EPiC Series in Built Environment

Volume 5, 2024, Pages 939–947

Proceedings of 60th Annual Associated Schools of Construction International Conference

# Key Factors and Representative Scenarios in Work Zone Safety: Indiana Crash Data Analysis

**Hongyue Wu, Yunfeng Chen, Jiansong Zhang, and James L. Jenkins**
Purdue University
West Lafayette, Indiana

Work zones are unique construction sites having multiple elements and complex scenarios. Existing studies have utilized crash data to explore the key factors contributing to work zone crashes. However, studies covering various elements comprehensively are missing, especially in Indiana. Also, there is a lack of studies exploring how to identify the representative scenarios from a large number of work zone crashes. Thus, this study aims to investigate the key factors contributing to work zone crashes covering multiple elements and propose a procedure to select representative scenarios based on crash data. Three Indiana crash data resources and the Natural Language Processing method were applied. There are several key findings. The key factors from multiple resources were summarized, including road conditions, work zone elements, and environment. Traffic activities were the top factors, followed by construction vehicles and weather conditions. Then, a seven-step procedure to extract the representative scenarios from crash records was proposed. Applying the procedure, two representative cases were identified from Indiana crash data. The study enhances the comprehensive understanding of work zone crashes based on large-scale data and helps improve safety for workers, drivers, and pedestrians.

**Key Words:** Crash data, Key factors, Representative scenarios, Safety, Work Zone

## Introduction

A work zone is an area in a road section with construction, maintenance, or utility work activities (Indiana Department of Transportation, 2016). It is different from a normal construction site due to the potential conflicts between construction activities and traffic activities, making work zone scenarios more complex and hazardous (Garber & Zhao, 2002; Zhang et al., 2018). The multiple elements, such as workers, trucks, channelizing devices, equipment, and private vehicles, also make work zones more complicated. Safety issues in work zones have broader impacts than normal construction sites because they influence not only workers and properties on construction sites but also drivers, pedestrians, and private properties. There are 28,636 fatalities in work zone crashes from 1982 to 2019 in the U.S. (NIOSH, 2022), which even increased by 1.4% from 2019 to 2020 (Federal Highway Administration, 2022). In Indiana, there were 6,357 construction-related crashes in 2022 (Indiana State Police, 2022). The complexity of various scenarios and elements makes it challenging to understand and address the work zone safety issues. Thus, it is necessary to explore the key factors covering different elements and representative scenarios within work zones to help improve safety.

Crash data has been identified as a critical resource for analyzing work zone safety issues (Imprialou & Quddus, 2019). Key factors contributing to work zone crashes were identified, such as traffic conditions, traffic control, time of the day, road conditions, and weather conditions (Clark & Fontaine, 2015; Mokhtarimousavi et al., 2021; Sze & Song, 2019; Zhang et al., 2018). However, limited studies covered various elements comprehensively, especially in Indiana. More importantly, there is a lack of studies exploring how to identify the representative scenarios from a large number of crash records to better understand work zone crashes. To address the above gaps, this study aims to investigate the key factors contributing to work zone crashes covering multiple elements and propose a procedure to select representative work zone scenarios based on Indiana crash data. The findings could contribute to a comprehensive understanding of work zone crashes based on large-scale data and help improve safety for workers, drivers, and pedestrians.

