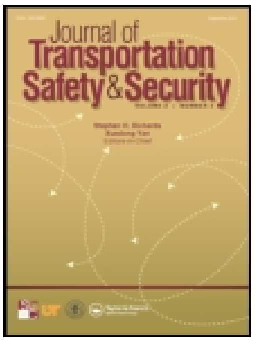

Taylor & Francis
Taylor & Francis Group

Journal of Transportation Safety & Security

ISSN: 1943-9962 (Print) 1943-9970 (Online) Journal homepage: http://www.tandfonline.com/loi/utss20

Work zones versus nonwork zones: Risk factors leading to rear-end and sideswipe collisions

Claire Silverstein, Justin Schorr & Samer H. Hamdar

To cite this article: Claire Silverstein, Justin Schorr & Samer H. Hamdar (2016) Work zones versus nonwork zones: Risk factors leading to rear-end and sideswipe collisions, Journal of Transportation Safety & Security, 8:4, 310-326, DOI: 10.1080/19439962.2015.1036332

To link to this article: http://dx.doi.org/10.1080/19439962.2015.1036332

Accepted author version posted online: 13 Jun 2015.
Published online: 13 Jun 2015.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
2016, VOL. 8, NO. 4, 310–326
http://dx.doi.org/10.1080/19439962.2015.1036332

Work zones versus nonwork zones: Risk factors leading to rear-end and sideswipe collisions

Claire Silverstein, Justin Schorr, and Samer H. Hamdar

Department of Civil and Environmental Engineering, The George Washington University, Ashburn, Virginia, USA

ABSTRACT

Tens of thousands of vehicular collisions occur annually in work zones with nearly double the fatality risk as compared to all collisions (work zone and nonwork zone). Due to this increase in risk, this study's objective is to investigate the possible causes of work zone collisions. After reviewing previous studies, the authors examined behavioral, environmental, and roadway geometric factors to understand their influence on fatal collision type for work zones and nonwork zones. Data from the National Highway Traffic Safety Administration Fatality Analysis Reporting System database was used for years 2010 to 2012. To analyze the data, negative binomial regression and multinomial logit models were utilized. A binary probit model directly compares work zone and nonwork zone data. Results demonstrate that rear-end and sideswipe collisions are more likely to cause fatalities in work zones compared to nonwork zones. Clear conditions, daylight, and straight roads increase the likelihood of these two collision types when compared to other types such as single-vehicle collisions. These findings suggest that Intelligent Transportation Systems countermeasures (speed harmonization, vehicle-to-vehicle communications) should be investigated to encourage safer car-following and lane-changing behaviors rather than to mitigate work zone–related infrastructure challenges.

KEYWORDS
work zones; fatalities; collisions; negative binomial regression; multinomial logit; binary probit

1. Introduction and motivation

Work zones lead to the alteration of roadway geometric characteristics and traffic-flow conditions. Such alterations have the potential to create an unsafe driving environment and require a heightened level of alertness and awareness in response to dynamically changing surroundings (Milton & Mannering, 1998). The resulting behavior is a manifestation of complex (and often unsafe) interactions between drivers, infrastructure components, control measures, and construction workers.

CONTACT Claire Silverstein csilvs@gwmail.gwu.edu Department of Civil and Environmental Engineering, Center for Intelligent Systems Research, Traffic and Networks Research Laboratory, The George Washington University, Exploration Hall, 20101 Academic Way, Ashburn, VA 20147, USA.

To understand these unsafe interactions, previous research has focused on understanding the causes of work-zone collisions (Duffy & McAvoy, 2009; Mathes, 2012). The corresponding studies have used a variety of data sets, variables, and modeling techniques to address work-zone collision injury and fatality-risk levels. Many of these studies utilize only one, often basic, statistical model, and output metrics are typically presented as composite indices indicating risk level.

Owing to the importance of this research topic, the objective of this article is to develop and compare a multinomial logit (MNL; using LIMDEP software) model and a negative binomial (NB - using SAS 9.3 software) regression model to identify different variables affecting fatal collision type for work-zone and nonwork zone collisions. Specifically, the authors aim to identify and focus on fatal collision types that feature the most dramatic increase when comparing nonwork zones to work zones (mainly rear-end and sideswipe collisions [RESS]). Data used to validate both modeling techniques was provided by the Fatality Analysis Reporting System (FARS) and features a number of environmental (time of crash, atmospheric conditions), behavioral (alcohol involvement and driver distraction), and roadway geometric (speed limit, number of lanes, road alignment) characteristics. Because every collision in the FARS database involves a fatality, utilizing this data allows for the exploration of the relationships between the different exogenous parameters/variables featured in the most severe collisions. This comparative framework provides a better indication of the models' validity and may lead to the identification of specific Intelligent Transportation Systems (ITS) countermeasures to be implemented in an effort to "neutralize" the causes of the most severe types of work-zone collisions.

To realize the stated objective, the specific research tasks to be achieved in this study are as follows: (1) study the manner in which certain variables affect fatal collision type by formulating two types of statistical models (MNL and NB), (2) look into the validity of the MNL models by using the NB models' results, (3) compare the models with similar ones developed for nonwork zone collisions, and (4) based on this comparative framework, identify potential areas in which countermeasures can be developed to reduce the propensity for fatalities in work zones.

