Safety Models for Rural Freeway Work Zones

Duplication for publication or sale is strictly prohibited without prior written permission of the Transportation Research Board

Authors
Shyam Venugopal, Graduate Research Assistant
Andrzej P. Tarko, Assistant Professor
Purdue University
1248, Civil Engineering Building
West Lafayette, IN 47906

Transportation Research Board
79th Annual Meeting
January 9-13, 2000
Washington, D.C
Safety Models for Rural Freeway Work Zones (CD-ROM)

by

Shyam Venugopal
Graduate Research Assistant
venugops@ecn.purdue.edu
(765) 494-7381

and

Andrzej Tarko
Assistant Professor of Civil Engineering
tarko@ecn.purdue.edu
(765) 494-5027

Purdue University
1284, Civil Engineering Building
West Lafayette, IN 47907
Fax: (765) 496-1105

A paper presented
at the 79th Annual Transportation Research Board Meeting,
Washington D.C., January 2000
ABSTRACT

Construction and maintenance work zones have traditionally been hazardous locations within the highway environment. Numerous studies show that the accident rates during road construction are generally higher than during periods of regular traffic operations. The increase in number of crashes may be attributed to these factors: (i) General disruption to the flowing traffic due to sudden discontinuities caused by closed lanes, (ii) Improper lane merging maneuvers, (iii) The presence of heavy construction equipment within the work area, (iv) Inappropriate use of traffic control devices, and (v) Poor traffic management.

The objective of the presented research was to develop regression models to predict the expected number of crashes at work zones on rural, two lane freeway segments. Crashes on approaches to work zones and those inside the work zones were analyzed separately. For developing these models, an extensive database was obtained which included freeway data, crash data and work zone characteristics.

Negative binomial models were developed with ADT (Average Daily Traffic), the length of the work zones and the duration of the work projects as exposure-to-risk variables. The cost of the various work projects was found to be a good substitute for some of the exposure-to-risk variables. The investigated variables include the number on and off ramps, both on approaches and inside the work zones, type of work and the intensity of the road work involved. The models may be used for prediction purposes to evaluate before hand, the expected number of crashes on the work zone given, the work zone characteristics.

Keywords: Work zones, Work zone approaches, Expected crashes, Negative Binomial models

ACKNOWLEDGEMENTS
The research has been funded by the Indiana Department of Transportation through the Joint Transportation Research Project.
INTRODUCTION

Safety on roadways has always been a deep concern of transportation engineers. Considerable effort has been done to increase safety on road networks. Gunnarson (5) states that roadway safety may be defined as the acceptability of risk, where risk is further described as the consequence of a crash or the probability that a crash will happen. A roadway is judged safe if its risks are judged to be acceptable. A crash is defined as an undesirable, suddenly occurring event that may result in human and material losses. Motor vehicle crashes create significant delays and also negatively impact road safety by often leading to secondary crashes as well. Several factors including degree of congestion, facility type, geometric characteristics and weather conditions may contribute to the number and severity of crashes.

Freeways form the most important segment in the road network system in the United States. Therefore, an attempt to increase safety on roads should include the free ways. As statistics indicate, work zones on freeways have always been potential sites for crashes. Studies indicate that a greater number of crashes occur at locations with work zones than at those without work zones. A number of factors have been cited as being responsible for the increase in the number of crashes at locations in the work zones.

Most of the research on work zone safety concentrated on the segment inside the work zone. Approaches to work zones have so far been neglected. But, problems with the disturbed flow of traffic occur on approaches to work zones, especially at sections immediately before work zone locations, where one or more lanes are discontinued. Thus, work zone entry points seem to be more dangerous due to the aggressive behavior of certain drivers. Some drivers try to avoid the congested traffic conditions on the continuous lanes by approaching the work zone in the discontinued lane up to the point where a lane change maneuver becomes difficult and risky. Such aggressive lane changes maneuvers create turbulence in the traffic stream, which negatively affects performance. The effects are shock waves in the continuous lane and development of road rage. All these culminate in a potentially dangerous situation both at the merge point and within the work zone, which continues far beyond the point at which the aggressive behavior took place.

The influence of approaches to safety in work zones has been realized by the Indiana Department of Transportation and appropriate steps have been taken to improve safety in work zone locations. In addition to traditional traffic management, special traffic control devices are being installed on approaches to the work zones. An important advancement in this direction has been the development and installation of Indiana Lane Merge System (ILMS), Tarko et. al. (10) on approaches to work zones. ILMS developed by Indiana Department of Transportation (INDOT) and being evaluated by the Purdue University team is an advanced dynamic traffic control system for devising merging locations depending on congestion levels on approaches to work zones. The presented research is a part of a project aimed at studying the performance and efficacy of the ILMS in work zone locations. The crash prediction models, presented here, are important
components of the project. The present research continues some of the earlier work done in this field by trying to identify factors responsible for crashes in freeway work zone locations and their approaches, trying to fill in the gap present in studies conducted in the past. Crash prediction models for sections inside the work zones and on approaches to work zones, provide a comprehensive tool for evaluating safety of construction zones.

