TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–15

DOI: 10.1177/03611981231152254
journals.sagepub.com/home/trr
SAGE

Diagnosis of Encroachment-Related Work-Zone Crashes by Applying Pattern Recognition

Subasish Das[1], Mahmood Tabesh[2], Bahar Dadashova[2], and Chiara Dobrovolny[2]

Abstract

Work-zone safety is one of the critical goals of transportation agencies. Vehicles are required to change travel paths and lanes over a short length of a road section at work zones. Distracted drivers, unable to see advanced warning signals and pavement markings delineating the work-zone travel paths, could increase the likelihood of a crash. Recent statistics showed that fatal collisions in work zones had increased by 46% in 2019 compared with 2011. The frequency of roadway departures at work zones, the higher risk of fatalities, and little insight into encroachment types at work zones underscored the need for a thorough study. This study aimed to examine the vehicle encroachment conditions associated with work-zone locations and focused on 4 years (2016 to 2019) of crash data from the Texas Department of Transportation by applying a unique data-mining method known as cluster correspondence analysis. This method identified four clusters in both "no-injury" and "fatal and injury" crash data. Major factors contributing to vehicle encroachment were identified. Three dominating clusters were median-related crashes on two-lane divided high-volume roadways; single-vehicle overturning collisions on two-way divided roadways with unprotected median; and overturning crashes on two-lane undivided roadways in controlled traffic. The findings of this study will be useful for safety engineers to contribute to reducing encroachment-related work-zone crashes.

Keywords

work zone, encroachment, road departure, pattern recognition, safety, cluster correspondence analysis

A work zone is an area on a roadway network where roadwork occurs and may involve lane closures and detours with additional signs and warning devices. Work-zone safety is a vital goal of transportation agencies. According to the Fatality Analysis Reporting System (FARS) (*1*), fatal collisions in work zones have grown over the last decade, from 521 in 2010 to 762 in 2019. In 2010, work zones accounted for 1.7% of all fatal crashes, whereas in 2019, they accounted for 2.3%. In 2019, there were over 25,000 work-zone collisions reported in the state of Texas, averaging one every 20 min. Researchers cite several aspects that contribute to the safety of vehicles passing through work zones (*2*). Work-zone equipment, traffic control devices, and workers are typically positioned near travel lanes, which increases the risk of vehicles colliding with equipment or striking roadway personnel.

Furthermore, work zones require vehicles to change their travel paths and lanes over a short section of road. If drivers are distracted and fail to see the advanced warning signals and pavement markings, this could increase the likelihood of a crash. Intrusion, or entering the work-zone area, and road departures are among the major crash types at work zones (*2*). During the last decade, more than 1,300 workers died in work-zone crashes (*3*). Therefore, both the general public and work-zone personnel are in danger of work-zone crashes.

[1]Ingram School of Engineering, Texas State University, San Marcos, TX
[2]Texas A&M Transportation Institute, Texas A&M University System, College Station, TX

Corresponding Author:
Subasish Das, subasish@txstate.edu

Although several studies have attempted to characterize the factors associated with rear-end collisions at work zones (*4–6*), work-zone encroachment crashes have not been fully investigated to date. Roadway encroachment is defined as the passage of a vehicle on roadway areas outside the limits of the designated lane(s) of travel. Encroachment in a work zone indicates this condition when such a vehicle inadvertently traverses the boundaries of a work zone. A few studies have investigated crashes that involve the encroachment of vehicles into work-zone areas (*7–9*). The potential path a vehicle might take during an encroachment is affected by numerous factors, such as speed limit, various roadway and roadside characteristics (e.g., horizontal and vertical alignment, lane/shoulder width), weather and lighting conditions, and driver behavior. The frequency of roadway departures at work zones, the higher risk of fatalities, and the very little information that is available about encroachment in work zones emphasized the need for a comprehensive study in this area.

The objective of this research was to investigate vehicle encroachment conditions associated with work-zone locations. The Texas Department of Transportation (TxDOT) crash database provides information on crash, vehicle, driver, and road characteristics. This database was used to analyze the encroachment events at work-zone areas. The researchers applied statistical cluster analysis techniques to identify relevant properties in crash and safety problems (*10–13*). Cluster correspondence analysis (CA) (*14*) was used in this study to generate a meaningful grouping of crash observations comparable to a set of crash, driver, and roadway characteristics. With that, major factors contributing to vehicle encroachment were identified.

The paper is organized as follows. The next sections provide a literature review, followed by an overview of the data preparation, determination of encroachment, exploratory data analysis, and methodology used in this study. The results of the cluster CA are subsequently explained, and the last portion of the paper presents the overall results, discussion, and conclusions.

Literature Review

Roadways play a vital role in the transportation infrastructure in the United States. However, these critical lifelines have deteriorated, and over 44% of the U.S. system is currently in poor or mediocre condition (*15*). The need for repair, resurfacing, or replacement of sections of the roadway or roadway structures is increasing. Usually, maintenance and rehabilitation projects are performed on an existing roadway with traffic in proximity, since in many cases complete closure of the roadway is not feasible. Work zones have evolved through time from simple layouts that may or may not have helped promote safety to modern designs that strive to maximize safety whenever possible. However, notwithstanding the best efforts to improve work-zone safety, collisions continue to occur (*16*). Roadwork activities increase both the rates and severity of the crashes (*17*, *18*).

NCHRP Report 869 (NCHRP Project 17-61) studied factors affecting the safety of traffic approaching and passing through work zones (*2*). Based on the analyses undertaken, work-zone characteristics such as signing and warning layouts, ingress and egress of workspace access, visibility issues, barricades and channelizing devices, and access-point designs have been found to affect work-zone crashes, necessitating new studies on work-zone safety. By comparing crash rates on the approach to and within work zones, Garber and Woo discovered a 57% rise on multilane highways and a 168% increase on two-lane, urban highway crashes (*19*). Among the most significant work-zone crash-contributing factors, distracted driving, failing to yield, and traveling at unsafe speeds had an impact on work-zone crash severities. From an analysis of work-zone configurations, major contributing factors that increase crash severity include roadway facility type, access control type, number of lanes, surface conditions, posted speed limit, work-zone component area, the presence of workers, the time of day, truck involvement, and the number of vehicles involved (*20*). For a driver involved in a crash following an maneuver such as speeding or following too closely, the chances of being injured increase by 10% (*21*). According to data from the 2014 FARS (*22*), 71.4% of fatal work-zone crashes are caused by speeding, whereas speeding only causes about 30% of all fatal crashes. This indicates that driver behavior is a significant contributor to the safety of work zones, when compared with that at non-work-zones.

Work-zones can be divided into specific areas, including the transition area, activity area, termination zone, and advanced warning. Garber and Zhao found that rear-end collisions were the most common form of crash across all work zones, and that they were considerably more common in advance warning zones (*23*). The most reported locations for crashes in that study were the work-zone activity areas, followed by the transition-, advance warning-, and buffer area. However, an analysis of crash reports on a sample of work zones in Illinois indicated that approximately 40% of crashes occurred outside the working area, in the taper and approach areas. The most severe issues were incorrect merging behavior on the approach and drivers approaching bottlenecks at unsafe speeds (*24*). Other studies report the activity area as the "hottest" section for work-zone crash rate and severity (*25*, *26*). According to multiple studies, rear-end collisions are the most common form of crash in work zones (*6*, *25*, *27*, *28*). Work-zone encroachment and single-vehicle crashes (SVCs) such as roadway departure, in which the driver inadvertently traverses the boundaries of a work zone, are other significant crash types in work zones (*29*).

Based on FHWA statistics, roadway departure crashes resulted in an average of 18,275 fatalities from 2013 to 2015, accounting for 54% of all traffic fatalities in the United States during that time (*30*). The driver, the vehicle, the highway, and the environment all play a role in causing roadway departure crashes. This type of crash has mainly been studied by researchers at non-work-zone sections. Alshatti studied encroachment crash characteristics using second Strategic Highway Research Program (SHRP2) data (*31*). The results indicate that drivers' pre-incident maneuvers, judgment maneuvers, road alignment, driver distraction, driver education, and average driven mileage were significant factors that affected roadway departure crashes. McLaughlin et al. identified factors related to road departure using naturalistic driving study data (*32*). Low visibility, low friction, changes in roadway boundaries, and driver distraction were among the contributing factors to roadway departure.

