Work Zone Safety Analysis and Modeling: A State-of-the-Art Review

Hong Yang, Ph.D.
(Corresponding author)
Post-Doctoral Fellow
Department of Civil & Urban Engineering, NYU-Poly
Center for Urban Science + Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor
Brooklyn, NY 11201
Tel: +1-(646)-997-0548 Fax: +1-(646)-997-0560
Email: hong.yang@nyu.edu

Ozgur Ozturk
Ph.D. Candidate, Graduate Assistant,
Rutgers Intelligent Transportation Systems (RITS) Laboratory
Department of Civil and Environmental Engineering
Rutgers University
96 Frelinghuysen Road, Piscataway, NJ 08854, USA
Phone: (732) 445-0579 x127, Fax: (732) 445-0577
Email: ozturko@eden.rutgers.edu

Kaan Ozbay, Ph.D.
Professor,
Department of Civil & Urban Engineering, NYU-Poly
Center for Urban Science + Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor
Brooklyn, NY 11201
Tel: +1-(646)-997-0552 Fax: +1-(646)-997-0560
Email: kaan.ozbay@nyu.edu

Kun Xie
Graduate Assistant,
Department of Civil & Urban Engineering
Polytechnic Institute of New York University (NYU-Poly)
One MetroTech Center, 19th Floor
Brooklyn, NY 11201
Phone: +1-(646)-997-0547 Fax: +1-(646)-997-0560
Email: kun.xie@nyu.edu

Word count: 6246 Texts + 4 Table + 4 Figures = 8246
Abstract: 139
Submission Date: August 1, 2013
Resubmission Date: November 15, 2013

TRB Paper ID: 14-1692
Transportation Research Board’s 93rd Annual Meeting, Washington, D.C., January 2014
ABSTRACT

Prevention of work zone crashes is one of the top priorities for transportation agencies. To make more informed decisions on initiating appropriate programs and countermeasures, more precise information on the underlying mechanisms of work zone crashes is needed. Considerable research effort has been directed to examine work zone crash characteristics and possible causal factors in the last five decades. This study is designed to provide a thorough review of existing research focused on work zone crash-related analysis and modeling. Literature from multiple sources is carefully combed to determine about the state-of-knowledge related to the work zone safety problem. Experience and lessons learned from these studies are presented to highlight critical gaps in the knowledge of the safety of work zones. Future challenges are also discussed to address some missing information needs related to work zone safety analysis and modeling.
INTRODUCTION
The aging of highway infrastructure and the need for constant capacity expansion lead to an increasing number of work zones on roads. Many construction, maintenance and rehabilitation activities are occurring under demanding traffic levels on the roadways. As mentioned in a recent report (1), approximately twenty percent of the U.S. highways are under construction during the peak construction season, which involve over three thousand work zones. About twelve billion vehicle miles of travel a year will be through active work zones and travelers can expect to encounter an active work zone one out of every hundred miles driven on the highway system. Consequently, about 24 percent of the non-recurring delays and 10 percent of overall delays on freeways were attributed to work zones and 87,606 crashes occurred in work zones that led to over 37,000 injuries and 576 fatalities during 2010 (2). These statistics will be more alarming when more work zones present on freeways, state roads, and local roads.

Transportation agencies continuously strive to lower the operational and safety impacts of work zones on traveling public and workers through optimized planning, scheduling, and operating mechanisms. As such, knowledge of conditions/factors contributing to work zone crashes is of great importance. A thorough understanding of the potential risk factors associated with work zone crash occurrences and their consequence is needed to make informed decisions on deploying appropriate safety strategies and countermeasures. Despite the developing in the large body of research literature in recent years, work zone safety issues still need more investigations as many of the associated analysis approaches remain in their infancy.

The purpose of this paper is to provide a state-of-the-art review of current work zone safety-related analysis and modeling studies and issues. It aims to show how the analysis, methods and findings have evolved over the last five decades. To achieve the goal, the descriptive analyses performed by existing studies are examined and several key findings are discussed. Critical reviews of modeling studies associated with work zone crash occurrence and crash severity are also presented. Challenges associated with the work zone safety-related analysis and modeling are then discussed. The paper is concluded with recommendations of key areas which require further study and development.