**LITERATURE REVIEW**

An extensive literature survey was conducted to overview previous studies conducted on the subject. A number of papers were found that studied characteristics of construction zone crashes. A paper by Pigman and Agent (9) indicates the following factors as being responsible for increase in the number of crashes at work zone sites: (a) inappropriate use of traffic control devices, (b) poor traffic management, (c) inadequate overall layout of the work zone, and (d) general misunderstanding of the unique problems associated with construction and maintenance work zones. Hall and Lorenz (6) conducted a similar study on highway construction work zones in New Mexico for a three-year period. The authors came to the conclusion that some accident characteristics are typical for construction zones. They also ruled out the influence of adverse weather condition as being responsible for the increase in construction zone crashes. The authors noted that proper traffic control is able to reduce the number of crashes and perhaps, to influence the pattern of collisions. A study conducted by Casteel and Ullman (3) looks into the influence of entrance ramps in long-term construction work zones.

Some literature was found on modeling of crashes on freeways. A study by Madanat et al. (7) used binary logit models to predict the likelihood of vehicle crash incidents on the Borman Expressway. Studies conducted by Zeeger et al. (12) and Cleveland and Kitamura (4) investigated the relationships between accident frequency, roadway geometry, and roadside conditions. Benekohal and Hashmi (1) developed crash prediction models as a part of their attempt to estimate accident reduction factors on highways. Vogt and Bared (11) developed and analyzed accident models for two lane rural roads: segments and intersections.

The authors have found even less research on safety relationships for work zones. Since a preliminary study conducted by the authors has proven that approaches to work zones are as dangerous as work zones themselves, this study attempts to develop safety models for sections on approaches and inside work zones. Crash prediction models have been developed for PDO and fatal and injury crashes.

**DATA COLLECTION**

The data used in this study is classified into three categories.
1. Work zone characteristics,
2. Crash characteristics,
3. Road and traffic characteristics.
Work zone data

The required data regarding work zones is:
1. The freeway identifier (example, I-64, I-65 etc.),
2. The work zone location (mile markers at the starting and the ending points of the work zone),
3. The cost of the project,
4. The work code,
5. The duration of the work zone,
6. The length of the work zone.

The data obtained from the INDOT provided information about 393 construction projects that took place between 1993 and 1997.

Crash Data

Comprehensive data for work zone crashes was provided by INDOT. The crash database included not only crashes inside the work zone, but, also crashes on approaches to work zones. Only crashes indicated by the investigating police officers as that occurred due to the construction activities were included in the analysis. The final comprehensive database had a total of 5025 crashes for the period from 1993 to 1997. The final database had the following data:

1. The interstate identifier,
2. The exact mile marker of the crash (distance in miles from the state line to the crash location),
3. The time of the day, the day of the week, the month and the year of the crash,
4. Number of injuries, fatalities etc,
5. Number of vehicles involved,
6. Type of collision,
7. Weather and light conditions,
8. Road conditions,
9. Other vehicle and driver information.

Almost all the data obtained from INDOT were in a convenient format. Only the location of the crash required further processing to determine the distance between the crash location and the beginning of the work zone.

Road and Traffic Data

The authors retrieved relevant freeway data from a GIS database created by INDOT. Using TRANSCAD built-in filters, the authors obtained the following data.

1. The number of on and off ramps inside the work zone,
2. The number of ramps on the approach at less than 2 miles from the beginning of the work zone,
3. The number of ramps on the approach between 2 and 10 miles from the beginning of the work zone,
4. The Average Daily Traffic (ADT),
5. Percent of heavy vehicles.

The ADT values were used instead of Annual Average Daily Traffic (AADT) values to account for the seasonal and daily fluctuations in traffic volumes. The database provided by INDOT contained the AADT values and volume adjustment factors that were used to convert AADT into ADT. Percentage of heavy vehicles data was also available. Since, these are predictive models, the authors use the data available (average volume, AADT). It would be better to use actual volume during work zone conditions. But, this model was supposed to predict expected number of crashes before the work actually started. Therefore, keeping the practicality aspect in mind and also due to the unavailability of data, the authors decided to use the volume during normal periods as a substitute. This approximation is very appropriate with respect to this study since it is confined to rural freeway work zones. In rural freeways, the option of finding major alternate routes are rare and so the reduction in demand will be minimal. Therefore even if some bias is present it is expected to be limited.

Final Database

An initial analysis was conducted to understand the nature of the problem that was being studied and if possible, to identify some responsible factors. As mentioned earlier, the number of rural freeway work projects available for the study numbered 393 for the period 1993-1997. Some of the minor work projects were done in conjunction with the major projects. So, the actual number of work zones were actually lesser than the number of the work projects commissioned for the period. The number of work zones available for study numbered 243. The first attempt was to check the completeness of data for the set of work zones. All work zones with missing critical data like the work zone location, duration and costs were identified and attempts were made to obtain this missing data. The work zones for which no data could be traced had to be discarded from the study.

Since the study was confined to rural freeway work zones, the next step was to remove all urban freeway work zones. It was assumed that rural freeway roads are always two lane divided highways and hence, any highway with 3 or more lanes were removed from the data base. The final database chosen for detailed analysis comprised of about 117 work zones.

