01-07-2010	PROPERTY DAMAGE ONLY	3100000835	
E	21.62600	PARKING OR PARKED CAR	01-07-2010	PROPERTY DAMAGE ONLY	3100001691	
E	21.72100	OUT OF CONTROL	01-07-2010	PROPERTY DAMAGE ONLY	3100000545	
E	22.02600	OUT OF CONTROL	01-08-2010	PROPERTY DAMAGE ONLY	3100000655	
E	23.89400	OUT OF CONTROL	01-07-2010	PROPERTY DAMAGE ONLY	3100000960	

Figure 3.3.1 MoDOT TMS Accident Browser example



☑ Fit To Window ☐ Inset Map ☐ IRI Graph ☑ Use Current SS Pavement Job Number: J610947

Year	Desg	Name	Dir	Tway Id	%Cov	Min Log	Max Log
2013	IS	44	W	10	100	19.415	24.415
2013	IS	44	E	9	100	268.703	273.691
2012	IS	44	W	10	100	19.417	24.396

IS 44 W Travelway 10 Log 19.415 ARAN Year 2013

11 14 2013 10:21:56 AM 3BE00200.3M0

SECTION_BEGINLOG
19.247
SECTION_ENDLOG
19.451
SECTION_TRACKER
GOOD
SECTION_IRI
70.69
SECTION_COND_IDX
9
ARAN_IRI
83.935
NUMBER_OF_LANES
3
AADT
50545

Figure 3.3.2 MoDOT TMS ARAN Viewer example

Table 3.3.1 Examples of Work Zone Characteristics

work zone ID	Duration (days)	Length (miles)	urban/ rural	during work zone			without presence of work zone					
				AADT (vehicle/ day)	number of crashes		AADT (vehicle/ day)	number of crashes		AADT (vehicle/ day)	number of crashes	
					injury	non-injury		injury	non-injury		injury	non-injury
261880	17	1.502	0	13910	0	0	14883	0	2	14591	0	0
207528	25	0.999	0	19869.36	1	3	19479.85	2	1	20887.72	1	4
273479	29	0.858	1	67224.61	1	5	43024.63	1	7	43024.63	2	4
287337	19	0.88	0	13687	0	1	14537	0	0	14684	0	0
282216	16	0.894	1	37674	1	0	37674	1	2	52132	1	1
282232	16	3.038	0	19911	0	0	17686	0	0	15461	0	0
282309	17	3.274	0	13185.35	0	0	14186.19	0	0	14783.43	0	0
274227	16	2.647	0	3619	0	0	4967	0	0	6315	0	1
259231	25	3.092	1	23562.46	2	0	14834.16	0	0	14687.14	0	1
267594	21	3.128	1	28967.82	0	0	29416.26	0	0	29310.22	1	1
287585	25	4.847	1	55738.99	2	9	54042.54	2	7	54588.29	3	7
287584	25	4.878	1	56457.99	2	7	54357.43	2	8	54906.27	1	10
268674	20	5.332	1	85441.55	2	12	86305.89	5	9	87178.48	0	6

3.4 Data Descriptive Statistics

Table 3.4.1 shows the descriptive statistics of the sample used in this study. The sample covered work zone lengths ranging from 0.86 miles to 187.7 miles with an average of 9.47 miles. The weighted AADT through the work zone ranged from 1990 veh/day to 88017 veh/day with an average of 21691 veh/day. Work zone duration averaged at 105 days with the shortest being 16 day and the longest being 590 days. Short-term work zones which are under 16 days are not included in the sample because there were seldom any crashes that occurred within these work zones, which makes a model biased towards zero. A separate model could be built specifically for short-term work zones. Half of the work zones were classified as rural and the other half as urban to balance the sample.

Table 3.4.1 Descriptive Statistics of the Work Zone Sample (n=162)

Variable		Value
Length of work zone segments (mile)*	Average	9.24
	Min	0.76
	Max	187.74
AADT (vehicles per day)	Average	21789
	Min	1990
	Max	88017
Work Zone duration (days)	Average	105
	Min	16
	Max	590
Urban/rural percent		50/50

* The length of pre-work zone was the same as the work zone

Table 3.4.2 shows the average, minimum and maximum number of crashes before and after work zone. The average number of pre-work zone and during work zone crashes were very close, being 16.74 and 16.76 crashes/site, respectively, for all crashes. For non-injury crashes, there was 0.21 more crashes/site when a work zone was present, but the difference was not substantial enough to indicate negative impacts of work zones on traffic safety.

