International Journal of Transportation Science and Technology 10 (2021) 167–183

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Exploring the influential factors of roadway departure crashes on rural two-lane highways with logit model and association rules mining ☆

M. Ashifur Rahman [a,*], Xiaoduan Sun [a], Subasish Das [b], Sushmita Khanal [a]

[a] University of Louisiana, Lafayette, LA, United States
[b] Texas A&M Transportation Institute, San Antonio, United States

A R T I C L E   I N F O

Article history:
Received 14 June 2020
Received in revised form 31 October 2020
Accepted 2 December 2020
Available online 7 December 2020

Keywords:
Roadway departure crashes
Rural two-lane highways
Crash contributing factor
Association rule mining
Logit model

A B S T R A C T

Roadway departure (RwD) crashes are a major contributor of rural two-lane (R2L) highway crashes and fatalities. For targeted reduction of crashes and fatalities due to roadway departure, a thorough understanding of factors associated with RwD crashes is necessary. This study quantitatively assessed the available pre-crash factors that might influence the RwD crashes by developing a logit model comparing roadway, crash environment, and the vehicle and driver-related characteristics of 122,978 crashes that occurred in Louisiana over thirteen years. With a high prediction accuracy (81.9% area under the receiver operating characteristics curve), the model presented significant individual associations across crash characteristics with RwD crashes on R2L highways, for example – animals on roadways, snow/sleet/hail, 50–55 mph speed limit, AADT of 1,001 to 5,000 vehicles per day, drug intoxication, motorcycles, driving during 12 am to 6 am, curve radius of 501–1,000 ft., absence of streetlight, alcohol intoxication. Investigation on these top factors using association rules mining reveals findings such as – a higher likelihood of RwD crashes can be strongly associated animal presence coupled with the absence of streetlights, male drivers during the early morning (12 am to 6 am), male drivers driving with no passengers, drivers being intoxicated by both drugs and alcohol, etc. Findings from this study are expected to help highway safety specialists not only in identifying and predicting RwD crashes but also in an improved understanding of associated contributing factors leading to the application of proper countermeasures for the strategic reduction of RwD crashes.

1. Introduction

A roadway departure (RwD) crash is a non-intersection crash, which occurs after a vehicle crosses an edge line or a centerline, or otherwise leaves the traveled roadway (FHWA, 2017a). RwD crashes are the result of drivers running off the road to the right, crossing the centerline/median into an oncoming lane of traffic (head-on or opposite-direction-sideswipe crashes), or running off the road to the left. Vehicles running off the road may also involve a rollover, an immersion, or the hitting of a fixed object. The FHWA mentions four key reasons for roadway departure from the drivers' perspective – roadway condition, collision avoidance, vehicle component failure, and driver error (FHWA, 2019). Because vehicles involved

Peer review under responsibility of Tongji University and Tongji University Press.
* Corresponding author at: 131 Rex St, Madison Hall Rm 223, Lafayette, LA, United States.
E-mail address: ashifur@louisiana.edu (M.A. Rahman).

https://doi.org/10.1016/j.ijtst.2020.12.003
2046-0430/Published by Elsevier B.V.

in RwD crashes often end up hitting moving vehicles or fixed rigid structures (bridges, poles, guardrails, etc.), the outcome of RwD crashes tends to be severe.

Roadway departure (RwD) crashes are considered to be a major contributor of highway fatalities in the United States. During 2015–2017, more than half of all roadway fatalities occurred due to roadway departure (FHWA, 2017a). RwD crashes are a serious concern in the Louisiana State, specifically on rural two-lane (R2L) highways. According to the Louisiana crash data, between 2005 and 2017, 29.5 percent of all non-intersection crashes are caused by roadway departure on the state-controlled highways (LADOTD, n.d.). RwD crashes on the R2L highways consisted of 72.2 percent of all R2L non-intersection crashes over the course of thirteen years. In the same period, 79.7 percent of total fatal non-intersection crashes (4,903 of 6,151) were reported as having been caused by the roadway departure, including 37.4 percent from R2L highways. The general descriptive statistics indicate that RwD crashes require serious attention.