It was also decided to classify the work zones based on the number of on / off ramps on the approaches to the work zone locations. Ramps were considered important to the study, because they might affect traffic pattern. For example, an off ramp close to the work zone may encourage some drivers to divert from the congested freeway to save time spent in the queue. Ramps beyond a distance of 16 kilometers (10 miles) were assumed to have no effect on the traffic. The study conducted by the authors indicated that traffic backup longer than 16 kilometers is very rare. Ramps on the approaches to work zones are classified into three categories.

i) Ramps at a distance of 16 kilometers or greater,
ii) Ramps at a distance between 3.2 kilometers and 16 kilometers,
iii) Ramps at a distance less than 3.2 kilometers.

The ramps at distances less than two miles and those between two and 10 miles are put in separate categories because it seems plausible that the closer the ramp the greater is its effect on the traffic.

**CRASH ASSIGNMENT TO WORK ZONES**

Safety inside a work zone may differ from safety on approach to work zones. Barriers, reduced dimensions of cross sections, construction activities, and passing restrictions may influence traffic inside the work zone. Shock waves of congestion and aggressive lane changes influence traffic safety on the work zone approach. That is why, the work zone-related crashes were separated between work zone segments and work zone approaches to enable separate safety analyses.

It must be pointed out that the crashes included in the analysis had been classified by the investigating police officers as work zone related. The task that the authors faced was to distribute the crashes between the work zone segments and work zone approaches and not to determine which crashes had been caused by the construction activities and which ones were not.

Work zone crashes are defined henceforth as crashes that take place inside the work zone. A certain crash was assigned to a work zone if (1) the crash location fell between the beginning and end of a particular work zone and (2) the time of the crash coincided with the work zone presence.

Approach crashes are crashes that take place upstream of the work zone and are caused by the work zone presence. A certain crash was assigned to a work zone approach if (1) the crash location fell within the estimated congested segment upstream of the work zone beginning and (2) the time of the crash coincided with the work zone presence.

The length of the congested segment can be calculated using the following equation, (Tarko et al., 10):

\[ L_c = 0.92 \times Q_{of} (D_c - D) \]  

where:
- \( L_c \) = length of congested segment, km,
- \( Q_{of} \) = maximum overflow queue, veh,
- \( D_c \) = average congested density (in-queue density), veh/km,
- \( D \) = average unaffected density, veh/km.

The maximum overflow queue is estimated using:
\[ Q_{of} = \sum_{i=1}^{n} (V_i - C) \]  

(2)

\( V_i \) is the demand for interval \( i \) and \( C \) is the capacity of the road segment under construction. This is summed over all intervals with overflows.

An approximate estimate of density for two lane rural freeways in Indiana (Tarko et al., 1998) is used for the purpose,

\[ D = 46.3 - \sqrt{2150 - 0.742V} \]  

(3)

The average congestion density on two lane approaches to work zones observed at two sites in Indiana is 71.6 veh/km.

**Initial Safety Study**

Before the model was developed, the authors decided to do a preliminary study of crashes in work zone locations for all the interstates in Indiana. As explained earlier, the total number of work zones selected for study was 117. The total number of crashes associated with these work zones numbered 2035. This included 696 injury crashes and 33 fatal crashes. Separate analyses were conducted for approach crashes and for work zone crashes. The crashes were classified by their severity levels to show the gravity of the problem. The results have been shown in Table 1.

Table 1 reveals the magnitude of safety problem on work zone approaches. On an average, a freeway work zone in Indiana experiences about 18 crashes, 9 on the approach to the work zone and 9 inside the works zone. It is obvious that the frequency of crashes on approaches to work zones is very similar to that inside work zones. Unlike the frequencies, the severity of crashes on approaches seems to be greater than that inside the work zone. Percentage of injury plus fatal crashes on approaches represents about 40% of crashes on approaches to the work zone while inside the work zone it is about 30%. These results indicate a very serious problem on approaches to work zones. This situation can be caused by typically higher speeds on work zone approaches and sudden lane change maneuvers from the discontinued lane into the continuous lane.

A cursory look at the average number of crashes per work zone reveals the effect of traffic volumes on crashes. The highest average number of crashes are for interstates I-65 and I-94 which also have the highest volumes of traffic in the state. The present study hopes to identify factors that are responsible for these crashes, to construct crash prediction models and also, to suggest some measures by which safety on work zones can be enhanced.

**MODELING**

The authors decided to develop a regression model for predicting the number of crashes at a work zone location. The models predict the expected number of crashes for a
single work zone, given the work zone characteristics such as ADT and duration of work. Predictive models were deemed suitable for the purpose, since our aim was to determine “crash potential” of work zones even before the work starts. Therefore, the aggregate crash prediction model was considered as the best choice available. The following variables were identified as potential factors for crashes at work zone locations.

1. **Cost of work**, \( C \), is the total contract amount paid in $ 1000s by INDOT to the contractors for the project. It is expected to be correlated with duration and magnitude of work. This variable is used to calculate the intensity of work. The intensity of work is estimated as \( \text{cost} / (\text{duration} \times \text{length of work zone}) \). The intensity is believed to have a positive correlation with crashes, i.e., the crashes are supposed to increase as the intensity of work increases.