## Literature Review

Previous studies have utilized crash data to explore the key factors influencing work zone safety. In the U.S., Virginia crash reports were used to identify the activities impacting the work zone crashes, such as congestion, changing lanes, and flagging control (Clark & Fontaine, 2015). Michigan crash data revealed that environmental and occupant characteristics are contributory factors to work zone crashes, such as speed limit, weather, and number of lanes (Weng, Zhu, Yan, & Liu, 2016). Florida crash data was also utilized to explore factors contributing to work zone crashes, covering work-zone-specific and crash-specific characteristics (Mokhtarimousavi et al., 2021). In addition, the risk factors (e.g., time of the day, number of vehicles, and road condition) contributing to work zone crashes in New Zealand were examined using 453 crash data (Sze & Song, 2019). The main factors influencing work zone crashes were extracted from Egyptian crash data, including work zone, road, and vehicle information, visibility and traffic conditions, and weather conditions (Zhang et al., 2018). Although existing work has identified those key factors, there is a lack of studies covering various elements comprehensively, especially in Indiana, as well as studies exploring how to identify the representative scenarios from a large amount of crash data. This paper aims to address the two gaps by identifying key factors covering multiple elements and selecting representative scenarios based on crash data.

## Methodology

Figure 1 shows the research framework of this study, including two major parts: identifying key factors using Natural Language Processing (NLP) method and determining representative scenarios based on a seven-step procedure. Three major data sources were utilized, including the Indiana Crash Fact Book 2019 (Thelin, Rukes, & Palmer, 2020), the Indiana Automated Reporting Information Exchange System (ARIES) dataset (Indiana State Police, 2022), and the Indiana Safety Occurrence System dataset (Indiana Department of Transportation, 2023).

To identify the key factors of work zone crashes, key information from the three data sources was extracted. For the Indiana Crash Fact Book 2019, which summarizes the information on work zone crashes in Indiana, nine key parameters were used as key criteria. The values that caused the most crashes were selected. For the other two datasets covering thousands of crash records, NLP was applied to extract key information. First, narratives related to work zone crashes were extracted, resulting in 1835 narratives from the ARIES dataset and 857 narratives from the Safety Occurrence System. Second, the narratives were cleaned, including removing punctuations, numbers, special characters, and extra spaces, splitting the texts into tokens (e.g., words, phrases, and symbols), removing stop words (such as "the", "a", "an"), and stemming the tokens to reduce words to root forms (Kwayu et al., 2020). Third, three types of n-grams were generated from the narratives by the Python NLTK package (Kwayu et al., 2020; Zhang & El-Gohary, 2016), including unigrams (i.e., a single word), bigrams (i.e., two consecutive words, letters, or syllables), and trigrams (i.e., three consecutive words, letters, or syllables). Fourth, after counting the frequencies of those n-grams, developing the correlation matrixes and network graphs, as well as reviewing raw narratives related to those n-grams manually, key information was summarized. Finally, information from three data sources was combined to obtain the key factors of work zone crashes in Indiana.

To extract the representative work zone scenarios, 2634 crash records related to construction from the ARIES dataset 2019 – 2022 were further analyzed. The same nine parameters identified in the Indiana Crash Fact Book were used as major criteria to narrow down crash reports. Whether crashes caused injuries and fatalities or not was also a criterion. The identified crash records were located and measured in Google Maps to obtain road geometry information that cannot be determined from the ARIES dataset. Finally, representative work zone scenarios in Indiana were identified.