2. Background

According to the Federal Highway Administration (FHWA; 2013), 87,606 collisions occurred in work zones across the United States in 2010. Of these work zone–related collisions, more than 500 were fatal (nearly double the fatality rate when observing all collisions, work zone or not; National Highway Traffic Safety Administration [NHTSA], 2010) and more than 25,000 resulted in injuries (FHWA, 2013). Due to the large number of work zone–related collisions—specifically those causing fatalities—it is necessary to look into ways to assess work-zone safety and develop effective countermeasures. To realize such goal, detailed data sets must be analyzed and pertinent variables must be identified. In many previous studies, specific state Department of Transportation (DOT) work-zone crash data was utilized (Chen, 2008; Garber & Zhao, 2002; Ha & Nemeth, 1995; Harb, Radwan, Yan, Pande, & Abdel-Aty, 2008; Khattak, Khattak, & Council, 2002; Khattak & Targa, 2004; Meng, Weng, & Qu, 2010; Schrock, Ullman, Cothron, Kraus, & Voigt, 2004; Li & Bai, 2008). Although a statewide analysis has advantages in terms of implementation, results are specific to the area of analysis. This study aims at using national data in the FARS database to address work-zone collision fatalities on a broader scope. FARS also has a more standardized and consistent data set compared to some of the other data sets used; other research has used data taken directly from site reviews and narratives conducted by researchers that vary depending on the researcher observing the scene and writing the report (Schrock et al., 2004). Within the data sets selected, many combinations of variables can be examined in relation to collisions. Some studies focus on a select number of variables (five or fewer) (Chen, 2008; Garber & Zhao, 2002; Khattak & Targa, 2004), whereas others focus on a larger set of variables (six or more) (Chen, 2008; Harb et al., 2008; Khattak, Khattak, & Council, 2002; Meng et al., 2010; Schrock et al., 2004; Li & Bai, 2008). The most common variables analyzed include injury severity, lighting condition, vehicle type, roadway type, type of work zone, and presence of countermeasures in the work zone. It is important to note that a number of previous studies suggested that further investigations into visibility (Garber & Zhao, 2002) and geometric factors (Khattak, Khattak, & Council 2002) be conducted. Moreover, alcohol involvement and driver distraction have been looked into by researchers as well (Bai & Li, 2006; Meng et al., 2010; Schrock et al., 2004). Based on these suggestions and given the scope of this study, the following variables are included by the authors from the FARS database: crash type, lighting condition, atmospheric condition, roadway alignment, number of lanes, speed limit, alcohol involvement, and driver distraction.

A critical component of any transportation safety study is the modeling technique. Many work-zone studies have used simple regression techniques (Chen, 2008; Garber & Zhao, 2002; Harb et al., 2008; Khattak & Targa, 2004; Meng & Weng, 2011; Li & Bai, 2008) showing the relationship between dependent and independent variables. Some additional modeling techniques include ordered probit models (Khattak & Targa, 2004), Ordinary least squares (OLS) log-transformed models (Khattak & Targa, 2004), NB models (Chen. 2008; Khattak, Khattak, & Council 2002), chi-squared statistics (Li & Bai, 2008), Cochran-Mantel-Haenszel (CMH) statistics (Li & Bai, 2008), conditional logistic regression (Harb et al., 2008; Li & Bai, 2008), multiple logistic regression (Harb et al., 2008), stepwise regression (Meng & Weng, 2011), and Poisson binomial regression (Chen, 2008). Other studies use more qualitative methods for assessing work-zone collision risk (Meng et al., 2010; Schrock et al., 2004). A qualitative approach can give insight into collision hazard but should be used as a second form of analysis as opposed to the primary form. To confirm the statistical significance and accuracy of the results in this article, two types of statistical models (MNL and NB) are used to look into different collision-type probabilities. The findings are compared and validated against each other. Furthermore, a binary probit model is used to compare the effects of the exogenous variables on work zones and nonwork zones. The utilization of multiple modeling techniques not only allows for comparison and confirmation of results, but also demonstrates the types of models that generate statistically significant results for the FARS database.

One major difference between the work-zone and nonwork zone driving scenario is the frequency at which certain collision types occur. Consequently, in their *Traffic Safety Evaluation of Nighttime and Daytime Work Zones* Report, the National Cooperative Highway Research Program (NCHRP) classified a set of work-zone collision data by collision type (National Research Council, 2008). For the 17,228 collisions occurring in California, North Carolina, Ohio, and Washington states, daytime rear-ends accounted for 46.9% and 54.4% of the collisions in active work zones with and without lane closures, respectively. Furthermore, sideswipes accounted for 13.6% and 14.8% of the same type of daytime collisions, respectively. These two types of collisions were shown to be the most common to occur in daytime work zones. Collision type is the main focus of only a few work-zone studies (Chen, 2008; Garber & Zhao, 2002; Meng et al., 2010; Li & Bai, 2008). Of these, a couple focus on the severity of injury caused by different collisions (Chen, 2008; Meng et al., 2010; Li & Bai, 2008); others look into the location of the collision types (Garber & Zhao, 2002); and there has even been analyses on the primary factors leading to the cause of the various types of collisions (Hill, 2003). With this said, there is a need for research that identifies conditions conducive to the various types of fatal collisions that occur in work zones. By identifying these conditions for the work-zone and nonwork zone scenarios, the influences of specific exogenous factors can be isolated, and a more detailed understanding of these hazardous situations can be achieved.

The remainder of this study is organized as follows: a description of the data set and preliminary analysis are provided in Section 3. This is followed by an explanation of the modeling techniques presented in Section 4. Results of all mathematical models for both scenarios are presented and analyzed in Section 5, and conclusions are discussed in Section 6.

3. Data analysis

After defining the variables used in this study and the corresponding data library, a basic exploratory data analysis is conducted in this section. The resulting preliminary findings lead to the use of the two models described in Section 4.

3.1. *Definition of variables*

Data for this study was provided by the FARS and included fatal collisions occurring in 2010, 2011, and 2012 on all roadways in the United States (U.S. Department of Transportation, 2014). Only collisions where all variables were provided

Table 1. Exogenous variable description.