2. **Average daily traffic volume** \((Q)\), The ADT volumes are used instead of AADT volumes to take into account the daily and seasonal variability of traffic. The ADT volumes were obtained for both the approaches and inside the work zone. These values can differ due to the presence of entry or exit ramps. The volume is assumed to be an exposure-to-risk variable; any increase in volume with other parameters remaining the same will increase the number of crashes.

3. Ramps on work zone segments and on work zone approaches are represented by the following variables:
   - \( R_2 \), the number of ramps at a distance of 3.2 kilometers or shorter to the work zone.
   - \( R_{10} \), the number of ramps at distances between 3.2 and 16 kilometers on approaches to the work zone.
   - \( R_{\text{ON}} \) and \( R_{\text{OFF}} \), the number of on and off ramps on approaches or inside the work zone.

4. **Work zone length** \((L)\). This measure of exposure-to-risk can be used to estimate the vehicle miles traveled. The work zone length is measured in kilometers. For all bridge works only the starting point have been given in the data, indicating very short work areas. To rectify this problem, we assume a fixed value of 0.008 kilometers (0.005 miles) for all bridge work zones and also for all work zones for which only the starting points have been specified. The length of work zone is an exposure-to-risk variable for crashes inside work zones. However for approach crashes, the effect of length is not very apparent. The length maybe used to substitute for some other variables not be included in the model.

5. **Duration of work** \((T)\), This is the number of days when construction zone was present. The duration of work should also have an almost positive linear influence on crashes given other parameters constant. The objective of the study was to develop practical crash prediction models. The data that was available to the authors was the starting date and the ending date of the construction. But, no specific information was available about periods with actual work force in place. Since the idea was to develop practical models for predictive purpose, the authors decided to use the total duration of construction as it is easier to predict.
6. **Type of work (W)**. This indicates the type of work using letter codes as follows:

- Road rehabilitation, resurfacing, other road works, J
- Bridge Rehabilitation and repair, C
- Bridge replacement works, E
- Roadside maintenance, landscaping etc., N
- Sign painting, signal installation etc., V
- Interchange work, R
- Other type, X

Works differ from each other by the visual and physical distraction to the traffic, by the construction equipment present in the work site, and by the number of people involved. For example, a road rehabilitation work/road re-surfacings may be expected to have a greater effect to traffic than painting signs. The one reason for this is that resurfacing is very frequently associated with lane closures whereas the other works usually are not. Since lane closure is believed to be a significant safety factor especially on approaches, the road works have been classified into two categories, those with lane closure (type J) and those without (other types). Since a mathematical model was to be developed, it was necessary that the independent variables had numerical values. All J works were given the value W=1 and the rest of the works were given the value W=0.

The models are being developed to predict expected number of crashes on typical rural freeway work zones. The past data base used by the authors is assumed to be representative of typical rural work zones and hence they represent the typical past police presence in such locations. Of course, unusual police activity in some work zones could cause biases in results. But, unfortunately the authors did not have the data to incorporate these factors.

The dependent variables are as follows:

a) Number of crashes (all types included), A
b) Number of pdo crashes, D
c) Number of fatal and injury crashes, I

It is reasonable to assume that crashes occurring on a particular roadway segment are independent of one another and that a certain mean number of crashes is characteristic of the given location and of other locations with the same properties. This particular property makes Poisson or negative binomial models a reasonable choice. Poisson and negative binomial models, seem to be a better method of modeling discrete rare events such as roadway accidents, Miaou and Lum (8). The negative binomial model is superior to Poisson since the negative binomial model allows for extra variation caused by other variables not included in the model. This variation is represented by the overdispersion parameter. Several alternative forms of the model were tested. The form of the model with the best overall feature has been:

$$A = K(Q)^{\beta_1}(T)^{\beta_2}(L)^{\beta_3} \exp\left(\sum_i \delta_i X_i \right)$$
where:

\[ A = \text{number of crashes on the approaches to or inside the work zone}, \]
\[ Q = \text{average daily traffic, both on approaches and inside the work zone} \]
\[ L = \text{length of the work zone}, \]
\[ T = \text{duration of work} \]
\[ K = \text{slope parameter}, \]
\[ X_i = \text{explanatory variable (R}_2, R_{10}, R_{ON}, R_{OFF}, C/LT, W \text{ etc.}), \]
\[ \beta_i, \delta_i = \text{coefficients of factors}. \]

Variables \( Q, T \) and \( L \) and their products represent the exposure to risk. Since it seems plausible that for crashes on approaches to the work zone, the length \( L \) is not an exposure-to-risk factor, it was moved inside the exponential function for approach crashes and were used as exposure-to-risk variable in safety models for work zone crashes. Also, Cost was dropped and Duration of work was included in its place. Several models were tested changing the positions of variables, but the following forms of models were found to be the most suitable:

- **Work zone crashes**:
  \[ A = K(Q)^{\beta_1} (T)^{\beta_2} (L)^{\beta_3} \exp(\gamma_1 \left( \frac{C}{LT} \right) + \gamma_2 W + \gamma_3 R_{ON} + \gamma_4 R_{OFF}) \]