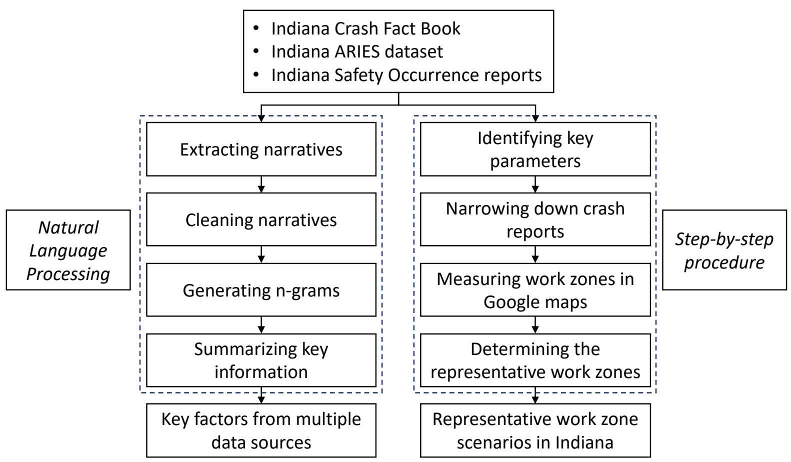


Figure 1. Research framework

## Results and Discussion

### *Key Factors Influencing Work Zone Safety*

Table 1 shows the comprehensive list of key factors contributing to work zone crashes in Indiana from three data sources. There were three major categories: (1) road conditions including the geometry information and work site settings, (2) work zone elements including traffic devices, vehicles, and workers and drivers as well as their behaviors, and (3) environment showing the weather, season, and time of the day information. Overall, the findings support that work zone safety is a complex issue compared with general construction sites because of the combination of various elements. In particular, it was illustrated that lane closure/changing/narrowing and traffic control/lane control were the top two factors that were supported by all three data sources. The result indicated that traffic activities play a critical role in work zone safety instead of construction activities, revealing the importance of traffic control and management in work zone safety (Clark & Fontaine, 2015; Li & Bai, 2009; Schrock et al., 2004; Weng et al., 2016). Trucks and other work zone vehicles, clear weather, and dry surface were also important factors mentioned by two data sources. This finding support the common features of work zones and general construction sites, that is, safety is impacted by weather conditions and construction vehicles (Mokhtarimousavi et al., 2021; Schuldt et al., 2021).

Table 1

*Key factors influencing work zone safety from Indiana crash data*

| Category | Key factors | Crash Fact Book | Safety occurrence reports | ARIES crash reports |
|---|---|---|---|---|
| Road conditions | Lane closure/changing/narrowing | √ | √ | √ |
| | Suburban area | √ | | |
| | Interstate road | √ | | |
| | Traffic control/lane control | √ | √ | √ |
| | Work on shoulder | | √ | |
| | Intersection with traffic lights | | | √ |
| | Turn directions | | √ | |
| Work zone elements | Employee/crew/workers | | √ | |
| | Truck and other work zone vehicles | | √ | √ |
| | Backing behavior | | √ | |
| | Entering the work zone | | √ | |
| | Concrete barrier wall/concrete barricade | | | √ |
| | Signs for road closure | | | √ |
| | Ground guide | | √ | |
| Environment | Daylight | √ | | |
| | Clear weather | √ | | √ |
| | Dry surface | √ | | √ |
| | Late fall and winter (October, November, and January) | √ | | |
| | 3:00 pm – 5:59 pm | √ | | |

To manage those key factors, several recommendations were provided. For road conditions, adding additional warning mechanisms for lane closures (Clark & Fontaine, 2015) and integrating sensing technology with traffic control plans (Fan et al., 2014) could be effective. For work zone elements, applying different colors of lights for Truck-Mounted Attenuators (Zhang et al., 2019) and utilizing an on-road or in-vehicle driver assistance system to warn the presence of workers (Sze & Song, 2019) would be helpful. For environment, emphasizing the illumination plan in dark work zones (Li & Bai, 2009) and adjusting work zone planning based on weather conditions and stages of projects (Ghasemzadeh & Ahmed, 2019) could support work zone safety.

## *Representative Scenarios from Crash Data*

Figure 2 shows the proposed seven-step procedure for determining representative work zone scenarios from crash data. Applying this step-by-step procedure, two representative cases were identified from more than 2600 Indiana crash records.

First, the crash records having injuries or fatalities were identified as more dangerous cases. “Number injured” and “Number dead” columns from the AREIS dataset were used. 454 out of 2634 crash records (17.24%) that had people injured or dead were identified, in which 439 crash records only had injuries, 8 crash records only had fatalities, and 7 crash records had both injuries and fatalities. Second, nine parameters from the ARIES dataset were used to further identify the dangerous work zones. They are also the same nine parameters in the Indiana Crash Fact Book. The top 3 values with more crash records for each of those parameters as well as some other critical values from references were selected. The results are shown in Table 2 (the selected values were *italicized*). For example, for Light Condition, Daylight, Dark (lighted), and Dark (not lighted) were the top three values with a greater number of crash records. Dawn/dusk was also selected because the Indiana Crash Fact Book showed that late afternoon (3:00 – 5:59 pm) was a more dangerous time slot, which is related to dusk.

