# Highway Work Zone Risk Factors and Their Impact on Crash Severity

Yingfeng Li, Ph.D.[1]; and Yong Bai, Ph.D., P.E., M.ASCE[2]

**Abstract:** Numerous factors may contribute to high-severity crashes in highway work zones. Identifying these factors and then alleviating their impact is a challenging task that traffic engineers and researchers have to confront. In this study, the work zone risk factors that could increase the probability of causing fatalities when severe crashes occur were examined using a comprehensive approach. The researchers first identified the significant risk factors based on a screening process that incorporates both statistical analyses and empirical research findings. They then systematically investigated these factors using logistic regression and frequency analysis techniques. The severe crashes including the fatal crashes between 1998 and 2004 and injury crashes between 2003 and 2004 in Kansas highway work zones were used in the study. The assessed risk factors included variables describing driver characteristics, environmental conditions, crash road conditions, and other crash information. The results of this study will help traffic engineers to understand these risk factors and how the factors could increase the likelihood of having fatalities when a severe crash occurs in a work zone. Consequently, effective safety countermeasures may be designed at the work zone planning and installation stages to prevent safety deficiencies.

**DOI:** 10.1061/(ASCE)TE.1943-5436.0000055

**CE Database subject headings:** Construction sites; Highway and road construction; Safety; Human factors; Errors; Risk management; Logistics; Statistics.

## Introduction

As the American highways age, an increasing number of projects have been funded to preserve, expand, and enhance the existing system. These projects result in a large number of highway work zones that interrupt regular traffic flows and create safety concerns. Improving safety without sacrificing the main function of highways in work zones has become a challenging task that traffic engineers and researchers have to confront. Nationally, significant attention has been devoted to work zone safety. Provisions on highway work zone safety and related issues have been included in the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (FHWA 2007).

Work zone safety is affected by a large variety of risk factors and many of them are not fully understood. Comprehensive knowledge of the risk factors discovered from crash data are critical for reducing risk levels and preventing severe crashes in work zones. In this study, the researchers examined a wide range of crash variables describing work zone settings, environmental conditions, driver characteristics, and crash information based on Kansas work zone crash data. Through the examination, the risk factors that could lead to high-severity crashes (injury and/or fatal crashes) were identified and their impact on crash severity was quantified.

The data analyses in this study involved two major steps. First, Chi-square statistics, Cochran-Mantel-Haenszel (CMH) statistics, and relevant historical findings were used to identify significant risk factors in work zones based on crash data from Kansas. Then, the impact of these risk factors on crash severity was investigated using frequency analyses and logistic regression statistics. The outcome knowledge will help traffic engineers to better understand the risk factors and how they could increase the likelihood of having fatalities when a severe crash occurs in a work zone. With this knowledge, more effective safety countermeasures may be developed during work zone planning and installation to better prevent safety deficiencies.

## Literature Review

Work zone safety has been a research focus for decades and many publications are available on work zone crash characteristics and traffic control effectiveness. Among the crash risk analyses documented in literature, many focused on various types of nonwork zone crashes. The researchers did not find in-depth analyses that assessed the impact of individual risk factors on work zone crash severity based on injury and fatal crashes. Nevertheless, some relevant studies and their findings are briefly summarized as follows.

Harb et al. (2008) analyzed work zone crashes on Florida freeways using multiple and conditional logistic regression methods in an effort to identify risk factors in freeway work zones. The study indicated that factors including roadway geometry, weather condition, age, gender, lighting condition, residence code, and influence of alcohol/drugs could increase crash risk in freeway work zones. A few crash characteristic studies indicated that a number of human errors, such as following too close, inattentive

[1]Assistant Research Scientist, Traffic Operations Group, Texas Transportation Institute, 3500 NW Loop 410, Suite 315, San Antonio, TX 78229 (corresponding author). E-mail: y-li@tamu.edu

[2]Associate Professor, Dept. of Civil, Environmental, and Architectural Engineering, Univ. of Kansas, 1530 West 15th St., 2150 Learned Hall, Lawrence, KS 66045-7609. E-mail: ybai@ku.edu

Note. This manuscript was submitted on March 26, 2008; approved on April 1, 2009; published online on September 15, 2009. Discussion period open until March 1, 2010; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering***, Vol. 135, No. 10, October 1, 2009. ©ASCE, ISSN 0733-947X/2009/10-694–701/$25.00.

**Table 1.** Fatal and Injury Work Zone Crashes by Year

| Year | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | Total |
|---|---|---|---|---|---|---|---|---|
| Number of fatal crashes | 9 | 11 | 9 | 13 | 14 | 11 | 18 | 85 |
| Number of injury crashes | — | — | — | — | — | 283 | 337 | 620 |

driving, and misjudging, could increase the risk of work zone crashes (Mohan and Gautam 2002, Chambless et al. 2002; Daniel et al. 2000). Some studies also indicated that speeding (Garber and Zhao 2002) and inefficient traffic control (Ha and Nemeth 1995) were two other factors contributing to crashes in work zones. Adverse environmental and road surface conditions, however, did not contribute more to work zone crashes than to nonwork zone crashes when comparing their characteristics (Garber and Woo 1990).