Exogenous Variable	Variable Description
Precipitation/visibility (X1)	Dummy variable: X1 = 0 clear; severe crosswinds, cloudy and X1 = 1 for rain; sleet, hail; snow; fog, smog, smoke; blowing sand, soil, dirt; blowing snow
Lighting (X2)	Dummy variable: X2 = 0 for daylight, dawn, dusk and X2 = 1 for dark, dark lighted, dark-unknown lighting
Alignment (X3)	Dummy variable: X3 = 0 for a straight roadway and X3 = 1 elsewhere
Alcohol involvement (X4)	Dummy variable: X4 = 0 for no alcohol involvement and X4 = 1 for alcohol involvement
Distraction (X5)	Dummy variable: X5 = 0 for no driver distraction and X5 = 1 for driver distraction
Speed limit (X6)	Speed limit on the roadway
Number of lanes (X7)	The number of lanes for the direction of travel in which the accident occurred

(i.e., complete data points) were considered for analysis; and in total this data set contained 856 fatal work-zone collisions along with 50,014 fatal collisions not occurring in work zones. One drawback of this data set is the challenge faced when trying to link collisions with the average annual daily traffic (AADT) specific to the roadway on which the collision occurred. The inclusion of a "linking" variable in the FARS data set would be helpful to researchers in the future. At this stage, the variables considered for analysis are shown in Table 1. Additional details on the description of collision type classification scheme (Table 2) may be found in the NHTSA FARS data description report (U.S. Department of Transportation, 2014).

3.2. *Preliminary data analysis*

Preliminary analysis began with a compilation of collision type frequency measured across both scenarios. These results are displayed in Table 3.

Observation of Table 3 indicates a marked increase in the percentage of fatal rear-end collisions (nearly 4 times as frequent) as well as fatal side-swipe collisions

Table 2. Endogenous variable description.

Crash Type Category	FARS Variable Description[a]
No impact (Y0)[b]	No impact
Single vehicle (Y1)[b]	Configuration A: Right roadside departure; Configuration B: Left roadside departure; Configuration C: Forward impact (Category I)
Rear-end (Y2)[b]	Configuration D: Rear end (Category II)
Frontal impact (Y3)[b]	Configuration E: Forward impact (Category II)
Sideswipe/angle (Y4)[b]	Configuration F: Sideswipe/angle (Category II)
Different directions (Y5)[b]	Configuration G: Head-on (Category III); Configuration H: Forward impact (Category III); Configuration I: Sideswipe/angle (Category III)
Turning (Y6)[b]	Configuration J: Turn across path; Configuration K: Turn into path (Category IV)
Other (Y7)[b]	Configuration L: Straight paths (Category V); Configuration M: Backing, etc. (Category VI)

[a]Refer to U.S. Department of Transportation (2011), 243–245 for a more detailed description of categories and configurations.
[b]1 if corresponding crash type occurred; 0 otherwise.

(nearly double the frequency) when comparing work-zone collisions to those not occurring in work zones. Results from a two sample t test conducted on the work-zone and nonwork zone frequency numbers shows that the observed differences between the numbers found for each collision type are statistically significant to the 95% level of confidence; the t statistic (–2.0256) and p value (0.03115) are less than $p = .05$. The increase in the occurrence of these specific collision types is an important observation for preliminary analysis. Considering the stop-and-go nature of traffic in work zones along with the inherent distraction of an atypical driving scenario, there is no surprise that RESS collisions are of paramount concern. Moreover, congestion can be associated with bottlenecks where the supply is greater than the demand: a condition that is faced in work-zone lane-closure scenarios leading to stop-and-go congestion dynamics with higher acceleration variance (Edara, Kianfar, & Sun, 2012). This illustrates the need for additional consideration to be given to the corresponding car-following and lane-changing maneuvering errors (i.e., related to rear-end and side-swipe collisions, respectively) associated with these fatal collision types. Analysis conducted in Section 5 will pay special attention to the factors surrounding these specific collision types.

Collision types were grouped together for analytical modeling purposes, leading to the following collision type categories: Single vehicle (SV) (includes only SV collisions) and RESS (includes only RESS collisions—the two types with the highest frequency increase in work zones); and different directions (DD) (includes all head-on and turning collisions—both categories where vehicles are initially traveling in different directions). Support for these groupings, specifically RESS collisions, is derived from the literature on collisions occurring at signalized intersections (Rodegerdts et al., 2004). Because work zones present a traffic-controlled scenario, parallels can be drawn to the conditions present at four-way stop controlled and signalized intersections. According to the FHWA report, rear-end and angle (or side-swipe) collisions are frequently caused by sudden or unexpected slowing, too much slowing or stopping due to turbulent traffic flow, and sudden or unexpected slowing due to inadequate capacity—all of which are present when dealing with a work-zone scenario (FHWA, 2013). Furthermore, suggested improvements to mitigate these collisions include improved traffic control and approach improvement—both of which are elaborated upon in the conclusion section of this study. This grouping scheme is used in the following sections.

4. Modeling techniques

Multiple attempts were made in an effort to model the occurrence of the selected collision types in terms of the previously identified variables. Due to the nature of the data (multiple dummy variables, a limited number of observations in the work-zone set, and a large number of different collision types), achieving a converging model for either a multinomial probit or a structural equation formulation became an exhaustive task that produced few meaningful results. In lieu of this, the data was modeled using a NB regression model and a MNL model for each case (work-zone and nonwork zone collisions).

A NB regression model was used to establish the relationship between collision type and various environmental, behavioral, and roadway geometric parameters. This technique allows for flexibility in the type of distributions that can be analyzed, separates the assumptions of the mean from the dispersion, and is able to model very skewed data (Hill, 2003). In Equation 1, it is assumed that the number of collisions (Y) is the dependent variable and is independently and negative binomially distributed with parameters a and λ_i given the independent variables x_1, x_2, …, x_i (corresponding to the exogenous covariates defined in Table 1). The probability of the occurrence of y_i number of collisions for collision type i is given as

$$P(Y=y_i)=\frac{\Gamma\left(\frac{1}{a}+y_i\right)}{\Gamma\left(\frac{1}{\mathrm{a}}\right)\Gamma\left(\mathrm{y_i}+1\right)}\left(\frac{1}{1+a\lambda_i(X_i,\beta)}\right)^{\frac{1}{a}}(1-\frac{1}{1+a\lambda_i(X_i,\beta)}^{y_i} \tag{1}$$

where, a = dispersion parameter of NB model and λ_i = mean number of crashes for collision type i and is defined as

$$\lambda_i=E(Y=y_i)=e^{\sum_{j=1}^{k}x_{ij}\beta_j} \tag{2}$$

The maximum likelihood function is used to estimate the β_{j} coefficients. Due to the fact that this model looks at each collision type separately and a comparative model is required to gain additional insight and confirm regression results.