- **Approach crashes**:
  \[ A = K(Q)^{\beta_1} (T)^{\beta_2} \exp(\gamma_1 L + \gamma_2 \left( \frac{C}{LT} \right) + \gamma_3 W + \gamma_4 R_2 + \gamma_5 R_{10} + \gamma_6 R_{ON} + \gamma_7 R_{OFF}) \]

where:

\[ \frac{C}{LT} = \text{Explanatory variable for representing the intensity of the work}, \]
\[ W = \text{Binary variable for work type, 1 for road works, 0 for non-road works}. \]

Table 2 summarizes the models by regression analysis for various severity levels and for two work zone components: approaches and crashes. The table also gives the \( R^2 (1 - (\log \text{likelihood/restricted log likelihood})) \) values and over dispersion parameters which help determine the goodness of the model. Tables 3 and 4 present the detail results of the various models. It gives us the calibration values of the various variables, their standard errors, t-ratios, and p-values. Figures 1-6 compare the observed and predicted crash counts for the investigated work zones. A discussion of these results follows the tables and the graphs.
DISCUSSION

General Discussion

The software used for regression analysis, in this study, is LIMDEP. As can be seen, Negative Binomial models were used for the purpose of developing aggregate crash prediction models. The overdispersion parameter $\alpha$ is highly significant for all the cases (Table 2) indicating that the selection of negative binomial model for regression was a good choice. The goodness of the models is measured with the $\rho^2$ value and the standard error of estimation of the expected count $Y$.

The $\rho^2$ value $(1 - (\log \text{likelihood/restricted log likelihood}))$ for all the models ranged from 0.20 to about 0.43 (Table 2) indicating reasonable results. The variance of the expected count estimate $Y$ is the variance of the crash counts at $Y$ reduced by the Poisson variance. In negative binomial model this variance is $= \alpha Y^2$, where $\alpha$ is the overdispersion parameter. Thus the estimation error is $\alpha^{1/2} Y$ and therefore, the relative error of estimation is $100 \alpha^{1/2}$. The relative error of estimation for the models developed varies between 75% and 99%. Such errors might be due to variables that have not been included in the model. However, the estimation error is comparable to the estimation errors obtained in other research.

To decide whether a particular variable is significant or not we use the p values (Tables 3 and 4). p-value is the probability that the estimated coefficient has the value shown or larger when the true coefficient is zero. If the p value is less than 10%, then that particular variable is accepted, otherwise rejected.

The graphs (1 through 6) indicate that the results are more or less unbiased. At lower values, the points are close together and as the observed value increases, the dispersion seems to be increasing. This is in complete agreement with the overdispersion values obtained from the regression analysis. Since overdispersion is greater than zero for all cases, dispersion will be small for low values of the observed dependent variable (in this case, the crashes) and it will increase with the increase in observed values. For injury and fatal crashes for both crashes on approaches and inside crashes, dispersion even at low values of observed crashes seems to be more than for other categories. This probably might be due to the fact that less factors responsible for injury crashes have been included into the model compared to the other categories.

Cost had a high correlation with duration, length and work type. But, the inclusion of correlated variables in the same model does not affect the contribution of these variables to the expected crashes. In fact, even when the variables are correlated, the slopes of the variables in the final model represent their true effect. Only, when the independent variable has a high degree of correlation with some missing variable, then the effect (slope in the model) of that variable has to be questioned. In such cases, the slope of the variable is representative of the variable as well as the one that has been omitted. The authors tested this on a set of correlated variables by using Monte Carlo method and LIMDEP.
The change in standard errors were tested by bringing in and taking out independent variables. The final variables which were included in the model had a significant impact on the model. Exclusion of these variables, caused significant increase in the overdispersion parameter. On the other hand, inclusion of non-significant variables in the model caused very insignificant increase in the over dispersion parameter hinting that these variables might not be very representative of crashes. The slopes of independent variables were fairly stable under co-variate inclusion and exclusion.

Discussion of Factors

As could be expected, traffic volumes turned out to be a primary factor affecting crashes during road construction. The regression parameter associated with volume is positive and very similar for both crashes inside and on the approaches to the work zone. The volume parameter had a value close to one for all crashes inside work zone segments and on approaches to the work zone and also for PDO crashes on the approaches to the work zone. This indicates a fairly linear relationship between traffic volumes and the crash numbers.

The length of the work zone was found to be a strong factor in all the cases for work zone crashes. The calibration parameter is close to 1 in all cases for work zone crashes indicating the dependence is almost linear. The calibration parameter is also positive which means that expected number of crashes increase with the length of the work zone. Thus, the work zone length turned out to be an exposure-to-risk factor for inside segments.

The length of the work zone was not expected to be a factor for crashes on approaches to work zones. In contradiction to our expectations, the variable turned out to be statistically very significant. It was found that shorter work zones had a larger number of approach crashes than longer work zones. The authors suspect that the work zone length may represent the effect some factors omitted in the model. One such omitted factor may be traffic management on work zones and that length can substitute for traffic management. For long work zones, usually because of their long time periods, the traffic management is more intensive than for short work zones.