This study used logistic regression technique to assess the impact of work zone risk factors. The significance of this technique in traffic safety-related studies has been recognized for years. For example, Li and Bai (2008a, 2009) used this technique in the analysis of traffic control devices and overall risk level in work zones. Applications of this technique in the analyses of nonwork zone crashes were found in a number of studies (Lu et al. 2006; Chang and Yeh 2006). In addition, Dissanayake and Lu (2002) used sequential binary logistic regression to analyze the contributing factors of single-vehicle, fixed-object crashes involving young drivers and found that factors including restraint device usage and being a male clearly reduced the tendency of high severity.

## Data Description

This study focused on the severe crashes including 85 fatal crashes between 1998 and 2004 and 620 injury crashes between 2003 and 2004 in Kansas highway work zones, as shown in Table 1. Including the fatal cases between 1998 and 2002 enriched the fatal crash information and minimized the analysis error caused by data sparseness. The crash data were obtained from the Kansas Department of Transportation (KDOT) database. The researchers identified the at-fault drivers in the original crash data and compiled their characteristics and other necessary crash information into spreadsheets. For the cases with missing or unclear information, the original crash reports, including detailed crash scene descriptions and sketches, were examined to ensure the data accuracy.

The collected crash and related information was organized into six categories. Each category included various variables with specific observations. Some low-frequency observations that were similar in nature were combined into more general observation groups so that the frequencies of the cross-categorized observations were increased. The increased data frequencies would reduce the errors caused by data sparseness in statistical tests and logistic regression analyses. Table 2 summarizes the variable categories, observations, and a preliminary comparison of fatal and injury crash frequencies.

## Work Zone Risk Factor Determination Methodology

The collected crash variables were examined in a comprehensive manner to identify those that had significant impact on the severity of crashes upon occurrence. An approach combining both statistical methods and empirical research findings was developed for the data screening. Chi-square statistics and CMH statistics were used to ensure the accuracy of risk factor identification. Briefly introduced below is the mathematical theory of the CMH statistics.

### *Cochran-Mantel-Haenszel (CMH) Statistics*

The mathematical theory of CMH statistical technique is briefly introduced herein based on (SAS 2004); detailed description of CMH can be found in Agresti 1996 and example applications in crash analyses can be found in Chira-Chavala and Mak (1986) and Chen and Jovanis (2000). As in a typical three-way contingency table, suppose the control variable $Z$ has $q$ strata, indexing each of them by $h=1,2,\ldots,q$. Each stratum contains a two-way contingency table with $X$ representing the test variable and $Y$ representing the outcome variable. For table $h$, denote the cell frequency in row $i$ and column $j$ by $n_{hij}$, with corresponding row and column marginal totals denoted by $n_{hi}$ and $n_{hj}$, and the overall stratum total by $n_h$. To test the conditional association between $X$ and $Y$, the null hypothesis, $H_0$, is that there is no association between $X$ and $Y$ in any of the strata. Thus, the respective expected value and covariance matrix of the frequencies are calculated as

$$m_h = E[n_h|H_0] = n_h(P_{h*} \otimes P_{h*})$$

and

$$\mathbf{Var}[n_h|H_0] = \frac{n_h^2}{n_h - 1}[(D_{P_{h\cdot *}} - P_{h\cdot *}P'_{h\cdot *}) \otimes (D_{P_{h*\cdot}} - P_{h*\cdot}P'_{h*\cdot})]$$

where

$$n'_{hi} = (n_{hi1}, n_{hi2}, \ldots, n_{hiC})$$

$$n'_h = (n'_{h1}, n'_{h2}, \ldots, n'_{hR})$$

$$p_{hi\cdot} = n_{hi}/n_h$$

$$p_{h\cdot j} = n_{h\cdot j}/n_h$$

$$P'_{h*\cdot} = (p_{h1\cdot}, p_{h2\cdot}, \ldots, p_{hR\cdot})$$

$$P'_{h*} = (p_{h\cdot 1}, p_{h\cdot 2}, \ldots, p_{h\cdot C})$$

and $\otimes$=Kronecker product multiplication and $D_a$=diagonal matrix with elements of $a$ on the main diagonal.