Existing work zones are among the factors associated with roadway departure and encroachment crashes. Work-zone encroachments are not just a problem for driver safety, they are also a significant concern for construction workers. A few studies have attempted to investigate and characterize work-zone encroachments that have resulted in crashes. Bryden et al. characterized work-zone intrusion crashes and compared them with other traffic and construction crashes (*7*). Speeding, vehicle failure, roadway conditions, driver inattention, and -impairment were among the greatest contributing factors. Ullman et al. investigated the contributing factors of work-zone encroachment crashes during daytimes and nighttimes (*8*). The results indicated that working at night did not increase the probability of work-zone encroachment crashes. Furthermore, crashes that occurred at night were not always more severe than those that occurred during the day in the same operating work-zone area.

In another study by Ullman et al. on the New York State Department of Transportation work-zone crash database, four work-zone types, including lane closures, flagging operations, mobile operations, and traffic control setup and removal activities, were investigated (*9*). Depending on the type of work zone, 25% to 40% of the encroachments were the result of a deliberate action by the driver. For nondeliberate cases, driver inattention and abrupt changes in traffic flow were among the main influencing factors that yielded work-zone encroachment incidents. Currently, there are minimal studies addressing encroachment conditions in work zones because existing encroachment datasets are not representative of a wide range of roadway characteristics. This lack of studies on the topic of work-zone encroachment highlighted the need to develop a work-zone encroachment crash database and a comprehensive study characterizing the factors that influence this type of crash.

Methodology

Data Preparation

TxDOT's crash database was used as the main source of data for this study. The collated 4-year (2016 to 2019) TxDOT data include information on crash specifications, persons, and units. The crash specification file contains general information on crash type, harmful event and location, crash severity, contributing factors, as well as roadway and roadside characteristics. However, the sequence of events leading to the crash is not available. The unit file contains information on vehicle type, model, year, damaged areas, and crash severity. The person file provides information on driver and passenger age and gender, driver impairment, and driver alcohol- and drug test-results. The data preparation and filtering steps shown in Figure 1 were applied to TxDOT's crash database. The data preparation steps follow:

- The crash, unit, and person files were merged through a unique crash identification number to create one single file. For the vehicle and driver data, the vehicle contributing to the first harmful event was chosen, whereas from persons involved in the crash, only the driver information was selected.
- Work zone and work-zone-related crashes were selected from the dataset. Work-zone specifications, however, were not available. Owing to the complexity of the contributing factors to intersection- and junction-related crashes, these types of crashes were removed from the dataset. With that, the analysis focused on encroachment and roadway departure crashes regardless of the intersection- or junction location's specifications.
- Encroachment is considered to be one of the significant crash types in SVCs. Thus, to determine encroachment crashes, SVCs were first filtered from data on collision type, then harmful events were investigated. Events that involved encroachment were then selected, such as run-off road, collision with a fixed object on the roadside, -with maintenance equipment, -with a barrier/guardrail.
- Among multiple vehicle crashes (MVCs), head-on, opposite sideswipe, and overturn crashes result from or can result in a vehicle encroachment.

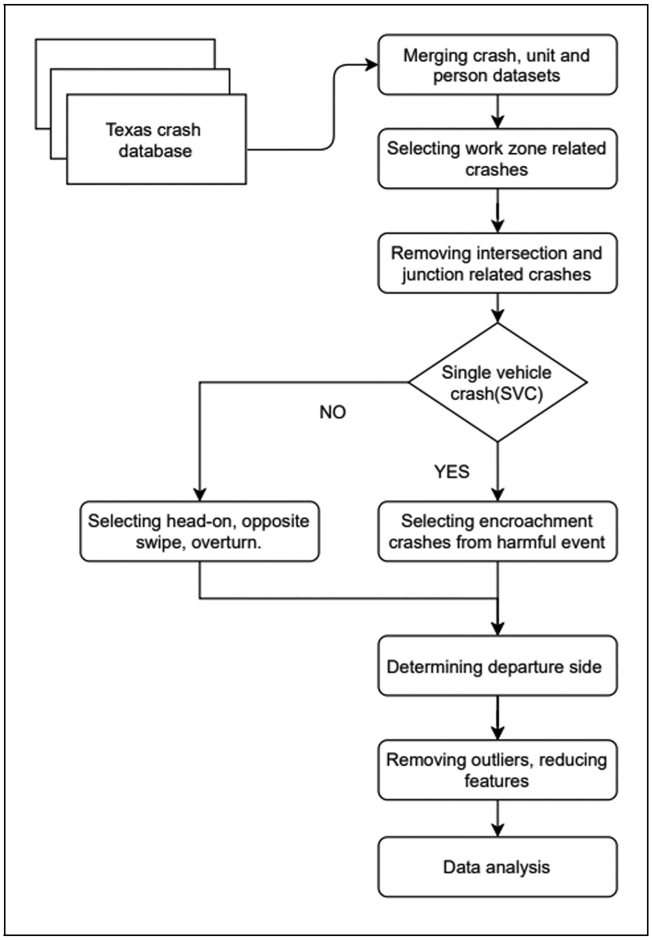

Figure 1. Data preparation and encroachment identification process.

Thus, only these types of MVC were selected for the analysis.

- Left- or right-side encroachment are vital features of such an analysis. The direction of the encroachment was determined from the available data on the location of the harmful event, type of crash, and type of fixed object included in the crash events.
- After determining encroachment crashes in the dataset, each variable was investigated, and events with outlier values were removed. Also, unrelated and redundant features were removed from the dataset.

Following the filtering and data preparation process, a total of 10,677 encroachment cases were identified in the data. The resulting work-zone crash database had 50 attributes (i.e., potential crash-contributing factors). In the following section, an exploratory data analysis of the prepared dataset is presented.

Exploratory Data Analysis

Table 1 describes the data on encroachment-related work-zone crashes, including no-injury and fatal and injury categories. From 2016 to 2019, there were a total of 10,677 encroachment-related work-zone crashes in Texas. Out of these, 3,678 resulted in a fatality or injury (34%), whereas 6,998 resulted in no injuries (i.e., property damage only). The rows provide the odds ratio (OR) of each category of attributes included in the dataset. The measure "*p*.ratio" provides the *p*-value, which indicates the statistical significance of the attribute-level OR measures (for example, "curve" is an attribute of "alignment"). The measure "*p*-overall" provides the *p*-values of the variables.

With respect to left-side encroachments, the OR was greater than 1 for right-side encroachment and unknown encroachments in these crashes. The unknown encroachments were also marginally significant ($p < 0.05$). For the first harmful event, the OR for off-roadway was greater than 1 when compared with the median. However, it was not statistically significant. The OR was less than 1 for work zones with no control (TRAF_CTRL), indicating that the odds of being involved in a crash decreased at work zones with no control, which might seem counterintuitive. Under "trafficway," the ORs for two-way (divided), two-way (undivided), two-way (undivided) with continuous left-turn lane (CLTL), and unknown were higher than 1 when compared with "other." Of these, only two-way undivided was statistically significant. This finding is in line with several other studies. Under "lighting conditions," the ORs for dark not-lighted and daylight were higher than 1 when compared with dark-lighted. Both these attributes were statistically significant. Under "weather condition," the OR for "other" was greater than 1 when compared with clear, but was not statistically significant. Under "Driver 1 condition," the OR for "not normal" was greater than 1 when compared with "apparently normal." This was statistically significant, and this finding was intuitive. Under "Driver 1 distraction," the OR for "not distracted" was greater than 1 when compared with "distracted" and was statistically significant. Under "Vehicle 1 action" the OR for violation was equal to 1 when compared with improper driving, but it was not statistically significant. Under "Vehicle 1 important point," the ORs for variables front, left, other, overturn, and right were greater than 1 when compared with "back." These attributes were all statistically significant.