For this reason, a MNL model was developed. MNL models explain and predict discrete choices using estimation that is based on the utility theory (Ben-Akiva & Lerman, 1985). This model compares the probability of each of the j = 1, 2,…,$\mathrm{J}_{\mathrm{i\text{-}1}}$ categories (collision type for this study) to the probability of the baseline category J_i; for this analysis, the J_i is single-vehicle collisions (SV). MNL was used to give the likelihood of each collision type in terms of the environmental and roadway parameters. The main advantage of this model is that it is easy to compute (as compared to a multinomial probit model); but it is important to keep in mind that the associated error term is logistically distributed and the model assumes the independence of irrelevant alternatives (IIA) (Hilbe, 2011).

In the MNL model there is one vector of characteristics describing each choice case and a set of J parameter vectors (corresponding to the collision type defined earlier). Each J alternative has a set of K attributes (x_{ij}) (corresponding to the exogenous covariates defined in Table 1). In each case, respondent i' (corresponding to the driver who is involved in the collision) makes a choice from the J alternatives. MNL has one parameter vector, β (weights corresponding to the exogenous variables). The random utility for the model is given as

$$U(choice\ j\ for\ individual\ i) = U_{ij} = \beta x_{ij} + \varepsilon_{ij}\ ,\ j = 1, \ldots, J_i \tag{3}$$

where, the random individual error terms (ε_{i1}, ε_{i2},…, ε_{iJ}) are assumed independently distributed, each with a value distribution described as extreme.

The general form of the MNL model is

$$Prob(choice\ j) = \frac{\exp\left(\beta_j x_i\right)}{\sum_{q=0}^{J} \exp\left(\beta_j x_i\right)}, j = 0, \ldots, J. \tag{4}$$

where, *Prob(choice j)* = probability of outcome *j*, x_i = vector of exogenous variables, and β_j = coefficients found using the maximum likelihood estimation (Washington, Karlaftis, & Mannering, 2011).

Lastly, a third modeling technique is employed to further demonstrate the consistency of results as well as the techniques that can be applied to the data set. A binary probit model determines the probability for the choice of one of two outcomes given a set of independent variables. Here, the binary probit model is used to compare work-zone collisions and nonwork zone collisions based on the variables presented earlier. This model assumes that its error follows a normal distribution, making it easier to interpret (Dow & Endersby, 2004). The probability that a given collision type occurs in a work zone is given as

$$\Pr(y = 1 \mid x_i) = F(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k) \tag{5}$$

where, x_i = vector of exogenous variables, β_j = coefficients estimated using the maximum likelihood function, and *F(.)* follows a normal cumulative distribution. The probability that a given collision type occurs in a nonwork zone is given as

$$\Pr(y = 0 \mid x_i) = 1 - \Pr(y = 1 \mid x_i) \tag{6}$$

5. Numerical results and analysis

Results from the NB regressions (completed using the SAS 9.3 software) are presented in Tables 4 and 5.

In the tables mentioned above, *p* values less than 0.05 are considered to be statistically significant (and these instances have are recognized by a superscript letter). Given that there were many more nonwork zone collisions, it is not surprising that results for the nonwork zone scenario are more statistically significant.

Looking first at the coefficient estimates for the work-zone scenario there are a number of interesting observations that can be made. For the RESS collisions, increases in the number of lanes and the speed limit lead to additional instances of

Table 3. Collision type frequency.

	No Impact	Single Vehicle	Rear End	Frontal Impact	Sideswipe	Head On	Turning	Other
Work zones								
Frequency	18	366	211	3	42	139	46	31
Percent	2.10	42.76	24.65	0.35	4.91	16.24	5.37	3.62
Nonwork zones								
Frequency	759	26198	3482	25	1205	9235	5113	3997
Percent	1.52	52.38	6.96	0.05	2.41	18.46	10.22	7.99

this collision type whereas the presence of a curve, precipitation, or alcohol lead to a decrease in this collision type. Contrary to this finding, the estimates for SV collisions indicate that the presence of a curve and alcohol lead to increases in the frequency of fatal collisions. Another article found that SV collisions are more likely to occur with the presence of alcohol (Bai & Li, 2006). Similar results were obtained in an additional study that found that having a curve in the roadway increases the chances of a severe SV collision (Daniel, Dixon, & Jared, 2000). Furthermore, for the DD collision type, estimates indicate that the presence of precipitation as well as a decrease in the speed limit, and the absence of alcohol and driver distraction will increase the collision type frequency.

Comparing these results with the regression model for the nonwork zone scenario, the first observation that can be made is the consistency across all estimates of the sign for each significant coefficient. This indicates that for both scenarios, the same variables influence collision occurrence in the same manner (either positive or negative) with the only change being the magnitude of this influence. For example, though increasing the number of lanes will lead to more RESS collisions in both scenarios, this increase has less of an effect in the work-zone scenario (0.1185) than it does in the nonwork zone scenario (0.3450).

Next, a MNL was computed for both scenarios. Results for the MNL model (completed using the LIMDEP software) are presented in Tables 6 and 7 and the goodness-of-fit statistics are presented in Table 8.

Table 4. Negative binomial regression: work-zone collisions.