Table 3.4.2 Crashes during Work Zone and Before Work Zone

		All crashes	Injury crashes	Non-Injury crashes
Pre work zone (N=276)	Crashes	4621	1161	3460
	Average	16.74	4.21	12.53
	Min/max	0/293	0/79	0/214
During work zone (N=138)	Crashes	2313	554	1759
	Average	16.76	4.01	12.75
	Min/max	0/281	0/63	0/218

4. METHODOLOGY

Negative binomial model is the most commonly used model in work zone crash frequency modeling. Crashes can be considered as a result from a series of Bernoulli trials. Using Bernoulli terminology, the occurrence of a crash is considered a “success” and the alternative a failure. The use of this statistical terminology does not mean that crashes are positive phenomena.

For Y_i independent trials or crashes, there are y_i observed crashes, a negative binomial distribution is appropriate and is given the form of:

$$\mathrm{P}(Y_i = y_i) = \binom{Y_i}{y_i} p^{y_i}(1-p)^{Y_i - y_i}$$

When Y_i is large enough, let $\mathrm{p} = \lambda_i / Y_i$, and the negative binomial distribution can be approximated as a Poisson distribution (Lord 2004):

$$\mathrm{P}(Y_i = y_i) \cong \frac{\lambda_i}{Y_i\square} e^{-\lambda_i}$$

If *i* represents a work zone with a specific duration and length, then λ_i is the expected crash frequency of that work zone i. Y_i and y_i are all natural numbers.

Then explanatory variable x_i is introduced into λ_i (Khattak and Council 2002; Ozturk et al. 2013):

$$\lambda_i = e^{(\beta x_i + \varepsilon_i)}$$

where ε_i is error term, and is used to account for errors such as an omitted explanatory variable. For the negative binomial model, e^{ε_i} is assumed to have a gamma distribution with mean 1 and variance α^2.

In Generalized Linear Models, overdispersion is a situation where the variance of the crash frequency data exceeds the mean (Ismail and Jemain 2007). If the overdispersion condition exists, then the negative binomial model form should be used instead of the Poisson. With additional parameter α, the natural form of overdispersion is:

$$\mathrm{Var}[y_i] = \mathrm{E}[y_i]\{1 + \alpha \mathrm{E}[y_i]\}$$

The overdispersion rate is:

$$\alpha = \frac{\mathrm{Var}[y_i]}{\mathrm{E}[y_i]^2} - \frac{1}{\mathrm{E}[y_i]} = \frac{\mathrm{Var}[y_i] - \lambda_i}{{\lambda_i}^2}$$

and should not be zero for the negative binomial model to be applicable.

There were 6 different models that were created and investigated, with the final models using few variables and yielding good prediction results.

The final models were:

All crashes: $N = AADT^{B_1} D^{B_2} L^{B_3} e^{B_4 Urban} e^{B_5 Injury} e^{B_6 WZ} e^{B_7}$

I**njury and non-injury:** $N = AADT^{B_1} D^{B_2} L^{B_3} e^{B_4 Urban} e^{B_5 WZ} e^{B_6}$

and the variables are as follows:

- *N* - crash frequency
- *AADT* – Annual Average Daily traffic

- *D* – duration of observation
- *L* –segment length
- *Urban* – dummy variable for work zone location, 1 for urban and 0 for rural.
- *Injury* – dummy variable for crash severity, 1 for injury and 0 for non-injury
- *WZ* – dummy variable for work zone presence, 1 for with work zone and 0 for without work zone

5. MODEL ESTIMATION

5.1 Modeling All Crashes (Model 1)

A negative binomial model similar to the HSM model (9) was developed to model crashes with and without the presence of work zones. The dependent variable in this model was crash frequency. There were 828 (=138*6) data rows for presenting injury and non-injury crash frequencies in both pre-work zone and during work zone situations. The maximum likelihood method was used to estimate parameters using these 828 data points. The resulting model is:

$$N = AADT^{B_1} D^{B_2} L^{B_3} e^{B_4 Urban} e^{B_5 Injury} e^{B_6 WZ} e^{B_7}$$

Table.5.1.1 shows the estimated parameters of the combined model. All explanatory variables were highly statistically significant (i.e. $p<0.05$). R-square is a measure of goodness of fit of a function fitting data points (Elvik, 2009). The R-square is 0.9 which shows a good fit for the model. From this crash modification function, crash modification factors can be derived for all explanatory variables. A 1% increase in AADT, length and duration leads to the number of crashes changing by 0.96%, 0.58% and 1.01% respectively. Exp (0.7051) = 2.024 which means urban road segments have 2.024 times the crashes in comparison to rural roads. Frequency of injury crashes is 32.56% of non-injury crashes as Exp (-1.1221) = 0.3256. The presence of work zone increases the number of crashes by 21.51%, Exp (0.1948) = 1.2151.

Table 5.1.2 shows the CMFs for this SWZDI research and the HSM. The SWZDI CMF for duration was around 10% smaller, and the SWZDI CMF for length was 15.5% smaller. This difference between SWZDI and HSM could be due to geographical and driver differences between the Midwest and California, from where the HSM data was obtained. This difference could also be due to the fact that the HSM data involved only high impact work zones while the SWZDI data used a more diverse set of work zones.