Cluster Correspondence Analysis

Cluster CA is one of the approaches that has been developed for categorical data analysis (*24*). The goal of

Table 1. Percentage Distribution of Key Attributes by Crash Severity

Variable category	No injury (*N* = 6,998)	Fatal and injury (*N* = 3,678)	Odds ratio (OR)	*P*.ratio	*P*.overall
FIRST_DEP_DIR (direction of departure)	na	na	na	na	0.072
Left	3,154 (45.1%)	1,583 (43.0%)	Ref.	Ref.	na
Right	1,644 (23.5%)	865 (23.5%)	1.05 [0.95;1.16]	0.365	na
Unknown	2,200 (31.4%)	1,230 (33.4%)	1.11 [1.02;1.22]	0.022	na
RDWY_ALIGN (alignment)	na	na	na	na	0.066
Curve	1,326 (18.9%)	766 (20.8%)	Ref.	Ref.	na
Other	46 (0.66%)	25 (0.68%)	0.94 [0.57;1.54]	0.818	na
Straight	5,626 (80.4%)	2,887 (78.5%)	0.89 [0.80;0.98]	0.020	na
FRST_HARM_LOC (location of first harmful event)	na	na	na	na	0.112
Median	1,904 (27.2%)	1,011 (27.5%)	Ref.	Ref.	na
Off roadway	4,227 (60.4%)	2,270 (61.7%)	1.01 [0.92;1.11]	0.810	na
On roadway	832 (11.9%)	382 (10.4%)	0.86 [0.75;1.00]	0.046	na
Other	35 (0.50%)	15 (0.41%)	0.81 [0.43;1.47]	0.500	na
RDWY_GRADE (grade)	na	na	na	na	0.630
Grade	1,312 (18.7%)	725 (19.7%)	Ref.	Ref.	na
Hillcrest/uphill/downhill	358 (5.12%)	194 (5.27%)	0.98 [0.80;1.19]	0.848	na
Level	5,282 (75.5%)	2,734 (74.3%)	0.94 [0.85;1.04]	0.208	na
Other	46 (0.66%)	25 (0.68%)	0.99 [0.59;1.61]	0.957	na
TRAF_CTRL (traffic control)	na	na	na	na	0.017
Controlled	635 (9.07%)	324 (8.81%)	Ref.	Ref.	na
No controls	1,320 (18.9%)	615 (16.7%)	0.91 [0.77;1.08]	0.279	na
Others	5,043 (72.1%)	2,739 (74.5%)	1.06 [0.92;1.23]	0.388	na
SUR_TYPE (surface type)	na	na	na	na	<0.001
Asphalt	2,253 (32.2%)	1,257 (34.2%)	Ref.	Ref.	na
Concrete	3,039 (43.4%)	1,654 (45.0%)	0.98 [0.89;1.07]	0.595	na
Unknown	1,706 (24.4%)	767 (20.9%)	0.81 [0.72;0.90]	<0.001	na
TRAF_WAY (trafficway)	na	na	na	na	<0.001
Other	11 (0.16%)	2 (0.05%)	Ref.	Ref.	na
Two-way, divided, unprotected median	4,319 (61.7%)	2,299 (62.5%)	2.76 [0.73;19.5]	0.148	na
Two-way, undivided	704 (10.1%)	500 (13.6%)	3.68 [0.97;26.0]	0.057	na
Two-way, undivided, with TLWTL	258 (3.69%)	110 (2.99%)	2.21 [0.57;15.8]	0.277	na
Unknown	1,706 (24.4%)	767 (20.9%)	2.33 [0.61;16.5]	0.237	na
SEC_T_ADTC (AADT)	na	na	na	na	<0.001
1,001–4,000 vpd	237 (4.48%)	189 (6.49%)	Ref.	Ref.	na
4,001–40,000 vpd	64 (1.21%)	50 (1.72%)	0.98 [0.64;1.49]	0.925	na
401–1,000 vpd	3,476 (65.7%)	1,807 (62.1%)	0.65 [0.53;0.80]	<0.001	na
GT 40,000 vpd	26 (0.49%)	19 (0.65%)	0.92 [0.49;1.71]	0.789	na
LT 400 vpd	64 (1.21%)	50 (1.72%)	0.98 [0.64;1.49]	0.925	na
MED_WIDTHC (median width)	na	na	na	na	0.392
11–20 ft	338 (7.82%)	194 (8.43%)	Ref.	Ref.	na
GT 20 ft	2,809 (65.0%)	1,450 (63.0%)	0.90 [0.75;1.09]	0.268	na
LT 10 ft	1,172 (27.1%)	655 (28.5%)	0.97 [0.80;1.19]	0.793	na
No median	4 (0.09%)	1 (0.04%)	0.48 [0.02;3.50]	0.506	na
RD_SRFC_COND (road surface condition)	na	na	na	na	<0.001
Dry	5,011 (71.6%)	2,828 (76.9%)	Ref.	Ref.	na
Not dry	1,987 (28.4%)	850 (23.1%)	0.76 [0.69;0.83]	<0.001	na
LIGHT_COND (lighting condition)	na	na	na	na	<0.001
Dark-lighted	1,743 (24.9%)	814 (22.1%)	Ref.	Ref.	na
Dark not-lighted	1,724 (24.6%)	900 (24.5%)	1.12 [1.00;1.26]	0.060	na

(continued)

Table 1. (continued)

Variable category	No injury (N = 6,998)	Fatal and injury (N = 3,678)	Odds ratio (OR)	P.ratio	P.overall
Daylight	3,233 (46.2%)	1,842 (50.1%)	1.22 [1.10;1.35]	<0.001	na
Other	298 (4.26%)	122 (3.32%)	0.88 [0.70;1.10]	0.254	na
EVNT_WTHR_COND (weather condition)	na	na	na	na	<0.001
Clear	4,418 (63.1%)	2,426 (66.0%)	Ref.	Ref.	na
Other	1,194 (17.1%)	698 (19.0%)	1.06 [0.96;1.18]	0.246	na
Rain/snow	1,345 (19.2%)	550 (15.0%)	0.74 [0.67;0.83]	<0.001	na
Unknown	41 (0.59%)	4 (0.11%)	0.18 [0.05;0.46]	<0.001	na
D1_DR_COND (Driver 1 condition)	na	na	na	na	<0.001
Apparently normal	5,832 (83.3%)	2,972 (80.8%)	Ref.	Ref.	na
Not normal	788 (11.3%)	562 (15.3%)	1.40 [1.24;1.57]	<0.001	na
Unknown	378 (5.40%)	144 (3.92%)	0.75 [0.61;0.91]	0.003	na
D1_SEX (Driver 1 gender)	na	na	na	na	<0.001
Female	2,019 (28.9%)	1,420 (38.6%)	Ref.	Ref.	na
Male	4,448 (63.6%)	2,246 (61.1%)	0.72 [0.66;0.78]	<0.001	na
Unknown	531 (7.59%)	12 (0.33%)	0.03 [0.02;0.06]	0.000	na
D1_DR_DSTR (Driver 1 distraction)	na	na	na	na	<0.001
Distracted	3,822 (54.6%)	1,846 (50.2%)	Ref.	Ref.	na
Not distracted	1,804 (25.8%)	958 (26.0%)	1.10 [1.00;1.21]	0.053	na
Unknown	967 (13.8%)	393 (10.7%)	0.84 [0.74;0.96]	0.009	na
V1_FRST_DR_ACTN (Vehicle 1 action)	na	na	na	na	0.001
Improper driving	1,148 (16.4%)	642 (17.5%)	Ref.	Ref.	na
No contributing action	2,963 (42.3%)	1,521 (41.4%)	0.92 [0.82;1.03]	0.144	na
Other	201 (2.87%)	77 (2.09%)	0.69 [0.52;0.90]	0.007	na
Unknown	364 (5.20%)	139 (3.78%)	0.68 [0.55;0.85]	0.001	na
Violation	2,322 (33.2%)	1,299 (35.3%)	1.00 [0.89;1.13]	0.996	na
V1_POINT_IMP (Vehicle 1 important point)	na	na	na	na	<0.001
Back	298 (4.26%)	75 (2.04%)	Ref.	Ref.	na
Front	4,042 (57.8%)	2,083 (56.6%)	2.04 [1.59;2.67]	<0.001	na
Left	807 (11.5%)	355 (9.65%)	1.75 [1.32;2.33]	<0.001	na
Other	154 (2.20%)	63 (1.71%)	1.62 [1.10;2.39]	0.015	na
Overturn	482 (6.89%)	747 (20.3%)	6.14 [4.67;8.16]	0.000	na
Right	924 (13.2%)	322 (8.75%)	1.38 [1.05;1.85]	0.022	na
Unknown	291 (4.16%)	33 (0.90%)	0.45 [0.29;0.70]	<0.001	na

Note: AADT = annual average daily traffic; TLWTL = Two way left turn lane; GT = greater than; LT = less than; vpd = vehicles per day; na = not applicable.

cluster analysis is to establish a meaningful grouping of data based on a set of observable variables. Cluster CA aims to establish cluster allocation and scaling values for categorical variables in a way that maximizes intergroup variance. In the following, the mathematical algorithm applied in the cluster CA in this study will be explained.

