

# Freeway Work-Zone Crash Analysis and Risk Identification Using Multiple and Conditional Logistic Regression

Rami Harb ; Essam Radwan, Ph.D., P.E. ; Xuedong Yan, Ph.D. ; Anurag Pande, Ph.D. ; and Mohamed Abdel-Aty, Ph.D., P.E.

**Abstract:** Work-zone safety continues to be a priority and a concern for the Federal Highway Association as well as most state departments of transportation. The main objective of this study is to uncover work-zone freeway crash characteristics to help develop countermeasures that limit work-zones' hazards. The Florida Crash Records Database for years 2002, 2003, and 2004 was utilized for this study. Conditional logistic regression along with stratified sampling and multiple logistic regression models were estimated to unveil work-zone freeway crash traits. According to the models' results, roadway geometry, weather condition, age, gender, lighting condition, residence code, and driving under the influence of alcohol and/or drugs are significant risk factors associated with work-zone crashes.

## Introduction

Work-zone safety continues to be a priority and a concern for the Federal Highway Association (FHWA) as well as most state Departments of Transportation (DOTs). In fact, the concurrent climb in roadway work-zone activity nationwide, especially in the state of Florida, has produced an increase in work-zone crashes and fatalities. According to the Fatality Analysis Reporting System (FARS), Florida fatal work-zone crashes have risen 334% since 1999, ranking Florida the second highest state in fatal work-zone crashes in 2004 after the state of Texas (FARS 2006). Several studies were undertaken to assess the safety of highway construction zones (Hall and Lorenz, 1989; Ha and Nemeth 1995; Garber and Woo 1990; Roupail et al. 1988; Wang et al. 1996) in numerous states in the United States. These studies corroborate that work zones produce a significantly higher rate of crashes under certain conditions when compared to nonwork-zone locations. In

particular, Hall and Lorenz (1989) stated that work zones are responsible for a 26% increase in motor vehicle crashes during construction or roadway maintenance. Moreover, Roupail et al. (1988) showed that crash rates during construction activities augmented by 88% compared to the before period at work zones. Garber and Woo (1990) stated that, on average, work-zone accident rates increased approximately by 57% on two-lane urban highways. Zhao (2001) investigated the characteristics of work zone crashes in Virginia for years 1996–1999 and concluded that work zones involve a higher proportion of fatal crashes than nonwork-zone locations. These facts underscore the urgent need to develop a substantive understanding about how work-zone crashes occur and their corresponding risk factors.

Studies on work-zone crashes have typically inspected a combination of injury, fatal, and property damage crashes to discover aspects that contribute to unsafe conditions within work zones. Daniel et al. (2000) focused only on the analysis of fatal crashes within work zones in Georgia since their database did not identify work zones unless there was a fatal injury. This study examined the difference between fatal crashes within work zones compared with fatal crashes in nonwork-zone locations. The overall findings of the study indicate that work zones influence the manner of collision, lighting conditions, truck involvement, and roadway functional classification under which fatal crashes occur. Ming and Garber (2001) conducted research to uncover work-zone crash attributes accounting for the location of each crash within the work zone and its surroundings in Virginia. However, their study strictly presented statistical summaries and basic inferential statistics of these crashes and their attributes without relating to interactions and confounding effects. This study concluded that work-zone crashes are predominant in the activity area and that there is a higher rate of multivehicle accidents in work-zone locations compared to nonwork-zone locations. Benekohal et al. (1995) considered exclusively the effect of trucks and their involvement in work-zone crashes. Their study indicated that the accident experiences were significantly related to the experience of bad driving situations but not other driver/truck characteristics. However, other studies showed that heavy vehicles were over-represented in work-zone areas (Hall and Lorenz 1989; Pigman

and Agent 1990; Nemeth and Rathi (1983)). Garber and Zhao (2002a,b) suggested that a major causal factor for work-zone crashes is speed related. The accidents are mainly caused by speed differentials resulting in a speed variance. Raub et al. (2001) indicated that distraction from work in progress, failure to yield at the taper point, and excessive speed are overrepresented causes for work zone crashes.

The lack of literature concerns the overall aspect of the crash traits at work zones such as environment, vehicle, and driver characteristics and their interactions. Therefore, this study aims to evaluate freeway single-vehicle and two-vehicle crashes in work zones to identify their drivers/vehicles/environment traits accounting for interactions and confounding factors. For that purpose, the Florida Traffic Crash Records Database for years 2002, 2003, and 2004 is employed. The first section of this paper describes in detail the methodology used in conducting the analysis. The second section elaborates on the statistical modeling for the single and the two-vehicle crashes at work zones. The third part summarizes the findings of this analysis.

## Methodology

### ***Accident Database and Work-Zone Risk Factors Identification***

The Florida Traffic Crash Records Database for years 2002, 2003, and 2004 was utilized in this study and was obtained from the Office of Management Research and Development in Florida. The database consists of seven main files: events file, drivers file, passengers file, pedestrians file, property file, vehicles file, violation file, and DOT file. The events (containing information about the characteristics and environment of the crash), vehicles (containing the information about the vehicles' characteristics and vehicles actions in the traffic crash), and drivers (containing information about drivers' characteristics) files were the subject of interest in this study. It should be mentioned that the work-zone classification variable was first incorporated in the Florida database in 2002. Table 1 lists the variables included in each model and the number of observations in each model in addition to the percentage of each level under each variable.

### ***Comparison Methodology***

The purpose of this study is to identify the characteristics and risk factors (drivers, vehicles, and environment) that classify work-zone crashes solely on freeways. The first part of this study (Model 1) focuses on single-vehicle crashes at work zones and the second part (Models 2 and 3) highlights two-vehicle crashes at work zones. The single-vehicle crashes are defined as any vehicle that crashes with a fixed object (or pedestrian/worker) contained by the work zone or any vehicle that runs off the road within a work-zone area.

For the single-vehicle crash analysis, freeway work-zone single-vehicle crashes were compared to freeway nonwork-zone (exposure) single-vehicle crashes as shown in Fig. 1. As for two-vehicle crashes and as shown in Fig. 2, first (Model 2) a comparison between at-fault drivers and not-at-fault drivers (quasi-induced exposure analysis) was conducted which unveiled drivers/vehicles attributes using multiple logistic regression. Second (Model 3), similarly to single vehicle analysis, a conditional multiple logistic regression revealed the two-vehicle work-zone

crash environments' characteristics. It should be mentioned that comparing freeway work-zone and nonwork-zone crashes (exposure) could be problematic due to the nonhomogeneity with the exposure distributions. To illustrate this, Fig. 3 shows that the highest frequency for crashes in work zones occurs at a speed limit varying between 72 and 105 km/h (45 and 65 mi/h) and nonwork zone at a speed limit varying between 89 and 113 km/h (55 and 70 mi/h). This is due to the reduced speed limit for the duration of the work zone. Therefore, a comparison between crashes with different speed distributions is erroneous and misleading. To overcome this issue, the within-stratum analysis (or stratified sampling) was implemented. As mentioned previously and as shown in Fig. 1 (Model 1) and Fig. 2 (Model 3), the stratification criteria for these models were speed limit, number of lanes and time of day (a.m. or p.m.). For example, a within stratum analysis characterized by a 89 km/h (55 mi/h) speed limit, three lanes, and a.m. time, will be performed to classify the risk factors associated with work-zone crashes.