Table 5.1.1 Combined Model for Injury and Non-Injury Crashes (Model 1)

Parameter	Variable	Estimate	t value	p-value
B_1	AADT	0.9613	55.45	<0.0001
B_2	D	1.0116	89.69	<0.0001
B_3	L	0.5802	113.93	<0.0001
B_4	Urban	0.7051	6.52	<0.0001
B_5	Injury	-1.1221	-40.11	<0.0001
B_6	WZ	0.1948	17.23	<0.0001
B_7	Constant	-13.3878	-59.39	<0.0001
Log likelihood	-2668			
R^2	0.9079			
Root mean squared error	6.144			

Table 5.1.2 Comparison of SWZDI and HSM CMF

CMF	SWZDI	HSM
Duration	1.01	1.11
Length	0.58	0.67

As mentioned before, 24 work zones were used to test the prediction power of the model and to compare it with the model used by HSM. From 24 work zone, 144 data points were generated. Figures 5.1.1 and 5.1.2 show the fitted line between the predicted and the observed number of crashes for Model 1 and HSM model respectively. To compare the prediction power of the models, the observed data is shown as the y axis (Piñeiro *et al.* 2008).

In Figures 5.1.1 and 5.1.2, the red line presents the 1:1 line, or a perfect match between the predicted and observed values. Thus the closer the fitted line slope is to 1, the better. Model 1's slope is 0.95 while the HSM model's slope is 3.8. Therefore Model 1's predictions were closer to observed values, and the HSM model heavily underestimates crash frequencies. On the other hand, the intercept for Model 1 and HSM model were 0.07 and -1.03 respectively. The closer the intercept is to 0 the better the model.