Given the expected value and covariance matrix of the frequencies, the generalized CMH statistic is defined as

$$Q_{\text{CMH}} = G' V_G^{-1} G$$

where

$$G = \sum_h B_h(n_h - m_h)$$

**Table 2.** Data Categories, Variables, and Crash Frequencies

| Category | Variable | Observation | Fatal crash frequency | Injury crash frequency |
|---|---|---|---|---|
| Driver at fault | Age | 15–19 | 8.2% | 15.5% |
| | | 20–24 | 7.1% | 16.1% |
| | | 25–34 | 15.3% | 20.3% |
| | | 35–44 | 25.9% | 17.4% |
| | | 45–54 | 15.3% | 13.7% |
| | | 55–64 | 4.7% | 6.5% |
| | | ≥65 | 23.5% | 7.4% |
| | Gender | Male | 75.3% | 64.2% |
| | | Female | 24.7% | 35.8% |
| Time | Time of day | Morning peak hours: 6:00–10:00 | 14.1% | 16.9% |
| | | Daytime nonpeak hours: 10:00–16:00 | 35.3% | 36.5% |
| | | Afternoon peak hours: 16:00–20:00 | 16.5% | 21.8% |
| | | Nighttime: 20:00–6:00 | 34.1% | 24.8% |
| | Day of week | Monday | 16.5% | 12.1% |
| | | Tuesday | 10.6% | 12.6% |
| | | Wednesday | 11.8% | 15.3% |
| | | Thursday | 15.3% | 13.4% |
| | | Friday | 14.1% | 19.8% |
| | | Saturday | 20.0% | 16.0% |
| | | Sunday | 11.8% | 10.8% |
| Environmental conditions | Light condition | Good condition, i.e., daylight | 52.9% | 68.5% |
| | | Fair conditions including dawn, dusk, and dark with streetlights | 15.3% | 15.3% |
| | | Poor condition, i.e., dark without streetlights | 31.8% | 15.8% |
| | | Other unfavorable light conditions | 0.0% | 0.3% |
| | Weather condition | Good condition, i.e., no adverse conditions | 90.6% | 87.9% |
| | | Poor conditions including rain, mist, drizzle, sleet, snow, fog, smoke, strong winds, and other | 9.4% | 12.1% |
| | Road surface condition | Good condition, i.e., dry surface | 89.4% | 87.6% |
| | | Fair conditions including wet, mud, dirt, sand, and debris | 9.4% | 11.0% |
| | | Poor conditions including snow, slush, ice, and snow packed | 1.2% | 1.5% |
| Road conditions | Road class | Interstates and other freeways and expressways | 30.6% | 57.3% |
| | | Other principal arterials and minor arterials | 64.7% | 40.8% |
| | | Low-classification roads including major collectors, minor collectors, and local roads | 4.7% | 1.9% |
| | Road character | Straight and level | 52.9% | 61.6% |
| | | Straight on grade | 27.1% | 22.4% |
| | | Curve and level | 9.4% | 7.3% |
| | | Curve on grade | 7.1% | 6.1% |
| | | Other geometric alignments | 3.5% | 2.6% |
| | Number of lanes | Two | 62.4% | 19.5% |
| | | Four | 30.6% | 46.5% |
| | | Six | 7.1% | 30.8% |
| | | Eight or more | 0.0% | 3.2% |
| | Speed limit | ≥61 mph | 56.5% | 23.9% |
| | | 51–60 mph | 35.3% | 40.8% |
| | | 41–50 mph | 3.5% | 8.2% |
| | | ≤40 mph | 4.7% | 24.5% |
| | Crash location | Nonintersection areas | 64.7% | 58.9% |
| | | Intersection or intersection related areas | 16.5% | 17.1% |
| | | Other areas including interchange areas, crossover areas, and other | 18.8% | 24.0% |
| | Surface type | Concrete | 27.1% | 49.7% |
| | | Blacktop | 71.8% | 49.8% |
| | | Other | 1.2% | 0.5% |

**Table 2.** (*Continued.*)

| Category | Variable | Observation | Fatal crash frequency | Injury crash frequency |
|---|---|---|---|---|
| | Road special feature | No special feature impact | 82.4% | 79.4% |
| | | Special features including bridge, overhead bridge, railroad bridge, railroad crossing, interchange, ramp, and other | 17.6% | 20.6% |
| | Area information | Urban area | 16.5% | 14.0% |
| | | Rural area | 83.5% | 86.0% |
| Crash information | Vehicle body type | Truck[a] involved | 42.4% | 15.5% |
| | | Nontruck involved | 57.6% | 84.5% |
| | Number of vehicles | One | 28.2% | 26.8% |
| | | Two | 50.6% | 54.4% |
| | | Three or more | 21.2% | 18.9% |
| Driver error | No driver error | 0 (Not present) | 89.4% | 86.9% |
| | | 1 (Present) | 10.6% | 13.1% |
| | Drug or alcohol impairment | 0 (Not present) | 90.6% | 90.2% |
| | | 1 (Present) | 9.4% | 9.8% |
| | Disregarded traffic signs, signals, and markings | 0 (Not present) | 80.0% | 92.1% |
| | | 1 (Present) | 20.0% | 7.9% |
| | Exceeded posted speed limits or too fast for conditions | 0 (Not present) | 84.7% | 80.2% |
| | | 1 (Present) | 15.3% | 19.8% |
| | Following too closely | 0 (Not present) | 97.6% | 75.8% |
| | | 1 (Present) | 2.4% | 24.2% |
| | Inattentive driving[b] | 0 (Not present) | 49.4% | 53.1% |
| | | 1 (Present) | 50.6% | 46.9% |

[a]The term truck in this study refers to such heavy vehicle types as single-unit large trucks, truck and trailers, tractor-trailers, and buses.