### ***Quasi-Induced Exposure Technique***

The quasi-induced exposure technique (Carr 1970; Haight 1973; Stamatiadis and Deacon 1997) is used in traffic safety research to explore traffic crash databases by comparing at-fault drivers' characteristics to not-at-fault drivers (exposure) traits. The at-fault drivers are those who are blamed by the police officer for the crash occurrence and the not-at-fault drivers are those found not responsible for the crash occurrence. The fundamental conjecture of this method is that the distribution of the not-at-fault drivers characterizes (or pseudoduplicates) the distribution of all drivers (drivers' population) exposed to crash hazards. Several studies (Stamatiadis and Deacon 1997; Albridge et al. 1999) applied the quasi-induced exposure technique where the determination of at-fault drivers strictly depended upon whether the driver was issued a citation. According to Jiang and Lyles (2007), this could be problematic. Jiang and Lyles (2007) stated that a police officer may be likely to assign responsibility and issue a ticket to a driver once he determines an indication of another violation (e.g., drinking and driving, revoked license, etc.) regardless of the hazardous driving related to the accident itself. According to De Young et al. (1997) this would inflate the involvement ratio of these groups and result in biased data and results. To overcome this issue in our analysis, the at-fault drivers were selected if they match two criteria; they were issued a citation, and they contributed (e.g., careless driving, speeding, etc.) to the crash occurrence.

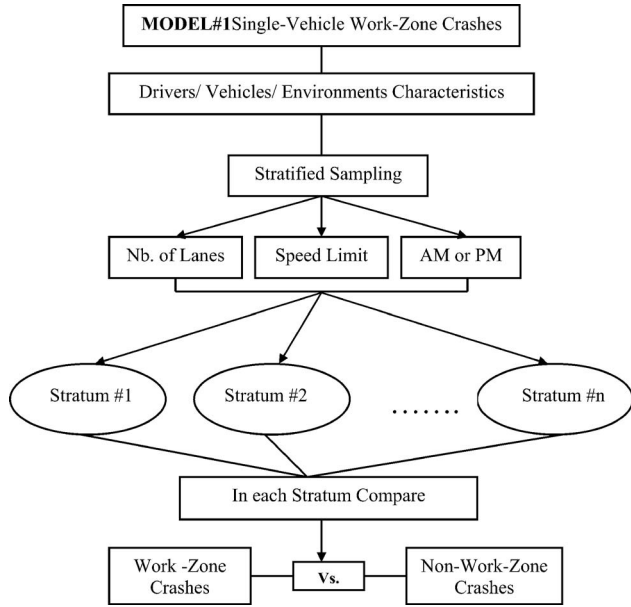
Yan et al. (2005) were some of the few researchers to focus on the investigation of nondriver/vehicle-related (road environment) factors as exclusive main effects on the traffic safety. To introduce the road environment factors into the statistical model and test their exclusive main effects on crashes, Yan et al. extended the application of the quasi-induced exposure technique. In their study, they modeled rear-end collisions at signalized intersections. First, two-vehicle crashes occurring at signalized intersections were identified. Then, they were categorized into two groups: rear-end crashes and nonrear-end crashes (exposure) instead of at-fault and not-at-fault (exposure) drivers. By doing so, Yan et al. were able to compare the environment distributions in the rear-end group and the nonrear-end group to investigate crash propensities, which indicate whether specific traffic conditions increase the likelihood of rear-end crashes at signalized intersections. Similarly to Yan et al.'s approach, this research extends the quasi-induced exposure technique to examine work-zone traffic crash susceptibility. For the single-vehicle crash analysis, a comparison

**Table 1.** Variables Description

Type	Variables	Categories	Model 1 (Single vehicle)		Model 2 (2 vehicles)		Model 3 (2 vehicles)		
			Work zone percent of each level (%)	Nonwork zone percent of each level (%)	W.Z. at fault percent of each level (%)	W.Z. not at fault percent of each level (%)	W.Z. at fault percent of each level (%)	N.W.Z. at fault percent of each level (%)	
Driver characteristics	Age (years)	<25	32.35	36.41	29.72	19.51	29.72	32.11	
		26–35	23.02	23.21	24.35	23.60	24.35	25.37	
		36–45	20.97	18.45	19.71	24.44	19.71	16.20	
		46–55	12.66	11.43	13.06	17.65	13.06	5.21	
		56–65	6.27	6.11	7.15	9.68	7.15	11.10	
		66–75	3.20	3.00	4.81	3.72	4.81	1.33	
		>75	1.53	1.39	1.20	1.40	1.20	8.68	
	Gender	Male	68.09	65.89	50.03	64.32	50.03	62.71	
		Female	31.91	34.11	49.97	35.67	49.97	37.29	
	Driving under the influence	Not under the influence	Alcohol/drugs/both	84.11	87.29	91.34	98.80	91.34	74.58
Live in the state of the accident			15.89	12.71	3.57	1.20	3.57	25.42	
Live outside the state of the accident			86.84	88.67	88.26	86.33	88.26	86.30	
Vehicle characteristics	Speed (mi/h)	<25	13.16	11.33	11.74	13.67	11.74	13.70	
		26–35	2.22	2.75	3.22	4.21	3.22	3.14	
		36–45	0.14	2.25	2.10	1.99	2.10	1.90	
		46–55	3.83	5.20	4.26	5.20	4.26	3.40	
		56–65	15.20	9.60	31.01	27.88	31.01	31.22	
		66–75	50.31	20.93	40.23	42.04	40.23	39.89	
		>75	22.50	49.42	18.20	17.89	18.20	16.50	
	Vehicle type	Passenger car light trucks (SUV)	5.80	9.85	0.98	0.79	0.98	3.95	
		Trucks/large truck	86.21	93.11	82.85	84.57	82.85	86.32	
			13.79	6.89	17.15	15.43	17.15	13.68	
Environment characteristics	Speed limit (mi/h)	<56 km/h (<35 mi/h)	1.20	2.50	2.00	2.00	2.00	1.90	
		72 km/h (45 m/h)	3.56	9.56	10.31	10.31	10.31	7.89	
		89 km/h (55 mi/h)	51.62	14.84	60.05	60.05	60.05	65.22	
		105 km/h (65mi/h)	36.43	17.84	22.72	22.72	22.72	21.10	
		113 km/h (70mi/h)	7.19	45.57	4.91	4.91	4.91	3.89	
	Road surface condition	Normal surface condition	72.74	66.41	65.37	65.37	65.37	71.20	
		Wet/slippery surface condition	27.26	33.59	34.63	34.63	34.63	28.80	
	Rural/urban	Rural area	50.70	62.12	37.36	37.36	37.36	44.48	
		Urban area	49.30	37.88	62.64	62.64	62.64	55.52	
	Road characteristics	Straight level	Straight level	69.95	63.25	75.36	75.36	75.36	74.97
			Straight upgrade/downgrade	14.62	15.73	14.89	14.89	14.89	16.81
		Curve level	Curve level	7.08	10.48	5.38	5.38	5.38	4.50
			Curve upgrade/downgrade	8.35	10.53	4.37	4.37	4.37	3.72
	Event location	Bridge	Bridge	83.65	79.41	88.79	88.79	88.79	86.39
			Entrance ramp	5.97	4.70	3.11	3.11	3.11	3.08
		Exit ramp	Exit ramp	3.46	6.45	3.88	3.88	3.88	4.28
			Straight segment	6.92	9.44	4.22	4.22	4.22	6.26
Weather	Clear	Clear	53.49	55.27	62.53	62.53	62.53	65.30	
		Cloudy/rainy/foggy	46.51	44.73	37.47	37.47	37.47	34.70	
Lighting condition	Dark with lighting	Dark with lighting	50.76	56.60	63.61	63.61	63.61	66.45	
		Dark without lighting	3.85	3.97	3.39	3.39	3.39	3.90	
		Dusk/drawn	23.22	21.56	19.99	19.99	19.99	19.59	