In the first step, a normal data matrix, $\boldsymbol{X}$, with n observations containing q number of categorical variables needs to be converted to a new matrix, $\boldsymbol{Z}$, referred to as the super indicator matrix. This matrix will be generated through a one-hot encoding process, which transforms each categorical variable into a binary matrix, that is, $\boldsymbol{Z} = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, ..., \boldsymbol{Z}_q]$, where $\boldsymbol{Z}_j$ is an $n \times p_j$ matrix of the encoded j- categorical variable with $\boldsymbol{p}_j$ number of categories. The indicator matrix, $\boldsymbol{Z}$, has the same n number of rows and $Q = \sum_{j=1}^{q} Pj$ number of columns. One can define $\boldsymbol{Z}_k$ as an $n \times K$ binary matrix, indicating the memberships of each observation into K number of clusters. To take into account the clusters' relationship with the categorical variables, cross-tabulation of the indicator matrix and membership matrix are constructed as a $K \times p_j$ matrix, that is, $\boldsymbol{F} = \boldsymbol{Z}_K^{\mathrm{T}}\boldsymbol{Z}$. By applying cluster CA to contingency matrix $\boldsymbol{F}$, the scaling values corresponding to the clusters can be categorized in such as way that it will maximize the intergroup variances (*25*). In the optimal condition, the clusters maximize variances from

Table 2. Centroids and Size of the Clusters

Cluster	No-injury crashes				Fatal and injury crashes			
	Dim 1	Dim 2	Within-cluster sum of squares	Size	Dim 1	Dim 2	Within cluster sum of squares	Size
Cluster 1	−0.0080	−0.0025	0.0198	3,114	−0.0117	−0.0013	0.0197	1,548
Cluster 2	0.0138	−0.0077	0.0156	1,706	−0.0010	−0.0060	0.0193	869
Cluster 3	−0.0031	0.0034	0.0176	1,464	0.0102	0.0200	0.0152	767
Cluster 4	0.0084	0.0225	0.0127	714	0.0226	−0.0166	0.0137	494

Note: Dim = dimension or axis.

distributions over categorical variables and simultaneously obtain the distributions of categories in each variable.

In brief, the cluster CA process involves

- Randomly assigning observations to clusters and creating an initial membership matrix, $\boldsymbol{Z_k}$, and contingency matrix, $\boldsymbol{F}$;
- Applying the corresponding analysis to $\boldsymbol{F}$ and finding the category quantifications matrix, $\boldsymbol{B}$;
- Calculating object coordinate matrix $\boldsymbol{Y} = \frac{1}{q}\left(\boldsymbol{I_n} - \frac{1_n 1_n'}{n}\right)\boldsymbol{ZB}$;
- Applying k-means clustering to $\boldsymbol{Y}$ and updating $\boldsymbol{Z_k}$ and repeating the process until the $\boldsymbol{Z_k}$ values converge to a fixed number.

The resulting $\boldsymbol{Z_k}$ gives the optimum cluster centroid matrix, $\boldsymbol{G}$, and category quantification matrix, $\boldsymbol{B}$. The coordinate matrix, $\boldsymbol{G}$ $\boldsymbol{B}$, can be used to present a biplot of the clusters and categories. However, to facilitate the interpretation of the biplots, the two matrices are scaled by a constant value of $\gamma = \left(\frac{K}{Q} \cdot \frac{TrB^TB}{TrG^TG}\right)^{1/4}$. The new measures $\boldsymbol{G_s} = \gamma\boldsymbol{G}$, and $\boldsymbol{B_s} = \frac{1}{\gamma}\boldsymbol{B}$ have the same average squared deviation from the origin that are used for biplot presentations of analysis. The advantage of cluster CA among the CA variants is its formation of automated clusters. Other CA variants require formal generation of clusters based on cooccurrence patterns, which can generate biased results.

Results

Based on the exploratory data analysis, it was evident that patterns of categorical contributing factors vary by injury types. Cluster CA was applied separately to both no-injury crash data and fatal and injury crash data. Determination of the optimal clusters was conducted based on the convergence properties. Finally, four clusters for each of the datasets were developed. The final number of clusters was determined by trial and error. Optimization algorithms are not always sufficient for justifying the trends identified in each cluster and engineering judgment is required in such cases for the determination of the required number of clusters. Table 2 shows the locations of the centroids and relevant cluster-based properties (i.e., size of the clusters and within-cluster sum of squares). The first two clusters of each of the datasets contained 70% or above of the data used for the analysis.

No-Injury Crash Data

The no-injury crash data from encroachment-related work-zone crashes were used to develop clusters, which are illustrated in Figure 2 as a two-dimensional (2D) biplot. The locations of the cluster centroids are denoted by red texts (C1, C2, C3, and C4) and these texts are shown inside ellipses. As the biplot illustrates the locations of all the attributes, it was difficult to explore individual attributes. This plot is provided to show the nature of attributes present and their locations in 2D space. Additional cluster-level bar plots are subsequently shown and explained in detail. Cluster 1 (C1), which explained 44.5% of the data, primarily includes annual average daily traffic (AADT) greater than 4,000 vehicles per day (vpd), concrete surface type, two-way divided roadways, unprotected median as the trafficway, six lanes, a median width of less than 10 ft or greater than 20 ft, a left departure direction, the median being the location of the first harmful event, and the first harmful event being a collision with a barrier as the crash-contributing factors. This cluster was located mostly in Quadrant 3 and partially in Quadrant 2. Cluster 2 (C2; 24.4% of the data) primarily includes the first harmful location being on a roadway, along with several other positively associated but unknown categories such as surface type, trafficway, total lane count, crash severity, and injury severity. This cluster was mostly located in Quadrant 4. Cluster 3 (C3; 20.9%), which primarily relates to an AADT of 4,001 to 40,000 vpd, four lanes, a

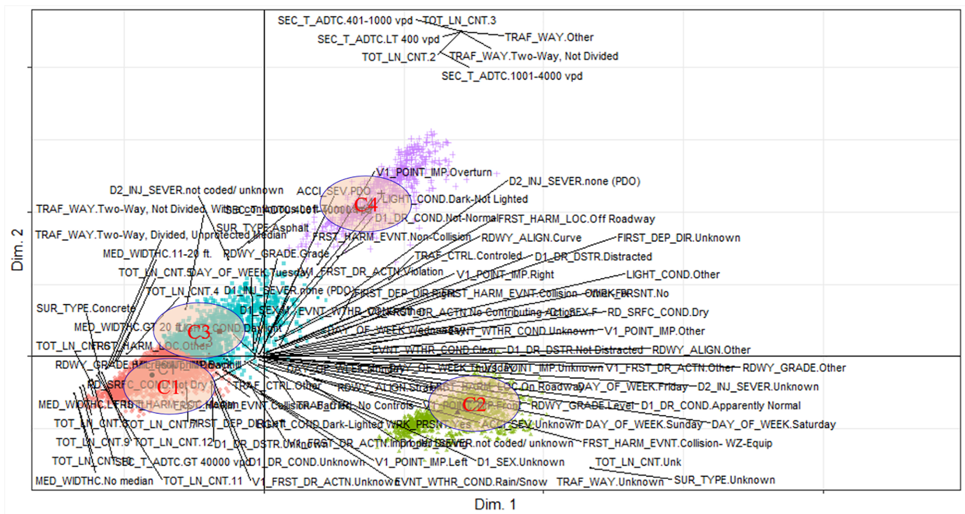

Figure 2. Clusters developed from no-injury crash data.
Note: C = cluster number.

two-way undivided roadway, with a CLTL trafficway or a two-way divided roadway, with unprotected median trafficway, asphalt surface type, median width of greater than 20 ft, noncollision or collision with other fixed object as the first harmful event, and overturning, was located mostly in Quadrants 1 and 2, and partially in Quadrant 3. Cluster 4 (C4; 10.2%), which primarily relates to two-way undivided roads as the traffic type, total lane count of two, surface type of asphalt, AADT less than 40,000 vpd, controlled traffic, noncollision being the first harmful event, and an important aspect of Vehicle 1 being overturn, was located mostly in the first quadrant. C3 was closest to the center of the biplot.

Cluster 1 (C1)—Median-Related Crashes on High-Volume, Two-Lane Divided Roadways. In C1 (representing 44.5% of the data) of encroachment-related no-injury work-zone crash data, AADT greater than 40,000 vpd had the longest bar on the positive side (see Figure 3*a*). It indicates that it has the strongest association with the other categories in this cluster, including concrete surface type, two-way divided roadways, unprotected median as the trafficway, six lanes, a median width of less than 10 ft or greater than 20 ft, a left departure direction, the median being the location of the first harmful event, and the first harmful event being a collision with a barrier. This cluster indicates patterns of median- or barrier-related encroachment-related crashes in work-zone area. The crashes mostly occur on high-volume two-lane divided (unprotected median) roadways with concrete pavement. Therefore this cluster is defined as "median-related crashes on high-volume, two-lane highways".