[b]Inattentive driving includes such errors on the KDOT accident reports as "fell asleep," "inattention," "other distraction in or on vehicle," "distraction-cell phone," and "distraction-other electronic devices."

$$V_G = \sum_h B_h[\mathbf{Var}(n_h|H_0)]B_h'$$

and where $B_h$=Kronecker product of the column scores $C_h$ and row scores $R_h$. When the null hypothesis is true, the CMH statistic has an asymptotic chi-square distribution with degree of freedom equal to the rank of $B_h$.

The SAS software outputs three CMH statistics: the nonzero correlation statistic, the row mean scores statistic, and the general association statistic. These statistics test the $H_0$ of no association against different alternative hypotheses ($H_1$). The alternative hypotheses for these three statistics are

1. The nonzero correlation statistic: there is a linear association between $X$ and $Y$ in at least one stratum.
2. The row mean scores statistic: for at least one stratum, the mean scores of the $R$ rows are unequal.
3. The general association statistic: for at least one stratum, there is some kind of association between $X$ and $Y$.

### *Risk Factor Determination Procedure*

As shown in Fig. 1, the procedure of identifying work zone risk factors included three steps.

Step 1. The variables that are statistically associated with crash severity were selected first as principal risk factors through chi-square statistics. Pearson chi-square and Likelihood Ratio chi-square tests were used in this step. A variable was selected when at least one of the two tests supported its relationship with the crash severity at a 0.05 level of significance. The principal risk factors selected in this step are listed in Table 3.

Step 2. The insignificant variables from the previous step were further examined by CMH statistics at 0.05 level of significance to detect the ones that affected work zone crash severity interactively with the principal risk factors selected in Step 1. The direct impact of these variables may not strong enough to be statistically detected through chi-square tests. However, when combined with other factors, they could yield significant impact and thus need to be considered when analyzing the principal risk factors. CMH statistics test the relationships between the unselected variables and the crash severity outcome in a three-way contingency table by using the selected risk factors as control variables. The signifi-

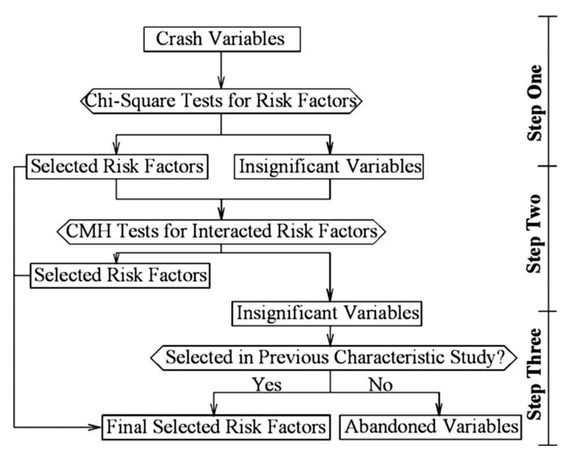


**Fig. 1.** Risk factor selection flowchart

**Table 3.** Risk Factors Identified at Step 1

| Variable category | Risk factor | $\chi^2$ test $p$-value LR[a] | $\chi^2$ test $p$-value Pearson |
|---|---|---|---|
| At-fault Driver | Age | <0.01 | <0.01 |
| | Gender | *0.04* | *0.04* |
| Envir. condition[b] | Light condition | <0.01 | *0.01* |
| Crash information | Vehicle type | <0.01 | <0.01 |
| Road condition | Road class | <0.01 | <0.01 |
| | Number of lanes | <0.01 | <0.01 |
| | Speed limit | <0.01 | <0.01 |
| | Surface type | <0.01 | <0.01 |
| Driver error | Disregarded traffic control | <0.01 | <0.01 |
| | Followed too closely | <0.01 | <0.01 |

Note: A $p$-value less than or equal to 0.05 indicates a significant relationship tested by the statistic at 0.05 level of significance and is italic.

[a]Likelihood ratio.

[b]Environmental condition.

cant variables supported by CMH statistics in this step were also selected as (second-level) risk factors, as shown in Table 4.

