PERGAMON

Accident Analysis and Prevention 34 (2002) 19–29

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Effects of work zone presence on injury and non-injury crashes

Asad J. Khattak [a,*], Aemal J. Khattak [b,1], Forrest M. Council [c]

[a] Department of City and Regional Planning, University of North Carolina, 3140 New East Building, Chapel Hill, NC 27599, USA
[b] Department of Civil Engineering, University of Nebraska, W 348 Nebraska Hall, Lincoln, NE 68588-0531, USA
[c] Highway Safety Research Center, University of North Carolina, 730 Airport Road, Chapel Hill, NC 27599-3430, USA

Received 1 September 1999; received in revised form 10 August 2000; accepted 17 August 2000

Abstract

Work zones in the United States have approximately 700 traffic-related fatalities, 24 000 injury crashes, and 52 000 non-injury crashes every year. Due to future highway reconstruction needs, work zones are likely to increase in number, duration, and length. This study focuses on analyzing the effect of work zone duration mainly due to its policy-sensitivity. To do so, we created a unique dataset of California freeway work zones that included crash data (crash frequency and injury severity), road inventory data (average daily traffic (ADT) and urban/rural character), and work zone related data (duration, length, and location). Then, we investigated crash rates and crash frequencies in the pre-work zone and during-work zone periods. For the freeway work zones investigated in this study, the total crash rate in the during-work zone period was 21.5% higher (0.79 crashes per million vehicle kilometer (MVKM)) than the pre-work zone period (0.65 crashes per MVKM). Compared with the pre-work zone period, the increase in non-injury and injury crash rates in the during-work zone period was 23.8% and 17.3%, respectively. Next, crash frequencies were investigated using negative binomial models, which showed that frequencies increased with increasing work zone duration, length, and average daily traffic. The important finding is that after controlling for various factors, longer work zone duration significantly increases both injury and non-injury crash frequencies. The implications of the study findings are discussed in the paper.

Keywords: Work zones; Crash rates; Negative binomial model; Injury severity

1. Introduction

Highway work zones present a hazardous roadway environment to drivers. The presence of workers, construction machinery, roadside construction barriers, and other paraphernalia associated with work zones create a high degree of conflict that leads to hazardous conditions. According to the fatal accident reporting system (FARS), about 700 fatalities occur in work zones across the United States each year. The economic cost of a motor vehicle crash involving a fatality was reported to be $2 854 500 in 1994 (Blincoe, 1996). Based on this estimate, the annual cost of work zone fatalities in a year amounts to more than two billion dollars. Furthermore, there are approximately 24 000 non-fatal injury crashes and 52 000 property damage only (PDO) crashes costing additional billions of dollars in damages annually. Work zones in the United States are likely to increase in number, duration, and length due to emphasis on repair and highway reconstruction — a significant portion of all federal-aid highway funds are now geared toward highway rehabilitation.

Highway agencies use several strategies to minimize the adverse effects of work zones. In particular, this study focuses on the effect of work zone duration, which is an important policy-sensitive variable that is often used to minimize road user and work zone worker exposure. Another policy-sensitive variable explored in this study is work zone length. Agencies may control work zone length to reduce exposure by undertaking the work in increments. Though not investigated in this study, another strategy is to disseminate infor-

* Corresponding author. Tel.: + 1-919-9624760; fax: + 1-919-9625206; http://www.unc.edu/ ~ khattak/res951.htm.

E-mail addresses: khattak@unc.edu (A.J. Khattak), akhattak@unl.edu (A.J. Khattak), council@claire.hsrc.unc.edu (F.M. Council).

[1] Tel.: 1-402-4728126; fax: 1-402-4728934.

0001-4575/01/$ - see front matter
PII: S0001-4575(00)00099-3

mation about the existence of a work zone to the public. This can reduce traffic on segments while they are part of the work zone.

The objectives of this study are to (a) investigate changes in total and average crash rates during work zones compared with pre-work zone periods; and (b) empirically estimate and compare the effects of work zone duration on injury and non-injury crashes, while controlling for other factors. To achieve these objectives, we combined crash data (crash frequency and injury severity), road inventory data (average daily traffic (ADT) and urban/rural character), and work zone data (duration, length, and location) that spans considerable time and space. The study utilizes a 'pre-work zone and during-work zone' design to control for certain exogenous factors, while other factors are controlled statistically through modeling. We are unaware of any similar effort where a unique and extensive dataset that included crash and injury information from numerous major work zones along with details of roadway inventory was comprehensively analyzed/modeled.

A literature review follows Section 1, which is followed by a section describing the research approach adopted for this study. The next section describes the sample and provides crash rate comparisons between the pre-work zone and work zone periods. The next section describes the negative binomial model specification and discusses the modeling results. The last section provides conclusions and implications of the findings.

2. Literature review

The literature provides significant information on different aspects of crashes in work zones. With respect to the size of the problem, Stammer (1988) reported that the annual fatalities in work zones across USA increased from about 500 in 1982 to about 700 in 1987. Fatalities in work zones located on Interstate highways doubled during this 5-year period. According to Bryden (1993) the safety record in work zones was declining. Work zone crashes account for 2–3% of all police-reported crashes (Hargroves and Martin, 1980; Wang et al., 1996).

Several studies indicate that crash rates increase in work zones. Juergens (1972) reported increased crash rates of 7.0–21.4% relative to the pre-construction period for ten long-term construction projects. Liste et al. (1976) reported an increase of 119% in crash rate during work zone operation compared with the pre-work zone period. Graham et al. (1977) reported on pre-work zone and during-work zone crashes for 79 long-term construction projects in seven states. Their analysis indicated an average increase of 7.5% in crashes during the work zone period. Nemeth and Migletz (1978) reported an increase of 7% in work zone crash rates relative to the pre-work zone period. Rouphail et al. (1988) studied the difference between long-term and short-term work zones. For one long-term work zone site, they reported an 88% increase in crash rate relative to the pre-work zone period. For short-term sites, Rouphail et al. (1988) found a nearly constant rate of 0.8 crashes per mile per day. This rate was independent of the length and duration of the work activity.