METHODOLOGY
A systematic literature search of work zone safety studies is conducted first. The National Work Zone Crash Information Clearinghouse (http://www.workzonesafety.org) is served as the basis. In addition, the Transport Research International Documentation (TRID) database (http://trid.trb.org) is also used as a major source to track down related studies. In addition, other public digital libraries and search engines such as Google Scholars and CiteSeerX are also used to locate scientific and academic papers on the related topics. The work zone safety-related subject headings identify approximately 342 candidate papers. After excluding reviews, non-English papers and papers focused on work zone traffic control devices, 179 papers available from January 1962 to 2013 are initially considered. These include 47 technical reports of the transportation agencies as well research articles from other countries such as Singapore, Japan, and Canada. The left diagram of FIGURE 1 shows the number of studies that analyzed/modelled work zone crash data in each state (Alaska and Hawaii are not included). In addition, yearly distribution of these studies is illustrated in the right diagram of FIGURE 1. Each of these identified studies is carefully reviewed and the most relevant eighty one studies on work zone crash data analysis and modeling are specifically examined and discussed in present study.
DESCRIPTIVE ANALYSIS OF WORK ZONE SAFETY

Researchers made great efforts to examine the safety impact of work zones over the past five decades. The majority of the previous research focused on the descriptive analysis of crash data to explore the long-term work zone characteristics such as crash severity, crash rate, type, location, and other factors. In addition, crash characteristics occurred at before-, during-, and after-work zone periods were frequently compared to highlight both the similarity and differences of safety between work zone and non-work zone locations. The findings from these studies varied significantly because of various reasons such as the differences in locations, data size and quality and analysis approach. TABLE 1 provides an overview of these studies. Some of the important findings identified by previous studies are reviewed and discussed in following sub-sections.

### TABLE 1 Work Zone Safety Studies based on Descriptive Statistics

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Work Zones (WZ) Scope</th>
<th>WZ Sites</th>
<th>Crash Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3)</td>
<td>1978</td>
<td>construction zones on rural interstate systems in OH</td>
<td>21</td>
<td>3-year</td>
</tr>
<tr>
<td>(4)</td>
<td>1978</td>
<td>highway construction zones in seven states (no name information)</td>
<td>79</td>
<td>1 year +duration</td>
</tr>
<tr>
<td>(5)</td>
<td>1981</td>
<td>highway work zones in VA</td>
<td>statewide</td>
<td>1-year</td>
</tr>
<tr>
<td>(6)</td>
<td>1988</td>
<td>short- and long-term urban freeway work zones in IL</td>
<td>26</td>
<td>6-year</td>
</tr>
<tr>
<td>(7)</td>
<td>1989</td>
<td>construction-zone accidents on rural state highways in NM</td>
<td>3</td>
<td>3-year</td>
</tr>
<tr>
<td>(8)</td>
<td>1990</td>
<td>construction and maintenance work zones in KY</td>
<td>20</td>
<td>4-year</td>
</tr>
<tr>
<td>(9)</td>
<td>1990</td>
<td>construction and maintenance work zones in VA urban areas</td>
<td>7</td>
<td>4-year</td>
</tr>
<tr>
<td>(10)</td>
<td>1995</td>
<td>construction and maintenance work zones in OH</td>
<td>60</td>
<td>5-year</td>
</tr>
<tr>
<td>(11)</td>
<td>1998</td>
<td>construction work zones in NY</td>
<td>statewide</td>
<td>3-year</td>
</tr>
<tr>
<td>(12)</td>
<td>2000</td>
<td>work zones with fatal crashes in GA</td>
<td>statewide</td>
<td>3-year</td>
</tr>
<tr>
<td>(13)</td>
<td>2001</td>
<td>work zones in VA</td>
<td>statewide</td>
<td>4-year</td>
</tr>
<tr>
<td>(14, 15)</td>
<td>2002</td>
<td>work zones in AL, MI, TN</td>
<td>statewide</td>
<td>5,3,2-year</td>
</tr>
<tr>
<td>(16)</td>
<td>2004</td>
<td>work zones with fatal crashes in TX</td>
<td>77</td>
<td>15-month</td>
</tr>
<tr>
<td>(17)</td>
<td>2005</td>
<td>work zones in the Canada</td>
<td>countrywide</td>
<td>5-year</td>
</tr>
<tr>
<td>(18, 19)</td>
<td>2005</td>
<td>work zones on interstate highways in OH (fatal &amp; injury crashes)</td>
<td>statewide</td>
<td>3-year</td>
</tr>
<tr>
<td>(20)</td>
<td>2007</td>
<td>highway construction work zones with fatal crashes in IL</td>
<td>statewide</td>
<td>6-year</td>
</tr>
<tr>
<td>(21)</td>
<td>2008</td>
<td>NY work zones &amp; projects in other CA, NC, OH, WA</td>
<td>NY+64 projects</td>
<td>multi-year</td>
</tr>
<tr>
<td>(22)</td>
<td>2008</td>
<td>road construction projects on different types of highways in UT</td>
<td>202</td>
<td>4-year</td>
</tr>
<tr>
<td>(23)</td>
<td>2008</td>
<td>highway work zones with fatal and injury crashes in KS</td>
<td>statewide</td>
<td>13-year</td>
</tr>
<tr>
<td>(24, 25)</td>
<td>2009 &amp; 11</td>
<td>work zones in IA, KS, MO, NE, WI</td>
<td>statewide</td>
<td>5-year</td>
</tr>
<tr>
<td>(26)</td>
<td>2010</td>
<td>work zones on an expressway with lane closure in Japan</td>
<td>132</td>
<td>5-year</td>
</tr>
<tr>
<td>(27)</td>
<td>2011</td>
<td>highway work zones (vehicle intrusion crashes) in NY</td>
<td>statewide</td>
<td>6-year</td>
</tr>
<tr>
<td>(28)</td>
<td>2012</td>
<td>work zones in the MA</td>
<td>statewide</td>
<td>3-year</td>
</tr>
<tr>
<td>(29)</td>
<td>2012</td>
<td>highway work zones (truck-related crashes) in KS</td>
<td>statewide</td>
<td>9-year</td>
</tr>
<tr>
<td>(30)</td>
<td>2013</td>
<td>work zones in the IL</td>
<td>statewide</td>
<td>10-year</td>
</tr>
</tbody>
</table>
Crash Severity
Crash records provide information about the consequences of accidents—namely, whether the crash results in property damage, injury and fatality or not. This type of information is commonly used in the literature to identify crash severity. There was no consensus in the literature on whether the work zones are a reason for more severe crashes. Schrock et al. (16) analyzed fatal crashes in Texas, and found that 8 percent of the 77 fatal crashes were directly, and 39 percent were indirectly influenced by work zones. Some studies reported that work zone crashes were significantly more severe than non-work zone crashes (8, 31). In contrast, others also concluded that there is no significant difference between work zone crashes and regular crashes in terms of severity (7, 14); in other words, injury and fatal crashes in work zones do not differ significantly from injury and fatal crashes in non–work zones. In addition, there were some cases in which work zone crashes were found to be less severe than non-work zone crashes (5, 6, 9, 32).