Step 3. To identify all potential risk factors, the results of the previous characteristic comparisons between fatal and injury crashes (Li and Bai 2008b) were examined. Risk factors that were identified based on the comparisons yet not detected in the previous two steps were also selected. As unveiled in the comparison study, factors including road character, alcohol/drug impairment, and too fast for conditions/speeding had significant impact on crash severity outcomes but were not selected in the first two steps. These factors were selected in Step 3 as risk factors. Using this risk determination procedure, 15 out of the 23 variables were selected as risk factors.

## Impact of Work Zone Risk Factors on Crash Severity

The identified work zone risk factors were studied using frequency analysis and logistic regression methods for their impact on crash severity. The impact of the work zone risk factors was assessed by comparing the odds or conditional probabilities of causing fatalities when a severe crash occurred. The study results are organized by crash variable categories. These results will benefit work zone traffic control design and provide necessary knowledge for controlling high-risk factors in work zones and consequently mitigating crash severity.

### *At-Fault Driver*

Both age and gender of the at-fault drivers had significant impact on the probability of causing fatalities in a severe work zone crash. The statistical tests showed that crash-time variable could interactively affect crash severity with the age variable. Listed in Table 5 are the proportions of fatal crashes by crash time and at-fault-driver age. Note that these percentages are only for comparison purpose and do not reflect the true proportions because of the different time scopes of fatal and injury crashes. Comparisons showed that drivers older than 64 years of age and drivers aged between 35–44 generally had higher probabilities of causing fatal crashes in each time period. In particular, a large proportion of the

**Table 4.** Risk Factors Identified at Step 2

| Variable Category | Second-level risk factor | Associated risk factor | CMH $p$-value NC[a] | CMH $p$-value [b]RMS | CMH $p$-value [c]GA |
|---|---|---|---|---|---|
| At-fault driver | Crash time | Age | *0.05* | *0.05* | *0.04* |
| Road condition | Area information | Number of lanes | <0.01 | <0.01 | <0.01 |
| | Area information | Speed limit | <0.01 | <0.01 | <0.01 |

Note: A $p$-value less than or equal to 0.05 indicates a significant relationship tested by the statistic at 0.05 level of significance and is italic.

[a]Nonzero correlation statistic.

[b]Row mean scores statistic.

[c]General association statistic.

**Table 5.** Proportions of Fatal Crashes for Each Category by Age and Crash Time

| Age | Percentage in driving population[a] | Crash time 6:00–10:00 | Crash time 10:00–16:00 | Crash time 16:00–20:00 | Crash time 20:00–6:00 |
|---|---|---|---|---|---|
| 15–19 | 7% | 13%[b] | 6% | 0% | 9% |
| 20–24 | 9% | 13% | 0% | 0% | 10% |
| 25–34 | 17% | 8% | 10% | 4% | 14% |
| 35–44 | 18% | 13% | 18% | 7% | 26% |
| 45–54 | 19% | 5% | 14% | 6% | 22% |
| 55–64 | 14% | 0% | 9% | 17% | 0% |
| ≥65 | 14% | 22% | 24% | 42% | 40% |

[a]Source: FHWA 2005. Highway Statistics 2005, Section III: Driver Licensing. Federal Highway Administration, Washington D.C., ⟨http://www.fhwa.dot.gov/policy/ohim/hs05/driver_licensing.htm⟩ (Sept. 15, 2008).

[b]The percentages of fatal crashes were calculated as the proportion of fatal crashes in the total severe crashes (fatal and injury) of each category cell. For instance, during 6:00–10:00, drivers between 15 and 19 years of age caused 3 fatal crashes and 21 injury crashes. Thus, the corresponding percentage for this category was $3/(3+21)\approx 13\%$. These percentages are only for comparison purpose since they do not reflect the true proportional constitution of fatal crashes considering the different time spans for the two types of crashes.

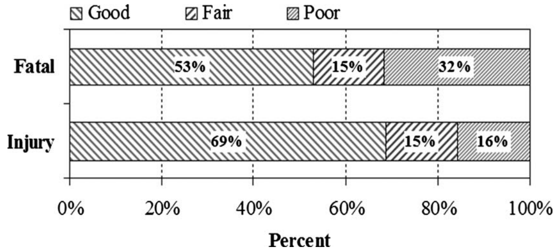


**Fig. 2.** Crash comparison by light conditions

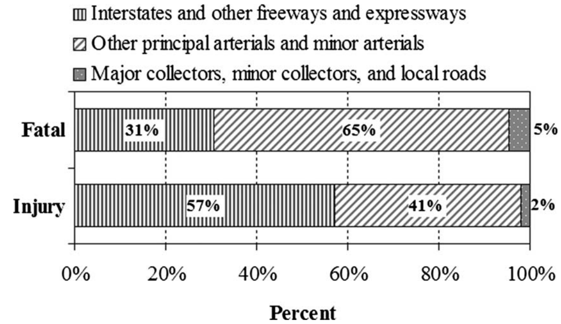


**Fig. 3.** Crash comparison by road class

severe crashes that were caused by senior drivers (≥65) during afternoon peak hours (16:00–20:00) involved fatalities. The comparisons generally implied that being a driver between 35 and 44 or older than 64 could increase the risk of causing fatalities when a severe crash occurred. Table 5 also included the percentages of the licensed drivers in Kansas by age.