With respect to crash severity in work zones, Rouphail et al. (1988) reported a 20% decrease in fatal and injury crash proportions and close to 50% proportional increase in rear-end crashes during long-term construction projects. Ha and Nemeth (1995) found that work zone crashes were slightly less severe than crashes in non-work zones. According to Wang et al. (1996), rear-end collisions were a large percentage of work zone crashes and this percentage was higher than non-work zone crashes involving a rear-end collision. Furthermore, the percentage of sideswipe collisions in work zones was higher than the percentage of sideswipe collisions elsewhere.

Other studies were aimed at testing the crash performance of traffic control procedures (Dudek and Richards, 1982). Reporting on work zone traffic control, Rouphail et al. (1988) found correlation between speed variations and compliance with standards for traffic control devices. Garber and Tzong-Shiou (1991) studied the effects of traffic control devices on multi-lane and two-lane highway work zones. They report that increase in crash rate depends on the type of traffic control devices used at the site. Ha and Nemeth (1995) mention case studies where better traffic control could prevent many crashes.

Whether or not drivers make behavioral adjustments while passing through work zones was the subject of a study conducted by Benekohal et al. (1993). Based on a survey administered to drivers after travelling though a work zone, the study found that the majority (77.5%) of the drivers paid more attention to work zone signage and thought speed limits were posted correctly (97.0%). Respondents reported increased nervousness while driving through a work zone. The study did not report whether the people, who felt the speed limit was correct for the work zone, actually complied with the limit.

In summary, it appears that crash rates increase when work zones are introduced. The amount of the increase varies across studies. As in non-work zone locations, a majority of the crashes in work zones involve no injury. The injury crashes in work zones seem less severe than injury crashes in non-work zones. Rear-end and sideswipe collisions occur more frequently in work zones than in non-work zones and traffic controls in work zones influence the crash rates. While the literature is insightful, there is a lack of major work zone datasets that span considerable time and space. Furthermore, the effects of work zone duration on frequency of injury and non-injury crashes, while controlling for other factors, have not been quantified.

3. Research approach

Using a unique set of freeway work zone data, this study provides information on the change in crash rates accompanied with the establishment of work zones and the effects of work zone duration, work zone length, and traffic on both injury and non-injury crashes. To investigate work zone crashes, we examined the safety performance of highway segments before the introduction of a work zone and during the work zone period. Thus a 'pre-work zone and during-work zone' design was used in this study. Although a 'pre-work zone and during-work zone control group' design can provide stronger control, identifying equivalent control segments was problematic (see below).

This study required information on work zone start and end dates, location of work zones, crashes during the pre-work zone and work zone periods, and other information such as ADT. The study used data from the Highway Safety Information System (HSIS). Developed and maintained by the Federal Highway Administration (FHWA), HSIS contains crash, roadway inventory, and traffic data from selected states. We were able to assemble the required information in a single data file by merging data resident in multiple HSIS files. Although no statewide construction zone files exist in the HSIS, there is a file containing information on major projects on freeways and expressways in California. This file, originally developed by the Construction Division staff of the California Department of Transportation, includes projects with significant traffic control plans. That is, projects with traffic control budgets that equal or exceed approximately 5% of the total project cost. The file contains segment-level information on work zone location (mileposting) and time duration for 36 projects undertaken in 1993. Though this sample size is relatively small, several other researchers have reported research based on fewer work zone observations. For example, Ha and Nemeth (1995) reported their findings based on observing nine work zones. Findings reported by Rouphail et al. (1988) were based on observing three long-term and 25 short-term work zones; Juergens (1972) research findings were based on observing ten construction projects. However, Graham et al. (1977) reported on 79 long-term construction projects in seven states.

The quality of the location and time data in the file is relatively high because this file was developed electronically and checked manually. Note that the work zone includes 0.8 km (0.5 mile) on each end of the actual work location. This would include mileage around the work area where warning signs and possible traffic queues might affect driver behavior and crash propensity. Unfortunately, the file does not contain any details of the project itself, or of the traffic control plans. The presence of such information could provide additional insights into work zone crashes.

We merged the work zone file with California traffic and crash data files for 1992 and 1993. The 1992 traffic and crash data provides the bulk of the pre-work zone statistics. The 1993 traffic and crash files provide part of the pre-work zone (in case when projects started after January 1, 1993) and the during-work zone data. The segment-based traffic and crash file was aggregated to the project level to form a work zone-level file (our unit of analysis). Thus, the file provides 36 observations each on pre-work zone and during-work zone periods in California during 1992–1993. The variables in this unique file relate to crashes, work zone duration, work zone length, traffic and terrain. Overall, the quality and accuracy of the California traffic and crash data is relatively high (Council and Williams, 1995).

The California traffic data file did not contain work zone-specific traffic counts. It is likely that traffic decreases during the life of a work zone compared with the pre-work zone period. This analysis assumes that traffic did not vary during the period of a work zone. This is a conservative assumption because the actual traffic (presumably less) during the work zone would probably result in fewer crashes. From a statistical viewpoint, if traffic volume is actually lower during work zones due to diversion of vehicles, then our analysis/modeling will indicate a lesser effect of traffic on crashes than the true effect. Other researchers in previous studies have also used this assumption in their analysis (see Rouphail et al., 1988).