As shown in FIGURE 2, the synthesis of existing studies (3, 5-10, 12, 14, 16, 31, 33-46) that directly/indirectly compared work zone crashes with non-work zone crashes imply that there is no clear evidence that work zone crash is more severe than non-work zone crashes, and vice versa. Given such inconsistent finding among the existing studies, the severity of work zone crashes deserves more investigation.

![FIGURE 2. Work zone crashes vs. non-work zone crashes in terms of crash severity and rate](image)

Crash Rate
Crash rate is frequently used as an indicator to examine the impact of work zone presence on safety. It is generally calculated as the total number of crashes in a given time period divided by the vehicle miles of travel through the work zone section. (38). Juergens (47) reported an increase of 7.0 to 21.4 percent in crash rate for ten work zone sites. Roupahil et al. (6) determined that the crash rate increased by an average of 88 percent in the presence of work zones in comparison to the before period and decreased by an average of 34 percent in the after period compared to the construction period. For short-term work zone sites, a constant accident rate of 0.80 crashes per mile-day of construction or maintenance was observed in the same study. Hall and Lorenz (7) found that the crash rate increased by 26 percent during the construction period. Garber and Woo (9) reported that the crash rates at work zones on multilane highways and two-lane urban highways in Virginia increased on average by 57 percent and 168 percent.
respectively. Contrary to aforementioned studies, the increase in crash rate is not found at all work zones. The study by Pigman and Agent (8) showed only 14 out of 19 work zone sites experienced increasing crash rates compared to the before period. Similarly, Graham et al. (4) reported 79 work zones experienced an overall increase in crash rate of 6.8 percent whereas 31 percent of the projects experienced crash rate decreases during construction. Jin et al. (22) observed lower crash rates during construction periods on urban non-interstate highways in Utah. Interestingly, they observed lower rates especially for severe crashes during construction periods at two case sites. Tsyganov et al. (48) also noted that some of the work zones have lower crash rates.

Despite some work zones experienced more crashes, as shown in FIGURE 2, the synthesis of existing studies (3, 4, 6-9, 12, 32, 37-39, 44, 45, 48-52) generally imply that the crash rate is higher during the presence of work zones, particularly in light of limited time periods and sample sizes for work zone crash data. However, there are considerable differences in crash rate changes between studies. Such variability in the findings suggests that the impact of work zones on crash rates is still not well understood as the changes are likely associated with factors such as work zone types and traffic conditions. One important question is the accuracy of exposure data (i.e., traffic volume and work zone length) used to calculate the crash rate. Current practices tend to either assume the same traffic volume or estimate AADT before and during work zone presence for calculating crash rate. Such as assumption and/or estimation is questionable as the presence of work zone can significantly change the traffic condition in the work zone area. In addition, the direct use of work zone duration and work zone length in the calculation equation is also not reliable. This is because existing work zone crash rate estimation seldom address the effects of the active work period and non-active work period as well as dynamic change of work zone. Without measuring these actual exposure data, the calculated changes in crash rates should be used with caution.