In term of driver gender, the following logistic regression equation was developed:

$$\text{logit}\{\text{CRASH\_SEVERITY} = \text{Fatal} | \text{GENDER}\} = -1.30 - 0.53\ \text{GENDER}$$

According to this equation, the odds ratio between a fatal crash caused by a male driver and a fatal crash caused by a female driver was 1.70. In another word, the odds of involving fatalities in a severe crash caused by a male driver were 1.7 times as high as those for a severe crash caused by a female driver.

### Environmental Condition

Statistical tests showed that light condition was a risk factor affecting crash severity. By comparing the proportions of fatal crashes among the total severe crashes occurred in different light conditions, Fig. 2 clearly illustrates that poor light conditions contributed to a much larger percent of fatal crashes, which indicates that poor light conditions could increase the probability of causing fatalities when a severe crash occurred. In this analysis, good light condition refers to the daylight condition, fair condition refers to the dawn, dusk, or dark-with-streetlights condition, and poor condition refers to the dark-without-streetlights condition.

### Crash Information

Statistical tests showed that vehicle type was directly related to the severity of crashes. The following logistic regression equation was developed to model the conditional probability of causing fatalities in a severe crash in terms of vehicle type:

$$\text{logit}\{\text{CRASH\_SEVERITY} = \text{Fatal} | \text{VEHICLE\_TYPE}\} = 0.41 - 1.39\ \text{VEHICLE\_TYPE}$$

The ratio of the odds of causing fatalities in a truck-involved severe crash and the odds of causing fatalities in a nontruck-involved severe crash was estimated as 4.0. Equivalently, based on the modeling results, the odds of causing fatalities in a truck-involved severe crash was four times as high as the odds in a nontruck-involved severe crash. The term "truck" refers to such heavy vehicle types as single-unit large trucks, truck and trailers, tractor-trailers, and large buses.

### Road Condition

As identified through the risk factor selection procedure, road-condition variables such as road class, number of lanes, speed limit, surface type, and road character had impact on the probability of involving fatalities when severe crashes occurred. Fig. 3 exhibits the distributions of both types of crashes on different classes of roadways. Comparing to injury crashes, a much higher percentage of fatal crashes occurred on the arterials other than interstate highways or other freeways and expressways. A higher proportion of fatal crashes were also observed on the low-class roads such as collectors and local roads. In terms of road surface type, 72% of the fatal crashes occurred on roadways with asphalt pavement, while the corresponding percent on concrete roads was about 50%. Fig. 4 illustrates the frequencies of both fatal and injury crashes by road character. Unfavorable road characters including straight on grade, curve and level, curve on grade, and other unfavorable alignments contributed to 9% more (47% versus 38%) fatal crashes than to injury crashes. This fact may indicate that unfavorable alignments increased the involvement rate of fatalities in severe crashes.

Statistical tests showed that the number-of-lanes variable and the area-information variable interactively affected the probability of having fatalities in severe crashes. The following equation is the logistic regression model for the conditional probability of having fatalities in a severe crash in terms of number of lanes and area information:

$$\begin{aligned}\text{logit}\{&\text{CRASH\_SEVERITY} \\ &= \text{Fatal} | \text{NO\_LANES},\ \text{AREA\_INFO}\} \\ &= 2.52 - 0.80\ \text{NO\_LANES} - 0.90\ \text{AREA\_INFO}\end{aligned}$$

where NO LANES = number of lanes; and AREA INFO = area information.

The conditional probabilities calculated based on the regression model are listed in Table 6. Comparisons of these probabili-

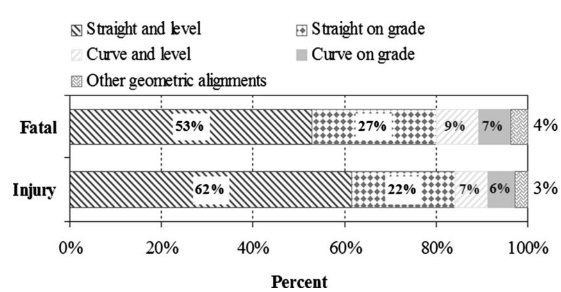


**Fig. 4.** Comparison of crash frequencies by road characters

**Table 6.** Conditional Probabilities of Involving Fatalities by Number of Lanes and Area Information

| Number of lanes (both directions) | Area information | |
|---|---|---|
| | Urban area | Rural area |
| 2 Lanes | 0.50 | 0.29 |
| 4 Lanes | 0.17 | 0.08 |
| 6 Lanes | 0.04 | 0.02 |
| 8 Lanes | 0.01 | 0.00 |

ties indicated that severe crashes in work zones on two-lane highways, especially urban two-lane highways, were more likely to involve fatalities than the crashes in the work zones at other locations.