Many work zones ended at different times in 1993. The crash data after a work zone ends (in 1993) were excluded from this analysis because often the maintenance and/or construction activity alters significantly the roadway characteristics. The beginning and ending of certain work zones at different times in 1993 makes the identification of equivalent control sites very difficult. Although a design that has pre-work zone and during-work zone data with controls may provide better control for some confounding factors, we chose the simpler pre-work zone and during-work zone design due to the problems associated with identifying equivalent controls.

4. Crash rate comparisons

Table 1 provides the summary statistics for the work zone data. There are 72 observations (36 each for the pre-work zone and during-work zone periods) on limited-access highways for which crash information is

Table 1
Summary statistics for California work zone crash data ($N = 72$)

Variable	Value
Total length of segments observed, km (mile)	242.67 (151.67)
Work zone length, km (mile)	
Average	6.73 (4.21)
Minimum/maximum	0.83 (0.51)/19.53 (12.20)
Urban/rural split, km (mile)	117.66 (73.54)/125.01 (78.13)
Percent	48.50/51.50
Flat/rolling/mountainous, km (mile)	145.36 (90.85)/71.76 (44.85)/25.52 (15.95)
Percent	59.90/29.57/10.53
ADT (vehicles per day)	
Average	100 983
Minimum/maximum	4329/237143
Total crashes	8090
Mean	112.36
Variance	29 664

available. The total length of highway segments observed was 242.67 km (151.67 mile) for each time period, with an average work zone length of 6.73 km (4.21 mile). Of the 242.67 km observed in each period, 117.66 km (73.54 mile, 48.5%) were located in urban areas and 125.01 km (78.13 mile, 51.5%) located in rural areas. Furthermore, 145.36 km (90.85 mile, 59.9%) were on flat terrain, 71.76 km (44.85 mile, 29.57%) on rolling terrain, and 25.52 km (15.95 mile, 10.53%) were located on mountainous terrain. The ADT spans a considerable range with average ADT equal to 100 983 vehicles per day. The ADT range indicates that the data includes lightly traveled as well as heavily traveled freeways. The total number of police-reported crashes during the study period was 8090 with a mean of 112.36 crashes per highway segment. The variance of crash frequency is 29 664. Part of the reason for high variance is that the duration of observation ranges from 16 to 714 days.

Table 2 provides information on the safety performance of the pre-work zone and during-work zone periods. In the pre-work zone period, a total of 6052 crashes were reported (74.80% of the total crashes). Of these, 3887 (64.23%) were non-injury crashes and 2165 (35.77%) involved injuries (fatality, A-type, B-type, and C-type or collectively KABC injuries). The total crash rate (CR) was calculated using the formula,

$$CR = \frac{\sum T}{\sum (A \times L \times D)/10^6} \tag{1}$$

where, T is total number of crashes in a work zone/segment; A stands for average daily traffic; L denotes work zone/segment length in kilometers; and D is duration of observation in days.

The total crash rate in the pre-work zone period is 0.65 crashes per million vehicle kilometers (MVKM). This crash rate consists of the non-injury and injury producing crash rates, which are 0.42 and 0.23 crashes per MVKM, respectively.

During the time that the segments were work zones, 2038 crashes were reported (25.20% of total collisions). Of these, 1329 (65.21%) were non-injury and 709 (34.79%) involved injuries. The total crash rate during this period was 0.79 crashes per MVKM, which is 21.53% greater than the rate on the same highway

Table 2
Details of non-injury and injury crashes in the pre-work zone and work zone periods

Crash details		Total	Non-injury crashes	Injury crashes	Average duration of observation (days)
Pre-work zone ($N = 36$)	Crashes	6052	3887	2165	504.06
	Percent		64.23	35.77	
	Minimum/maximum	3/819	2/530	1/289	
	Mean crashes	168.11	107.97	60.13	
	Crash variance	38433.87	17734.31	4273.90	
	[a]Crash rate	0.65	0.42	0.23	
During work zone ($N = 36$)	Crashes	2038	1329	709	125.44
	Percent		65.21	34.79	
	Minimum/maximum	1/712	0/464	0/248	
	Mean crashes	56.61	36.91	19.70	
	Crash variance	15348.30	6303.90	1824.27	
	[a]Crash rate	0.79	0.52	0.27	
[b]Percent change in crash rate between pre-work zone and work zone periods		+21.53	+23.80	+17.40	

[a] Total crash rate = Σ(crashes)/Σ(ADT × segment length × duration of observation/10^6).
[b] Percent change = (work zone rate − pre-work zone rate) × 100/(pre-work zone rate).

segments in the pre-work zone period. The crash rates during the work zone period for non-injury and injury crashes are 0.52 and 0.27 per MVKM, respectively. These two rates show an increase of 23.8 and 17.4% when compared with the non-injury and injury rates of the pre-work zone period. Thus, the total crash rates increased during the work zone period compared with the pre-work zone period and there was a larger increase in the rate of non-injury crashes than injury crashes.

Average crash rates were analyzed by calculating crash rates for each work zone separately (note that the total crash rates discussed above do not vary across the 36 work zones). The average crash rate during the work zone period was 0.72 crashes per MVKM compared with a pre-work zone average crash rate of 0.68 crashes per MVKM (the respective ranges were 0.07–2.28 and 0.25–2.22). To test if the average crash rates were statistically higher during work zones, a paired samples t-test was conducted. The null hypothesis that the two crash rates are statistically different at the 10% level could not be rejected (t-statistic = 0.67).

Consistent with the literature, the total crash rates were higher during work zones. However, analysis of average crash rates did not provide strong statistical evidence in this regard, which could be due to a relatively small sample size ($N = 36$ work zones). Interestingly, there is an increase in both non-injury and injury total crash rates during the work zone period. Also, the relative increase in (total) non-injury crash rate during work zones is higher than the increase in injury crash rate.