**Table 1.** (Continued.)

Type	Variables	Categories	Model 1 (Single vehicle)		Model 2 (2 vehicles)		Model 3 (2 vehicles)	
			Work zone percent of each level (%)	Nonwork zone percent of each level (%)	W.Z. at fault percent of each level (%)	W.Z. not at fault percent of each level (%)	W.Z. at fault percent of each level (%)	N.W.Z. at fault percent of each level (%)
		Day light	22.17	17.87	13.00	13.00	13.00	10.06
	Number of lanes	1-2-3L	7.23	15.41	43.14	43.14	43.14	35.40
		4>4L	92.77	84.60	56.86	56.86	56.86	64.60
Number of observations				950.00	7,100.00	3,353.00	3,353.00	8,300.00



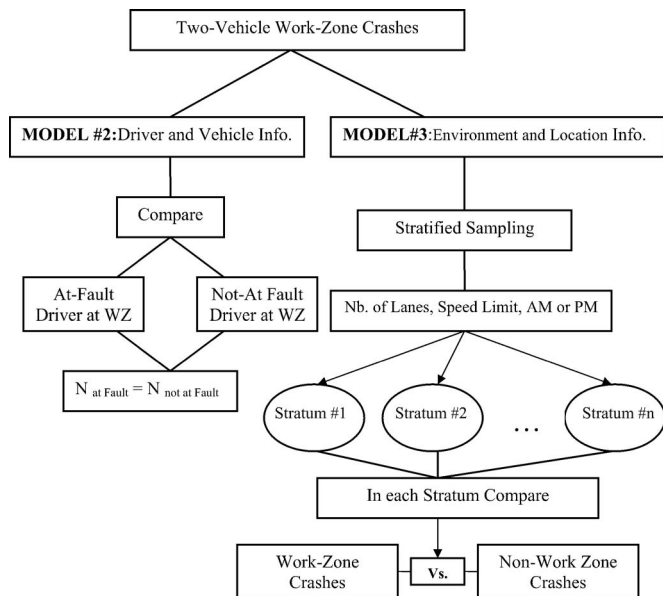
**Fig. 1.** Single vehicle work zone crashes comparison methodology

between work-zone single-vehicle crashes and nonwork-zone (exposure) single vehicle crashes is conducted. This comparison is explained in detail in the next section. As for two-vehicle work-zone freeway crashes, first, we categorize vehicles/drivers into at-fault and not-at-fault drivers. Second, comparing at-fault and not-at-fault drivers unveils drivers/vehicles attributes. To extend the quasi-induced exposure technique into exploring the environment characteristics for work-zone two-vehicle crashes, we compare at-fault work-zone drivers and at-fault nonwork-zone drivers. This comparison is further explained in the next section.

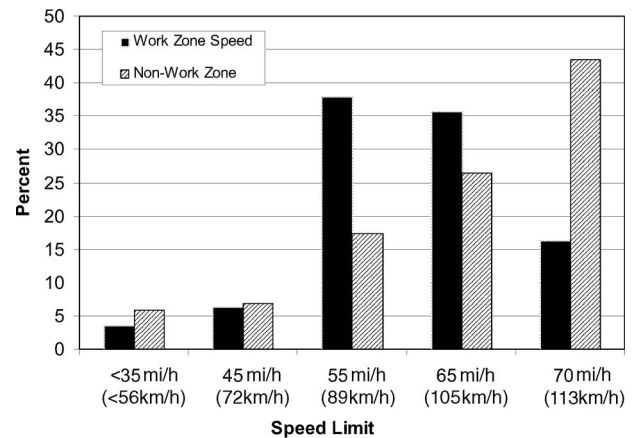
Based on the above categorization, three types of relative accident involvement ratios (RAIRs) are calculated to test the main effect of driver, vehicle, and environment factors related to work-zone crashes for each of the three models. Using the RAIR formula developed by Stamatiadis and Deacon (1997), the relative crash involvement ratio is defined as Eq. (1)

$$RAIR_i = \frac{D1_i}{\sum D2_i} \quad \text{or} \quad RAIR_i = \frac{V1_i}{\sum V2_i} \quad \text{or} \quad RAIR_i = \frac{E1_i}{\sum E2_i} \quad (1)$$

$RAIR_i$  = relative accident involvement ratio for type  $i$  drivers/vehicles/environments. For instance, in the comparison of work-zone at-fault drivers and nonwork-zone at-fault drivers,  $D1_i$  = number of at-fault drivers of type  $i$  in work-zone crashes,  $D2_i$  = number of at-fault drivers in nonwork-zone crashes;



**Fig. 2.** Two-vehicle work zone crashes comparison methodology



**Fig. 3.** Speed limit comparison work zone versus nonwork zone

$V1_i$ =number of at-fault vehicles of type  $i$  in work-zone crashes;  $V2_i$ =number of at-fault vehicles of type  $i$  in nonwork-zone crashes;  $E1_i$ =number of work-zone crashes involving environment type  $i$ ; and  $E2_i$ =number of nonwork-zone crashes involving environment type  $i$ .

Furthermore, to test the interaction between type  $i$  drivers/vehicles/environments and type  $j$  drivers/vehicles/environments, the RAIR can be defined as Eq. (2)

$$\text{RAIR}_{i,j} = \frac{\frac{N1_{i,j}}{\sum \sum N1_{i,j}}}{\frac{N2_{i,j}}{\sum \sum N2_{i,j}}} \quad (2)$$

where  $\text{RAIR}_{i,j}$ =relative accident involvement ratio types  $i$  and  $j$  drivers/vehicles/environments. For example, in the comparison of work-zone at-fault drivers and nonwork-zone at-fault drivers,  $N1_{i,j}$ =number of work-zone crash drivers, vehicles, or the related environments of type  $i$  and  $j$  in work-zone collisions; and  $N2_{i,j}$ =number of at-fault drivers, vehicles, or the related environments of type  $i$  and  $j$  in nonwork-zone crashes.

### Multiple Logistic Regression Modeling

Previous studies had properly applied logistic regression analysis to test the significance of traffic crash risk factors based on the technique of induced exposure (Hing et al. 2000; Stamatiadis and Deacon 1995). Logistic regression belongs to the group of regression methods for describing the relationship between explanatory variables and a discrete response variable. It is a powerful alternative to classical discrimination and regression methods and it applies to a large family of parametric distributions, involving both discrete and continuous variables (Cox 1966; Day and Kerridge 1967; Anderson 1972). A binary logistic regression is proper to use when the dependent variable is dichotomous (i.e., the dependent variable is binary) and can be applied to test association between a dependent variable and the related potential risk factors. Binary logistic regression is used to model at-fault and not-at-fault drivers at work zones. The dependent variable  $Y$  (crash classification) can only take two values:  $Y=1$  for at-fault drivers, and  $Y=0$  for not-at-fault drivers. The probability that a driver is at fault or not is modeled as logistic distribution in Eq. (3)

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (3)$$

The logit of the multiple logistic regression model (link function) is given by Eq. (4)