Cluster 2 (C2)—Encroachment on Roadway. In C2 of the encroachment-related no-injury work-zone crash data, several of the attributes were listed as unknown (see Figure 3). This indicated that around 25% of the no-injury work-zone-related data had missing information. One of the outstanding features of this cluster was "on roadway." As indicated earlier, encroachments usually take place outside of the travel lane, therefore one can assume that this cluster represented a special case in which the encroachment occurred on the roadway, where the vehicle was not running off the road.

Cluster 3 (C3)—Single-Vehicle Overturning Collisions on Two-Way Divided Roadways With Unprotected Median. In C3 of the encroachment-related, no-injury, work-zone crash data, the longest bar on the positive side was AADT of 4,001 to 40,000 vpd (see Figure 4). The other key category in this cluster included four lane, two-way undivided, with CLTL trafficway or a two-way divided roadways, with unprotected median trafficway, asphalt surface type, median width of greater than 20 ft, noncollision or collision with other fixed objects as the first harmful event, and overturning. The first departure direction also had a positive association, but the direction type was unknown. This cluster also indicated median-related collisions on two-lane divided roadways with low to moderate traffic-volume roadways. Safety barriers were the prospective countermeasure to reduce these crashes.

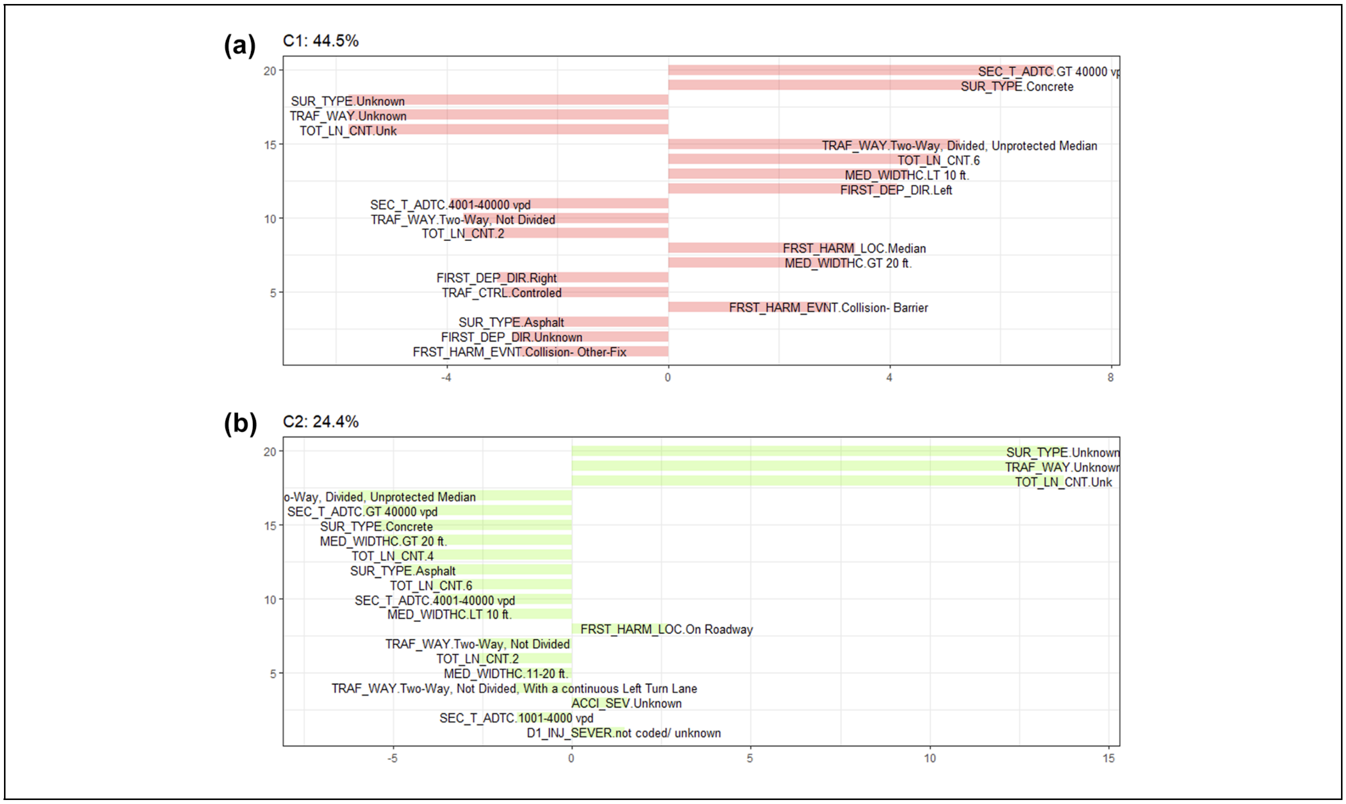

Figure 3. No-injury crash data: (*a*) Cluster 1 (C1) and (*b*) Cluster 2 (C2).

Cluster 4 (C4)—Overturning Crashes on Two-Lane Undivided Roadways in Controlled Traffic. In C4 of the encroachment-related no-injury work-zone crash data, the longest bar on the positive side was two-way and undivided roads as the traffic type, meaning that it had the highest association with other categories within this cluster (see Figure 4). This included two lanes, asphalt pavement, AADT less than 40,00 vpd, controlled traffic, noncollision being the first harmful event, and overturning. This cluster indicated encroachment-related work-zone crashes on two-lane undivided roadways with different traffic volumes (ranging from <400 to 40,000 vpd). The surface type in this cluster was asphalt.

Fatal and Injury Crash Data

The fatal and injury crash data from encroachment-related work-zone crashes were used to develop clusters (portrayed in Figure 5). The locations of the cluster centroids (denoted with red texts) are shown inside ellipses. As stated, the biplot as a visual representation renders exploring individual attributes difficult. This plot is provided to show the nature of attributes of fatal and injury crash data present and their locations in 2D space. Additional cluster-level bar plots are shown and explained in detail. Cluster 1 (42.1%), located mostly in Quadrants 2 and 3, primarily relates to roadways with AADT greater than 40,000 vpd, concrete surface, six lanes, the first departure direction being left, the trafficway being a two-way divided roadway with unprotected median, a median width of less than 10 ft, the first harmful location being the median, and the first harmful event being collision with a barrier. Cluster 2 (23.6%), which primarily relates to an AADT of 4,001 to 40,000 vpd, four lanes, the trafficway either being two-way undivided with CLTL traffic or two-way divided with an unprotected median, a median width of greater than 20 ft, an asphalt surface type, the first harmful event being noncollision with other vehicles, overturning, fatality, and the first departure direction being right or unknown, is mostly located in Quadrants 3 and 4. Cluster 3 (20.9%), which has several unknowns, but includes positive associations with collision with other fixed objects as the first harmful event, the first harmful location being on the roadway, the first departure direction being right, and the lighting conditions being dark but lighted, and the unknown categories including surface type, trafficway, and total lane count, is mostly located in Quadrant 1. Cluster 4 (13.4%), which primarily relates to two-way undivided trafficways, two-lanes, AADT less than 40,000 vpd, an asphalt surface type, controlled traffic, and an important aspect of Vehicle 1 being overturn, is located mostly in Quadrant 4. Cluster 2 is located closest to the center of the graph.