5. Modeling injury and non-injury crashes

The effects of duration of observation, work zone length, traffic and location (urban vs. rural) on the number of crashes in the pre-work zone and during-work zone period are investigated in this section. Poisson regression is appropriate for modeling crash counts. However, this model requires that the mean equals the variance of the count data. The negative binomial regression is appropriate for modeling count data when the mean and variance of the data differ significantly. Table 1 indicates that the variance of crashes is substantially higher than the mean. The appropriate model in this case is the negative binomial. The negative binomial model arises from the Poisson model (Greene, 1997; Cameron and Trivedi, 1998).

Let Y_i denote the number of crash occurrences for the ith of N work zones in a given time, $Y_i = 0, 1, 2, \ldots$. Then the number of crash occurrences in an interval of a given length can be Poisson distributed with probability density,

$$P[Y_i = y_i] = \frac{e^{\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{2}$$

where λ_i is work zone i's expected crash frequency; $y_i = 0, 1, 2, \ldots$ (realized value of the crash frequency); $i = 1, 2, \ldots, N$ and $y_i!$ denotes the factorial of y_i. The mean and variance of Y_i equals λ_i. To incorporate explanatory variables x_i the parameter λ_i is specified to be,

$$\lambda_i = \exp(\beta' x_i) \tag{3}$$

where β' is the vector of estimated parameters; and x_i is work zone i's explanatory variables (e.g. duration and length).

The exponential function ensures the non-negativity of y_i. One way to relax the mean–variance equality assumption of the Poisson model is to allow for unexplained randomness in λ_i by specifying,

$$\ln \lambda_i = \beta' x_i + \varepsilon_i \tag{4}$$

where ε_i is error term, which can reflect a specification error such as omitted explanatory variables and/or intrinsic randomness. For the negative binomial model, $\exp(\varepsilon_i)$ is assumed to have a gamma distribution with mean 1 and variance α^2. The derivation of the probability distribution for this model is given in Cameron and Trivedi (1998), Greene (1998). Compared with the Poisson model, this model has an additional estimable parameter α, such that,

$$\text{Var}[y_i] = E[y_i]\{1 + \alpha E[y_i]\} \tag{5}$$

This is a natural form of overdispersion and the overdispersion rate is,

$$\frac{\text{Var}[y_i]}{E[y_i]} = 1 + \alpha E[y_i] \tag{6}$$

The model can be estimated by the standard maximum likelihood methods. If α is not statistically different from zero, then the simple Poisson model is more appropriate.

The goodness-of-fit for the count data models are discussed in Greene (1997). In this paper, we will consider an informal goodness-of-fit statistic that measures the fraction of a restricted log-likelihood explained,

$$\rho^2 = 1 - \frac{L(\beta)}{L(0)} \tag{7}$$

where $L(\beta)$ is the log-likelihood at convergence and $L(0)$ is the restricted log-likelihood.

Unlike linear regression, the estimated parameters of independent variables in a negative binomial model do not indicate the effect of a unit change in the jth independent variable. To interpret the coefficients, consider the conditional mean,

$$E[y_i] = \exp(\beta' x_i) \quad (8)$$

Differentiating with respect to x,

$$\frac{\partial E[y_i]}{\partial x_j} = \beta_j \exp(\beta' x) \quad (9)$$

For example, if $\beta_j = 0.5$ and $\exp(\beta' x) = 2.0$, then a unit change in the jth regressor increases the expectation of y by 1.0 unit. For a dummy variable the (conditional) mean is $\exp(\beta_1)$ times larger if the indicator variable is 1 compared with when it is 0. Inclusion of the natural logarithmic transformations of the work zone duration and length variables in the negative binomial model specification can test if the number of crashes is directly proportional to the duration of observation and the length. A priori, if duration of observation and lengths of work zones were longer, then more crashes will occur. Furthermore, if crash frequency is directly (one-to-one) proportional to the duration and length, then the estimated log-transformed parameters in the model should be unity (McCullagh and Nelder, 1983). For example, if a 1% increase in work zone duration results in a 1% increase in crashes, then the estimated parameter for duration should be unity (note that both the dependent and independent variables are undergoing natural log transformation). Parameter values that are much larger than one imply that the crash frequency is very responsive to small changes in the independent variable. In short, when the regressors enter the equation logarithmically, the parameter is the elasticity, giving the percentage change in $E(y)$ for a 1% change in x. Using our notation,

$$E[y_i] = \exp(\beta_1 \ln(x_1) + \beta' x_i) = x_1^{\beta_1} \exp(\beta' x_i) \quad (10)$$

In this paper, we also compute confidence intervals for the parameter estimates using the following formula,

$$\beta = \hat{\beta} \pm Z \times \text{S.E.} \quad (11)$$

where the value of Z for a 95% confidence interval and a large sample size is 1.96. For a sample size of 36, the Z-value is very close to 2. The hat symbol denotes the estimated beta parameter. S.E. denotes the standard error. Finally, in this study, we generally use a 5% significance level (or 95% confidence interval) to argue for including variables in model specification. However, in some instances a more moderate 10% significance level is also referred to.

5.1. Model specification

Five negative binomial models are reported in this paper. First, a combined negative binomial model (model 1) for both injury producing and non-injury crashes was estimated. The dependent variable in this model is the reported crash frequency with pooled data. There are $36 \times 4 = 144$ rows in the data set with each row representing injury or non-injury crash frequency in either pre-work zone or during-work zone periods. Besides other independent variables (given below), the specification for model 1 includes two indicator variables. These are (a) an indicator variable for injury-producing crashes (injury-producing crashes = 1; non-injury crashes = 0); and (b) an indicator variable for crashes in the during-work zone period (crashes during work zone period = 1; pre-wok zone period = 0). Model 1 provides overall information on the factors affecting crash frequency.