	Single Vehicle		Rear-End/Sideswipe		Different Directions	
Work Zones	Estimate	*p* Value	Estimate	*p* Value	Estimate	*p* Value
Constant	**−0.7464**	0.0086	**−2.4855**	<.0001	−0.1876	0.6202
PRECIP	0.1267	0.4738	**−1.4155**	0.0018	**0.5738**	0.007
LIGHT	0.1555	0.1966	−0.2411	0.0984	−0.0082	0.9602
CURVE	**0.3874**	0.0011	**−1.3144**	<.0001	−0.006	0.9745
ALCOHOL	**0.5682**	<.0001	**−0.8348**	<.0001	**−0.5442**	0.0071
DISTRACT	0.0425	0.7563	0.166	0.2535	**−0.4885**	0.0324
SPDLMT	−0.0045	0.3251	**0.0256**	<.0001	**−0.0208**	0.0009
LANES	−0.0841	0.1407	**0.1185**	0.0391	−0.0143	0.8564

PRECIP = Presence of precipitation; LIGHT = daylight conditions; CURVE = presence of horizontal and/or vertical curve in the roadway; ALCOHOL = presence of driver under the influence of alcohol; DISTRACT = presence of driver distraction; SPDLMT = speed limit; LANES = number of lanes. The estimates in bold have $p < 0.05$, showing that they are statistically significant.

The base case for the MNL model was that of SV collisions. Once again, p values less than 0.05 were considered to be significant, and these instances have been highlighted with a superscript letter. Additional fit criteria were considered for the MNL model—for the individual variables standard error should be less than 2.00 (demonstrates that colinearity does not exist) and b/SE should be greater than 2.00. Additionally, for the model as a whole the absolute value of the log likelihood should be less than that of the restricted log likelihood for good data (as is the case in both models), and the chi-squared statistic should be greater than 36.12 (the cutoff for significance on the 0.001 level for a model with 14 degrees of freedom).

Looking at the MNL model for the work-zone scenario, results indicate that for RESS collisions the presence of darkness and a curve in the roadway have a negative correlation with the event of a collision in comparison to SV; whereas increasing the number of lanes and speed limit are positively correlated with RESS collisions when compared to SV. Additionally, there are inverse relationships for RESS collisions and presence precipitation, daylight, the presence of a curve in the roadway, and alcohol involvement. Moving to DD collisions, results indicate that the presence of a curve in the roadway, alcohol, and driver distraction are less likely to lead to a DD collision when compared to SV. These results are consistent with those from Harb et al. (2008), who found that when alcohol is involved drivers are more likely to have a ***SV*** collision in a work zone. The positive relationship between increasing the number of lanes and increasing DD collisions is questionable given the fit statistics.

Studying the nonwork zone scenarios, results for RESS collisions indicate that the presence of precipitation, darkness, and a curve in the roadway are less likely to be associated with this collision type in comparison to SV and increasing the number of lanes and speed limit are more likely to be associated with a RESS collision than a SV. Additionally, the presence of a curve in the roadway and alcohol decreases the likelihood of RESS collisions in relation to SV. It should be noted that in another study, the likelihood of SV collisions in work zones compared to nonwork zones increases with the occurrence of darkness (Harb et al., 2008). This is consistent with the larger coefficient values for lighting variable in the work-zone data compared to the nonwork zone data. Looking at DD collisions, results demonstrate that the presence of precipitation and increasing the number of lanes are more likely to result in a DD collision whereas the presence of darkness, a curve in the roadway, alcohol, driver distraction, and increasing speed limit are less likely to result in these collisions in comparison to SV.

As was the case for the NB regression, there is a consistency in terms of the sign of each statistically significant variable across both scenarios in the MNL model. For example, in work-zone and nonwork zone scenarios, an increase in the number of lanes is indicative of an increase in the number of RESS collisions (when compared to SV); but this increase in the number of lanes has less of an effect (0.2543) for work-zone collisions than it does for nonwork zone collisions (0.5672).

5.1. *Critical collision type: Rear end and sideswipes*

As mentioned earlier, examination of frequency data for work-zone collision type indicates a marked 20.21% increase in the number of fatal RESS collisions when compared to nonwork zone scenarios. This finding is reinforced by the results of another analysis that found that the percentage of RESS collisions in work zones was considerably higher than the percentage of the same type of collisions in non-work zones subject areas in Ohio (Ha & Nemeth, 1995). Furthermore, it has been suggested that RESSs are the most common collision types in work zones (Garber & Zhao, 2002). Identifying conditions that are conducive to these fatal collision types is helpful in creating a safer driving environment.

Looking first at results from the NB model, when compared to the other fatal collision type categories, RESS collisions are most sensitive to changes in precipitation, roadway curvature, alcohol involvement, speed limit, and the number of lanes in the work-zone scenario and changes in precipitation, roadway curvature, alcohol involvement, and the number of lanes in the nonwork zone scenario. Supporting these results, another study found that the likelihood of a severe collision in a work zone increases with the speed limit (Khattak & Targa, 2004). Results indicating that rear-end collisions in work zones increase with higher speed limits were also found by Meng and Weng (2011). For both scenarios, RESS collisions occur more frequently when there is no precipitation and the roadway is straight as well as with an increase in the number of lanes and speed limit.

Results from the MNL model for RESS collisions demonstrate complete consistency when compared with the validation results of the NB model as the sign of all variables is the same for all variables across both scenarios. Furthermore, significant values for lighting conditions and alcohol involvement indicate that RESS collisions occur more frequently during the day and without the presence of alcohol when compared to SV collisions in work-zone and nonwork zone scenarios.

Table 5. Negative binomial regression: nonwork zone collisions.