**Location of Crash Occurrence**

A work zone commonly consists of five different sections: (a) advance warning area, (b) transition area, (c) buffer area, (d) activity area and (e) termination area (see the left diagram in FIGURE 3). A number of studies specifically examined the crash distributions within these sections (3, 5, 8, 16, 19, 25, 31, 34, 40, 53, 54). The majority of these studies identified the activity (work) area as the most predominant crash location (5, 8, 16, 19, 25, 31). For example, Garber and Zhao (31) and Pigman and Agent (8) found that 70 percent and 80 percent of crashes respectively occurred in the activity area. However, results inconsistent with these are also reported. For example, Jin and Saito (51) concluded that the transition area upstream of the activity area is the most crash prone than the activity area itself. Nemeth and Migletz (3) found that more crashes (39.1%) occurred in buffer area whereas Srinivasan et al. (53) reported a greater percentage (34.65%) of crashes occurred in advance warning area. The right diagram in FIGURE 3 presents the synthesized results of the findings in literature (3, 5, 8, 16, 19, 25, 31, 34, 40, 53, 54). The synthesis indicates that on average about 55 percent of crashes are likely to occur in buffer and activity areas together.

However, whether or not activity area is indeed the most unsafe element of a work zones needs further careful interpretation. This is because existing studies hardly captured the effect of the great variability in lengths of each work zone component of different projects. For example, buffer zones are generally quite short, though close to the work environment, while advance warning areas are much longer, with potential crash risk increasing nearer to the transition and activity area. In addition, the length of each work zone component can vary in accordance with the progress of the entire project. Last but not least, almost all studies rely on crash reports and few of them clearly provide the accurate spatial information for the work zone crashes. Given the difficulties in providing accurate location information of work zone crashes, current synthesis in FIGURE 3 deserves further exploration when reliable data become available.
Other than investigating crash occurrence within work zones, a number of studies also examined crashes at different types of facilities and reported inconsistent results. For example, rural interstate and state highways are found to be more prone to work zone crashes by Pigman and Agent (14) and Chambless et al. (14). In contrast to this finding, Garber and Zhao (31) claimed that urban highways have a much higher percentage of work zone crashes than rural highways. Jin et al. (22) concluded that there is no clear relationship between highway class and work zone crash distributions. Therefore, more comparisons among work zones at different facilities would be beneficial to identify work zone hot spots.

### Time of Crash Occurrence

Time is considered as an important factor that affects crash occurrence in work zones. Understanding whether nighttime or daytime is more prone to have work zone crashes provides valuable information for transportation agencies to select appropriate work zone deployment strategies. The research by Arditi et al. (20) concluded that nighttime construction is nearly five times more hazardous than daytime construction in terms of the number of frequency of fatal accidents. In contrast, Dissanayake and Akepati (24) reported that most work zone crashes occur in clear daylight conditions and good weather conditions. Bai and Li (41) found that daytime off-peak period (10 am - 4 pm) is the most dangerous time interval for injury crashes in work zones. In addition, Ullman et al. (21) examined the safety of daytime and nighttime work zones and stated that there is no significant difference in terms of crash risk for an individual driver in daytime and nighttime construction zones. During temporal lane closure conditions total crashes increased by 66% for daytime and by 61% for nighttime.

### Crash Types

Crash types are identified in most of the state crash reports and there is strong agreement in the literature that rear-end crashes are the most predominant work zone crash type (6-9, 14, 31, 32). Hall and Lorenz (7) concluded that rear-end crashes proportion increases from 9 percent to 14 percent during construction. Rouphail et al. (6) stated that there is 50 percent increase in rear-end crashes during work zone period. Wang et al. (32) found that percentage of rear-end crashes within the work zone is significantly higher than rear-end percentage in non-work zone crashes. Daniel et al. (12) found that single-vehicle crashes, angle, and head-on crashes are the leading cause of fatal work zone crashes. Another important issue is truck involvement in work zone crashes. Researchers found that heavy truck-related crashes increased the probability of multiple vehicle involvement and fatalities in a work zone crash (8, 16). Garber and Zhao (31) investigated crash types by work zone locations. They concluded that most frequent crash types for advanced warning and transition area are rear-end and side-swipe crashes. However, the proportion of
these crashes decreases when going through activity area and termination area. In addition, angular and fixed object crashes proportion increased at activity area and termination area. Regarding the prevalence of rear-end crashes, special countermeasures such as advanced variable message signs to warn approaching vehicles should be investigated to prevent this type of crashes in work zones.