Second, we estimated four additional negative binomial models (models 2–5) to discern meaningful relationships between the independent variables and injury and non-injury crash frequencies reported during the pre-work zone and work zone periods. The dependent variables for these four models are, model 2, the frequency of non-injury crashes in pre-work zone period ($N = 36$); model 3, the frequency of non-injury crashes in the work zone period ($N = 36$); model 4, the frequency of injury-producing crashes in the pre-work zone period; and model 5, the frequency of injury-producing crashes in the work zone period. The estimation of these four models allows us to study separately the effects of the independent variables on non-injury and injury-producing crashes when work zones are established. They also overcome the potential problem of correlation among observations of the same work zone segments and avoids the complexity of interpretation resulting from several interaction terms that must be included in an equivalent one model specification. The explanatory variables tested in the model specification were,

1. duration of observation in days, i.e. the period of the pre-work zone and work zone operation. A natural logarithmic transformation of the duration of observation was used in the model specification;
2. work zone length. The natural logarithmic transformation of work zone length measured in kilometers was used in the model;
3. the ADT for the work zone. Each work zone consists of one or more highway segments. For those with two or more segments, the average ADT of the relevant segments is used in the model specification;
4. an exposure (interaction) term equal to ADT × segment length × duration/10^6 (MVKM). The exposure term was intended to capture any interactive effects that work zone duration, length, and ADT may have on crash frequency; and
5. an indicator variable for urban locations. The coding for the model specification was urban = 1 and rural = 0.

Table 3
Combined negative binomial model for both injury and non-injury crash frequency (model 1)

Explanatory variable	Parameter	S.E.	z-Statistic	Mean
ln of ADT (vehicle per day)	1.2659	0.0668	18.93	11.07
ln of duration (days)	1.1149	0.0959	11.62	5.37
ln of length (km)	0.6718	0.0539	12.44	1.55
Urban indicator	−0.2257	0.1828	−1.23	0.69
Injury indicator (injury crash, 1; non-injury crash, 0)	−0.5126	0.0924	−5.54	0.50
Work zone indicator (work zone, 1; pre-work zone, 0)	0.1988	0.1455	1.36	0.50
Constant	−17.7748	0.9018	−19.71	–
α	0.1789	0.0282	6.32	–
Summary statistics				
Number of observations	144			
log Likelihood function $L(\beta)$	−536.20			
Restricted log likelihood $L(0)$	−1139.51			
ρ^2	0.5294			
χ^2	1206.62			

5.2. Modeling results

5.2.1. Combined model for both injury and non-injury crash frequency

Model 1 (Table 3) shows the results of the negative binomial model estimated for both injury producing and non-injury crashes with indicator variables for injury (relative to non-injury) crashes and for workzone periods (relative to non-work zone periods). The ρ^2 value, which provides a measure of the model fit, indicates a reasonable fit and the model is significant overall (5% level), based on the χ^2 statistic. A positive sign of the estimated parameters implies increased crash frequency with increase in the value of the independent variable. The α parameter estimate in the model is positive and statistically significant at the 5% confidence level, indicating that the data are overdispersed. This also confirms the appropriateness of the negative binomial model compared with the Poisson regression model. The model estimations indicated that logarithmic transformations improved the model statistics compared with their untransformed counterparts.

The model can be rewritten as,

$$Y = (x_1)^{1.2659}(x_2)^{1.1149}(x_3)^{0.6718}\exp(-0.2257x_4) \times \exp(-0.5126x_5)\exp(0.1988x_6)\exp(-17.7748)$$

where Y is expected number of total crashes in a given duration on work zone segments; x_1 is average ADT of the work zone (vehicles per day); x_2 is duration of observation (days); x_3 is length of the work zones (km); x_4 is 1 if the roadway is in urban area; 0 otherwise; x_5 is 1 if injury producing crash; 0 otherwise; and x_6 is 1 if crashes recorded during work zones; 0 otherwise (pre-work zone).

The relative effects of variables can be calculated from the above equation. The estimated log-transformed parameters of ADT, duration and segment length are positive and statistically significant (at the 5% level), indicating that higher ADT, longer duration and longer segment lengths are strongly associated with higher crash frequency. The natural logarithmic transformation of these variables provides evidence of correspondence with the dependent variable. In particular, the estimated parameter for work zone duration is slightly higher than unity, indicating a greater than 'one-to-one' correspondence between crash frequency and duration of observation. According to the model results, a 1% increase in duration of observation will result in a 1.1149% increase in crash frequency; similarly, a 1% increase in ADT and segment length will increase crash frequency by 1.2659% and 0.6718%, respectively.

The model shows that the urban location variable is not statistically significant, though it is retained in the model for demonstration purposes. The interpretation is that the relative effect of urban locations on crash frequency is $\exp(-0.2257) = 0.7979$, implying that the mean crash frequency in urban locations is 20.21% lower than other locations. The exposure term was not statistically significant (even at the 10% level), perhaps because exposure effects are already captured individually by the ADT, duration and length variables and it was dropped from the model.