	Single Vehicle		Rear-End/Sideswipe		Different Directions	
Nonwork Zones	Estimate	*p* Value	Estimate	*p* Value	Estimate	*p* Value
INTERCEPT	**−0.5486**	<.0001	**−4.1871**	<.0001	**−0.4175**	<.0001
PRECIP	−0.0397	0.0581	**−0.4058**	<.0001	**0.187**	<.0001
LIGHT	**0.1638**	<.0001	**0.0755**	0.0159	**−0.3373**	<.0001
CURVE	**0.3185**	<.0001	**−1.4005**	<.0001	**−0.4018**	<.0001
ALCOHOL	**0.3408**	<.0001	**−0.5282**	<.0001	**−0.6593**	<.0001
DISTRACT	**0.1109**	<.0001	**0.13**	0.0013	**−0.2868**	<.0001
SPDLMT	0.0004	0.3874	**0.0267**	<.0001	**−0.0086**	<.0001
LANES	**−0.1583**	<.0001	**0.345**	<.0001	**0.0502**	<.0001

PRECIP = Presence of precipitation; LIGHT = daylight conditions; CURVE = presence of horizontal and/or vertical curve in the roadway; ALCOHOL = presence of driver under the influence of alcohol; DISTRACT = presence of driver distraction; SPDLMT = speed limit; LANES = number of lanes. The estimates in bold have $p < 0.05$, showing that they are statistically significant.

Table 6. Multinomial logit model: work zone collisions.

	Rear-End/Sideswipe		Different Directions	
Work Zones	Estimate	*p* Value	Estimate	*p* Value
Constant	−**2.2005**	0.0001	0.4767	0.3366
PRECIP	−**1.7779**	0.0005	0.3946	0.1711
LIGHT	−**0.4936**	0.0166	−0.1933	0.3512
CURVE	−**1.9244**	0.0000	−**0.4556**	0.0438
ALCOHOL	−**1.5304**	0.0000	−**1.0997**	0.0000
DISTRACT	0.1573	0.4865	−**0.5342**	0.0500
SPDLMT	**0.0381**	0.0000	−0.0150	0.0673
LANES	**0.2543**	0.0059	0.0886	0.3834
Fit Statistics	Standard Error (SE)	*b*/SE.	SE	*b*/SE
Constant	0.5469	−4.0240	0.4962	0.9610
PRECIP	0.5077	−3.5020	0.2883	1.3690
LIGHT	0.2061	−2.3950	0.2074	−0.9320
CURVE	0.3311	−5.8110	0.2260	−2.0160
ALCOHOL	0.2507	−6.1050	0.2340	−4.7000
DISTRACT	0.2261	0.6960	0.2726	−1.9600
SPDLMT	0.0088	4.3180	0.0082	−1.8290
LANES	0.0924	2.7510	0.1017	0.8720

PRECIP = Presence of precipitation; LIGHT = daylight conditions; CURVE = presence of horizontal and/or vertical curve in the roadway; ALCOHOL = presence of driver under the influence of alcohol; DISTRACT = presence of driver distraction; SPDLMT = speed limit; LANES = number of lanes. The estimates in bold have $p < 0.05$, showing that they are statistically significant.

Analysis to this point has demonstrated that, for both scenarios, RESS collisions occur more commonly during the most basic driving conditions—daylight hours, no precipitation, and less roadway curvature. Two previous studies have concluded that fatal work-zone crashes are less influenced by horizontal and vertical curves

Table 7. Multinomial logit model: nonwork zone collisions.

	Rear-End/Sideswipe		Different Directions	
Nonwork Zones	Estimate	*p* Value	Estimate	*p* Value
Constant	−**3.80873**	0.00000	0.08970	0.12780
PRECIP	−**0.37456**	0.00000	**0.23879**	0.00000
LIGHT	−**0.14177**	0.00010	−**0.54465**	0.00000
CURVE	−**1.79340**	0.00000	−**0.78870**	0.00000
ALCOHOL	−**0.96898**	0.00000	−**1.04456**	0.00000
DISTRACT	−0.00897	0.84810	−**0.43345**	0.00000
SPDLMT	**0.02792**	0.00000	−**0.00899**	0.00000
LANES	**0.56725**	0.00000	**0.24894**	0.00000
Fit Statistics	Standard Error (SE)	*b*/SE.	SE	*b*/SE
Constant	0.0936	−40.7000	0.0589	1.5230
PRECIP	0.0616	−6.0780	0.0348	6.8700
LIGHT	0.0361	−3.9320	0.0241	−22.5550
CURVE	0.0547	−32.7580	0.0247	−31.9300
ALCOHOL	0.0430	−22.5420	0.0271	−38.5410
DISTRACT	0.0468	−0.1920	0.0336	−12.8900
SPDLMT	0.0014	19.6340	0.0009	−10.0010
LANES	0.0167	34.0370	0.0138	18.0980

PRECIP = Presence of precipitation; LIGHT = daylight conditions; CURVE = presence of horizontal and/or vertical curve in the roadway; ALCOHOL = presence of driver under the influence of alcohol; DISTRACT = presence of driver distraction; SPDLMT = speed limit; LANES = number of lanes. The estimates in bold have $p < 0.05$, showing that they are statistically significant.

Table 8. Multinomial logit model: goodness-of-fit statistics.

Likelihood Estimates	Work Zones	Nonwork Zones
Degrees of freedom	14	14
Chi-squared value	238.1741	8988.157
Log-likelihood function	−736.7301	−36970.61
Restricted log likelihood	−855.8171	−41464.69

compared to nonwork zone crashes (Daniel et al., 2000; Harb et al., 2008). One of these studies also found that collisions occur more frequently without the presence of precipitation in work zones compared to nonwork zones (Harb et al., 2008). Additionally, increases in speed limit and the number of lanes also increase the frequency of this collision type.

Now that the conditions conducive to RESS collisions have been identified, the next step is to compare coefficient values for all variables in both models across both scenarios. This comparison is presented in Table 9.

Results were again consistent for both models as the variables more influential for the work-zone scenario were identical in both statistical models (inherently the same is true for variables influencing nonwork zone collisions).