National transportation agencies have identified lowering the frequency of roadway departure crashes as a national priority (AASHTO, 2008; Julian, 2013). To prevent RwD crashes by keeping vehicles on the roadway, the FHWA and AASHTO (American Association of State Highway Transportation Officials) recommended several countermeasures, such as – pavement friction, alerting drivers with rumble strips, enhancing delineation along horizontal curves, and improving nighttime visibility, etc. (AASHTO, 2008; FHWA, 2017a). In line with the nationwide urgency to cut down RwD crashes, Louisiana Department of Transportation and Development (DOTD) also reported preventing RwD crashes as one of the top priorities in obtaining the goal of halving traffic fatalities and severe injuries from 2009 to 2030 (LADOTD, 2018). The Louisiana DOTD has already implemented several countermeasures on a large scale in recent years, notably centerline rumble strips and shoulder rumble strips on R2L highways. However, a data-driven approach with an improved understanding of the multitude of factors associated with the RwD crash is required for the application of safety countermeasures aiming towards its targeted reduction and prevention.

The approach of this study is twofold. First, the research team utilized a conventional binary logistic regression model to analyze roadway, driver, vehicle, crash and environmental characteristics from crash data of thirteen years. In the logit model, the crashes essentially form one element of the binary outcome of a crash (RwD versus non-RwD) to estimate the strength of association each crash characteristic carries in the occurrence of RwD crashes. This prediction model also helps to estimate the probability of an RwD crash as a function of a specified group of characteristics.

Second, aiming at gaining more knowledge on the associative contributing factors, we applied 'Association Rule Mining' (ARM) – a non-parametric unsupervised data mining algorithm – that offers analysts the flexibility to explore interconnections among factors without prior knowledge of them. Widely used in business disciplines, this method has been popular among transportation safety researchers (Rahman et al., 2020). Pande and Abdel-Aty (2009) suggested that ARM could be very useful to uncover unknown patterns in the crash data obtained from large jurisdictions and potentially be a decision support tool for traffic safety administrators. The ability of this technique in detecting interdependencies among crash factors has been reiterated in a number of studies in the last decade, especially in recent years (e.g. Das et al., 2020a, 2018; Weng et al., 2016). We utilized this data mining tool to identify patterns of crashes that are related to the selected key contributing factors of RwD crashes from the logit model and to demonstrate how interconnections of multiple crash contributing factors of RwD crashes can be explored further.

2. Literature review

To gain initial insights on the contributory factors of RwD crashes, the research team delved into the previous studies (Table 1), dredging up different characteristics that might influence roadway departure. Lord et al. (2011) performed a statewide thorough investigation on the contributing factors related to roadway departure on R2L highways in Texas using traffic, roadway geometry, and crash data between 2003 and 2008. Simple descriptive analysis was used in addition to the negative binomial regression model on RwD crashes that examined the effect of lane width, shoulder width, traffic volume, curve density, number of driveways on RwD crash frequency. A follow-up study was conducted by Avelar et al. (2019) in the recent years.

Practically, the majority of the RwD crashes are single vehicle crashes or more specifically single vehicle run-off-road (SVROR) crashes – as for Louisiana, more than 80% of RwD crashes are single vehicle crashes. Studies are available on the factors related to single vehicle or SVROR crashes. Using Fatality Analysis Reporting System (FARS) data on fatal crashes involving passenger vehicles during 1991–2007, a nationwide study analyzed factors related to single vehicle run-off-road (SVROR) crashes (Cejun and Subramanian, 2009) by developing a logit model of run-off-road crashes with regards to on-road crashes. Das et al. (2020b) developed rule-based regression models for run-off-road crashes.

Using 557 randomly selected fatal crash data between 1997–1998 in four southern states in the USA, Zhu et al. (2010) applied binary logit models to predict the probability that a fatal crash is a single-vehicle run-off-road crash or not. The study employed roadway design characteristics, roadside environment features, and traffic conditions proximal to the crash site as explanatory variables that could critically influence single-vehicle run-off-road crashes. An investigation on run-off road crashes or near-crash events, based on the 100-Car Naturalistic Driving Study, explored the relationship between frequency of ROR event per million vehicle miles traveled with various driving conditions (Shane et al., 2009). Hallmark et al. (2011) employed naturalistic driving data to model left-side and right-side departure separately using 'non-departure' crashes as 'control'.

Table 1
Findings from key studies.