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n \quad (4)$$

where  $\pi(x)$ =conditional probability of at-fault work-zone drivers, which is equal to the number of at-fault drivers divided by the total number of drivers; and  $x_n$ =independent variables (driver/vehicle/environment factors). The independent variables can be either categorical or continuous, or a mixture of both. Both main effects and interactions can be accommodated.  $\beta_n$ =model coefficient, which directly determines the odds ratio involved in the at-fault drivers. The odds of an event are defined as the probability of the outcome event occurring divided by the probability of the event not occurring. The odds ratio is equal to  $\exp(\beta_n)$  and

tells the relative amount by which the odds of the outcome increase (OR greater than 1.0) or decrease (OR less than 1.0) when the value of the predictor is increased by 1.0 units (David and Lemeshow 1989). Previous studies (Stamatiadis and Deacon 1995; Hing et al. 2000) clearly expressed the relationship between logistic regression and RAIR in the quasi-induced exposure analysis. In fact, for a specific type of drivers/vehicles/environments, the odds generated from the logistic regression model are analogous to the corresponding RAIRs, and the odds ratio from the model are equivalent to the comparisons among those RAIRs. In this paper, the RAIRs were based on the univariate analysis rather than the network analysis which clarifies the small differences between the models' odds ratios and the RAIRs. Furthermore, a significant  $p$  value (e.g.,  $P \leq 0.05$ ) for a Wald  $\chi^2$  statistic is evidence that a regression coefficient in the model is nonzero, which also indicates the statistical importance of those RAIRs' comparisons between different types of drivers/vehicles/environments. The SAS program procedure, LOGISTIC, was used for the model development and the hypothesis testing was based on the 0.05 significance level.

### Conditional Logistic Regression Modeling (Matched Work-Zone Nonwork-Zone Crashes)

For modeling at-fault work-zone drivers and at-fault nonwork-zone drivers, a matched work-zone nonwork-zone analysis is implemented. The purpose of the proposed matched work-zone nonwork-zone analysis is to explore the effects of traffic characteristics variables while controlling for the effects of other confounding variables through the design of the study. This modeling is called conditional logistic regression. It is used in this study to model single-vehicle work-zone crashes against single vehicle nonwork-work-zone crashes and two-vehicle work-zone at-fault drivers versus two-vehicle nonwork-zone at-fault drivers.

In a matched work-zone nonwork-zone crash study, first crashes are selected. For each selected crash, some nonenvironment variables such as number of lanes, time of day, speed limit, etc., associated with each crash are selected as matching factors. A subpopulation of work-zone crashes is then identified using these matching factors. For example, for freeway work-zone crashes, with a specific number of lanes, speed limit, and time of day, a subpopulation of work-zone crashes is identified based on the matching criteria. A total of  $m$  nonwork-zone crashes are then selected at random from each subpopulation of work-zone crashes. Within stratum differences between work-zone and nonwork-zone characteristics are utilized in the development of the statistical model. This is done under the conditional likelihood principle of statistical theory.

Abdel-Aty et al. (2004) employed this modeling technique to predict freeway crashes based on loop detector data. Similarly to them, we assumed that there were  $N$  strata with  $n$  work-zone crashes and  $m$  nonwork-zone crashes in stratum  $j$ ,  $j=1, 2, \dots, N$ . We also assumed that  $p_j(x_{ij})$  was the probability that the  $i$ th observation in the  $j$ th stratum is a crash where  $x_{ij}=(x_{1ij}, x_{2ij}, \dots, x_{kij})$  was the vector of  $k$  traffic characteristics variables  $x_1, x_2, \dots, x_k$ ;  $i=0, 1, 2, \dots, m+n-1$ ; and  $j=1, 2, \dots, N$ . This crash probability  $p_j(x_{ij})$  may be modeled using a linear logistic model as follows

$$\text{logit}(p_j(x_{ij})) = \alpha_j + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_k x_{kij} \quad (5)$$

The intercept term  $\alpha$  is different for different strata. It summarizes the effect of variables used to form strata on the probability of the crash. In order to take into account the stratification in the analysis of the observed data, one constructs a conditional likelihood. This conditional likelihood function is the product of  $N$  terms, each of which is the conditional probability that the crash in a particular stratum says the  $j$ th strata, is the one with explanatory variables  $x_{0j}$ , conditional on  $x_{0j}, x_{1j}, \dots, x_{mj}$  being the vectors of explanatory variables in the  $j$ th stratum. The mathematical derivation of the relevant likelihood function is quite complex and is neglected here. The reader may consult Collett (1991) for full derivation of the conditional likelihood function that can be expressed as [Abdel-Aty et al. (2004)]

$$L(\beta) = \prod_{j=1}^N \left[ 1 + \sum_{i=1}^m \exp \left\{ \sum_{u=1}^k \beta_u (x_{uij} - x_{u0j}) \right\} \right]^{-1} \quad (6)$$

where  $\beta$  = same as in Eq. (5). The likelihood function  $L(\beta)$  is independent of the intercept terms  $\alpha_1, \alpha_2, \dots, \alpha_N$ . So the effects of matching variables cannot be estimated and hence Eq. (5) cannot be used to estimate crash probabilities. However, the values of the  $\beta$  parameters that maximize the likelihood function given by Eq. (6) are also estimates of  $\beta$  coefficients in Eq. (6). These estimates are log odds ratios and can be used to approximate the relative risk of a crash.

The SAS procedure PHREG gives these relative risks (termed hazard ratio under PHREG). The log odds ratios can also be used to develop a prediction model under this matched crash-noncrash analysis.

## Data Analysis

### Statistical Modeling for Single-Vehicle Work-Zone Crashes

Based on the model for single-vehicle work-zone crash analysis, the conditional logistic regression identified the risk factors associated with work-zone crashes. As shown in Table 1, the numbers of observations for work-zone and nonwork-zone crashes were 950 and 7,100, respectively. The reader should be cautious that the identified risk factors imply that these factors have higher sensitivity to workzones than to nonwork-zone locations. The hazard ratio is analogous to the odds ratio. A hazard ratio (odds ratio) of one implies that the event is equally likely in both groups. A hazard ratio (odds ratio) greater than one implies that the event is more likely in the first group. A hazard ratio (odds ratio) less than one implies that the event is less likely in the first group. Table 1 lists the model estimation and the hazard ratios (or odds ratios) properly adjusting other factors for significant independent variables.

Fig. 4 illustrates the univariate comparisons of relative crash involvement ratios between different conditions for drivers/vehicles/environment characteristics prior to the application of the stratified sampling. The listed graphs in Fig. 4 stand for the variables found significant at the 0.05 significance level in the univariate analysis. The RAIRs show a trend for each of the drivers/vehicles/environment factors. For instance, the RAIR of trucks is clearly higher than the RAIR of passenger cars/SUVs/vans. The weather graph shows that the RAIR of cloudy weather is higher than RAIR of clear weather and the RAIR of rainy weather is undoubtedly lower than the RAIR of clear weather.

The conditional logistic regression previously compares drivers/vehicles/environment characteristics associated with work-zone versus nonwork-zone crashes. The final model's results shown in Table 2 illustrate the model's significant variables and goodness of fit. The Log likelihood, AIC, and SBC criteria show that the final model has a good fit. This statistical modeling accounts for the confounding effects and interactions between the factors from the univariate analysis. The model shows that large trucks have additional risk at work-zone locations compared to nonwork-zone locations ( $p$  value=0.0005). Trucks and large trucks are 44.6% more likely to be involved in a work-zone single-vehicle crash compared to nonwork-zone locations. According to the model, roadway geometry including vertical and horizontal alignment is a significant risk factor. Within a work zone straight-level segments have an increased likelihood of single-vehicle crashes compared to straight upgrade/downgrade, curve level, and curve upgrade/downgrade. The hazard ratios (or odds ratios) are 0.749, 0.728, and 0.718, respectively, when compared to straight level. The corresponding  $p$  values are 0.0037, 0.0239, and 0.017 in that order (see Table 2). An explanation of this is that drivers are more likely to drive cautiously on vertical and horizontal curves. The lighting condition is also one of the risk factors associated with work-zone single-vehicle crashes. The model shows that with poor or no lighting during dark at work zones, motor vehicles are more prone (23.5%) to crashes compared to nonwork-zone locations ( $p$  value=0.0151). The weather condition is also one of the statistically significant risk factors. In fact, the model results illustrate that during rainy weather, drivers are less likely to be involved in work-zone single-vehicle crashes ( $p$  value=0.0476). This fact may be due to the vigilant driving pattern during rain, especially at work zones. Finally it should be mentioned that the work-zone presence was found to have no statistically significant effect on the gender and age factors.