Other Factors
A number of studies also explored other potential contributing factors to work zone crash occurrence. Among these, environmental conditions, road geometry, vehicle characteristics and human factors are extensively analyzed (13, 14, 16, 20, 21, 24, 31). For example, Bryden et al. (11) found that one-third of all work zone accidents involve impacts with work zone traffic control devices and safety features introduced into the roadway environment by construction activity; 37 percent of those accidents cause serious injury. Daniel et al. (12) found that fatal work zone crashes are affected by variables such as the type of collision, light conditions, truck involvement, and roadway functional classification. Qi et al. (55) found that there is a correlation between weather condition and rear-end work zone crashes occurrence. Schrock et al. (16) found that the key variables contributing to fatal work zone crashes are roadway type, weather and lighting conditions, alcohol or drug use, and truck involvement. In 8 percent of the investigated crashes, the presence of work zone has a direct influence (such as improper traffic plan layout or a missing traffic control device), whereas in other 39 percent the work zone had an indirect influence (such as a vehicle struck a traffic control device). In addition, work zones have no influence on 45 percent of the crashes. The study by Dissanayake and Akepati (24) analyzed several crash characteristics, such as injury severities, weather conditions, vehicle characteristics, roadway characteristics, vehicle maneuvers, and alcohol involvement. They reported that most of the crashes in work zones occur because of the absence of traffic control devices; besides, inattentive driving, speeding and failing to yield right of way are top three driver-contributing factors to work zone crashes.

WORK ZONE CRASH FREQUENCY MODELING
Crash counts are non-negative integers that are deemed to be the outcomes of a set of contributing factors. A number of crash frequency (CF) models were developed for examining traffic safety issues. Lord and Mannering (56) provided a comprehensive review of different types of crash frequency models. Statistical models such as Poisson and negative binomial (NB) are extensively used to capture the relationship between crash counts and the selected contributing factors (57-61). In addition, more advanced statistical methods such as zero inflated NB and Poisson, truncated regression, generalized additive model, Conway Maxwell Poisson model, and negative multinomial model are also considered (56, 62).

Despite the advances in CF models for traffic safety analysis, only a limited number of studies specifically focus on modeling work zone crash frequency (36-38, 53, 55, 63-67). Among these studies, several statistical modeling techniques have been used to investigate the CF at work zones. Most of them developed negative regression (NB) models to predict the expected number of crashes at work zones (36, 38, 64-66). Venugopal and Tarko (64) proposed two models for modeling crashes at approaches to work zones and inside the work zones. Srinivasan et al. (65) developed NB models for total crashes, injury and fatal and property damage only (PDO) crashes to estimate the traffic safety impacts of roadway construction. Similarly, Ozturk et al. (66) developed NB models for construction work zone crashes in New Jersey. Several studies also compared the performance of different models. For example, Pal and Sinha (36) investigated safety impact of lane closure strategies on interstate work zones in Indiana. Normal regression, negative binomial and Poisson regression models were developed and compared. They found that the linear regression model outperformed the classical NB and Poisson models as the data set was too small to improve the predictive power of these exponential models. Qi et al. (55) constructed the truncated NB regression and truncated Poisson regression models to examine the rear-end crashes at work zones in New York. The truncated NB regression model was found to have better predictive power. Despite the zero-inflated Poisson (ZIP) and zero-inflated NB (ZINB) models
were developed to model rear-end accident frequency by county level, no comparison was made to discuss the performance of these zero-inflated models.

Other than these models focusing on crash counts, some studies also used other types of approaches to examine the derived safety measurements. For example, Elias and Herbsman (63) developed a crash rate probability function by using the Monte Carlo (MC) simulation approach to account for the intrinsic scarcity of work zone crash data. Srinivasan et al. (53) attempted to model the location of crashes within work zones as a function of length of work zone segments, traffic volume and weather. Multinomial logit (MNL) model was used to capture crash probability per lane-mile for each work zone element.

TABLE 2 summarizes the aforementioned studies that focused on work zone crash occurrence. Despite model differences, factors most commonly found to significantly affect work zone crash occurrence include the length of the work zone, duration, and traffic volume. The modeling results generally indicate that traffic volume, work zone duration and length are positively correlated with work zone crash occurrence.