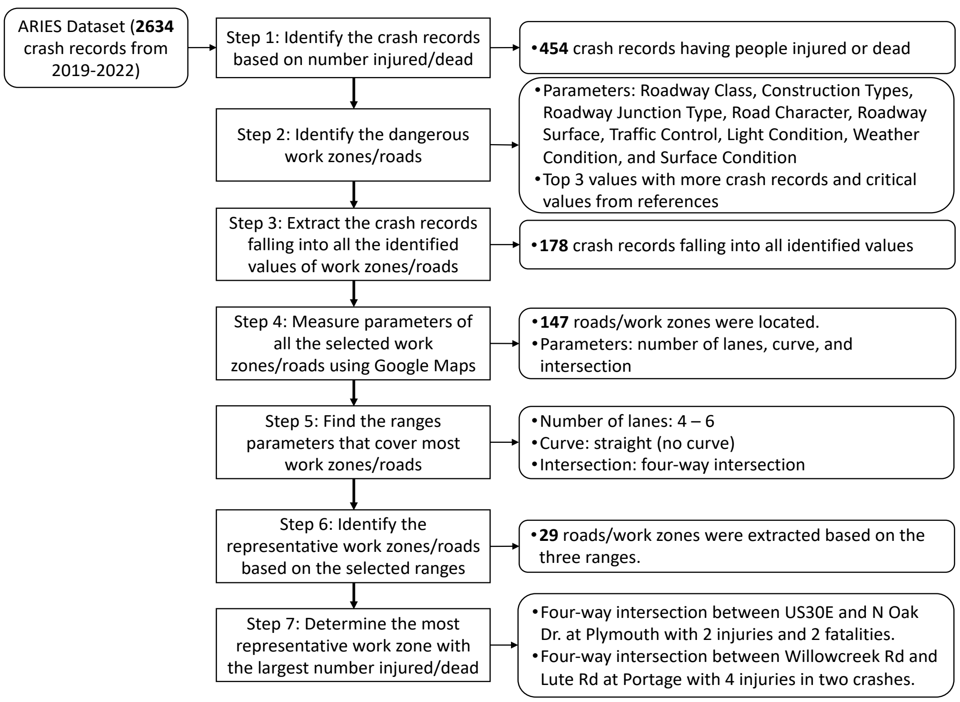

Figure 2. Process of selecting the representative work zone

Table 2

*Summary of identified values of nine parameters*

| Roadway Class | Number of crashes | Construction Type | Number of crashes | Roadway Surface | Number of crashes |
|---|---|---|---|---|---|
| County road | 15 | *Intermittent or moving work* | *73* | *Asphalt* | *378* |
| *Interstate* | *148* | *Lane closure* | *252* | *Concrete* | *73* |
| *Local/city road* | *133* | *Work on shoulder* | *88* | Gravel | 1 |
| *State road or US route* | *148* | X-over/lane shift | 41 | Other | 2 |
| Other or missing value | 10 | | | | |