### Statistical Modeling for Two-Vehicle Work-Zone Crashes

#### Drivers and Vehicles Characteristics

For two-vehicle crash analysis, the first multiple logistic regression model compares work-zone at-fault drivers versus work-zone not-at-fault drivers and unveils drivers/vehicles attributes. Table 1 shows the number of observations in this model (3,353 observation for at-fault drivers and 3,353 observations for not-at-fault drivers).

Fig. 5 illustrates the univariate comparisons of RAIRs between different conditions for each drivers/vehicles characteristic. The listed graphs in Fig. 5 show the variables found significant at the 0.05 significance level in the univariate analysis. The driving under influence (DUI) graph clearly shows that drivers under the influence of narcotics are more prone to accidents. The age graph illustrates that drivers at age 25 or less and 75 or more are the most sensitive to crashes at work zones. The graph also confirms that males are more at risk than females and that trucks are more sensitive to crashes than regular passenger cars at work zones. The last two graphs in Fig. 5 illustrate that local drivers have a higher relative crash involvement ratio than out-of-state drivers and that speeding [at >105 km/h (>65 mi/h)] in work zones produces a high crash hazard at work zones. Fig. 6 shows the interaction between age and gender. As illustrated by the graph, males of 25 years old and younger and females older than 75 years old have the highest relative crash involvement ratio.

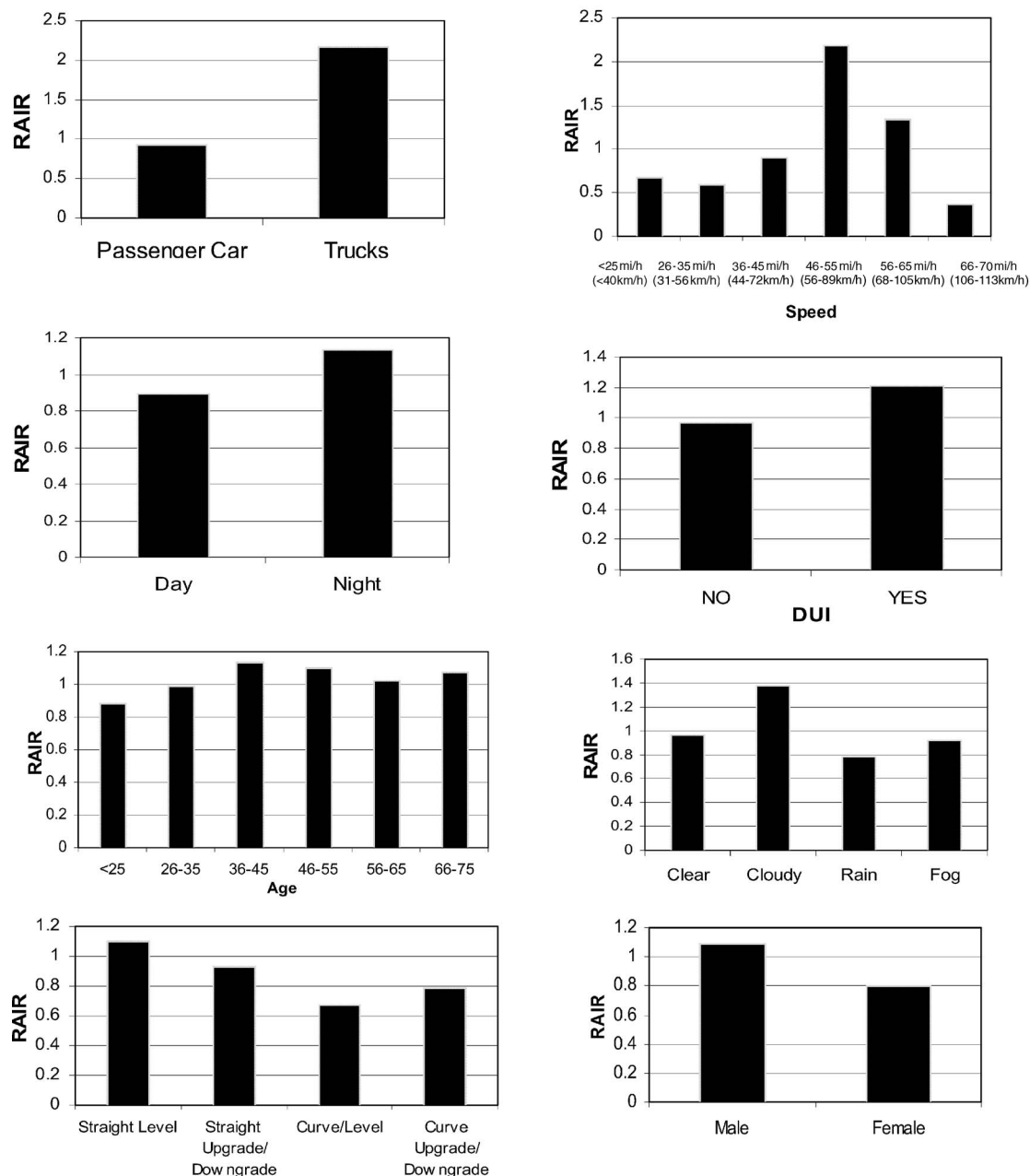


Fig. 4. Relative accident involvement ratios by road environment factors for single-vehicle crashes

The multiple logistic regression model accounting for interactions between terms and confounding effects is summarized in Table 3. The Log likelihood, AIC, and SC criteria show that the model has a good fit. According to the model, age constitutes a risk factor for work-zone crashes. Comparing 56–65, 46–55, 36–45, and 26–35 year old driver groups to <25 year old drivers group shows that drivers 25 years old or younger comprise the highest risk factor for work-zone crashes (Wald chi-square  $p$ -values: <0.0001). The odds ratios are 0.477, 0.444, 0.526, and 0.669, respectively. The model also shows that the crash likelihood for male drivers is significantly higher than female drivers ( $p$  value <0.0001). The odds ratio for females to be involved in a two-vehicle crash at a work zone is 0.714 compared to male drivers. This can be explained by the fact that male drivers are usually more aggressive in driving. The DUI factor is significant in the

final model. The model clearly shows that drivers under the influence of narcotics are 10.526 time more likely to cause crashes ( $p$  value <0.0001). The Rescode variable defines whether the driver lives in the state of where he was involved in the crash or not. The Final model shows that out-of-state drivers are less likely to be involved in a work-zone crash compared to local drivers ( $p$  value=0.0283). The model also illustrates that the odds ratio for foreign drivers to be involved in work-zone crashes is 0.979 compared to local drivers. This can be explained by the fact that foreign drivers are usually more careful on unfamiliar roads.