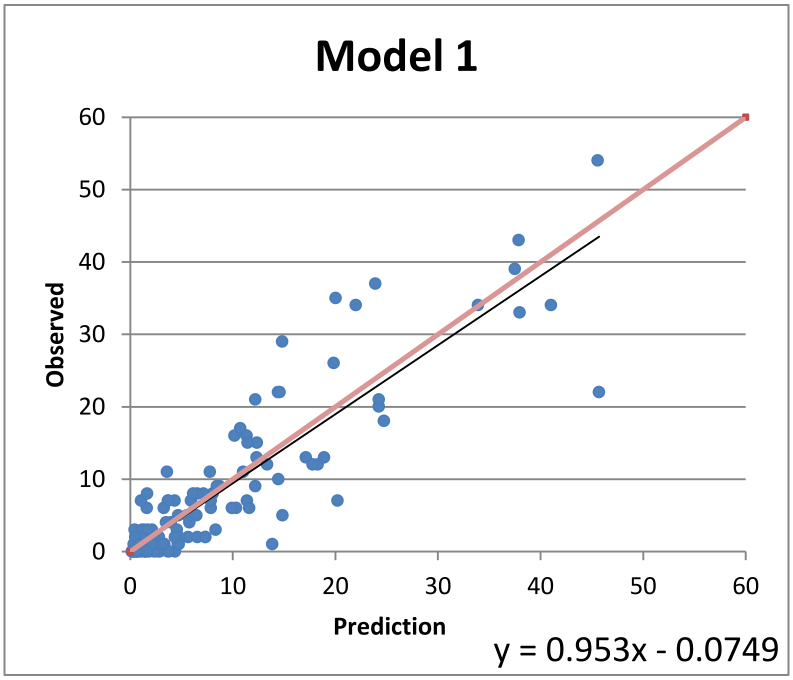


Figure 5.1.1 Model 1 versus observed number of crashes

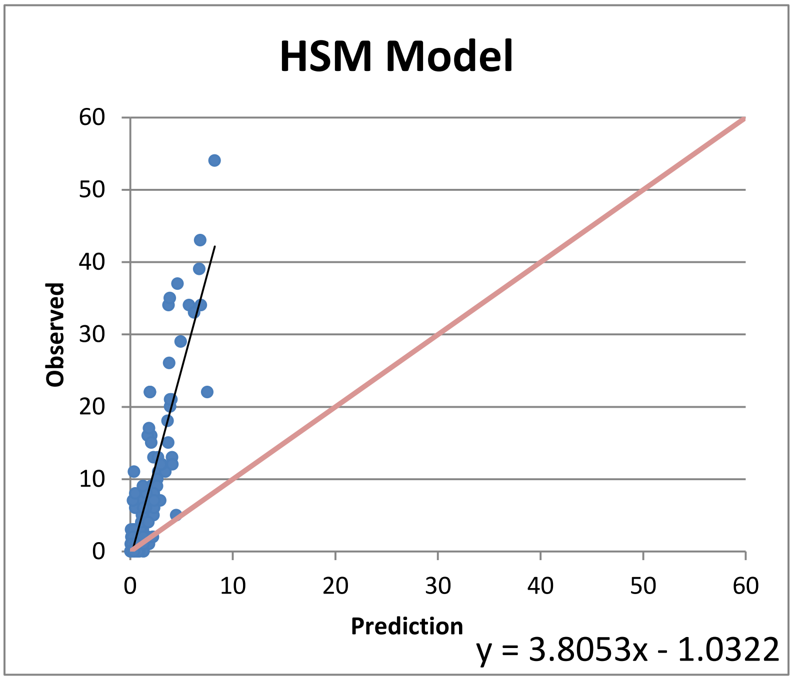


Figure 5.1.2 HSM model versus observed number of crashes

The HSM model for work zones performed poorly using Missouri test data because of several potential reasons. First, the HSM model was built using highly congested work zones data of California. Second, California driver behavior could different significantly from Missourian driver behavior. Last, this study used 828 data rows with a wide range of length, duration and AADT values, while the HSM model used 144 data rows.

In order to study separately the effects of the injury independent variable on work zone crashes, two additional models were developed. Model 2 focuses on injury crashes while Model 3 focuses on non-injury crashes. These separate models overcome a potential correlation problem among observations of the same work zone segments. This report recommends the use of Model 1 unless the practitioner focuses on injury or non-injury crashes only.

5.2 Modeling Fatal/Injury Crashes (Model 2)

A separate model was made for predicting fatal/injury crashes. The variables of this model, Model 2, were the same as Model 1, except for the lack of the Injury variable. As non-injury crashes were excluded, the training sample size was 414 (=138*3). The model structure is as follows:

$$N = AADT^{B_1} D^{B_2} L^{B_3} e^{B_4 Urban} e^{B_5 WZ} e^{B_6}$$

The variables are the same as the previous model. Table 5.2.1 shows the estimated parameters of Model 2. All explanatory variables, except Urban, were highly statistically significant ($p<0.05$). The Urban variable, though, had a small p-value of 0.0575 which was exceeded 0.05 slightly. The R-square is 0.88 and shows Model 2 has a good fit. From Model 2, injury crash modification factors can be developed for all explanatory variables. A 1% increase in AADT, length and duration, means the number of injury crashes is changed by 1.01%, 0.59% and 0.98% respectively. These quantities were close to Model 1 results. Exp (0.7462) = 2.109, which means urban road segments have 1.109 times more fatal/injury crashes in comparison to rural roads. Presence of work zone increases the number of fatal/injury crashes by 14.88% since Exp (0.1387) = 1.1488.

Table 5.2.1 Model for Fatal/Injury Crashes (Model 2)

Parameter	Variable	Estimate	t value	p-value
B_1	AADT	1.0051	32.00	<0.0001
B_2	D	0.9769	52.99	<0.0001
B_3	L	0.5882	59.86	<0.0001
B_4	Urban	0.7462	1.91	0.0575
B_5	WZ	0.1387	7.01	<0.0001
B_6	Constant	-14.8124	-27.94	<0.0001
Log likelihood	-1040			
R^2	0.8814			
Root mean squared error	9.1605			

Figure 5.2.1 shows the fitted line between predicted versus observed number of injury crashes for Model 2. The slope is 1.13 which is close to 1. The intercept is -0.7 thus it is not close to 0.

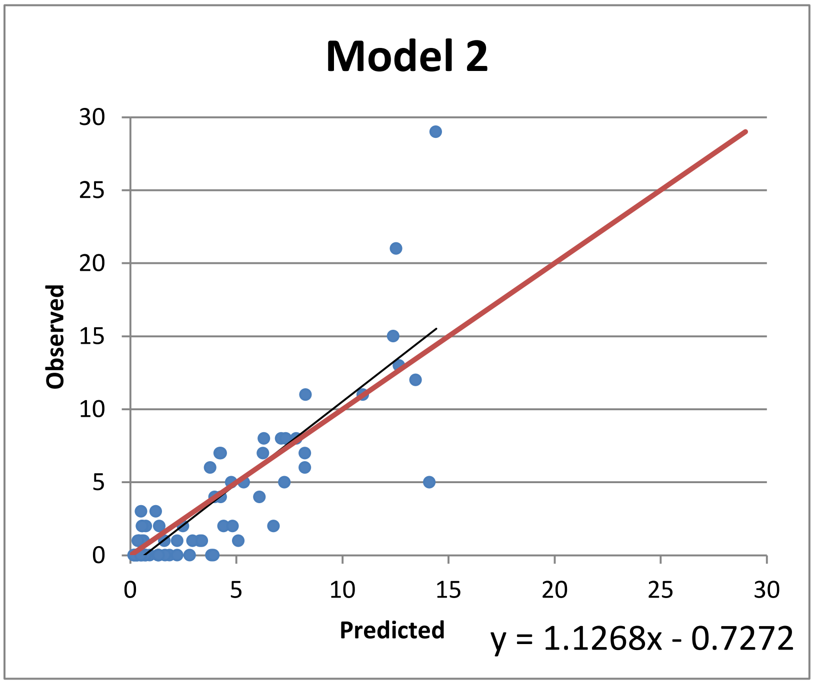


Figure 5.2.1 Model 2 versus observed number frequencies

5.3 Modeling Non-Injury Crashes (Model 3)

A model was built for predicting non-injury crashes. The variables of this model, Model 3, were the same as Model 2. Injury crashes were excluded to develop this model, so the sample size was 414 (=138*3). The model structure is as follows:

$$N = AADT^{B_1} D^{B_2} L^{B_3} e^{B_4 Urban} e^{B_5 WZ} e^{B_6}$$

The variables were defined the same as Model 2. Table 5.3.1 shows the estimated parameters of the Model 3.

Table 5.3.1 Model for Non-Injury Crashes (Model 3)

Parameter	Related Variable	Estimate	t value	p-valuen
B_1	AADT	0.9566	31.25	<0.0001
B_2	D	1.0148	50.87	<0.0001
B_3	L	0.5794	64.34	<0.0001
B_4	Urban	0.7009	3.70	0.0002
B_5	WZ	0.2007	10.08	<0.0001
B_6	Constant	-13.3528	-33.65	<0.0001
Log likelihood	-1450			
R^2	0.9062			
Root mean squared error	8.1658			

All explanatory variables were highly statistically significant. The R-square is 0.88 and the model's fit is good. Using Model 3, non-injury crash modification factors can be derived for all explanatory variables. A 1% increase in AADT, length and duration, means the number of non-injury crashes changed by 0.96%, 0.58% and 1.01%, respectively. These quantities are close to the Model 1 results. Exp (0.7009) = 2.015 which means urban road segments have 1.015 times more non-injury crashes in comparison to rural roads. The presence of work zones increases the number of non-injury crashes by 22.23% since Exp (0.0.2007) = 1.2223.

Figures 5.3.1 shows the fitted line between predicted vs observed number of non-injury crashes for Model 3. The Model 3 slope is 0.94 which is close to 1 and the intercept is also close to 0. Figure 5.3.1 shows Model 3 was able to predict non-injury crashes well.

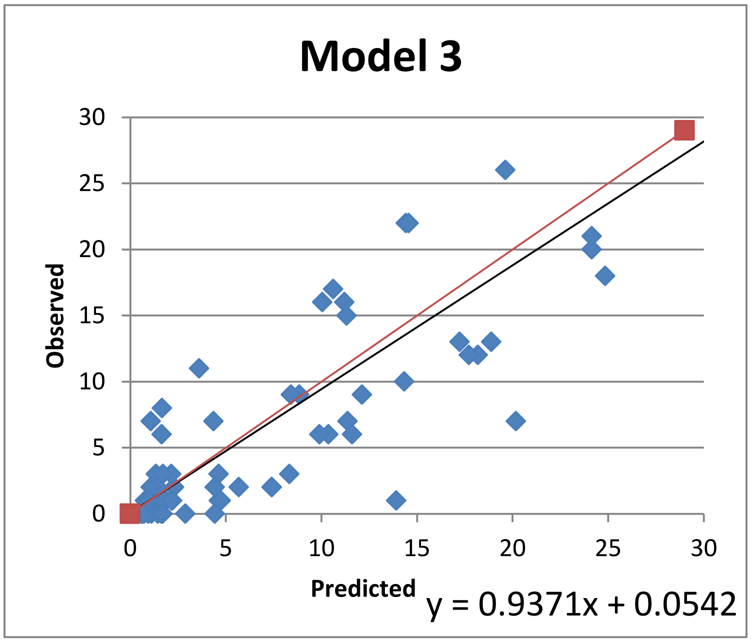


Figure 5.3.1 Model 3 versus observed number of crashes

6. SAMPLE APPLICATIONS

The application of the work zone crash modification functions (CMFcns) is illustrated in this section. A transportation agency can use the CMFcns to estimate the safety impact of alternative work zone plans. Two hypothetical examples illustrate the use of CMFcns to quantify the safety impacts of work zones for different alternative work zone plans.

6.1 Work Zone Length Example

A state transportation agency is considering a major rehabilitation of a 8-mile corridor of a major rural expressway. The expressway AADT is 50,000 vehicles per day. The agency has short-listed two alternatives based on preliminary analysis of traffic and safety data.

- *Alternative 1*: Complete the rehabilitation of the entire 8-mile corridor in one shot. The entire 8-miles will be an active work zone for a period of 100 days until the project completion.
- *Alternative 2*: Use a two-phase approach, rehabilitating 4-mile segments in each phase. Each phase is expected to take 50 days for completion. Thus, the entire duration to complete the two phases, 8 miles, is 100 days – same as Alternative 1.