The duration of work turned out to be a significant factor in all cases. For all crashes, the factor was much less than one, at around 0.6. This shows that, the number of crashes do not increase linearly with the duration, but tend to taper off after some time. In other words, the marginal increment in crashes with the increment in the duration of the work zone, tends to decrease as the duration of work zone keeps increasing. If the duration of a work zone is large, the familiarity of the drivers using the highway keeps increasing with time. More familiarity usually leads to safer driving and hence lesser number of accidents.
The effect of Ramps was found to be insignificant for both the work zone segments and the work zone approaches. This probably indicates an effective traffic management at ramp-freeway junctions during construction activities.

The type of work did not turn out to be a factor for approach crashes. This was a surprising result since the authors assumed that the lane closure will have a strong effect on approach crashes. However, this factor was highly significant for total and PDO crashes inside the work zone. This is a plausible result; since, the type of work will affect mainly vehicles inside the work zone. The type of work determines the construction equipment and personnel presence inside the work zone and the temporary lane closures. Usually long road work zones have, often one of the roadways closed and both traffic directions use the remaining roadway equipped with a separation. This solution causes discomfort to some drivers and may increase the risk of accidents.

The intensity of work represented by (Cost/(Duration x Length)) turned out to be a significant factor, only for work zone crashes. This is to be expected since, intensity of work affects the vehicles and drivers directly only inside the work zones. This is an important finding, because this shows that small costly work zones are the most dangerous. The previous result that shorter work zones have a greater number of approach crashes also seems to indicate the same fact. The intensity parameter did not turn out to be a significant factor for approach crashes.

**CLOSURE REMARKS**

Mathematical models were developed in this study to predict the expected number of crashes in freeway work zones, both inside and on approaches to the work zone. As expected, the traffic volume, length and duration of work turned out to be significant factors. In addition, the cost of work zone (as a measure of intensity of work) and the type of work were also critical factors of safety inside work zones. An important contribution of the study was that it developed separate models for crashes on approaches to work zones and for crashes inside the work zone segments. The results are very useful in predicting the crash potential of a work zone before work starts.

These models are very useful in optimizing work zone schedules and for deciding better pavement management strategies. The study shows that a long single work zone is better than two short work zones. Thus, if there are two similar work zones close to each other, it is safer to combine the two work zones than do them individually. The frequency of pavement rehabilitation is often decided after taking into account the benefits and costs of the project. The cost of crashes can be included as a cost in the benefit-cost analysis for more efficient pavement management strategies.

With the advent of new ITS technologies and user information systems like the Advanced Traffic Information Systems (ATIS), the user can be given information about the “crash potential” of a particular work zone, so that, a safer route may be adopted. The present study can be extended to predict the probability of a crash occurring in a work zone during a particular hour given the work zone and temporal traffic characteristics.
Such models will be useful in giving dynamic real-time information to the user about the crash potential of the work zone during that hour and thus will make re-routing strategies more fluent. Such models would give a more comprehensive evaluation of work zone safety to address safety issue for intelligent or smart work zones.

REFERENCES

### TABLE 1 Crash Statistics for Crashes on Approaches to and Inside the Work Zone

<table>
<thead>
<tr>
<th>Interstate</th>
<th>All crashes</th>
<th>Injury crashes</th>
<th>Fatal crashes</th>
<th>% of injury and fatal crashes</th>
<th>Number of work zones</th>
<th>Average per work zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-64</td>
<td>43</td>
<td>19</td>
<td>1</td>
<td>46.51</td>
<td>19</td>
<td>2.26</td>
</tr>
<tr>
<td>I-65</td>
<td>311</td>
<td>103</td>
<td>7</td>
<td>35.37</td>
<td>33</td>
<td>9.42</td>
</tr>
<tr>
<td>I-69</td>
<td>343</td>
<td>156</td>
<td>2</td>
<td>46.06</td>
<td>28</td>
<td>12.25</td>
</tr>
<tr>
<td>I-70</td>
<td>186</td>
<td>63</td>
<td>7</td>
<td>37.63</td>
<td>14</td>
<td>13.29</td>
</tr>
<tr>
<td>I-74</td>
<td>70</td>
<td>31</td>
<td>4</td>
<td>50</td>
<td>18</td>
<td>3.89</td>
</tr>
<tr>
<td>I-94</td>
<td>89</td>
<td>20</td>
<td>0</td>
<td>22.47</td>
<td>4</td>
<td>22.25</td>
</tr>
<tr>
<td>All</td>
<td>1042</td>
<td>392</td>
<td>21</td>
<td>39.64</td>
<td>116</td>
<td>8.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interstate</th>
<th>All crashes</th>
<th>Injury crashes</th>
<th>Fatal crashes</th>
<th>% of injury and fatal crashes</th>
<th>Number of work zones</th>
<th>Average per work zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-64</td>
<td>62</td>
<td>20</td>
<td>2</td>
<td>35.48</td>
<td>19</td>
<td>3.26</td>
</tr>
<tr>
<td>I-65</td>
<td>313</td>
<td>114</td>
<td>1</td>
<td>36.74</td>
<td>33</td>
<td>9.48</td>
</tr>
<tr>
<td>I-69</td>
<td>212</td>
<td>71</td>
<td>1</td>
<td>33.96</td>
<td>28</td>
<td>7.57</td>
</tr>
<tr>
<td>I-70</td>
<td>192</td>
<td>55</td>
<td>6</td>
<td>31.71</td>
<td>14</td>
<td>13.7</td>
</tr>
<tr>
<td>I-74</td>
<td>114</td>
<td>32</td>
<td>1</td>
<td>28.95</td>
<td>18</td>
<td>6.33</td>
</tr>
<tr>
<td>I-94</td>
<td>100</td>
<td>12</td>
<td>0</td>
<td>12</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>All</td>
<td>993</td>
<td>314</td>
<td>11</td>
<td>31.72</td>
<td>116</td>
<td>8.56</td>
</tr>
</tbody>
</table>
TABLE 2 Crash Prediction Models and their Overall Performance