| Roadway Junction Type | Number of crashes | Weather Condition | Number of crashes | Light Condition | Number of crashes |
|---|---|---|---|---|---|
| *Four-way intersection* | *69* | Blowing sand/soil/snow | 3 | *Dark (lighted)* | *44* |
| Interchange | 8 | *Clear* | *315* | *Dark (not lighted)* | *62* |
| *No junction involved* | *353* | *Cloudy* | *65* | *Dawn/dusk* | *17* (Thelin et al., 2020) |
| *Ramp* | *8* (Sun et al., 2013) | Fog/smoke/smog | 3 | *Daylight* | *330* |
| *T-intersection* | *13* | *Rain* | *54* | Unknown | 1 |
| Y-intersection | 3 | Sleet/hail/freezing rain | 1 | | |
| | | *Snow* | *13* (Thelin et al., 2020) | | |
| Road Character | Number of crashes | Traffic Control | Number of crashes | Surface Condition | Number of crashes |
| *Straight/grade* | *29* | *Lane control* | *81* | *Dry* | *357* |
| *Straight/hillcrest* | *10* | No passing zone | 2 | *Ice* | *7* |
| *Straight/level* | *186* | *None* | *82* | Loose material on road | 6 |
| *Curve/grade* | *7* | *Officer/crossing guard/flagman* | *11* (El-Rayes et al., 2014) | Muddy | 1 |
| *Curve/level* | *8* (Shen et al., 2021) | Other regulatory sign/marking | 11 | *Snow/slush* | *5* (Thelin et al., 2020) |
| Curve/hillcrest | 1 | Roundabout intersection | 1 | Water | 6 |
| Non-roadway crash | 1 | Stop sign and yield sign | 7 | *Wet* | *72* |
| Missing value | 212 | *Traffic control signal* | *43* | | |
| | | Other or missing value | 216 | | |

Third, the crash records falling into all identified values of nine parameters (*italic values* in Table 2) were extracted, resulting in 178 out of 454 crash records. Fourth, the 178 crash records were located in Google Maps to measure the actual road sections. Because 24 records had missing/inaccurate location information and five work zones had more than one crash record (i.e., three work zones had two crash records per work zone and two work zones had three crash records per work zone), 147 road sections from the total 178 records were finally identified. For the 147 road sections, three road-related parameters were measured, including the number of lanes, road character (whether it is a curve or straight and which type of curve it is), and intersection (such as four-way intersection, T-intersection, ramp, etc.). One thing to note is that even though the AREIS dataset already covered the curve and intersection information, there were some inconsistencies between the real road situations and the information from the AIRES dataset, which has also been mentioned by an expert from the Indiana Department of Transportation work zone safety division. Also, it was found that there were some curves or intersections near the exact crash locations, which may be related to the crash but were not considered in the ARIES dataset. Therefore, manually checking the road character and intersection is necessary. Fifth, after collecting the values of road-related parameters, the range of each parameter that covers most work zones was identified to further narrow down work zones. The 4 to 6 lanes (81 out of 147), straight roads without curves (101 out of 147), and four-way intersections (72 out of 147) were the most common values. Sixth, 29 representative work zones falling into the three ranges identified in Step 5 were extracted out of 147 work zones. Finally, to identify the most representative work zone from the 29 cases, the numbers of injuries and fatalities were considered. It was shown that 19 work zones only had one injury per work zone, five work zones had two injuries per work zone, three work zones had three injuries per work zone, one work zone had four injuries, and one work zone had two injuries and two fatalities. Thus, the two cases with four injuries and fatalities per work zone were recognized as the most representative work zones.

The US30 (near the four-way intersection with N Oak Dr.) at Plymouth had one crash record with two injuries and two fatalities. The Willowcreek Rd (near the four-way intersection with Lute Rd at Portage) had two crash records with four injuries in total. Between the two cases, the work zone with fatalities was recognized as a more dangerous one. US30 is a four-lane divided highway with asphalt surface and clear and dry situation. The crash happened on the morning of 8/24/2021, with a lane closure work zone. The narratives show that it is a three-vehicle collision, which is primarily due to the slowing and stopped traffic caused by construction and the high speed of the vehicle. In particular, one driver failed to perform proper stopping behavior when noticing the stopped traffic. It emphasizes the significance of alerting drivers to the potential slow or stopped traffic in work zones.