Study	Year	Key results
Lord et al. (2011)	2003–2008	• RwD crashes occurred more on curves and during nighttime, and driver-related prevalent factors for RwD crashes were distracted driving and speeding. • An increase in lane width and shoulder width was associated with a decrease in RwD crash frequency, whereas an increase in traffic volume and curve density were linked with its increase. • The presence of shoulders was associated with a decrease, and the number of driveways had little effect on RwD crash frequency
Hashemi and Archilla (2017)	2008–2011	Four key contributing factors with significant roles in distinguishing RwD from non-RwD crashes – crashes on curves, on straight segments with two lanes or less, during daylight condition, and on highways with a speed limit greater than 35 mph.
Zhu et al. (2010)	1997–1998	Lane width, horizontal curve, and lighting conditions are consistently found to critically influence single-vehicle run-off-road crashes for all four states.
Hallmark et al. (2011)		• An increase in lane width, radius, oncoming vehicle density, amount of time a driver traveling at 10 or more mph over the posted or advisory speed, highly visible markings, absence of shoulder etc. reduced the likelihood of 'right side' departure • 'Left side' departure was more associated with 20–30 year-old drivers, male drivers, moderately visible lane markings, nighttime driving etc.
Hashemi and Archilla (2016)	2008–2011	A higher probability of roadway departure crashes was found in urban areas (majority of roads are located in urban areas), 2-way undivided roads, curvy roads, hilly roads, dirt and gravel surfaces, oily and wet road surface, fatigue and medical medicines consumption by driver, hazy weather, and dark/no light condition.
Avelar et al. (2019)	2013–2018	Provides a systemic outline to identify high-risk sites for RwD crashes and suitable countermeasures for implementation.
Das et al. (2020b)	2010–2016	• Used seven years of crash to develop rule-based SPFs for rural two-lane roadways in Louisiana. • Cubist method shows better performance than other modes (negative binomial, random forest, and support vector regression).

Few studies investigated the dichotomy between RwD and non-RwD crash characteristics exclusively. Hashemi and Archilla (2017) explored roadway geometry, roadway inventory, and environmental characteristics based on 4-year crash data (2008 to 2011) from Oahu, Hawaii and differentiated RwD crashes from non-RwD crashes by a non-parametric methodology – classification and regression trees (CART). In a separate study, Hashemi and Archilla (2016) explored the spatial distribution of roadway departure crashes at the district level taking into account crash locality, type of collision, roadway design, and human and environmental factors. Several of the previous studies cited the issue that simultaneous presence of more than one factors may well be associated and restated the need to evaluate the combined effect of multiple contributing factors on RwD crashes (Cejun and Subramanian, 2009; Lord et al., 2011; Zhu et al., 2010).

The literature review shows that use of application of both logit model and unsupervised modeling such as ARM were not used before in identifying association patterns of the variables in RwD crashes. As conventional statistical model such as logit model provides odds of RwD likelihood based on individual variables, usage of ARM provides additional contexts in identifying the association patterns of different variable categories.

3. Methodology

Investigating the disparity of crash types or crash severity types provides researchers insight into driver, vehicle, roadway environment, and temporal factors that are only explorable at crash level. The analytic framework of this study has been presented in (see Fig. 1). In this study, we first estimated the individual association of the initially selected factors with RwD crashes in reference to non-RwD crashes. To get a clearer sight of the results, the existing literature was then discussed to interpret and compare with the estimates of the individual association of the factors (odds ratio) from our logit model. We verified the model by estimating the prediction accuracy by estimating the area under the curve (AUC) in the Receiver Operating Characteristics (ROC) Curve.

Using several criteria, the top key contributing factors were selected to explain how they can associate with other factors collectively only in RwD crashes. An association rule mining algorithm, apriori, was used to find associations of frequent itemsets (i.e. factors) represented by rules – where each rule has one or more antecedents with one specific top contributing factor as a consequent. A rule with multiple itemsets (e.g. 4-itemset rule means three antecedents and one consequent) may create a frequently occurring more than expected RwD crash scenario which is not supposed to be a prediction.