<table>
<thead>
<tr>
<th>References</th>
<th>(36)</th>
<th>(64)</th>
<th>(63)</th>
<th>(68)</th>
<th>(38)</th>
<th>(55)</th>
<th>(53)</th>
<th>(65)</th>
<th>(66)</th>
<th>(67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>NB, NLR, P</td>
<td>NB</td>
<td>MC</td>
<td>NB</td>
<td>NB</td>
<td>TNB, ZIP, ZINB</td>
<td>MNL</td>
<td>NB</td>
<td>NB</td>
<td>NB, MENB</td>
</tr>
<tr>
<td>Duration</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Traffic volume</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Operating lanes</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>WZ speed limit</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Cost of project</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane closure</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Shoulder width</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed reduction</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Urban indicator</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Road system</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Weather</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Crash rate</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Control device</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of work</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Intersection</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Ramp</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Daytime-Nighttime</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

Note: P--Poisson regression model; NB -- Negative Binomial regression model; NLR -- Normal Linear Regression, TNB -- Truncated NB; ZIP -- Zero-Inflated Poisson; ZINB -- Zero-Inflated Negative Binomial; MC -- Monte Carlo Simulation; MNL --Multinomial Logit Regression; MENB -- Measurement Error NB

Enormous progress has been made in work zone crash frequency modeling on the whole. However, there are a variety of hurdles that remain. In general, work zone crash frequency modeling is less explored compared to prevalent analytical methods for non-work zone crash-frequency data (56). More well-established methodological alternatives such as random-effects models, random-parameters models, and hierarchical/multilevel models deserves exploration in work zone crash data analysis. One of the major challenges that limited the use of more advanced models is attributed to the deficiencies of data associated with work zones (13, 32, 36, 67, 69, 70). Accurate work zone crash data are scarce. Most studies rely on the crash information derived from police crash reports. However, these archived crash records are frequently subject to a number of issues such as under reporting, missing data, and lack of
details about the work zones \((32, 67)\). In addition, data associated with those contributing factors such as average daily traffic (ADT), work zone length, and duration are also subject to measurement errors. For example, using the estimated ADT instead of the actual volume during the presence of work zone in CF models may bias the results, given the fact that a great number of traffic may shift to other routes \((37, 38, 64)\). Sometimes it is also very difficult to accurately define work zone length such as bridge works, those involving detours and projects with multiple stages and segments \((32, 64, 67)\). Moreover, the progress reports of work zone project may not be readily available to obtain the actual duration of each work zone \((64)\). Almost all the current modeling practices rely on estimated ADT and/or AADT as an exposure in work zones. Similarly, work zone lengths and work zone durations were also directed used without any consideration about their variability within and across studies. This leads to the concern about the accuracy of the these exposure data. If these data with deficiencies were used to develop predictive CF models, the results can be biased. As an example, Yang et al. \((67)\) recently examined the effect of work zone length to demonstrate the importance of accounting for the measurement errors associated with work zone data. They found that the CF model performance is greatly improved when the intrinsic deficiencies of data are addressed. With the advance of data collection methods, accurate traffic volume information through different sensors/detectors should be considered. Instead of using the total duration and fixed work zone length, the actual working periods and working stages documented by project profiles would provide more accurate information regarding the exposure data. In addition, lessons learned in modeling non-work zone safety also help tackle some of the modeling issues associated the deficiencies of work zone related data.

WORK ZONE CRASH SEVERITY MODELING
Other than crash frequency models, limited number of studies have specifically modeled links between work zone crash attributes and the severity levels road users have sustained \((40, 43, 55, 71-79)\). These studies have (partially) investigated the effect of user attributes, road conditions, environmental conditions, vehicle characteristics, crash characteristics, and work zone configurations on work zone crash severity. TABLE 3 provides a summary of these modeling studies. Depending on the objective, the unit of analysis varies across studies and includes crash level of severity, vehicle level of severity, driver severity, and occupant severity. In most modeling efforts, severity is categorized as an ordinal dependent variable of multiple levels (i.e., no injury, injury, and fatal). Advanced statistical techniques have been employed to analyze the links between crash severity and related independent variables. As seen from TABLE 3, the popular methods used in crash severity analysis concentrate on logistic regression (LR) for fatality analysis and ordered probit (OP) modeling for the multiple-level injury spectrum.

Crash injury severity in work zones are explored by several categories of factors summarized in TABLE 3. Findings from literature synthesis are, to a large extent, consistent. Factors most commonly found to increase work zone crash severity include high speed limit in the work zone \((40, 43, 71, 73, 74, 77-79)\), driving at night time \((40, 55, 73, 74, 79)\) driving under influence of alcohol or drugs \((55)\), vehicle age \((78, 79)\), the number of vehicles and people involved in the crash \((40, 55, 71)\), and truck-involved crashes \((55, 73, 74, 79)\). On the other hand, the deployment of safety equipment such as airbags and seat belts appears to significantly reduce the level of injury severity in work zones \((43, 78, 79)\). In addition, work zones with flagger control reduce the level of injury severity \((55, 73)\). Interestingly, adverse weather is also found to reduce the level of injury severity \((43, 71, 79)\).