The proposed procedure can be utilized to better understand work zone crashes based on a large amount of crash data. The critical values of parameters associated with more injuries and fatalities and road situations in dangerous work zones could assist policymakers in identifying work zones with high risks and implementing proper countermeasures. The representative scenarios identified from crash data can be used to further explore work zone safety and test countermeasures through field tests or driving simulations. When applying the procedure, it is important to adjust the parameters and ranges based on the characteristics of the crash data to achieve more accurate results.

## Conclusion

Work zone crashes influence workers, drivers, pedestrians, as well as private and public properties, due to the conflicts between construction activities and traffic activities and multiple elements within work sites. To help address this issue, the authors investigated the key factors contributing to work zone crashes covering multiple elements using the NLP method and proposed a procedure to select the representative scenarios based on crash data. The findings revealed 19 key factors contributing to work zone crashes in Indiana, covering road conditions, work zone elements, and environment. Traffic activities (lane closure/changing/narrowing and traffic control/lane control) were the top factors, supporting the importance of managing traffic activities in work zones. Construction vehicles and weather conditions were also important, illustrating the common features between work zones and general construction sites. Moreover, the proposed seven-step procedure showed a systematic process to identify representative scenarios from thousands of crash records. Using this procedure, two representative work zones were determined based on the Indiana ARIES dataset. This study contributes to new insights into work zone safety by incorporating multiple elements and utilizing large-scale data. In practice, the findings could help improve work zone safety by enhancing understanding of complex scenarios through identifying specific factors and elements. However, there are two major limitations. First, even though this study extracted crashes related to work zones from the ARIES dataset using parameters, there are still many narratives that did not mention specific work zone elements. Future studies could explore methods to identify more relevant crash data in work zones. Second, the proposed seven-step procedure was only applied to Indiana crash data. Future work should explore its application in other crash data to further validate and modify the procedure.

## References

Clark, J. B., & Fontaine, M. D. (2015). Exploration of work zone crash causes and implications for safety performance measurement programs. *Transportation Research Record*, *2485*, 61–69. https://doi.org/10.3141/2485-08

El-Rayes, K., Liu, L., El-Gohary, N., & Abdelmohsen, A. (2014). *Effect of flaggers and spotter in directing work zone traffic for Illinois expressways and freeways*. Retrieved from https://www.ideals.illinois.edu/items/50011

Fan, W., Choe, S., & Leite, F. (2014). Prevention of Backover Fatalities in Highway Work Zones: A Synthesis of Current Practices and Recommendations. *International Journal of Transportation Science and Technology*, *3*(4), 311–337. https://doi.org/10.1260/2046-0430.3.4.311

Federal Highway Administration. (2022). FHWA Work Zone Facts and Statistics. Retrieved September 24, 2022, from https://ops.fhwa.dot.gov/wz/resources/facts_stats.htm

Garber, N. J., & Zhao, M. (2002). Distribution and Characteristics of Crashes at Different Work Zone Locations in Virginia. *Transportation Research Record*, *1794*(02–2132), 19–25.

Ghasemzadeh, A., & Ahmed, M. M. (2019). Exploring factors contributing to injury severity at work zones considering adverse weather conditions. *IATSS Research*, *43*(3), 131–138. https://doi.org/10.1016/j.iatssr.2018.11.002

Imprialou, M., & Quddus, M. (2019). Crash data quality for road safety research: Current state and future directions. *Accident Analysis and Prevention*, *130*, 84–90. https://doi.org/10.1016/j.aap.2017.02.022

Indiana Department of Transportation. (2023). Access Indiana - Portal. Retrieved from https://entapps.indot.in.gov/SOSWeb/Admin/Home

Indiana State Police. (2022). ARIES Crash Data. Retrieved from https://hub.mph.in.gov/dataset/aries-crash-data-2007-2017/resource/c3fbac12-7d6b-4ec1-a3c2-9026c0cdffd8?inner_span=True