Alternative 1 Safety Impacts

Model 1 is used to estimate the number of injury and non-injury crashes for each alternative. The negative binomial equation of Model 1 is written as,

$$N = AADT^{0.9613} D^{1.1016} L^{0.5802} e^{0.7051 Urban} e^{-1.1221 Injury} e^{0.1948 WZ} e^{-13.3878}$$

The independent variable values for Alternative 1 are: AADT = 50,000, D = 100 days, L = 8 miles, Urban = 0 (since rural), Injury variable is set to 1 for injury crashes and 0 for non-injury crashes, and WZ = 1.

Total number of injury crashes is computed as,

$N_{inj} = 50000^{0.9613} 100^{1.1016} 8^{0.5802} e^{0.7051(0)} e^{-1.1221(1)} e^{0.1948(1)} e^{-13.3878}$ = **7 crashes**

Total number of non-injury crashes is computed as,

$N_{noninj} = 50000^{0.9613} 100^{1.1016} 8^{0.5802} e^{0.7051(0)} e^{-1.1221(0)} e^{0.1948(1)} e^{-13.3878}$ = **21.6 crashes**

Alternative 2 Safety Impacts

Again, Model 1 is used to estimate the number of injury and non-injury crashes for each alternative. The crashes occurring in the 8-mile segment during the course of the project, i.e., 100 days, must be estimated. The crash prediction model is used four times – 1) to compute crashes occurring in the work zone in phase 1 (work zone length of 4 miles), 2) to compute crashes occurring in the non-work zone in phase 1, 3) to compute crashes occurring in the work zone in phase 2 (work zone length of 4 miles), and 4) to compute crashes occurring in the non-work zone in phase 2.

Crashes occurring in the work zone in phase 1

The independent variable values for Alternative 2 are: AADT = 50,000, D = 50 days, L = 4 miles, Urban = 0 (since rural), Injury variable is set to 1 for injury crashes and 0 for non-injury crashes, and WZ = 1.

Total number of injury crashes is computed as,

$N_{inj} = 50000^{0.9613} 50^{1.1016} 4^{0.5802} e^{0.7051(0)} e^{-1.1221(1)} e^{0.1948(1)} e^{-13.3878} =$ **2.3 crashes**

Total number of non-injury crashes is computed as,

$N_{noninj} = 50000^{0.9613} 50^{1.1016} 4^{0.5802} e^{0.7051(0)} e^{-1.1221(0)} e^{0.1948(1)} e^{-13.3878} =$ **7.2 crashes**

Crashes occurring in the non-work zone in phase 1

All independent variables values are the same as in the previous step except the WZ variable is 0, since there is no work zone in the second 4-mile segment in phase 1.

Total number of injury crashes is computed as,

$N_{inj} = 50000^{0.9613} 50^{1.1016} 4^{0.5802} e^{0.7051(0)} e^{-1.1221(1)} e^{0.1948(0)} e^{-13.3878} =$ **1.9 crashes**

Total number of non-injury crashes is computed as,

$N_{noninj} = 50000^{0.9613} 50^{1.1016} 4^{0.5802} e^{0.7051(0)} e^{-1.1221(0)} e^{0.1948(0)} e^{-13.3878} =$ **5.9 crashes**

Since the work zone length and duration are the same in phases 1 and 2, the number of crashes occurring in each phase will also be the same. Thus, the total injury crashes in Alternative2 are 2 x (2.3+1.9), or **8.4 crashes**. Similarly, the total non-injury crashes are 2 x (7.2+5.9), or **26.2 crashes**.

Result: The crash frequency was estimated for both alternatives. They are summarized below,

- *Alternative 1*: 7 injury crashes and 21.6 non-injury crashes
- *Alternative 2*: 8.4 injury crashes and 26.2 non-injury crashes

Based on the estimated crash frequency, Alternative 1 results in 1.4 fewer injury crashes and 4.6 fewer non-injury crashes, or a total of 6 fewer crashes than Alternative 2. A monetary value can be ascribed to the difference in crashes. Societal costs of crashes are available in the literature. For example, a *Safety Handbook for Locals* (S-HAL) (Sun et al. 2014) has a compilation of the societal cost values from the Highway Safety Manual and other literature. The following table is reproduced from Sun et al (2014). From Table 6.1, injury costs of $158,200 and non-injury costs of $7,400 were used for monetizing the crash costs in Alternative 1. The total savings in crash costs are estimated as 1.4 x 158,200 + 4.6 x 7,400 = **$255,520**.