<table>
<thead>
<tr>
<th>Work Zone Approaches</th>
<th>Model</th>
<th>$\rho^2$</th>
<th>Overdispersion Parameter, $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A = 0.0636 \cdot (Q)^{1.7444} \cdot (T)^{0.6555} \cdot \exp(-0.0388 \cdot L)$</td>
<td>0.427</td>
<td>0.6809</td>
</tr>
<tr>
<td></td>
<td>$I = 0.006621 \cdot (Q)^{1.2786} \cdot (T)^{0.8651} \cdot \exp(-0.0586L)$</td>
<td>0.36</td>
<td>0.9839</td>
</tr>
<tr>
<td></td>
<td>$D = 0.01743 \cdot (Q)^{0.9810} \cdot (T)^{0.7931} \cdot \exp(-0.0381L)$</td>
<td>0.30</td>
<td>0.6740</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Work zone Segments</th>
<th>Model</th>
<th>$\rho^2$</th>
<th>Overdispersion Parameter, $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A = 0.00217 \cdot (Q)^{1.1588} \cdot (T)^{0.5126} \cdot (L)^{0.760} \cdot \exp(0.1615(C/LT) + 2.308(W))$</td>
<td>0.33</td>
<td>0.5593</td>
</tr>
<tr>
<td></td>
<td>$I = 0.00812 \cdot (Q)^{1.0497} \cdot (T)^{0.5263} \cdot (L)^{0.8531} \cdot (C/LT)^{0.3743}$</td>
<td>0.20</td>
<td>0.7940</td>
</tr>
<tr>
<td></td>
<td>$D = 0.0008 \cdot (Q)^{1.1901} \cdot (T)^{0.4952} \cdot (L)^{0.9956} \cdot \exp(0.1851(C/LT) + 2.3279(W))$</td>
<td>0.33</td>
<td>0.7003</td>
</tr>
</tbody>
</table>

Where:
- $A$ = Total number of crashes,
- $I$ = Number of injury and fatal crashes,
- $D$ = Number of PDO crashes,
- $Q$ = Average Daily Traffic, in 10000’s,
- $L$ = Length of the work zone segment in km,
- $T$ = Duration of the project in days,
- $C$ = Cost of the construction project in $ 1000’s,
- $C/LT$ = Proxy variable for intensity of work,
- $W$ = Work type; $W=1$ for J type work, $W=0$ for other types.
- $\rho^2 = (1 - (\text{log likelihood/restricted log likelihood}))$
TABLE 3 Calibration Parameters of Crash Prediction Models for Crashes on Approaches to Work Zones

<table>
<thead>
<tr>
<th>Total crashes on approaches to the work zone</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>-2.7540</td>
<td>1.4011</td>
<td>-1.9657</td>
<td>0.0493</td>
</tr>
<tr>
<td>Q</td>
<td>1.1444</td>
<td>0.2870</td>
<td>3.9867</td>
<td>0.0001</td>
</tr>
<tr>
<td>T</td>
<td>0.6555</td>
<td>0.2728</td>
<td>2.4025</td>
<td>0.0163</td>
</tr>
<tr>
<td>L</td>
<td>-0.0388</td>
<td>0.0219</td>
<td>-1.7665</td>
<td>0.0773</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.6809</td>
<td>0.1287</td>
<td>5.2906</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Injury and Fatal crashes on approaches to the work zone</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>-5.0175</td>
<td>2.2256</td>
<td>-2.2544</td>
<td>0.0242</td>
</tr>
<tr>
<td>Q</td>
<td>1.2786</td>
<td>0.4696</td>
<td>2.723</td>
<td>0.0065</td>
</tr>
<tr>
<td>T</td>
<td>0.8651</td>
<td>0.4219</td>
<td>2.0506</td>
<td>0.0403</td>
</tr>
<tr>
<td>L</td>
<td>-0.0586</td>
<td>0.0279</td>
<td>-2.1004</td>
<td>0.0357</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.9839</td>
<td>0.2938</td>
<td>3.3487</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PDO crashes on approaches to the work zone</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>-4.0496</td>
<td>1.2934</td>
<td>-3.1309</td>
<td>0.0017</td>
</tr>
<tr>
<td>Q</td>
<td>0.9810</td>
<td>0.2849</td>
<td>3.4432</td>
<td>0.0006</td>
</tr>
<tr>
<td>T</td>
<td>0.7951</td>
<td>0.2278</td>
<td>3.4899</td>
<td>0.0005</td>
</tr>
<tr>
<td>L</td>
<td>-0.0381</td>
<td>0.0187</td>
<td>-2.0374</td>
<td>0.0471</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.6740</td>
<td>0.1310</td>
<td>5.1420</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