#### Environment Characteristics

The second model (Model 3) conditional logistic regression previously mentioned compares the environments' characteristics

**Table 2.** Single-Vehicle Conditional Logistic Regression Model Estimation

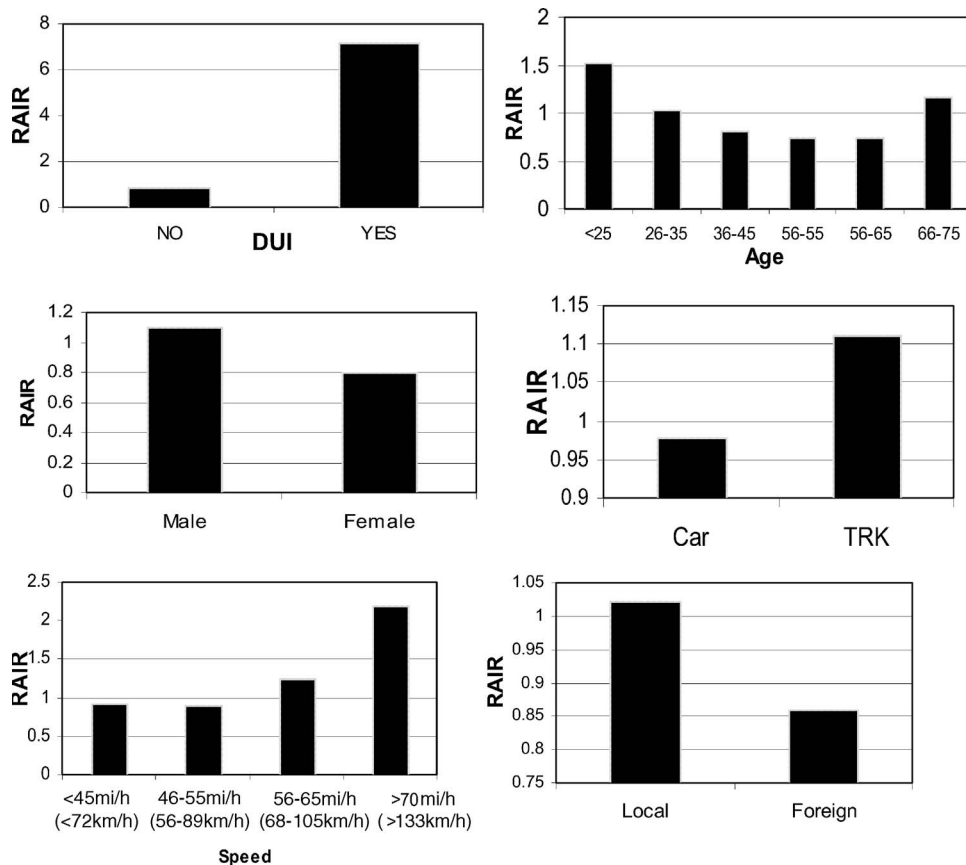
Variable	Parameter estimate	Standard error	Chi square	<i>P</i> value	Hazard ratio
Large truck versus passenger car/SUV/vans	0.36895	0.10573	12.17610	0.00050	1.44600
Straight upgrade/downgrade versus straight level	-0.28886	0.09955	8.41940	0.00370	0.74900
Curve level versus straight level	-0.31689	0.14034	5.09850	0.02390	0.72800
Curve upgrade/downgrade versus straight level	-0.33089	0.13865	5.69590	0.01700	0.71800
Dark with poor or no lighting versus day light	0.21098	0.08683	5.90440	0.01510	1.23500
Rainy weather versus clear weather	-0.17571	0.08869	3.92500	0.04760	0.83900
Model fit statistics					
Criterion	Without covariates		With covariates		
Log likelihood	-4,650.88000		-4,640.58000		
AIC	9,301.77500		9,293.60000		
SBC	9,301.77500		9,298.22800		

associated with work zone. In this model the strata had number of lanes, speed limit, and time of day (Am or Pm), and driver gender and age as matching criteria. Table 1 shows that the numbers of observations for work zone and nonwork-zone are 8,300 and 285,000, respectively.

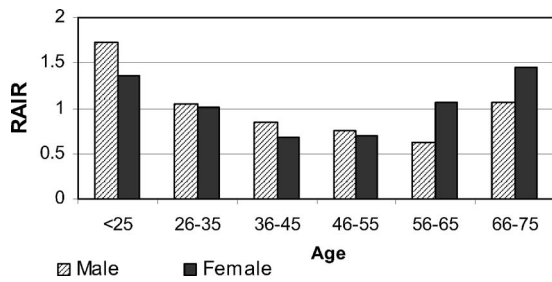
Fig. 7 demonstrates the univariate comparisons conducted prior to the statistical modeling of relative crash involvement ratios between different conditions for each driver/vehicle/environment characteristic before applying the stratified sampling technique. The graphs listed in Fig. 7 display the variables found significant at the 0.05 significance level in the univariate analysis. The weather graph in Fig. 7 clearly shows that the RAIR for

cloudy weather is higher than the RAIR for clear weather. The rural-urban graph confirms that the relative crash involvement ratio is higher for urban locations compared to rural locations. The lighting condition graph demonstrates that night time with poor or no lights could be a serious crash threat at work zones compared to nonwork-zone locations. The roadway characteristics graph shows that straight upgrades and straight downgrades have a lower likelihood for a crash at work zones compared to nonwork-zone settings.

A conditional logistic regression model identified the environmental factors associated with work-zone crashes. Table 4 recapitulates the final model parameter estimates. The Log likelihood,

**Fig. 5.** Relative accident involvement ratios by drivers/vehicles factors for two-vehicle crashes





**Fig. 6.** Relative accident involvement ratios: drivers' age and gender interaction for two-vehicle crashes

AIC, and SBC criteria show that the model has a good fit (see Table 4). Similarly to the single-vehicle model, the road geometry (upgrade/downgrade) had a negative effect on the crash likelihood on work zones compared to nonwork-zone locations. Similarly to the preceding model (single-vehicle crash), this fact can be clarified by the alertness of drivers on upgrades/downgrades compared to straight-level sections. The lighting condition factor is analogous to the previous model. Poor lighting or no lighting at all can cause a significantly ( $p$  value  $<0.0001$ ) higher crash hazard (35.2% increase, hazard ratio 1.352) on work zones compared to nonwork zones. The weather condition affects positively the work-zone crash likelihood. This model shows that foggy weather causes a significant ( $p$  value=0.0017) rise in work-zone crash risk (hazard ratio=1.161) compared to nonwork-zone locations. In addition to that, work zones located in rural areas have a higher crash potential than work zones located in urban areas.

## Conclusions and Discussion

The main objective of this study was to conduct a statistical analysis to unveil work-zone crash characteristics while accounting for confounding parameters. The Florida Traffic Crash Records Database for years 2002, 2003, and 2004 was employed and statistical models were assembled to draw drivers/vehicles/environment traits of work-zone crashes. Three models were developed to analyze single-vehicle and two-vehicle freeway work-zone crashes. The first model (conditional logistic regression model) compared work-zone versus non work-zone single-vehicle crashes and exposed the vehicles/drivers/environment attributes. The second model (multiple logistic regression model) compared two-vehicle work-zone at-fault versus not-at-fault drivers. This model revealed the drivers/vehicles characteristics. The third model (conditional logistic regression) compared at-fault work-zone versus at-fault nonwork-zone drivers for two-vehicle crashes and retrieved work-zone environment attributes. The hypotheses of Models 1 and 3 investigate whether the attributes (parameters included in the models) are significantly affected by the presence of work zones. The hypothesis of Model 2 assesses whether at-fault drivers' attributes are significantly different from the not-at-fault drivers' attributes at work zones.