Table 6.1 Societal Costs of Crashes

Crash type	**Crash costs**
Fatal	$4,008,900
Disabling injury	$216,000
Evident injury	$79,000
Fatal/injury	$158,200
Possible injury	$44,900
PDO	$7,400

6.2 Work Zone Duration Example

A state transportation agency is evaluating two alternatives for a maintenance project on an urban freeway. The first alternative requires 120 days to complete the entire work. On the other hand, the second alternative only requires 100 days to complete the project through the use of some advanced high-tech equipment. The agency is interested in knowing the safety benefits of completing the project 20 days earlier.

Alternative 1 Safety Impacts

Model 1 is used to estimate the number of injury and non-injury crashes for each alternative. The negative binomial equation of Model 1 is written as,

$$N = AADT^{0.9613}D^{1.1016}L^{0.5802}e^{0.7051Urban}e^{-1.1221Injury}e^{0.1948WZ}e^{-13.3878}$$

The independent variable values for Alternative 1 are: AADT = 25,000, D = 120 days, L = 3 miles, Urban = 1, Injury variable is set to 1 for injury crashes and 0 for non-injury crashes, and WZ = 1.

Total number of injury crashes is computed as,

$N_{inj} = 25000^{0.9613}120^{1.1016}3^{0.5802}e^{0.7051(1)}e^{-1.1221(1)}e^{0.1948(1)}e^{-13.3878}$ = **<u>5 crashes</u>**

Total number of non-injury crashes is computed as,

$N_{noninj} = 25000^{0.9613}120^{1.1016}3^{0.5802}e^{0.7051(1)}e^{-1.1221(0)}e^{0.1948(1)}e^{-13.3878}$ = **<u>15.3 crashes</u>**

Alternative 2 Safety Impacts

The values of all variables except the project duration, D, are the same as those in Alternative 1. The duration is now set to 80 days.

Total number of injury crashes is computed as,

$N_{inj} = 25000^{0.9613}80^{1.1016}3^{0.5802}e^{0.7051(1)}e^{-1.1221(1)}e^{0.1948(1)}e^{-13.3878}$ = **<u>3.3 crashes</u>**

Total number of non-injury crashes is computed as,

$N_{noninj} = 25000^{0.9613} 80^{1.1016} 3^{0.5802} e^{0.7051(1)} e^{-1.1221(0)} e^{0.1948(1)} e^{-13.3878}$ = **<u>10.1 crashes</u>**

<u>Result:</u> The crash frequency was estimated for both alternatives. They are summarized below,

- *Alternative 1*: 5 injury crashes and 15.3 non-injury crashes
- *Alternative 2*: 3.3 injury crashes and 10.1 non-injury crashes

Based on the estimated crash frequency, Alternative 2 results in 1.7 fewer injury crashes and 5.2 fewer non-injury crashes, or a total of 6.9 fewer crashes than Alternative 1. The total savings in crash costs are estimated as 1.7 x 158,200 + 5.2 x 7,400 = **$307,420**.

7. CONCLUSION

The arrival of the Highway Safety Manual (HSM) was hailed nationally as a great advance in safety, since it provided quantitative methods for analyzing highway safety. Thus the HSM complemented other national transportation standards such as the Highway Capacity Manual, the AASHTO Green Book (Geometric Design Manual), and the Manual of Uniform Traffic Control Devices. The HSM provides crash modification factors (CMFs) related to work zones. Specifically, it provided one crash modification function for work zone duration and another for work zone length. But because the data used for producing these functions were from high-impact work zones in California, there was a need to calibrate these functions for the Midwest.

This report documented the method, data, and results of the calibration of HSM CMFs for the Midwest. Obtaining useful and appropriate data for work zone safety modeling is a great challenge. This is because the majority of work zones have very short durations, thus no crashes occur at those work zones. And multiple sources of data need to be combined in order to produce the variables needed for modeling. One source is the work zone database that contains information on work zone characteristics such as duration, length, urban/rural, and location. Another source is the crash database that provides crash location, date/time, and severity. A third source is the traffic that travels through the work zones. These data are required for when the work zone is in place and also for normal conditions, i.e. before the work zone.

There were 10,973 Missouri freeway work zones that were examined in this research from which 536 work zones proved to contain useful modeling data. These work zones were then stratified according to duration and length. A stratified random sampling produced 162 work zones for use in modeling, 20% of which was set aside for model testing. These work zones varied in length from 0.76 mile to 9.24 miles, and they varied in duration from 16 days to 590 days. The traffic on these work zones varied between 1,990 vpd to 88,017 vpd. Before work zones were deployed, these locations resulted in 16.74 average crashes. During work zone deployment, the same locations resulted in 16.76 average crashes during same time period.

The before-after study was conducted using the aforementioned data. Due to overdispersion in crash frequency data, a negative binomial model form was used. The models included the variables of AADT, duration, length, urban/rural, injury, and work zone presence. A combined model along with a fatal/injury model and a non-injury model were produced. All models resulted in relative good fit and good prediction. The combined model had a R^2 fit of 0.9079 and a prediction slope of 0.963. This model produced crash modification factors of 1.01 for duration and 0.58 for length, which were smaller than the values in the HSM. The differences between this model and the HSM can be explained by data differences in terms of geography, driver population, and levels of work zone impact. Thus it was no surprising that the Midwest CMF values were smaller than the ones produced from California data.