The researchers found that speed limit and area information variables could interactively affect the conditional probability of having fatalities in a severe crash. The logistic regression model for the conditional probability in terms of speed limit and area information was developed as following:

$$\begin{aligned}&\text{logit}\{CRASH\_SEVERITY\\&\quad = \text{Fatal}|SPEED\_LIMIT,\ \ AREA\_INFO\}\\&\quad = 3.23 - 1.15\ \ SPEED\_LIMIT\text{-}1.63\ \ AREA\_INFO\end{aligned}$$

where AREA_INFO=area information.

Based on this model, the probabilities were estimated and listed in Table 7. According to these probabilities, the likelihood of causing fatalities when a severe crash occurred in urban high-speed work zones (speed limits $>60$ mph) was much higher than that in other work zones.

### *Driver Error*

Crash variable screening showed that some driver errors including disregarded traffic control, followed too closely, alcohol/drug impairment, and too fast for conditions/speeding could have significant impact on the probability of causing fatalities in severe crashes. According to the developed logistic regression models for the driver errors as listed in Table 8, the odds of causing fatalities in a severe crash when the disregarded-traffic-control error was present were almost three times as high as those in a severe crash that did not involve this driver error. On the other hand, the logistic regression model for following-too-closely driver error showed that the odds of involving fatalities when this error was present were much lower (by 92%) than those in the cases when the error was not present. This is consistent with the consensus that high volumes (when the following-too-closely driver error typically occurs) generally coincide with crashes of lower severity. As regards the alcohol/drug-impairment and too-fast-for-conditions/speeding driver errors, the logistic regression analysis could not establish significant models for the probability of having fatalities. However, the frequency analyses (Table 2) showed that the former contributed to about 10% of both the fatal and injury crashes and the latter contributed to 5% more (20% versus 15%) injury crashes than fatal crashes.

**Table 7.** Conditional Probabilities of Involving Fatalities by Speed Limit and Area Information

| Speed limit | Urban area | Rural area |
|---|---|---|
| $\geq 61$ mph | 0.61 | 0.24 |
| 51–60 mph | 0.33 | 0.09 |
| 41–50 mph | 0.14 | 0.03 |
| $\leq 40$ mph | 0.05 | 0.01 |

## Conclusion

Work zone safety is affected by many risk factors and some of them might have not been fully understood by traffic engineers. Comprehensive knowledge of the risk factors discovered from crash data therefore becomes critical for reducing risk levels and preventing severe crashes in work zones. Focused on the fatal and injury crashes in Kansas highway work zones, the researchers identified the risk factors from a wide range of crash variables using a variable screening procedure. This procedure ensured the capture of significant risk factors while eliminating the unimportant ones by incorporating both statistical techniques and previous research findings. The researchers thoroughly assessed the impact of the risk factors on the probability of having fatality in severe crashes based on the crash data in Kansas. The findings of this study are valuable to traffic engineers for developing countermeasures in work zones that can alleviate the safety risk resulted by a wide range of factors among which some could be overlooked when designing and setting up work zones. In addition, the knowledge is also beneficial for public education and information. Concluded below are the significant findings of this study.

In terms of at-fault driver characteristics, both age and gender had impact on the probability of causing fatalities when severe crashes occurred. Being a male diver could almost double the odds of having fatality in case of a severe crash. Severe Crashes caused by senior drivers (older than 64) during both afternoon peak hours and nighttime (16:00–6:00) and by drivers aged between 35 and 44 during nighttime (20:00–6:00) were more likely to involve fatalities. The findings indicate an immediate need for public education programs orienting these high-risk driver groups.

Light condition and vehicle type were significant risk factors in work zones as well. The poor light condition (i.e., dark without streetlights) contributed to a much higher proportion of fatal crashes than injury crashes. Involvement of heavy trucks in a severe crash increased the odds of causing fatalities by three times. The researchers therefore recommend that traffic engineers favorably weigh the needs of illumination and truck-oriented traffic control mechanisms in dark work zones and work zones with noteworthy truck traffic.