For the single-vehicle crashes, trucks and large trucks are 44.6% more likely to be involved in a work-zone single-vehicle crash compared to trucks and large trucks in nonwork-zone locations. This fact may be due to narrower lanes during maintenance or construction. Several studies agree that heavy vehicles are overrepresented in work-zone areas (Hall and Lorenz, 1989; Pigman and Agent 1990; Nemeth and Rath 1983). However, the main reason behind this issue is still obscure and a subject for

**Table 3.** Two-Vehicle Logistic Regression Model Estimation

Parameter	Estimate	Standard error	Wald chi square	$P$ value	Odds ratio	95% wald confidence limits		
Intercept	—	-1.1544	0.2554	20.4345	<0.0001	—	—	—
age	75 versus 25	-0.1744	0.1381	1.5952	0.2066	0.8400	0.6410	1.1010
	65 versus 25	-0.7405	0.1210	37.4426	<0.0001	0.4770	0.3760	0.6050
	55 versus 25	-0.8123	0.1005	65.3300	<0.0001	0.4440	0.3640	0.5400
	45 versus 25	-0.6420	0.0892	51.7665	<0.0001	0.5260	0.4420	0.6270
	35 versus 25	-0.4020	0.0860	21.8696	<0.0001	0.6690	0.5650	0.7920
Sex	Female versus male	-0.3384	0.0662	26.1291	<0.0001	0.7130	0.6260	0.8120
DUI	Yes versus no	1.9723	0.1947	102.6544	<0.0001	7.1870	4.9070	10.5260
Rescode	Foreign versus local	-0.2011	0.0917	4.8118	0.0283	0.8180	0.6830	0.9790
(a) Interactions								
Sex* age	75 versus 25	0.5472	0.3144	3.0301	0.0817			
	65 versus 25	0.6324	0.2591	5.9552	0.0147			
	55 versus 25	0.0316	0.2187	0.0209	0.8852			
	45 versus 25	0.0111	0.1917	0.0033	0.9540			
(b) Model fit statistics								
Criterion	Intercept only			Intercept and covariates				
Log likelihood	-3,346.5800			-3,198.3130				
AIC	6,695.1610			6,430.6260				
SC	6,701.6630			6,541.1540				

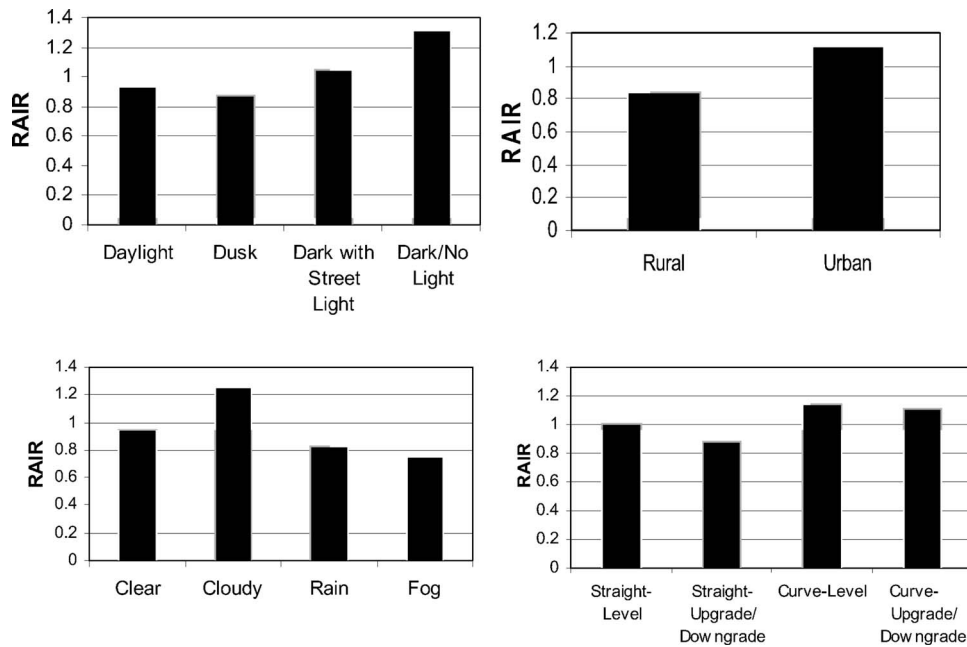


Fig. 7. Relative accident involvement ratios by environment factors for two-vehicle crashes

future investigations. Roadway geometry is also a significant risk factor associated with freeway single-vehicle work-zone crashes. Straight level has increased the likelihood compared to straight upgrade/downgrade, curve level, and curve upgrade/downgrade. In other words, straight level roadways are significantly affected by the presence of work zones compared to nonwork-zone locations. An explanation of this could be related to the fact that drivers may be more likely to drive cautiously on vertical and horizontal curves. In this context, Daniel et al. (2000) stated that fatal work-zone crashes are less influenced by horizontal and vertical alignment compared to nonwork-zone locations. The lighting condition is also one of the risk factors associated with work-zone single-vehicle crashes. The model shows that in work areas with poor or no lighting during the dark, motor vehicles are more prone (23.5%) to crashes compared to nonwork-zone locations with poor or no lighting during the dark. This fact may be due to the invisibility of the work-zone equipment during poor or no lighting which may lead to single-vehicle crashes. The weather condition is also associated with single-vehicle work-zone crashes. In fact, the first model shows that during rainy weather, drivers are less likely to be involved in work-zone crashes com-

pared to the same weather conditions in nonwork-zone locations. This fact may be due to the vigilant driving pattern during rain at work zones.

For two-vehicle crashes, the second model's results illustrate that drivers younger than 25 years old and drivers older than 75 years old have the highest risk of being the at-fault driver in a work-zone crash. Male drivers have significantly higher risk (approximately 40% higher) than female drivers of being the at-fault driver. The interaction between age and gender confirmed that younger ( $\leq 25$  years old) male drivers and older ( $\geq 75$  years old) female drivers are prone to be the at-fault driver in a work-zone crash. The age and gender trends in work-zone crashes are consistent with the general trend of age and gender in the overall crashes (National Highway Traffic Safety Administration 2000). This can be explained by the fact that young male drivers are usually more aggressive in driving and older females' alertness and reaction time decreases with age. The model noticeably shows that drivers under the influence of narcotics/alcohol are 10.526 times more likely to cause crashes (i.e., at-fault driver) at work zones. The second model finally shows that out-of-state drivers are slightly less likely to be the source (i.e., at-fault driver)

Table 4. Two-Vehicle Logistic Conditional Regression Model Estimation

Variable	Parameter estimate	Standard error	Chi square	P value	Hazard ratio
Straight upgrade/downgrade versus straight level	-0.26589	0.05725	21.56750	<0.00010	0.76700
Poor or no street light versus day light	0.30193	0.06567	21.14040	<0.00010	1.35200
Foggy weather versus clear weather	0.14943	0.04765	9.83450	0.00170	1.16100
Rural versus urban	0.25776	0.05014	26.42730	<0.00010	1.29400
Model fit statistics					
Criterion	Without covariates		With covariates		
Log likelihood	-2,524.04700		-2,519.67500		
AIC	5,048.09400		5,040.53500		
SBC	5,048.09400		5,047.36300		

of a work-zone crash compared to local drivers. This can be explained by the fact that foreign drivers are usually more careful on unfamiliar roads. The third model revealed the environment characteristics for two-vehicle work-zone crashes. Similarly to the single-vehicle model (first model), the road geometry and lighting conditions were significant risk factors for two-vehicle work-zone crashes. Freeways straight segments are more susceptible to crashes in work-zone areas. As explained before, this fact may be due to the alertness of drivers on a nonstraight segment. This finding is consistent with previous studies (Milton and Mannering 1998; Chang 2005). Poor lighting or no lighting at all during dark can lead to a significantly higher crash hazard (35.2% increase, hazard ratio 1.352) on work zones compared to nonwork zones. Analogously to this finding, Daniel et al. (2000) also concluded that poor or no lighting at night affects the increase of the likelihood of a fatal crash in work zones compared to nonwork zones. This third model shows that for two-vehicle crashes, foggy weather causes a significant amount in work-zone crash risk compared to nonwork-zone locations. In addition to that, work zones located in rural areas have a higher crashes potential than work zones located in urban areas.