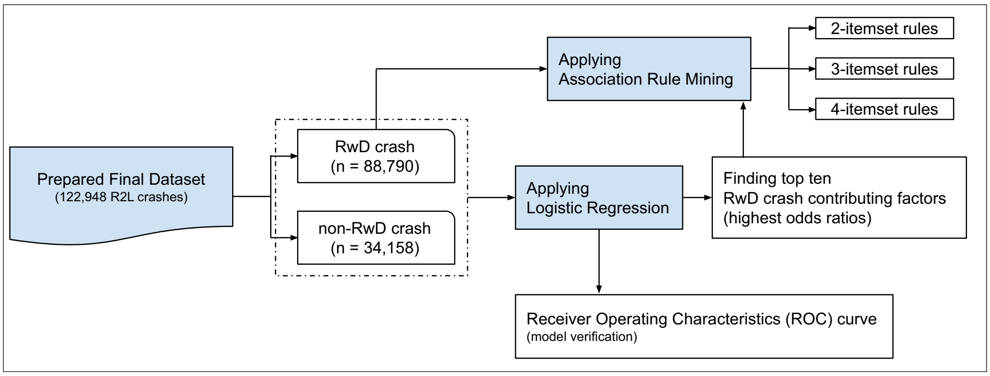

Fig. 1. Analytic framework of the study.

3.1. Logit model

A linear relationship between the n explanatory variables $x_1, x_2, \cdots.., x_n$ variables and the log-odds of the event $p = P(Y = RwD)$ can be expressed as, $\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$. The z-statistic is estimated as, $z = \frac{\widehat{\beta_i}}{S.E.\left(\widehat{\beta_i}\right)}$, where, $\widehat{\beta_i}$ = Estimated ith coefficient, and $S.E.\left(\widehat{\beta_i}\right)$ = Standard error of the coefficient. Odds can be estimated by $\frac{p}{1-p} = e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}$. Odds ratios in the logit model in this study will represent the constant effect of the contributing factors, on the likelihood that the RwD crash will occur.

3.1.1. Variable selection with AIC

The Akaike information criterion (AIC) estimates the relative quality of statistical models for a given set of data using out-of-sample prediction error (Akaike, 1974). If k is the number of estimated parameters in the model and $\widehat{L}$ is the maximum value of the likelihood function for the model, then the AIC value of the model is expressed as: $AIC = 2k - 2\ln\left(\widehat{L}\right)$. The *stepAIC* function from the *MASS* package (Ripley et al., 2019) of R software was used to select the best model by minimizing AIC in a stepwise algorithm. The *stepAIC* function uses with a full or null model to begin with, and direction arguments "forward" and "backward" is applied by eliminating and adding features to the progressed model. We applied both 'forward' and 'backward' selection and at the very last step *stepAIC* produced the optimal set of features with minimized AIC value.

3.1.2. Multicollinearity check with VIF

Multicollinearity to a high degree can overfit the model and substantially reduce the model's performance. Statistically, the variance inflation factor (VIF) is the quotient of the variance in a model with multiple terms by the variance of a model with one term alone. It provides an index that measures how much the variance (the square of the estimate's standard deviation) of an estimated regression coefficient is increased because of collinearity. The VIF, where Degrees of Freedom (Df) = 1, would be proportional to the inflation due to collinearity in the confidence interval for the coefficient. Since the log-odds is an unweighted linear model with more than 1 Df, we calculated Generalized Variance Inflation Factors (GVIFs) (Fox and Monette, 1992). The GVIF is a combined measure of collinearity for each group of predictors that should be considered together, i.e. multi-level categorical variables. To make GVIFs comparable, an adjustment for the dimension of the confidence ellipsoid, $\mathrm{GVIF}^{\frac{1}{2\times \mathrm{Df}}}$, is estimated (Fox, 2016).

3.2. Association rule mining

3.2.1. Theoretical background of ARM and key interest measures

Association rule mining (ARM) was intended to identify strong crash patterns in terms of a set of items that are present together during an RwD crash. Association rules are often implicated in the form $x \Rightarrow y$, in which x (antecedent) and y (consequent) are two separate subsets of all available items in the RwD crash dataset.

Two measures of significance are used in the generation of association rules – 'support' and 'confidence'. Support of a rule is a measure of how frequently the items (both antecedent and consequent) involved in it occur together in the dataset. Support of antecedent, consequent, and a rule can be expressed as: $s(x) = \frac{freq(x)}{n}, s(y) = \frac{freq(y)}{n}, s(x \Rightarrow y) = \frac{freq(x\cap y)}{n}$

Where,

$s(x) = \text{support of antecedent } x$
$s(y) = \text{support of consequent } y$

$s(x \Rightarrow y) = \text{support of the rule } x \Rightarrow y$
$freq(x) = \text{frequency of crashes with all itemset}(s) \text{ of } x \text{ in RwD crash dataset}$
$freq(y) = \text{frequency of crashes with consequent } y \text{ in RwD crash dataset}$
$freq(x \cap y) = \text{frequency of crashes with both } x \text{ and } y \text{ in RwD crash dataset}$
$n = \text{frequency of all crashes RwD crash dataset}$