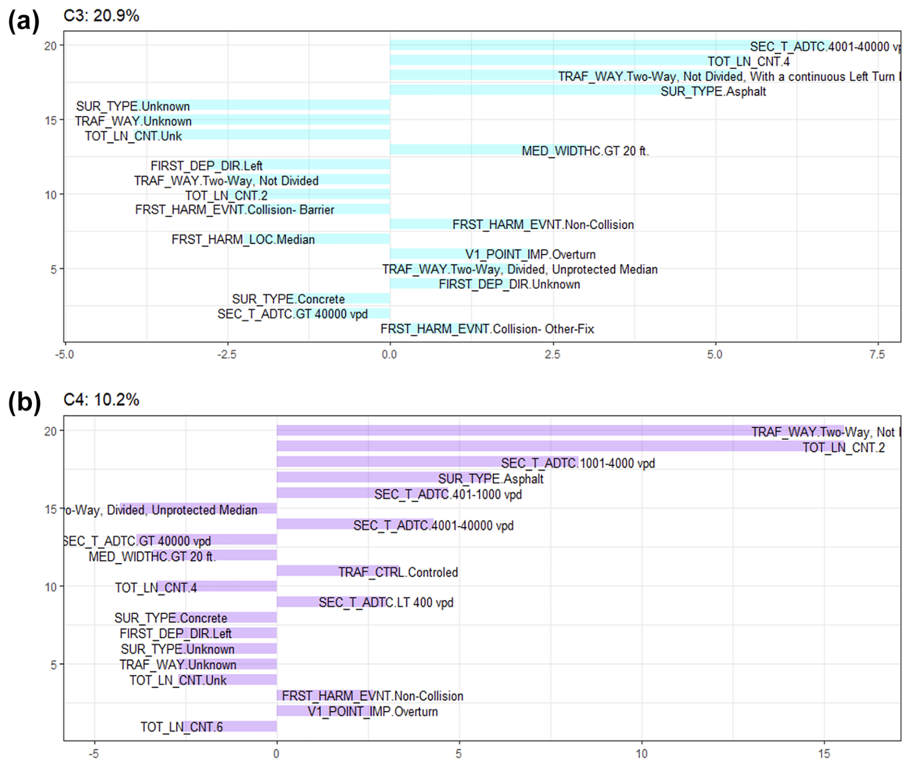

Figure 4. No-injury crash data: (*a*) Cluster 3 (C3) and (*b*) Cluster 4 (C4).

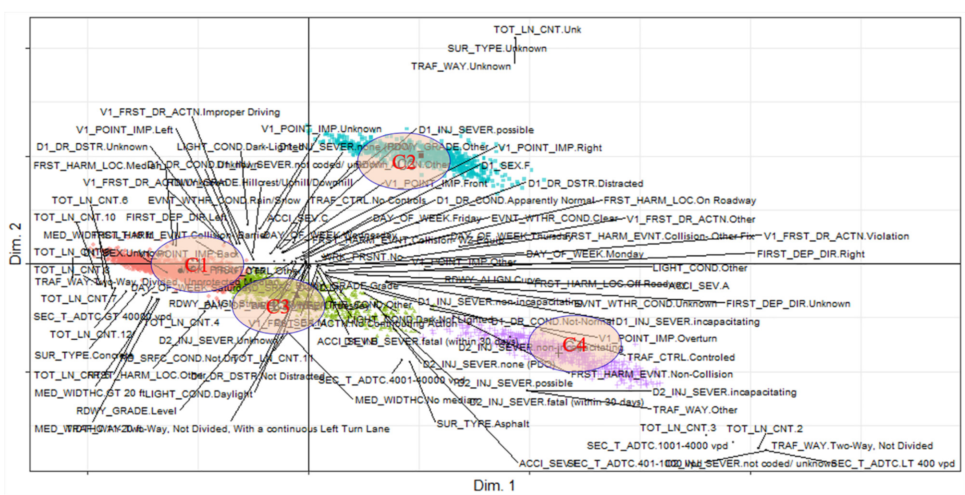

Figure 5. Clusters developed from fatal and injury crash data.

Cluster 1 (C1)—Median-Related Crashes on High-Volume Two-Lane Divided Roadways. In C1 of the encroachment-related work-zone fatal and injury crash data, the longest bar on the positive side was an AADT greater than 40,000 vpd, meaning it had the highest association with the other categories within this cluster (see Figure 6*a*). These include a surface type of concrete, six lanes, the first departure direction being left, the trafficway being a two-way, divided, with unprotected median, a median width of less than 10 ft, the first harmful location being

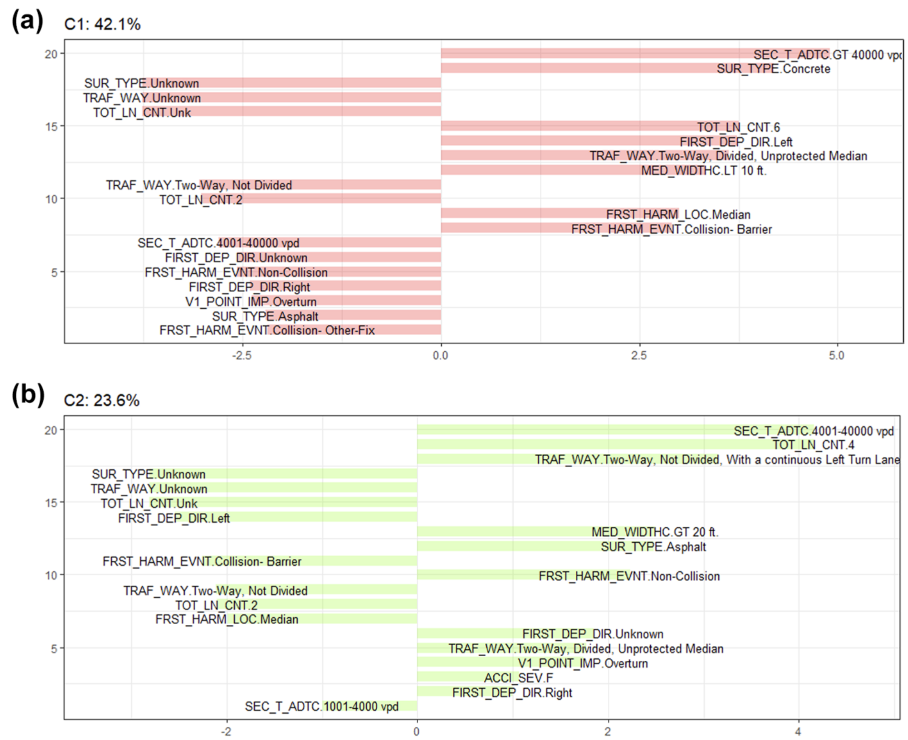

Figure 6. Fatal and injury crash data: (*a*) Cluster 1 (C1) and (*b*) Cluster 2 (C2).

the median, and the first harmful event being collision into a barrier.

Cluster 2 (C2)—Single-Vehicle Overturning Collisions on Two-Way Divided Roadways With Unprotected Median. In C2 of the encroachment-related work-zone fatal and injury crash data, the longest bar on the positive side was an AADT of 4,001 to 40,000 vpd, meaning it had the highest association with the other categories within this cluster (see Figure 6*b*). These include four lanes, the trafficway being either two-way, undivided, with a CLTL or two-way, divided, with an unprotected median, a median width of greater than 20 ft, asphalt surface type, the first harmful event being noncollision, an important point about Vehicle 1 being overturned, a crash severity of F, and the first departure direction being right or unknown.

Cluster 3 (C3)—Crashes on Roadways With Dark-Lighted Condition. C3 of encroachment-related work-zone fatal and injury crash data has several unknowns (see Figure 7*a*). However, the longest bar on the positive side that is known is the first harmful event, which is collision-other-fix. This means that it has strong positive associations with the other categories in this cluster, including the first harmful location being on roadway, the first departure direction being right, and the lighting conditions being dark but lighted. Other categories that are unknown but associated with it include surface type, trafficway, and total lane count. Although several key variable categories are unknown, the cluster indicates that work-zone crashes where the vehicle is encroaching to the right side of the road at dark (lighted condition). This cluster also shows that crashes took place on the main lane instead of shoulder or beyond.

Cluster 4 (C4)—Overturning Crashes on Two-Lane Undivided Roadways With Controlled Traffic. In C4 of the encroachment-related work-zone fatal and injury crash data, the longest bar on the positive side was a two-way undivided trafficway (see Figure 7*b*). This means that it had the highest association with the other categories within this cluster, two lanes, AADT less than 40,000 vpd, asphalt surface type, controlled traffic, and overturning. Cluster 4 represented 14% of the encroachment-related work-zone fatal and injury data. The attributes in this cluster were similar to Cluster 4 of the encroachment-related work-zone no-injury data.