Despite of some consistent findings, studies also reported conflicting findings about factors such as light condition, user age, gender, and number of lanes. For instance, Li and Bai \((73, 74)\) found that poor light condition increases the level of injury severity, while others indicated that good light conditions may increase it \((43, 79)\). Li and Bai \((73, 74)\) concluded that male drivers are associated with increased crash severity, while Weng and Meng \((79)\) suggested the opposite for construction and utility work zones. Akepati and Dissanayake \((43)\) observed that young drivers are associated with higher-severity crashes, but Li and Bai \((73, 74)\) and Weng and Meng \((79)\) both reported the opposite. Elhamrawy \((77)\) and Weng and Meng \((79)\) both found that crash injury severity positively correlate with the number of lanes whereas others \((73, 78)\) found a negative correlation.
The review of past studies indicates that significant gaps remain in understanding the relationship between work zone crash injury severities and potential risk factors. For instance, little has been done to compare the severity differences between different victims (driver vs. occupant) involved in work zone crashes. Different characteristics between a driver and occupants in a vehicle determine the dissimilarity of consequences they may suffer in a crash. Therefore, understanding the differences in risk factors between driver and occupants is valuable for constructing effective safety strategies toward specific victims of crashes. In addition, most analyses are performed based on the severity description in police crash reports. Few studies explored the actual level of severity recorded in hospital patient reports. More comparisons between work zone and non-work zone conditions will be helpful to highlight the special contributing factors of severe crashes in work zones. Existing modeling practices on work zone crash severity should be compensated by prevalent analytical methods used in non-work zone crash severity modeling. For instance, a recent study (80) analyzed and assessed a number of methods on non-work zone crash severity modeling. It provide many insightful suggestions that can be used to advance current practices in modeling work zone crash severity. For instance, other than logistic regression and ordered probit models, many other models such as those unordered (i.e., multinomial models), random parameter as well as mixed models can be considered to model work zone crash severity.
### TABLE 3 Summary of Work Zone Crash Injury Severity Modeling Studies

<table>
<thead>
<tr>
<th>Category</th>
<th>Reference</th>
<th>Methodology</th>
<th>Unit of analysis</th>
<th>Timeline</th>
<th>Environment</th>
<th>Road</th>
<th>Road user</th>
<th>Vehicle</th>
<th>Work zone</th>
<th>Crash</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OP, OLS</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OP</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LR</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LR</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OL</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OP, OLS</td>
<td>C/V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LR</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>QRA</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OP</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LR, GA</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OP</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Timeline**
- Time of day
- Day of week

**Environment**
- Light condition
- Weather condition
- Surface condition

**Road**
- Road class
- Road alignment
- Median
- Median width
- Road surface type
- Number of lanes
- Lane width
- Posted speed limit
- Area information
- Road special feature
- ADT

**Road user**
- Driver age
- Driver gender
- Driver race
- Vision obstruction
- Occupant age
- Occupant gender
- DUI
- Seat position
- Out-of-state driver

**Vehicle**
- Vehicle type
- Vehicle age
- Traveling speed

**Work zone**
- Type of work zone
- Traffic control
- Workers present
- Work zone activity
- Work zone duration
- Type of work activity
- Work effect on road

**Crash**
- Location within WZ
- No. of vehicles involved
- No. of persons involved
- Alcohol consumption
- Truck involved in crash
- Crash type
- Contributing factors
- Pre-crash actions
- 1st/most harmful event
- Incident location
- Restraint use
- Airbag deployment

**Note:**
- C -- crash level;
- D -- driver level;
- O -- Occupant level;
- V -- vehicle level;
- LR -- logistic regression;
- OP -- ordered probit model;
- OL -- ordered logit model;
- OLS -- ordinal least squares model;
- QRA -- quantitative risk assessment;
- GA -- genetic algorithm
CHALLENGES FACED BY WORK ZONE SAFETY STUDIES