Indiana Department of Transportation. (2016). *Indiana Manual on Uniform Traffic Control Devices (MUTCD)*. Retrieved from https://www.in.gov/dot/div/contracts/design/mutcd/mutcd.html

Kwayu, K. M., Kwigizile, V., Zhang, J., & Oh, J.-S. (2020). Semantic N-Gram Feature Analysis and Machine Learning–Based Classification of Drivers’ Hazardous Actions at Signal-Controlled Intersections. *Journal of Computing in Civil Engineering*, *34*(4), 04020015. https://doi.org/10.1061/(asce)cp.1943-5487.0000895

Li, Y., & Bai, Y. (2009). Highway Work Zone Risk Factors and Their Impact on Crash Severity. *Journal of Transportation Engineering*, *135*(10), 694–701. https://doi.org/10.1061/ASCETE.1943-5436.0000055

Mokhtarimousavi, S., Anderson, J. C., Hadi, M., & Azizinamini, A. (2021). A temporal investigation of crash severity factors in worker-involved work zone crashes: Random parameters and machine learning approaches. *Transportation Research Interdisciplinary Perspectives*, *10*. https://doi.org/10.1016/j.trip.2021.100378

National Institute for Occupational Safety and Health (NIOSH). (2022). Highway Work Zone Safety. Retrieved from https://www.cdc.gov/niosh/topics/highwayworkzones/default.html

Schrock, S. D., Ullman, G. L., Cothron, A. S., Kraus, E., & Voigt, A. P. (2004). *An Analysis of Fatal Work Zone Crashes in Texas*. Retrieved from http://www.ntis.gov

Schuldt, S. J., Nicholson, M. R., Adams, Y. A., & Delorit, J. D. (2021). Weather-Related Construction Delays in a Changing Climate: A Systematic State-of-the-Art Review. *Sustainability*, *13*(5), 2861. https://doi.org/10.3390/su13052861

Shen, J., Gao, W., Chen, Y., & Wu, H. (2021). *Cognitive Attention and Its Application in Countermeasures on a Curve Section*. Retrieved from https://g92018.eos-intl.net/eLibSQL14_G92018_Documents/19-12.pdf

Sun, C., Edara, P., & Zhu, Z. (2013). Evaluation of temporary ramp metering for work zones. *Transportation Research Record*, (2337), 17–24. https://doi.org/10.3141/2337-03

Sze, N. N., & Song, Z. (2019). Factors contributing to injury severity in work zone related crashes in New Zealand. *International Journal of Sustainable Transportation*, *13*(2), 148–154. https://doi.org/10.1080/15568318.2018.1452083

Thelin, R., Rukes, K., & Jamie Palmer. (2020). *Indiana Crash Facts 2019*. Retrieved from https://www.in.gov/cji/research/files/Indiana-Crash-Fact-Book-2019.pdf

Weng, J., Zhu, J. Z., Yan, X., & Liu, Z. (2016). Investigation of work zone crash casualty patterns using association rules. *Accident Analysis and Prevention*, *92*, 43–52. https://doi.org/10.1016/j.aap.2016.03.017

Zhang, J., & El-Gohary, N. M. (2016). Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking. *Journal of Computing in Civil Engineering*, *30*(2), 04015014. https://doi.org/10.1061/(asce)cp.1943-5487.0000346

Zhang, K., Hassan, M., Yahaya, M., & Yang, S. (2018). Analysis of Work-Zone Crashes Using the Ordered Probit Model with Factor Analysis in Egypt. *Journal of Advanced Transportation*, *2018*. https://doi.org/10.1155/2018/8570207

Zhang, S., Qing, Z., Brown, H., Sun, C., & Edara, P. (2019). Simulator and Field Study of Green Lights on Truck-Mounted Attenuators in Missouri during Mobile Operations. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(2), 769–778. https://doi.org/10.1177/0361198118823202