The estimated parameter for injury crashes indicator (injury crashes = 1 and non-injury crashes = 0) is negative and statistically significant. This simply implies that the frequency of injury-producing crashes is significantly lower than non-injury crashes. The indicator variable for during-work zone crashes is positive, as expected, though it is not statistically significant (at the 10% level). This result is interesting, given that total crash rates were higher by 21.5% during the work zone periods. The insignificance of the parameter is not surprising because the paired sample t-test for the before- and during-work zone average crash rates was not statistically significant either (at the 10% level). A key factor contributing to the statistical insignificance is the relatively small sample size of 36 work zones. Thus, after controlling for other factors in the pooled model (ADT, duration and length, etc.), the crash frequencies before- and during-work zone periods are not statistically significantly different at the 10% level. To further confirm the statistical insignificance of the work zone parameter, we estimated a total crashes (injury plus non-injury) pooled model ($N = 72$) for the before and during work zone data. The model provided similar results in that the work zone indicator was statistically insignificant, at the 10% level. So we could not conclude from these data that the crash frequency during the work zone period is statistically significantly higher than the pre-work zone crash frequency.

To further investigate differences in the effect of work zones (especially duration) on injury and non-injury crashes, we estimated four separate models, i.e. non-injury crashes in the pre-work zone and during-work zone periods and injury crashes in the pre-work zone and during-work zone periods. Four separate estimations are a relatively straightforward way of analyzing the work zone data. As stated previously, the estimation of four separate models overcomes the potential problem of correlation among observations of the same work zone segments and avoids the complexity of interpretation resulting from several interaction terms that must be included in an equivalent pooled model.

5.2.2. Non-injury crashes in the pre-work zone and during-work zone periods

Model 2 (Table 4) shows the factors that influence non-injury crash frequency in the pre-work zone period. The ρ^2 value indicates a reasonable fit. The α parameter estimate is positive and statistically significant at the 5% confidence level, indicating over-dispersed data and confirming the appropriateness of the negative binomial model, compared with the Poisson regression model.

The estimated parameter for the natural logarithm of duration is positive and statistically significant showing that longer duration of observation results in higher crash frequency. It is also very close to unity, implying a one-to-one correspondence with crash frequency. The estimated parameter for ADT is positive and statistically significant (at the 5% level) indicating that higher values of ADT result in higher non-injury crash frequency. The parameter for the natural logarithm of length is also positive and significant indicating that longer segments have more crashes. The estimated parameter for length is less than unity. This implies that the relationship between crash frequency and length are not directly (one-to-one) proportional. The model also indicates no statistically significant relationship between non-injury crashes and urban locations. The exposure term was also statistically insignificant (at the 10% level) and it was dropped from the model.

Model 3 (Table 5) provides information on non-injury crashes during the work zone. The model has reasonable summary statistics, although it has a relatively lower goodness-of-fit. This means that the variables explain pre-work zone non-injury crashes better than work zone non-injury crashes. The parameter for the natural logarithm of duration is positive and significant indicating that work zones of longer duration experience higher frequency of non-injury crashes. In contrast to the pre-work zone model, the estimated duration parameter is greater than unity. Longer length of the work zone is associated with higher frequency of non-injury crashes; the value of the estimated parameter is less than unity. Moreover, non-injury crashes occur more frequently when ADT is higher. The indicator variable for urban location is statistically non-significant (at 10% level).

Comparing the parameter magnitudes between models 2 and 3 can provide further insights. It indicates that duration has a marginally stronger effect on the frequency of work zone non-injury crashes compared with the frequency of pre-work zone crashes. A 1% increase in the duration of observation increases the pre-work zone non-injury crashes by 0.9919% ($\pm 2 \times 0.3866\%$), but increases during-work zone crashes by 1.2317% ($\pm 2 \times 0.1953\%$). The 95% confidence intervals for β reported in the parentheses are the Z-values (equal to 2

Table 4
Negative binomial model for non-injury crash frequency in the pre-work zone period (model 2)

Explanatory variable	Parameter	S.E.	z-Statistic	Mean
ln of ADT (vehicles per day)	1.3071	0.1856	7.04	11.07
ln of duration (days)	0.9919	0.3866	2.56	6.18
ln of length (km)	0.6291	0.1273	4.93	1.55
Urban indicator	−0.1219	0.5454	−0.22	0.69
Constant	−17.4966	2.8919	−6.05	–
α	0.2393	0.0686	3.48	–
Summary statistics				
Number of observations	36			
log Likelihood function $L(\beta)$	−169.28			
Restricted log likelihood $L(0)$	−490.85			
ρ^2	0.6551			

Table 5
Negative binomial model for non-injury crash frequency in the during-work zone period (model 3)

Explanatory variable	Parameter	S.E.	z-Statistic	Mean
ln of ADT (vehicles per day)	1.3331	0.1131	11.78	11.07
ln of duration (days)	1.2317	0.1953	6.30	4.56
ln of length (km)	0.6112	0.1691	3.61	1.55
Urban indicator	−0.5068	0.6914	−0.73	0.69
Constant	−17.4966	1.4991	−12.38	–
α	0.2280	0.0787	2.89	–
Summary statistics				
Number of observations	36			
log Likelihood function $L(\beta)$	−119.92			
Restricted log likelihood $L(0)$	−175.56			
ρ^2	0.3169			

Table 6
Negative binomial model for injury crash frequency in in the pre work zone period (model 4)

Explanatory variable	Parameter	S.E.	z-Statistic	Mean
ln of ADT (vehicles per day)	1.2392	0.0903	13.72	11.07
ln of duration (days)	0.7862	0.2411	3.26	6.18
ln of length (km)	0.6911	0.0644	10.72	1.55
Urban indicator	−0.2637	0.2631	−1.00	0.69
Constant	−15.9530	1.8424	−8.65	–
α	0.8024	0.0338	2.36	–
Summary statistics				
Number of observations	36			
log Likelihood function $L(\beta)$	−139.02			
Restricted log likelihood $L(0)$	−195.51			
ρ^2	0.2889			

for $N = 36$ work zones) multiplied by S.E. Given that the 95% confidence intervals overlap substantially, the results imply that reducing the work zone duration may reduce non-injury crashes only marginally. Similarly, a 1% increase in the length of segments that later became work zones can increase non-injury crashes by 0.6718% ($\pm 2 \times 0.0539$%) and the work zone length increases non-injury crashes by 0.6112% ($\pm 2 \times 0.1691$%). So the effect of length is largely unchanged between the pre-work zone and during-work zone periods, implying that reducing work zone length may not be a crucial consideration in reduction of work zone crash frequency.