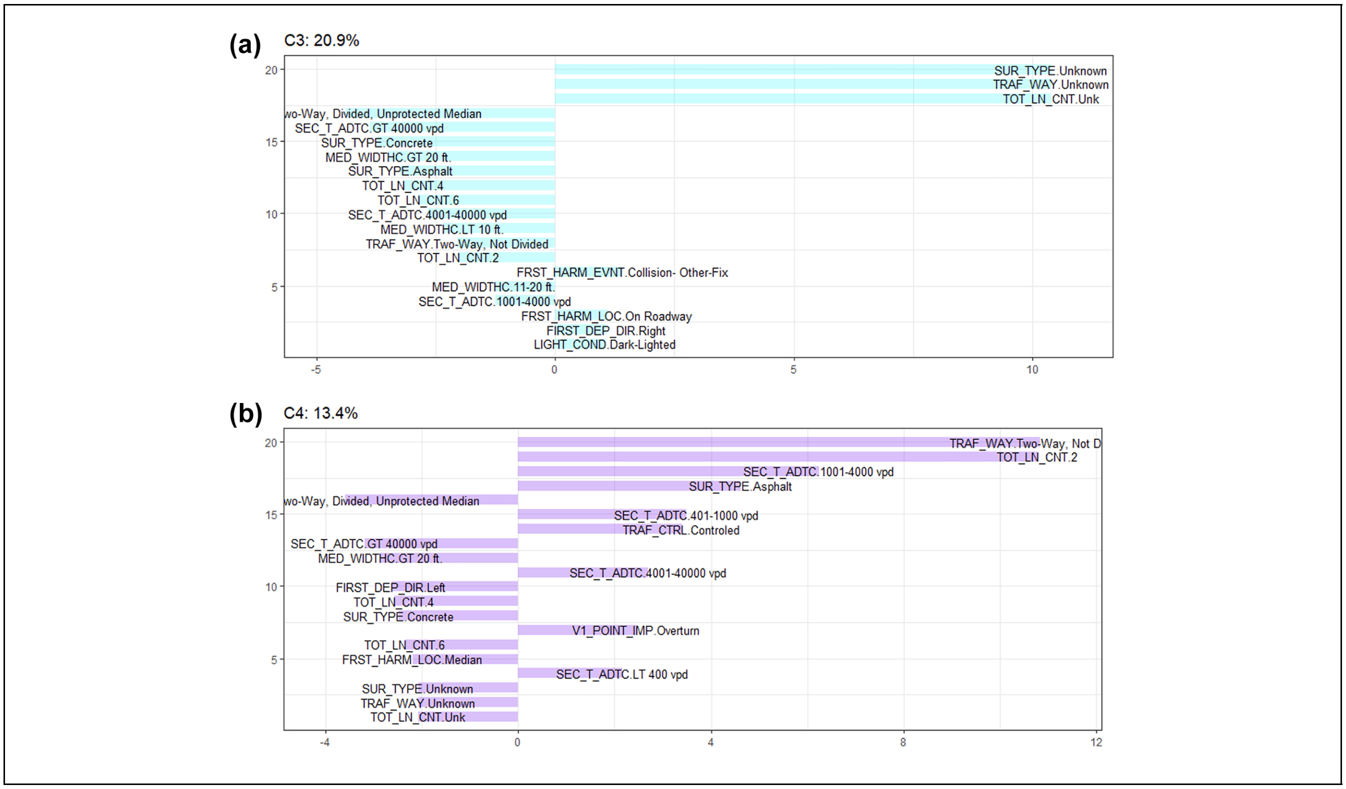

Figure 7. Fatal and injury crash data: Cluster 3 (C3) and Cluster 4 (C4).

Major Findings

The key findings from the cluster CA were following:

- The clusters and attributes in each cluster did not vary widely in the two datasets based on crash injury type.
- Encroachment-related work-zone no-injury data had a cluster that presented 25% of the data and some critical variable categories of this cluster were unknown. For fatal and injury data, a similar cluster covered around 21% of the data. This indicated that police reports often contain less information if crashes involve no injury.
- The three major clusters of both datasets were median-related crashes on two-lane divided high-volume roadways, single-vehicle overturning collisions on two-way divided roadways with unprotected median, and overturning crashes on two-lane undivided roadways in controlled traffic. The coverage by cluster varied in these two datasets. The results indicated a requirement for median protection to reduce encroachment-related work-zone crashes. Overturning crashes need to be examined in depth to identify the roadway geometry, posted speed limit, and curvature information. Although installation of median safety barriers could potentially increase the probability of errant vehicle crashes, implementation of such crash-worthy systems would help limit the severity of such crashes.
- As the datasets were analyzed based on injury type, individual injury type attributes were only seen in one cluster (i.e., Cluster 2 of fatal and injury crash data).
- Although median-related crashes dominated both datasets, overturning and right-encroachment crashes were also found to warrant careful attention from safety engineers.

Conclusions

Although there has been significant improvements in overall roadway safety in recent years, fatalities at work zones remain among the major concerns. In 2019 alone, there have been 752 fatal crashes at work zones. Work-zone crashes not only affect drivers but also construction workers and engineers, which makes this problem very personal to roadway safety researchers.

The overview of existing data and studies indicated that a major number of work-zone crashes resulted from vehicle encroachment. The vehicles exited the travel lane for a short amount of distance and time, indicating even the smallest mistake can have significant consequences. Unfortunately, there is little information available on work-zone encroachment conditions, making it challenging for safety researchers to identify crash-contributing factors to better understand the circumstance leading to crashes in work zones. This study used work-zone crash data from Texas to understand the various driver, roadway, and environmental conditions that may contribute to these crashes. The crash data included crash conditions and driver- and vehicle information. After matching all these variables, this study conducted a comprehensive filtering process to identify the work-zone crashes that may have resulted from encroachments. Note that such information is not readily available in crash databases and must be determined with the help of crash data experts.

After identifying the list of encroachment-related work-zone crashes, this study used cluster CA to identify the variable categories associated with fatal and injury and no-injury encroachment-related work-zone crashes. This method groups categories of variables into clusters, thus helping to better explore crash-contributing factors in encroachment-related work-zone crashes. Note that, in the recent years, several studies have applied cluster CA and other CA variants in traffic safety analysis (*33–39*). The results of this analysis indicated that encroachment-related work-zone crashes can be divided into several groups. Specifically, the fatal and injury work-zone crashes resulting from encroachment were divided into (1) median-related crashes on high-volume two-lane divided roadways; (2) single-vehicle overturning collisions on two-way divided roadways with unprotected median; (3) overturning crashes on two-lane undivided roadways with controlled traffic; and (4) work-zone crashes on roadways with dark-lighted conditions. No-injury crashes were divided into the following clusters: (1) median-related crashes on high-volume two-lane divided roadways; (2) single-vehicle overturning collisions on two-way divided roadways with unprotected median; (3) overturning crashes on two-lane undivided roadways in controlled traffic; and (4) encroachment on roadways. As observed, the first three clusters for both fatal and injury and no-injury crashes were the same. These clusters indicated that median presence and type, AADT, the number of lanes, as well as encroachment type (i.e., left or right) and crash type were some of the important factors that stood out in encroachment-related work-zone crashes. In addition to these factors, encroachments on travel lanes (i.e., roadways) and lighting conditions also played major roles in work-zone crashes occurring as a result of encroachments. This study could help roadway, roadside, and work-zone safety researchers better categorize and interpret the work-zone crashes resulting from encroachments. The work-zone clearinghouse (https://workzonesafety.org/) could include the findings of this study so that practitioners could have an easy access to pertinent data. Overall, the patterns identified in this study could be employed in developing the guidelines and manuals used by traffic engineers to make informed decisions when installing work zones at roadways locations, to prevent the most recurrent encroachment-related work-zone crashes.

The study was not without limitations. First, real work-zone-related information (for example, number of lane closure, speed limit change, length of the work zone) was missing from the database. Second, real encroachment-related information is not readily available. Incorporation of encroachment and work zone-related variables would make the findings more robust.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Subasish Das; data collection: Mahmood Tabesh; analysis and interpretation of results: Subasish Das and Mahmood Tabesh; draft manuscript preparation: Subasish Das, Mahmood Tabesh, Bahar Dadashova, and Chira Dobrovolny. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Subasish Das https://orcid.org/0000-0002-1671-2753
Mahmood Tabesh https://orcid.org/0000-0002-5187-4483
Bahar Dadashova https://orcid.org/0000-0002-4592-9118

References

1. NHTSA. *Fatality Analysis Reporting System (FARS)*. National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C, 2021.
2. Ullman, G. L., M. Pratt, S. Geedipally, B. Dadashova, R. J. Porter, J. Medina, and M. D. Fontaine. *Analysis of Work Zone Crash Characteristics and Countermeasures*. Publication 0309471281. Transportation Research Board, Washington, D.C., 2018.
3. The National Work Zone Safety Information Clearinghouse. Library of Resources to Improve Roadway Work Zone Safety for All Roadway Users. *The National Work Zone Safety Information Clearinghouse*. https://www.workzonesafety.org/. Accessed July 26, 2021.