K = Slope

$t$-ratio = Coefficient / standard error

p-value = The P-value is the probability that the estimated coefficient has the value shown or larger when the true coefficient is zero
## TABLE 4 Calibration Parameters of Crash Prediction Models for Crashes Inside Work Zones

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total crashes inside the work zone</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>-6.1332</td>
<td>1.3514</td>
<td>-4.5384</td>
<td>0.0000</td>
</tr>
<tr>
<td>Q</td>
<td>1.1588</td>
<td>0.2207</td>
<td>5.2501</td>
<td>0.0000</td>
</tr>
<tr>
<td>T</td>
<td>0.5126</td>
<td>0.2425</td>
<td>2.1138</td>
<td>0.0345</td>
</tr>
<tr>
<td>L</td>
<td>0.7601</td>
<td>0.1573</td>
<td>4.8331</td>
<td>0.0000</td>
</tr>
<tr>
<td>C/(LT)</td>
<td>0.1615</td>
<td>0.0558</td>
<td>2.8969</td>
<td>0.0038</td>
</tr>
<tr>
<td>W</td>
<td>2.3080</td>
<td>0.3128</td>
<td>7.3776</td>
<td>0.0000</td>
</tr>
<tr>
<td>α</td>
<td>0.5593</td>
<td>0.1437</td>
<td>3.8926</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Injury and Fatal crashes inside the work zone</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>-4.8139</td>
<td>1.8217</td>
<td>-2.6425</td>
<td>0.0082</td>
</tr>
<tr>
<td>Q</td>
<td>1.0497</td>
<td>0.3996</td>
<td>2.627</td>
<td>0.0086</td>
</tr>
<tr>
<td>T</td>
<td>0.5263</td>
<td>0.298</td>
<td>1.7664</td>
<td>0.0773</td>
</tr>
<tr>
<td>L</td>
<td>0.8531</td>
<td>0.2358</td>
<td>3.6177</td>
<td>0.0003</td>
</tr>
<tr>
<td>C/(LT)</td>
<td>0.3743</td>
<td>0.1832</td>
<td>2.0427</td>
<td>0.0411</td>
</tr>
<tr>
<td>α</td>
<td>0.7940</td>
<td>0.2311</td>
<td>3.4366</td>
<td>0.0006</td>
</tr>
<tr>
<td><strong>PDO crashes inside the work zone</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>-7.131</td>
<td>1.5828</td>
<td>-4.5053</td>
<td>0.0000</td>
</tr>
<tr>
<td>Q</td>
<td>1.1901</td>
<td>0.2211</td>
<td>5.3825</td>
<td>0.0000</td>
</tr>
<tr>
<td>T</td>
<td>0.4952</td>
<td>0.2872</td>
<td>1.7241</td>
<td>0.0847</td>
</tr>
<tr>
<td>L</td>
<td>0.9956</td>
<td>0.2363</td>
<td>4.213</td>
<td>0.0000</td>
</tr>
<tr>
<td>C/(LT)</td>
<td>0.1851</td>
<td>0.0833</td>
<td>2.223</td>
<td>0.0262</td>
</tr>
<tr>
<td>W</td>
<td>2.3279</td>
<td>0.5064</td>
<td>4.597</td>
<td>0.0000</td>
</tr>
<tr>
<td>α</td>
<td>0.7003</td>
<td>0.1936</td>
<td>3.617</td>
<td>0.0003</td>
</tr>
</tbody>
</table>
Figure 1: Predicted vs observed values for all crashes on work zone approaches

Figure 2: Predicted vs observed values for injury + fatal crashes on approaches to work zones
Predicted vs observed values for pdo crashes on approaches to work zones

Figure 3: Predicted vs observed values for pdo crashes on approaches to work zones

Predicted vs observed values for total crashes inside the work zone

Figure 4: Predicted vs observed values for all crashes in work zone segments
Figure 5: Predicted vs observed values for injury and fatal crashes in work zone segments

Figure 6: Predicted vs observed values for pdo crashes in work zone segments
List of tables
1. Crash Statistics for Crashes on Approaches to and Inside the Work Zone
2. Crash Prediction Models and their Overall Performance
3. Calibration Parameters of Crash Prediction Models for Crashes on Approaches to Work Zones

List of figures
1. Predicted vs observed values for all crashes on work zone approaches to work zones
2. Predicted vs observed values for injury and fatal crashes on approaches to work zones
3. Predicted vs observed values for pdo crashes on approaches to work zones
4. Predicted vs observed values for all crashes in work zone segments
5. Predicted vs observed values for injury and fatal crashes in work zone segments
6. Predicted vs observed values for pdo crashes in work zone segments