This comparative analysis further solidifies the finding that RESS collisions occur more frequently when there is no precipitation, during daylight conditions, with limited roadway curvature, no alcohol involvement, and with increasing number of lanes and speed limits. Additionally this comparison indicates that clear, daylight conditions with increasing speed limits are even more conducive to this collision type when considering work-zone conditions

Table 9. Rear end/sideswipe collisions: coefficient comparison.

Negative Binomial	Work Zone	Nonwork Zone	Greater Magnitude
Constant	**−2.4855**	**−4.1871**	—
PRECIP	**−1.4155**	**−0.4058**	Work zone
LIGHT	−0.2411	**0.0755**	Work zone
CURVE	**−1.3144**	**−1.4005**	Non–work zone
ALCOHOL	**−0.8348**	**−0.5282**	Work zone
DISTRACT	0.1660	**0.1300**	Work zone
SPDLMT	**0.0256**	**0.0267**	Non–work zone
LANES	**0.1185**	**0.3450**	Non–work zone
Multinomial logit	Work Zone	Non–work Zone	Greater Magnitude
Constant	**−2.2005**	**−3.8087**	—
PRECIP	**−1.7779**	**−0.3746**	Work zone
LIGHT	**−0.4936**	**−0.1418**	Work zone
CURVE	**−1.9244**	**−1.7934**	Work zone
ALCOHOL	**−1.5304**	**−0.9690**	Work zone
DISTRACT	0.1573	−0.0090	Work zone
SPDLMT	**0.0381**	**0.0279**	Work zone
LANES	**0.2543**	**0.5672**	Non–work zone

PRECIP = Presence of precipitation; LIGHT = daylight conditions; CURVE = presence of horizontal and/or vertical curve in the roadway; ALCOHOL = presence of driver under the influence of alcohol; DISTRACT = presence of driver distraction; SPDLMT = speed limit; LANES = number of lanes. The estimates in bold have $p < 0.05$, showing that they are statistically significant.

Table 10. Rear end/sideswipe collisions: binary probit (BP).

BP	Estimate	*p* Value	Standard Error (SE)	*b*/SE
Constant	−**1.9208**	0.0000	0.1710	−11.2340
PRECIP	−**0.6274**	0.0004	0.1787	−3.5100
LIGHT	−**0.1803**	0.0077	0.0676	−2.6670
CURVE	−**0.2894**	0.0276	0.1313	−2.2030
ALCOHOL	−**0.1931**	0.0366	0.0924	−2.0900
DISTRACT	**0.2607**	0.0004	0.0742	3.5130
SPDLMT	**0.0096**	0.0004	0.0027	3.5540
LANES	−**0.0564**	0.0470	0.0284	−1.9870
Likelihood estimates				
Degrees of freedom			7	
Chi-squared value			69.58208	
Log-likelihood function			–976.6529	
Restricted log likelihood			−1011.444	

PRECIP = Presence of precipitation; LIGHT = daylight conditions; CURVE = presence of horizontal and/or vertical curve in the roadway; ALCOHOL = presence of driver under the influence of alcohol; DISTRACT = presence of driver distraction; SPDLMT = speed limit; LANES = number of lanes. The estimates in bold have $p < 0.05$, showing they are statistically significant.

alone. Less roadway curvature and increasing number of lanes are more conducive to RESS collisions when considering the nonwork zone scenario.

In an effort to further confirm these findings, a binomial probit model was formulated (using the LIMDEP software) for all RESS collisions across work-zone and nonwork zone scenarios. Keeping the scenario type (either work zone or nonwork zone) as the dependent variable, model results are presented in Table 10.

Based on the likelihood estimates the model demonstrates statistical significant and, once again, variables with statistically significant results are displayed with a superscript letter. Observation of the variables found to be statistically significant in the NB and MNL models demonstrates complete consistency with the results discussed above.

From this analysis, it becomes increasingly apparent that there exists an opportunity to develop countermeasures to help prevent fatal RESS collisions in work zones. Considering the prominent conditions for these collisions, it appears that the major factors leading to such collisions are the surrounding traffic control conditions and the corresponding driver decision making rather than the uncontrollable environmental conditions or complicated roadway geometry. With an increasing speed limit and/or number of lanes on a roadway, there is inherently more traffic; and given the diminishing effects that these increases have on safety in work-zone scenarios, potential ITS applications should be targeted at creating a safer traffic flow conditions by encouraging safer driver maneuvers. Such ITS applications may include speed harmonization control methods and vehicle to vehicle communication driving assistance systems. Furthermore, results indicate that the use of common control measures such as speed enforcement cameras may need to be reevaluated given the marginal difference in the effect of speed limit when comparing work-zones to nonwork zones.

6. Conclusions and future work

In this article fatal collision types for work zones and nonwork zones were analyzed through a variety of mathematical modeling procedures. A NB regression was carried out to examine the individual effects of exogenous variables on each specific collision type; a MNL model was developed to confirm these results and gain perspective on how coefficient values vary when comparing one collision type to another; and finally a binomial probit was utilized to further solidify the findings of this study. Statistically significant results were achieved for the majority of variables in each model presented for the work-zone and nonwork zone scenarios. By utilizing multiple modeling techniques, the authors have worked to identify converging statistical models that can be achieved with FARS data. In general, results demonstrated that as environmental and roadway characteristics vary, so does the propensity for the different fatal collision types.