One of most difficult issues that impedes the investigation of work zone safety is data related (70). Most of studies used national or state level crash databases to extract work zone safety-related information. However, these data are subject to two major limitations: (a) correctness and (b) completeness. Although there are clear definitions of work zones, work zone crashes are usually classified based on the subjective judgment of police officers attending the incidents (10, 32). The accuracy of the records largely depends on their interpretation of each incident. Previous study (83) based on a limited investigation already showed that as many as 77 percent of crashed that occurred in work zones were not coded as work zones crashes. FIGURE 4 shows an example of all crashes occurred in a work zone in NJ. Notably, some crashes occur at a very similar time in the same work zone segment but they are classified into two different categories. In addition, many crashes which occurred at intersections were also treated as work zone crashes though they may be not related to work zones. Similar spatial misclassifications are expected to exist in many state crash reports (53). There is still lack of clear information about the classification standards used by police officers to determine whether or not a crash is a work zone crash. Even an explicit work zone crash label in the crash database may not ensure that all work zone crashes are identified. Any subjectivity in the determination of the precise location and category of a crash will affect the analysis results.

FIGURE 4. Example of all crashes occurred in a work zone

Regarding the completeness of information, more detailed description about work zone and associated crash is necessary to make better decisions. Some state departments of transportation such as Florida, Louisiana, Maryland developed work zone crash reporting forms to collect specific work zone crash information (21). However, most states' crash databases only document whether there is a work zone or not. However, detailed information about the work zone is hardly collected (84). This type of information can be collected if the crash report form is modified to reflect work zone-specific characteristics. Integration of other relevant data sources such as lane closure system data and police citation data (28, 85) may enhance the analysis of work zone safety. Typical information such as those shown in TABLE 4 is frequently advised to be included in crash reports to supplement existing data sets.
TABLE 4 Examples of Suggested Variables to be Included in a Crash Database

<table>
<thead>
<tr>
<th>Variable / Reference</th>
<th>(32)</th>
<th>(13)</th>
<th>(53)</th>
<th>(86)</th>
<th>(87)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration of work zone</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Law enforcement presence</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Location of crash</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Posted speed limit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Signs &amp; traffic control devices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Type of work zone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Near the work zone?</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worker involvement in a crash</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Worker presence activity</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Work equipment use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traffic info. (i.e., queue, volume)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

Once the available data sources are identified, all the necessary information for work zone crash analysis needs to be carefully processed. Each entry needs to be filled out carefully and officers should be instructed about the specific aspects of detailed information for work zone-related accidents to reduce the number of unclear records within the crash database (i.e., temporary traffic control zone, speed limit, and workers or equipment involved in crashes) (70, 88).

Another important question involves the modeling issues associated with work zone safety-related studies. As work zone crashes are the outcomes of a number of contributing factors, it is impossible to collect and incorporate all possible factors in the model due to data availability problems. For those usable variables, their measurement errors also have to be taken into account (67). In addition, when many factors are considered, selecting appropriate factors to avoid multicollinearity among variables is also a critical issue (70). The transferability of the developed models is also very important to support the development of safety performance functions (SPFs) for work zones.

SUMMARY AND CONCLUSION

Existing research on work zone safety-related analysis and modeling have been conducted since 1960s. A growing number of studies have been performed using aggregated crash data from a set of projects and/or statewide crash databases. A number of contributing factors associated with environment, road conditions, work zone features, and vehicles' and motorists' characteristics are also frequently explored. There are both conclusive and inconsistent findings about the effects of these factors. For example, work zone duration and length as well as other exogenous factors such as traffic exposure are generally found positively contributing to the crash occurrence whereas there is no agreement on the severity of crashes in work zones. In addition, enormous progress has been achieved in work zone safety modeling on the whole over the years. However, there are still a number of hurdles that remain. Most of the studies still rely on traditional models such as NB and logistic regression models to examine crash records reported by the police. The use of more advanced approaches are inherently restricted by the availability of accurate and complete data associated with work zones as well as crash records. Therefore, a detailed work zone data acquisition and analysis manual should be developed to guide "when, where, what, and how" to acquire high-quality work zone data/information. As shown in both modeling sections, previous studies attempted to incorporate as many parameters as possible in the models. However, the potential multicollinearity among variables are hardly discussed. In order to select appropriate factors, using other analytical approaches such clustering and feature selection would be beneficial. Besides, the model transferability has to be investigated over different work zones. Alternative analysis tools and/or surrogate safety measures should also be considered when investigating safety issues in work zones with relative short duration, because there might be few crash counts to perform before-duration comparison.
ACKNOWLEDGEMENTS

The contents of this paper only reflect views of the authors who are responsible for the facts and accuracy of the data and results presented herein. The contents of the paper do not necessarily reflect the official views or policies of sponsoring agencies. The author appreciate the New Jersey Department of Transportation for partially funding this study. The authors also thank for the insightful and constructive comments and suggestions by the anonymous reviewers to help enhance this paper.

REFERENCES