Regarding road condition, the study showed that being on

**Table 8.** Logistic Regression Results for Common Driver Errors

| Driver errors | Logistic regression model | Odds ratio (present: absent) |
|---|---|---|
| Disregarded traffic control (DISTC) | $\text{logit}\{CRASH\ \ SEVERITY = \text{Fatal}\ \ DISTC\}$ $= -2.12 + 1.06\ \ DISTC$ | 2.87 |
| Followed too closely (FOLCL) | $\text{logit}\{CRASH\ \ SEVERITY = \text{Fatal}\ \ FOLCL\}$ $= -1.73 - 2.58\ \ FOLCL$ | 0.08 |

"other principal arterials and minor arterials," rural two-lane highways, or urban highways with speed limits higher than 60 mph could increase the likelihood of causing fatalities in a severe crash. A severe crash occurring in work zones on highways with unfavorable geometric alignment features had a higher probability of involving fatalities as well. These facts indicate that there is room for improving the effectiveness of the traffic controls currently used in the high-risk work zones mentioned above. Notice that the findings indicated that severe crashes in work zones on asphalt-paved highways had a higher likelihood of involving fatalities. This result needs to be interpreted with caution and may require further exploration.

Some driver errors have clearly showed impact on crash severity in work zones. The odds of having fatalities in a severe crash contributed by disregarded traffic control tripled those for a severe crash not contributed by this driver error. However, the presence of followed-too-closely driver error actually decreased the odds of fatalities in severe crashes. These results indicate that there is a need to develop traffic control strategies that result in better compliance rates.

## Acknowledgments

The writers would like to thank KDOT for providing financial support for this study. The help from Mr. Anthony Alrobaire and Mr. Rex McCommon in KDOT during this study is greatly appreciated. In addition, this study would not be accomplished without the valuable advice from Dr. Yaozhong Hu in the Department of Mathematics at the University of Kansas.

## References

Agresti, A. (1996). "Cochran-Mantel-Haenszel methods." *An introduction to categorical data analysis*, Wiley, New York, 60–64.

Chambless, J., Chadiali, A. M., Lindly, J. K., and McFadden, J. (2002). "Multistate work zone crash characteristics." *ITE J.*, 72, 46–50.

Chang, H., and Yeh, T. (2006). "Risk factors to driver fatalities in single-vehicle crashes: Comparisons between nonmotorcycle drivers and motorcyclists." *J. Transp. Eng.*, 132(3), 227–236.

Chen, W., and Jovanis, P. P. (2000). "Method for identifying factors contributing to driver-injury severity in traffic crashes." *Transp. Res. Rec.*, 1717, 1–9.

Chira-Chavala, T., and Mak, K. K. (1986). "Identification of accident factors on highway segments: A method and applications." *Transp. Res. Rec.*, 1068, 52–58.

Daniel, J., Dixon, K., and Jared, D. (2000). "Analysis of fatal crashes in Georgia work zones." *Transp. Res. Rec.*, 1715, 18–23.

Dissanayake, S., and Lu, J. (2002). "Analysis of severity of young driver crashes, sequential binary logistic regression modeling." *Transp. Res. Rec.*, 1784, 108–114.

FHWA. (2007). "A summary of highway provisions in SAFETEA-LU." Program Analysis Team, Office of Legislation and Intergovernmental Affairs, ⟨www.fhwa.dot.gov/safetealu/summary.htm⟩ (Aug. 23, 2007).

Garber, N. J., and Woo, T. H. (1990). "Accident characteristics at construction and maintenance zones in urban areas." *Rep. No. VTRC 90–R12*, Virginia Transportation Research Council, Charlottesville, Va.

Garber, N. J., and Zhao, M. (2002). "Crash characteristics at work zones." *Rep. No. VTRC 02–R12*, Virginia Transportation Research Council, Charlottesville, Va.

Ha, T., and Nemeth, Z. A. (1995). "Detailed study of accident experience in construction and maintenance zones." *Transp. Res. Rec.*, 1509, 38–45.

Harb, R., Radwan, E., Yan, X., Pande, A., and Abdel-Aty, M. (2008). "Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression." *J. Transp. Eng.*, 134(5), 203–214.

Li, Y., and Bai, Y. (2008a). "Development of crash-severity-index models for the measurement of work zone risk levels." *Accid. Anal. Prev.*, 40(5), 1724–1731.

Li, Y., and Bai, Y. (2008b). "Comparison of characteristics between fatal and injury accidents in the highway construction zones." *Safety Sci.*, 46(4), 646–660.

Li, Y., and Bai, Y. (2009). "Effectiveness of temporary traffic control measures in highway work zones." *Safety Sci.*, 47(3), 453–458.

Lu, G., Noyce, D. A., and McKendry, R. J. (2006). "Analysis of the magnitude and predictability of median crossover crashes utilizing logistic regression." *Proc., TRB 85th Annual Meeting* (CD-ROM), Transportation Research Board, Washington, D.C.

Mohan, S. B., and Gautam, P. (2002). "Cost of highway work zone injuries." *Pract. Period. Struct. Des. Constr.*, 7(2), 68–73.

SAS. (2004). "Cochran-Mantel-Haenszel statistics. Chapter 2: The FREQ procedure." *Base SAS 9.1.3 procedures guide*, Vol. 3, SAS Institute Inc., Cary, N.C., 134–136.