4. Meng, Q., and J. Weng. Evaluation of Rear-End Crash Risk at Work Zone Using Work Zone Traffic Data. *Accident Analysis & Prevention*, Vol. 43, No. 4, 2011, pp. 1291–1300. https://doi.org/10.1016/j.aap.2011.01.011.
5. Weng, J., Q. Meng, and X. Yan. Analysis of Work Zone Rear-End Crash Risk for Different Vehicle-Following Patterns. *Accident Analysis & Prevention*, Vol. 72, 2014, pp. 449–457.
6. Zhang, K., and M. Hassan. Identifying the Factors Contributing to Injury Severity in Work Zone Rear-End Crashes. *Journal of Advanced Transportation*, Vol. 2019, 2019, p. 4126102. https://doi.org/10.1155/2019/4126102.
7. Bryden, J. E., L. B. Andrew, and J. S. Fortuniewicz. Intrusion Accidents on Highway Construction Projects. *Transportation Research Record: Journal of the Transportation Research Board*, 2000. 1715: 30–35.
8. Ullman, G. L., M. D. Finley, J. E. Bryden, R. Srinivasan, and F. M. Council. *Traffic Safety Evaluation of Nighttime and Daytime Work Zones*. NCHRP Report No. 627. Transportation Research Board of the National Academies, Washington, D.C., 2008.
9. Ullman, G. L., M. D. Finley, and L. Theiss. Categorization of Work Zone Intrusion Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2258: 57–63.
10. Sun, M., X. Sun, and D. Shan. Pedestrian Crash Analysis With Latent Class Clustering Method. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 50–57. https://doi.org/10.1016/j.aap.2018.12.016.
11. Chang, F., P. Xu, H. Zhou, A. H. S. Chan, and H. Huang. Investigating Injury Severities of Motorcycle Riders: A Two-Step Method Integrating Latent Class Cluster Analysis and Random Parameters Logit Model. *Accident Analysis & Prevention*, Vol. 131, 2019, pp. 316–326. https://doi.org/10.1016/j.aap.2019.07.012.
12. Nilsson, D., M. Lindman, T. Victor, and M. Dozza. Definition of Run-off-Road Crash Clusters—For Safety Benefit Estimation and Driver Assistance Development. *Accident Analysis & Prevention*, Vol. 113, 2018, pp. 97–105. https://doi.org/10.1016/j.aap.2018.01.011.
13. Li, Z., C. Chen, Y. Ci, G. Zhang, Q. Wu, C. Liu, and Z. (Sean) Qian. Examining Driver Injury Severity in Intersection-Related Crashes Using Cluster Analysis and Hierarchical Bayesian Models. *Accident Analysis & Prevention*, Vol. 120, 2018, pp. 139–151. https://doi.org/10.1016/j.aap.2018.08.009.
14. van de Velden, M., A. I. D'Enza, and F. Palumbo. Cluster Correspondence Analysis. *Psychometrika*, Vol. 82, No. 1, 2017, pp. 158–185. https://doi.org/10.1007/s11336-016-9514-0.
15. Feigenbaum, B., M. G. Fields, and S. Purnell. *24th Annual Highway Report*. Reason Foundation, Los Angeles, CA, August, 2019.
16. Schrock, S. D., G. L. Ullman, A. S. Cothron, E. Kraus, and A. P. Voigt. *An Analysis of Fatal Work Zone Crashes in Texas*. Report FHW A/TX-05/0-4028-1. Texas Department of Transportation, Research and Technology Implementation Office, Austin, 2004.
17. Whitmire, I. J., J. F. Morgan, T. Oron-Gilad, and P. A. Hancock. The Effect of in-Vehicle Warning Systems on Speed Compliance in Work Zones. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 54, No. 24, 2010, pp. 2023–2027.
18. Debnath, A., R. Blackman, and N. Haworth. A Review of the Effectiveness of Speed Control Measures in Roadwork Zones. *Proc., Occupational Safety in Transport (OSIT) Conference*, September 20–21, 2012, Gold Coast, Australia.
19. Garber, N. J., and T.-S. H. Woo. *Accident Characteristics at Construction and Maintenance Zones in Urban Areas*. Virginia Transportation Research Council, Charlottesville, 1990.
20. Osman, M., R. Paleti, and S. Mishra. Analysis of Passenger-Car Crash Injury Severity in Different Work Zone Configurations. *Accident Analysis & Prevention*, Vol. 111, 2018, pp. 161–172. https://doi.org/10.1016/j.aap.2017.11.026.
21. Liu, T., Y. Yang, G.-B. Huang, Y. K. Yeo, and Z. Lin. Driver Distraction Detection Using Semi-Supervised Machine Learning. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 4, 2016, pp. 1108–1120. https://doi.org/10.1109/TITS.2015.2496157.
22. NHTSA. Fatality Analysis Reporting System (FARS). https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars. Accessed July 19, 2021.
23. Garber, N. J., and M. Zhao. Distribution and Characteristics of Crashes at Different Work Zone Locations in Virginia. *Transportation Research Record: Journal of the Transportation Research Board*, 2002. 1794: 19–25.
24. Raub, R. A., O. B. Sawaya, J. L. Schofer, and A. Ziliaskopoulos *Enhanced Crash Reporting to Explore Workzone Crash Patterns*. Paper No. 01-0166. Northwestern University Center for Public Safety, Evanston, IL, 2001.
25. Hargroves, B. T. Vehicle Accidents in Highway Work Zones. *Transportation Engineering Journal of ASCE*, Vol. 107, No. 5, 1981, pp. 525–539. https://doi.org/10.1061/TPEJAN.0000946.
26. Nemeth, Z., and A. Rathi. Freeway Work Zone Accident Characteristics. *Transportation Quarterly*, Vol. 37, No. 1, 1983, pp. 145–159.
27. Abdel-Aty, M. Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models. *Journal of Safety Research*, Vol. 34, No. 5, 2003, pp. 597–603. https://doi.org/10.1016/j.jsr.2003.05.009.
28. Anastasopoulos, P. C., and F. L. Mannering. An Empirical Assessment of Fixed and Random Parameter Logit Models Using Crash- and Non-Crash-Specific Injury Data. *Accident Analysis & Prevention*, Vol. 43, No. 3, 2011, pp. 1140–1147.
29. Akepati, S. R., and S. Dissanayake. Characteristics of the Work Zone Crashes. *Proc., 1st Congress of Transportation and Development Institute (TDI)*, Chicago, IL, 2011.
30. McGee, H. W. *Practices for Preventing Roadway Departures*. Transportation Research Board, Washington, D.C., 2018.
31. Alshatti, D. A. *Examining Driver Risk Factors in Road Departure Conflicts Using SHRP2 Data*. Master's thesis. University of Dayton, OH, 2018.
32. McLaughlin, S. B., J. M. Hankey, S. G. Klauer, and T. A. Dingus. *Contributing Factors to Run-Off-Road Crashes and*

Near-Crashes. National Highway Traffic Safety Administration, Washington, D.C., 2009.

33. Rahman, M. A., S. Das, and X. Sun. Understanding the Drowsy Driving Crash Patterns From Correspondence Regression Analysis. *Journal of Safety Research*, 2022. https://doi.org/10.1016/j.jsr.2022.10.017.
34. Rahman, M. A., S. Das, X. Sun, M. Sun, and M. M. Hossain. Using Unsupervised Learning to Investigate Injury-Associated Factors of Animal-Vehicle Crashes. *International Journal of Injury Control and Safety Promotion*, 2022, pp. 1–10. https://doi.org/10.1080/17457300.2022.2125532.
35. Das, S., R. Tamakloe, B. Kutela, and A. Hossain Pattern Recognition From Injury Severity Types of Frontage Roadway Crashes. *Journal of Transportation Safety and Security*, 2022, pp. 1–22.
36. Das, S. *Artificial Intelligence in Highway Safety*. CRC Press, Boca Raton, FL, 2022.
37. Rahman, M. A., S. Das, and X. Sun. Single-Vehicle Run-Off Road Crashes Because of Cellphone Distraction: Finding Patterns With Rule Mining. *Transportation Research Record: Journal of the Transportation Research Board*, 2022.
38. Rahman, M. A., S. Das, and X. Sun. Using Cluster Correspondence Analysis to Explore Rainy Weather Crashes in Louisiana. *Transportation Research Record: Journal of the Transportation Research Board*, 2022. 2676: 159–173.
39. Das, S., M. M. Hossain, M. A. Rahman, X. Kong, X. Sun, and G. Mamun. Understanding Patterns of Moped and Seated Motor Scooter (50 cc or Less) Involved Fatal Crashes Using Cluster Correspondence Analysis. *Transportmetrica A: Transport Science*, 2021, pp. 1–21. https://doi.org/10.1080/23249935.2022.2029613.