5.2.3. Injury crashes in the pre-work zone and during-work zone periods

Model 4 (Table 6) presents the effects of the independent variables on injury-producing crashes in the pre-work zone period. Summary statistics for the model are reasonable. Crash frequency increases with higher ADT, longer duration of observation and longer length. The effect of duration variable is directly proportional (one-to-one) to the number of injury-producing crashes. Observe that a unit increase in duration of observation in the pre-work zone period can increase crash frequency from 0.3040 to 1.2684% ($\beta = 0.7862 \pm 2 \times 0.2411$), which clearly envelopes 1%. The length variable does not show such envelopment and one-to-one proportionality, with crash frequency increasing from 0.5623 to 0.8199% ($\beta = 0.6911 \pm 2 \times 0.0644$) with a unit increase in length. The ADT is statistically significant (at the 5% level) indicating that with increased ADT substantially higher injury crashes can be expected. The results for urban locations are not significant, at the 10% level.

Model 5 (Table 7) shows the effects of the independent variables on injury-producing crashes in work zones. The goodness of fit statistic for the model is relatively low implying that the independent variables are explaining comparatively less variation in the data. Nevertheless, the model provides useful information on injury crashes in work zones. As expected, such crashes increase with longer duration and length of the work zones, and with higher ADT. The urban location variable is statistically non-significant. Urban locations have a lower crash frequency compared with rural locations (at the 10% level but not the 5% level).

Importantly, the relative magnitudes of duration in the pre-work zone and during-work zone periods indicate a small relative increase in injury crashes during the work zone period. A 1% increase in duration of observation can increase injury crash frequency from 0.3040 to 1.2684% ($\beta = 0.7862 \pm 2 \times 0.2411$) in the pre-work zone period and from 0.9751 to 1.5347% ($\beta = 1.2549 \pm 2 \times 0.1399$) in the during-work zone period. This implies that shortening the work zone duration will probably reduce the frequency of injury crashes and do so at a slightly greater rate compared with the pre-work zone period. Another interesting comparison is between the duration parameters in models 3 and 5.

Table 7
Negative binomial model for injury crash frequency in the during-work zone period (model 5)

Explanatory variable	Parameter	S.E.	z-Statistic	Mean
ln of ADT (vehicles per day)	1.1984	0.1405	8.52	11.07
ln of duration (days)	1.2549	0.1399	8.96	4.56
ln of length (km)	0.7482	0.0981	7.63	1.55
Urban indicator	−0.6159	0.3435	−1.79	0.69
Constant	−17.7936	1.5065	−11.81	–
α	0.0667	0.0693	0.96	–
Summary statistics				
Number of observations	36			
log Likelihood function $L(\beta)$	−98.03			
Restricted log likelihood $L(0)$	−102.5			
ρ^2	0.0436			

The parameter of duration in the non-injury work zone crashes is about the same as the parameter for injury crashes ($\beta = 1.2317 \pm 2 \times 0.1953$ and $\beta = 1.2549 \pm 2 \times 0.1399$, respectively). This implies that lowering work zone duration is likely to reduce both types of crashes about equally.

Overall, the results show that work zone injury and non-injury crash frequency increases with longer duration and length. The duration effect seems relatively stronger in the during-work zone period compared with the pre-work zone period; the obvious implication is that by reducing work zone duration, both injury and non-injury crashes can be reduced.

6. Potential biases and limitations

There is potential for certain biases in the data analysis. As indicated earlier, potential bias may exist due to ADT counts for the pre-work zone and work zone periods. Work zone-specific counts were not available for this study. However, any hypothesized decrease in ADT (from what was estimated by the traffic staff as a non-work zone flow) would imply conservative results. That is, the calculated work zone crash rates may be lower than in reality, and the effects of ADT parameter estimates may appear lessened in the models.

If non-reporting of crashes varies systematically across pre-work zone and during-work zone periods, then it can bias the results. One could hypothesize both higher reporting in work zones (due to potentially higher police presence and less room to move damaged vehicles to the shoulders), and lower reporting (if vehicles in non-injury crashes are encouraged to leave the scene in order to decrease queues). However, given that all crashes in both parts of the sample are on freeways, where police presence is usually higher and where high numbers of passing traffic should increase calls to police, there should be less non-reporting, and less differential non-reporting than on other roadways.

The use of the same locations in the pre-work zone and during-work zone periods helps to control for certain potential biasing factors such as driver differences, grade, curvature near the zone, police agencies, and (to some extent) weather conditions. However, there may clearly be factors that we could not measure and thus control for in the analysis e.g. differences in weather between 2 years.

Some potential explanatory variables such as the number of lanes, number of ramps, type of work zone activity and posted speed limit could not be tested in the model specification due to their unavailability for the work zone period. Such exclusions can potentially give rise to model specification errors. Finally, the analysis presented in this study is based on 36 work zones. The use of a larger sample of work zones over a longer time period (when available) may enable exploration of year-to-year variations in work zone crashes.

7. Conclusions

To address concerns over work zone crashes and higher future work zone activity, this study explores the rate and frequency of injury and non-injury work zone crashes. A ‘pre-work zone and during-work zone’ study design is used to understand crash rates and the effect of work zone duration, while controlling for other variables. Our unique crash, inventory and work zone dataset consists of before and during work zone observations at 36 major work zone projects on California freeways. The important findings are.