Based on a simple frequency analysis as well as findings from previous research, RESS collisions were identified as particularly problematic in work-zone scenarios. For this reason, special attention was paid to fatal collisions of these two types specifically, and results demonstrated that clear, daylight conditions on straight roadways without the presence of alcohol are more conducive to fatal RESS collisions. Additionally, increasing the number of lanes and speed limit increases the propensity for these types of fatal collisions—indicating that the traffic flow conditions surrounding the work zone are the major contributing factor to these specific types of fatal collisions. Based on these findings, suggested countermeasures include speed harmonization, vehicle-to-vehicle communication logics, the use of dynamic message signs to update drivers on delays and alternative routes, and the implementation of temporary work-zone rumble strips (Texas Department of Transportation, 2013). Some other more conventional traffic engineering solutions include increasing the upstream distance at the start of the work zone where cones and barrels are placed, as well as reducing the speed limit with the help of flashing lights. By increasing the warning distance leading up to a work zone and reducing the number of lanes with barrels and cones, drivers will be more likely to slow down and be less likely to be involved in a collision (National Cooperative Highway Research Program, 2005). Additionally, adaptive traffic-calming techniques could be applied even further upstream of the work zone in an attempt to create a smoother transition and safer flow scenario. Future work should explore the implementation of such techniques as well as examine different ways to extend the FARS database such that collisions can be linked with existing traffic data (such as AADT).

Funding

This work was supported by the National Science Foundation under Grant No. 0927138. Any opinions, findings, conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Bai, Y., & Li, Y. (2006). Determining major causes of highway work zone accidents in Kansas (No. K-TRAN: KU-05-1). Kansas Department of Transportation.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: MIT Press.

Chen, Y. (2008). *A highway work zone design and traffic management decision system* (Unpublished doctoral dissertation). The University of Wisconsin – Madison. Retrieved from http://search.proquest.com/docview/304463256?accountid = 11243.

Daniel, J., Dixon, K., & Jared, D. (2000). Analysis of fatal crashes in Georgia work zones. *Transportation Research Record*, *1715*, 18–23.

Dow, J. K., & Endersby, J. W. (2004). Multinomial probit and multinomial logit: a comparison of choice models for voting research. *Electoral Studies*, *23*(1), 107–122. ISSN 0261-3794. http://dx.doi.org/10.1016/S0261-3794(03)00040-4.

Duffy, S., & McAvoy, D. S. (2009). *Demonstration of innovative techniques for work zone safety data analysis* (Tech. no. CSUTC-TR-09-02). Cleveland, OH: Cleveland State University, Transportation Center.

Edara, P., Kianfar, J., & Sun, C. (2012). Analytical methods for deriving work zone capacities from field data. *Journal of Transportation Engineering*, *138*(6), 809–818.

National Highway Transportation Safety Administration. (2010). *Fatality Analysis Reporting System*. Retrieved from www-fars. nhtsa. dot. gov. Accessed August 17, 2015.

Federal Highway Administration. (2013, July 2). *Facts and statistics - Work zone injuries and fatalities*. Work Zone Mobility and Safety Program, Federal Highway Administration. Retrieved from http://www.ops.fhwa.dot.gov/wz/resources/facts_stats/injuries_fatalities.htm.

Garber, N. J., & Zhao, M. (2002). *Crash characteristics at work zones* (Vol. 02-R12). Charlottesville, VA: Virginia Transportation Research Council.

Ha, T.-J., & Nemeth, Z. A. (1995). Detailed study of accident experience in construction and maintenance zones. *Transportation Research Record*, *1509*, 38–45.

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

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge, England: Cambridge University Press.

Hill, R. W. (2003). *Statistical analysis of fatal traffic accident data* (Master's thesis). Texas Tech University.

Khattak, A., & Targa, F. (2004). Injury severity and total harm in truck-involved work zone crashes. *Transportation Research Record*, *1877*(1), 106–116.

Khattak, Asad J., Khattak, A. J., & Council, F. M. (2002). Effects of work zone presence on injury and non-injury crashes. *Accident Analysis & Prevention*, *34*(1), 19–29.

Li Y., & Bai, Y. (2008). Development of crash-severity-index models for the measurement of work zone risk levels. *Accident Analysis & Prevention*, *40*(5), 1724–1731.

Mathes, J. A. (2012). *Integrated risk management for improving internal traffic control, work zone safety, and mobility during major construction*. Graduate Theses and Dissertation, Iowa State University. Paper 12734.

Meng, Q., & Weng, J. (2011). Evaluation of rear-end crash risk at work zone using work zone traffic data. *Accident Analysis & Prevention*, *43*(4), 1291–1300.

Meng, Q., Weng, J., & Qu, X. (2010). A probabilistic quantitative risk assessment model for the long-term work zone crashes. *Accident Analysis & Prevention*, *42*(6), 1866–1877.

Milton, J., & Mannering, F. (1998). The relationship among highway geometrics, traffic-related and motor-vehicle accident frequencies. *Transportation*, *25*, 395–413.

National Cooperative Highway Research Program. (2005). *Guidance for implementation of the AASHTO Strategic Highway Safety Plan* (Publication no. 500. Vol. 17). Washington, DC: Transportation Research Board.

National Research Council. (2008). *Traffic safety evaluation of nighttime and daytime work zones* (NCHRP Report 627). Washington, DC: National Academies Press.

Rodegerdts, L. A., Nevers, B., Robinson, B., Ringert, J., Koonce, P., Bansen, J., … Courage, K. (2004). *Signalized intersections: Informational guide* (Rep. no. FHWA-HRT-04-091). McLean, VA: Federal Highway Administration.

Schrock, D. S., Ullman, G. L., Cothron, A. S., Kraus, E., & Voigt, A. P. (2004). *An analysis of fatal work zone crashes in Texas* (Report FHWA/TX-05/0-4028-1). Texas Transportation Institute, TX.

Texas Department of Transportation. (2013). *TxDOT Rolls out new traffic tools to increase safety for motorists, highway construction workers*. TX: Texas Department of Transportation.

U.S. Department of Transportation. (2014). *Fatality Analysis Reporting System (FARS) analytical user's manual (1975-2012)* (Technical report DOT HS 811 855). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Washington, S., Karlaftis, M., & Mannering, F. (2011). *Statistical and econometric methods for transportation data analysis*. (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.