The research presented in this report can be expanded in several ways. First, even though this research more than quadrupled the number of samples used in the HSM, the sample size is relatively modest for safety studies. If more data were to be used, then other characteristics of work zones, beyond duration and length, could be better investigated. Second, Empirical Bayes or even full Bayes can be utilized to address regression-to-the-mean problem. This can be a significant undertaking as each work zone site would need to be calibrated and modeled using HSM Safety Performance Functions. Third, the fixed work zone impact threshold of 0.5 mile can be changed to better reflect traffic impact differences among different work zones. Fourth, non-freeway facilities, which were not discussed in this report, could be modeled. Last, data from other Midwest states could be used to account for geographical and driver differences within the Midwest.

REFERENCES

Carter, D., Srinivasan, R., Gross, F., and Council, F. (2012). Recommended Protocols for Developing Crash Modification Factors. NCHRP 20-7(314) Final Report. American Association of State Highway and Transportation Officials.

Chen, E., and Tarko, A.P. (2014). "Modeling safety of highway work zones with random parameters and random effects models." *Analytic Methods in Accident Research*, 1, 86-95.

Edara, P. (2009). *Evaluation of Work Zone Enhancement Software*. Missouri DOT Research Report. RI 07-062.

Edara, P., Sun, C., Zhu, Z. (2013). *Calibration of Work Zone Software Programs.* USDOT, MATC and Missouri DOT Final Report. December.

Elvik, R.. (2009). Developing Accident Modification Functions: Exploratory Study. *Transportation Research Record*, No. 2103. Washington, D.C. pp. 18-24.

FHWA (2009). "Work Zone Safety and Mobility Fact Sheet - 10th Annual National Work Zone Awareness Week." <http://safety.fhwa.dot.gov/wz/wz_awareness/2009/factsht09.cfm> (May 25, 2014)

FHWA (2010). A Guide to Developing Quality Crash Modification Factors. Federal Highway Administraction, Office of Safety, Washington, D.C.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer-Verlag, New York.

Ismail, N., & Jemain, A. A. (2007). Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. *Casualty Actuarial Society Forum*.

Khattak, A. J., and Council, F. M. (2002). "Effects of Work Zone Presence on Injury and Non-Injury Crashes." *Accident Analysis and Prevention*, 34(1), 19-29.

Lord, D. and Mannering, F. (2010). "The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives." *Transportation Research Part A* 44(5), 291–305.

Lord, D., Washington, S. P., and Ivan, J. N. (2005). "Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory." *Accident Analysis & Prevention*, 37(1), 35-46

Ozturk, O., Ozbay, K., Yang, H., and Bartin, B. (2013). "Crash Frequency Modeling for Highway Construction Zones." *Transportation Research Board 92nd Annual Meeting*, Washington, D.C.

Pal, R., and Sinha, K.C. (1996). "Analysis of Crash Rates at Interstate Work Zones in Indiana." *Transportation Research Record: Journal of the Transportation Research Board*, 1529, 43-53.

Piñeiro, G., Perelman, S., Guerschman, J.P., and Paruelo, J.M. (2008). "How to Evaluate Models: Observed vs Predicted or Predicted vs Observed?" *Ecological Modelling*, 216(3-4), 316–322

Qi, Y., Srinivasan, R., Teng, H., and Baker, R.F. (2005). “Frequency of Work Zone Accidents on Construction Projects.” *Report No: 55657-03-15.* University Transportation Research Center City. College of New York, New York.

Salkind, N. (2006). *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, California.

Srinivasan, R., Ullman, G.L., Finley, M.D., and Council F.M. (2011). “Use of Empirical Bayesian Methods to Estimate Temporal-Based Crash Modification Factors for Construction Zones.” *Transportation Research Board 90th Annual Meeting*, Washington, D.C.

Sun, C., Edara, P., Brown, H. (2014). *Safety Handbook for Local Agencies.* USDOT MATC and Missouri DOT Final Report. April 2014.

TRB (2010). *Highway Capacity Manual*. Fifth Edition, Transportation Research Board, Washington, D.C.

Venugopal, S., and Tarko, A. (2000). “Safety Models for Rural Freeway Work Zones.” *Transportation Research Record: Journal of the Transportation Research Board* , 1715, 1-9.

Yang, H., Ozagur Ozturk, O., Ozbay, K., and Xie, K. (2014). “Work Zone Safety Analysis and Modeling: A State-of-the-Art Review.” *Transportation Research Board 93rd Annual Meeting,* Washington D.C.