1. On the limited-access highways that later became work zones, the rate of total work zone crashes is higher (21.5%) than the pre-work zone crash rate. This finding is consistent with the literature and implies that major work zone projects on limited-access roadways can be more hazardous than those same segments in the pre-work zone period. However, when considering the average crash rates and crash frequency, statistical analysis showed that the increase in both average crash rates and crash frequencies during work zones were not statistically significant at the 10% level, compared with the before work zone period. Thus there is only limited evidence from these data that average crash rates and crash frequencies increase during work zones.

2. The increase in total crash rates is relatively larger in lower severity crashes; that is, the increase in non-injury total crash rate was 23.5% compared with 17.4% for injuries.
3. The crash frequency on limited-access roadways under study increased with higher values of work zone duration. It seems that reducing work zone duration can be a practical strategy to reduce work zone injury and non-injury crashes. When comparing non-injury and injury crashes in work zones, the effect of work zone duration is about the same. The implication is that by manipulating (reducing) work zone duration, reductions in both injury and non-injury crashes can be achieved.

Overall, by quantifying the effects of the work zone duration and other explanatory variables on injury and non-injury work zone crashes, this study provides useful information to policy makers. In particular, information on the relative effects of the work zone duration can be used to minimize the frequency of injury and non-injury crashes. There is increasing interest in work zone related policy options (e.g. contracting strategies — bonus/penalties for early/late project completion) and technological means of shortening work zone duration (e.g. faster construction sequence by using pre-cast structural entities). A mixture of policy and technology options was used in several cases including the San Francisco earthquake reconstruction, Salt Lake City and Atlanta Olympic games, and in the reconstruction of a segment of Interstate Highway 45 in Houston (the Pierce Elevated highway). Such situations raise important research issues, e.g. how should we evaluate the impacts of policy and technological options related to work zone duration reduction? The research presented in this paper strengthens the priority of current policy, research, and practical activities concerning shortening of work zone duration. Finally, it appears that major work zones on limited-access highways are less safe than non-work zones and efforts must focus on uncovering additional factors that may contribute to hazardous environments.

Acknowledgements

The authors appreciate FHWA and California DOT support in providing the data through the Highway Safety Information System database. The California DOT Accident Surveillance and Coding Branch provide the data to HSIS. Jack Carr, Caltrans Construction Safety Coordinator, provided the construction zone location information in HSIS. The authors recognize the efforts of Yusuf Mohamedshah of the FHWA in data reduction. Statistical software SAS was used for data reduction and LIMDEP for model estimation. The opinions, findings and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the FHWA or California DOT.

References

Benekohal, R., Hashmi, A., Orloski, R., 1993. Drivers' opinion on work zone traffic control. Transportation Quarterly 47 (1), ENO Foundation, pp. 19–38.

Blincoe, L.J., 1996. The Economic Costs of Motor Vehicle Crashes, 1994. National Highway Traffic Safety Administration, Washington, DC.

Bryden, J., 1993. Work zone traffic management. In: The Traffic Safety Toolbox: A Primer on Traffic Safety. Institute of Transportation Engineers, Washington, DC.

Cameron, C., Trivedi, P., 1998. Regression analysis of count data. In: Econometric Society Monograph No. 30. Cambridge University Press, Cambridge.

Council, F., Williams, C., 1995. Highway Safety Information System Guidebook for the California Data Files, vol. 1. University of North Carolina Highway Safety Research Center.

Dudek, C., Richards, S., 1982. Traffic capacity through urban freeway work zone in Texas. Transportation Research Record 869, TRB, National Research Council, Washington, DC.

Garber, N., Tzong-Shiou, H.W., 1991. Effectiveness of traffic control devices in reducing accident rates at urban work zones. Transportation Quarterly 45 (2), ENO Foundation, pp. 259–270.

Graham, J., Paulsen, R., Glennon, J., 1977. Accident and speed studies in construction zones. Report FHWA-RD-77-80. US Department of Transportation.

Greene, W., 1997. Econometric Analysis, third ed. Macmillan, New York, NY.

Greene, W., 1998. LIMDEP Version 7.0 User's Manual. Econometric Software.

Ha, T., Nemeth, Z., 1995. Detailed study of accident experience in construction and maintenance zones. Transportation Research Record 1509, TRB, National Research Council, Washington, DC.

Hargroves, B., Martin, M., 1980. Vehicle accidents in highway work zone. Report FHWA-RD-80-063. US Department of Transportation.

Juergens, W.R., 1972. Construction zone, detour and temporary connection accidents. Business and Transportation Agency, California Division of Highways.

Liste, F., Bernard, R., Melvin, B., 1976. Evaluation of Timber Barricades and Precast Concrete Traffic Barriers for Use in Highway Construction Areas. Virginia Highway and Transportation Research Council.

McCullagh, P., Nelder, J., 1983. Generalized Linear Models. Chapman & Hall, New York.

Nemeth, Z.A., Migletz, D.J., 1978. Accident characteristics before, during, and after safety upgrading projects on Ohio's rural interstate system. Transportation Research Record 672, TRB, National Research Council, Washington, DC.

Rouphail, N., Zhao, M., Yang, S., Fazio, J., 1988. Comparative study of short-and long-term urban freeway work zones. Transportation Research Record 1163, TRB, National Research Council, Washington, DC.

Stammer, R., 1988. Highway safety: at the crossroads. In: Proceedings of ASCE Conference. San Antonio, TX.

Wang, J., Hughes, W.E., Council, F.M., Paniati, J.F., 1996. Investigation of highway work zone crashes: what we know and what we don't know. Transportation Research Record 1529, TRB, National Research Council, Washington, DC.