Some recommendations can be drawn based on the findings of the work-zone crash analysis. First, for both single-vehicle and two-vehicle crashes, good lighting should be provided in the work areas and around them so drivers can be alerted ahead of time and to facilitate driving maneuvers during work-zone hazards at night. Trucks should be extra careful in the work zones, especially with lane closures and narrowing. A reduced speed limit could help the trucks better maneuver in work zones. The drivers' inattentiveness and hostile driving are overrepresented in work zones. This fact was illustrated by the age and gender factor, the road geometry factors, residence, and rainy weather. For that purpose, additional enforcement is recommended such as police cars, flashing signs, and double fining in work areas.

As a typical study based on traffic crash databases, some limitations may exist since some variables (or information) may not be available in these crash databases. For instance, the Florida Crash Records Database did not provide information about the work-zone duration and the work-zone design or configuration. These variables may be confounded or may interact with other variables in our models. Such data can be obtained and analyzed using driving simulation studies or field data collection.

## References

- Abdel-Aty, M., Uddin, N., Abdalla, F., Pande, A., and Hisa, L. (2004). "Predicting freeway crashes based on loop detector data using matched case-control logistic regression." *Transportation Research Record*. 1897, Transportation Research Board, Washington, D.C., 88–95.
- Albridge, B., Himmler, M., Aultman-Hall, L., and Stamatiadis, N. (1999). "Impact of passenger on young driver safety." *Transportation Research Record*. 1693, Transportation Research Board, Washington, D.C., 25–30.
- Anderson, J. A. (1972). "Separate sample logistic discrimination." *Biometrika*, 59, 15–18.
- Benekohal, R. F., Shim, E., and Resende, P. T. V. (1995). "Truck drivers' concerns in work zones: Travel characteristics and accident experiences." *Transportation Research Record*. 1509, Transportation Research Board, Washington, D.C., 55–64.
- Carr, B. R. (1970). "A statistical analysis of rural Ontario traffic crashes using induced exposure data." *Proc., Symp. on the Use of Statistical Methods in the Analysis of Road Accidents*, OECD, Paris, 86–72.
- Chang, L. (2005). "Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network." *Safety Sci.*, 43, 541–557.
- Collett, D. (1991). *Modeling binary data*, Texts in Statistical Science Series, Chapman and Hall/CRC, Boca Raton, Fla.
- Cox, D. R. (1966). "Some procedure associated with the logistic qualitative response curves." *Research papers in statistics*, D. F. David, ed., Wiley, New York, 55–71.
- Daniel, J., Dixon, K., and Jared, D. (2000). "Analysis of fatal crashes in Georgia work zones." *Transportation Research Record*. 1715, Transportation Research Board, Washington, D.C., 18–23.
- David, W. H., and Lemeshow, S. (1989). *Applied logistic regression*, Wiley, New York.
- Day, N. E., and Kerridge, D. F. (1967). "A general maximum likelihood discriminant." *Biometrics*, 23, 313–323.
- DeYoung, D. J., Peck, R. C., and Helander, C. J. (1997). "Estimating the exposure and fatal crash rates of suspended/revoked and unlicensed drivers in California." *Accid. Anal Prev.*, 29(1), 17–23.
- "Fatalities from fatality analysis reporting system (FARS)." (2006). ([http://wzsafety.tamu.edu/crash\\_data/fatal.stm](http://wzsafety.tamu.edu/crash_data/fatal.stm)) (March 2006).
- 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 Council, Charlottesville, Va.
- Garber, N. J., and Zhao, M. (2002a). "Distribution and characteristics of crashes at different work zone locations in Virginia." *Transportation Research Record*. 1794, Transportation Research Board, Washington, D.C., 19–25.
- Garber, N. J., and Zhao, M. (2002b). "Final report crash characteristics at work zones." *Rep. No. VTRC 02-R12*, Virginia Transportation Research Council, Charlottesville, Va.
- Ha, T.-J., and Nemeth, Z. A. (1995). "Detailed study of accident experience in construction and maintenance zones." *Transportation Research Record*. 1509, Transportation Research Board, Washington, D.C., 38–45.
- Haight, F. A. (1973). "Induced exposure." *Accid. Anal Prev.*, 5, 111–126.
- Hall, J. W., and Lorenz, V. M. (1989). "Characteristics of construction-zone accidents." *Transportation Research Record*. 1230, Transportation Research Board, Washington, D.C., 20–27.
- Hing, J. Y. C., Stamatiadis, N., and Aultman-Hall, L. (2000). "Evaluating the impact of passengers on the safety of old drivers." *J. Safety Res.*, 34, 343–351.
- Jiang, X., and Lyles, R. (2007). "Difficulties with quasi-induced exposure when speed varies systematically by vehicle type." *Accid. Anal Prev.*, 39, 649–656.
- Milton, J., and Mannering, F. (1998). "The relationship among highway geometrics, traffic-related and motor-vehicle accident frequencies." *Transp. J.*, 25, 395–413.
- Ming, Z., and Garber, N. (2001). "Crash characteristics at work zones." *Research Rep. No. UVACTS-15-0-48*, Center for Transportation Studies, at Univ. of Virginia, Charlottesville, Va.
- National Highway Traffic Safety Administration. (2000). "Traffic safety facts 2000: Young drivers (DOT HS 809 336)." ([http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/TSF 2000/ydrive.pdf](http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/TSF%2000/ydrive.pdf)) (Nov. 21, 2006).
- Nemeth, Z. A., and Rathi, A. (1983). "Freeway work zone accident characteristics." *Transp. Q.*, 37(1), 145–159.
- Pigman, J. G., and Agent, K. R. (1990). "Highway accidents in construction and maintenance work zone." *Transportation Research Record*. 1270, Transportation Research Board, Washington, D.C., 12–21.
- Raub, R. A., Sawaya, O. B., Schofer, J. L., and Ziliaskopoulos, A. (2001). "Enhanced crash reporting to explore work zone crash patterns." *Paper No. 01-0166*, Transportation Research Board, Washington, D.C.
- Rouphail, N. M., Yang, Z. S., and Frazio, J. (1988). "Comparative study of short-and long term urban freeway work zones." *Transportation Research Record*. 1163, Transportation Research Board, Washington D.C., 4–14.
- Stamatiadis, N., and Deacon, J. A. (1995). "Trends in highway safety:

- Effects of an aging population on accident propensity." *Accid. Anal Prev.*, 29, 37–52.
- Stamatiadis, N., and Deacon, J. A. (1997). "Quasi-induced exposure: methodology and insight." *Accid. Anal Prev.*, 27, 443–459.
- Wang, J., Hughes, W. E., Council, F. M., and Paniati, J. F. (1996). "Investigation of highway work zone crashes: What we know and what we don't know." *Transportation Research Record*. 1529, Transportation Research Board, National Research Council, Washington D.C., 54–62.
- Yan, X., Radwan, E., and Abdel-Aty, M. (2005). "Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model." *Accid. Anal Prev.*, 37, 983–995.
- Zhao, M. (2001). "Crash characteristics at work zones." *Research Rep. No. UVACTS-15-0-48*, U.S. DOT, Washington, D.C.