The confidence value of a rule, $x \Rightarrow y$, is the proportion of the instances in the dataset that contains x which also contains y. Confidence of a rule $x \Rightarrow y$ can be expressed as: $c(x \Rightarrow y) = \frac{s(x \Rightarrow y)}{s(x)}$. The most used measure of interestingness is 'lift'. Lift measures how many times more often x and y occur together than expected if they were statistically independent. For a rule $x \Rightarrow y$, lift can be expressed as: $l(x \Rightarrow y) = \frac{s(x \Rightarrow y)}{s(x) \times s(y)}$. A lift value more than 1 implies antecedent and consequent are dependent on one another and their simultaneous presence is more than expected. A lift value of less than 1 indicates the presence of item(s) has negative effect on presence of other item(s) and vice versa. A lift value of 1 indicates independence between x and y.

3.2.2. Issues with apriori rule generation

Interesting apriori rules were generated by using the top contributing factors as consequents and specifying the constraints on the key interest measures. 2-itemsets, 3-itemsets, and 4-itemsets were separated to provide clear details of associations. Considering the extent of this study, a total of fifty rules (top 5 rules for each top 10 contributing factors as a consequent) have been presented in this study. Only the generated rules with a lift value greater than 1 were selected.

In apriori algorithm, frequent subsets are sequentially generated one item at a time through a breadth-first search. The minimum support threshold is first used to find frequent (significant) itemsets to filter the dataset. For instance – with a dataset of 100 crashes, selecting $n\%$ support indicates that returning association rules could potentially explain at least n crashes. The minimum confidence criterion is then used in a second step to produce more reliable rules with specified consequents from the frequent itemsets.

Selecting the minimum threshold values for support and confidence could be challenging for researchers and may require domain knowledge. Very low values of support and confidence will produce a large number of rules, a lot of which will be circumstantial spurious rules and difficult to interpret. Very high values of support and confidence will generate a small number of rules, in which patterns may not be found. Applying complex optimization to estimate the effective values of thresholds might produce an optimal set of best rules (Das et al., 2018), however, user-interaction in lieu of a rigorous approach has been encouraged (Bayardo and Agrawal, 1999). Since the consequent was pre-selected (top contributing factors from the logit model) in this study, minimum support was selected based on the percentage of the contributing factor (in the consequent) in concern. Considering the percentage of the consequent in the RwD crash dataset (Table 2) and typical high sensitivity of confidence on consequent, an iterative process for simple convex optimization was used in this study. The apriori rules of ARM were generated using the 'arules' package (Hahsler et al., 2008) of R software (R Development Core Team, 2019).

4. Data

Identification of an RwD crash requires an understanding of the sequence of crash events and vehicle maneuvers during the crash (Kweon and Lim, 2019). The Louisiana DOTD research team derives whether each crash is an RwD or not based on careful consideration of collision diagrams, crash narratives comprised of evidence found at the scene, and findings from the interview of participants and witnesses on crash occurrence.

To develop a complete dataset for the analysis, the research team used crash data provided by Louisiana DOTD with a comprehensive crash environment, driver and vehicle data, and highway section databases containing annually updated roadway geometric information. After filtering out the 115 crashes with multiple missing information, a database of 122,978 crashes was obtained. For this study, the researchers selected a number of explanatory variables from the available databases that are perceived to presumably contribute to roadway departure. Initially, variables were selected based on reasonable engineering judgment, learning from previous studies, and availability in the Louisiana crash database. RwD crashes take a large share (72.2 percent) of total R2L crashes (122,978) during 2005–2017. The percentage can be considered very high compared to the percentage of RwD crashes (29.5 percent) on all roadways non-intersection crashes during the same period. The percentage distribution of the initially selected variables has been presented in Table 2, in which some key observations can be made:

- *Crash and Environmental Characteristics:* RwD crashes prevalently occurred during nighttime (50.7 percent, during 6 pm – 12 am and 12 am – 6 am) as opposed to non-RwD crashes (21.5 percent). Lord et al. (2011) also found a proportionately higher frequency of RwD crashes on Texas rural 2-lane highways during nighttime. The percentage of RwD crashes is higher during the weekend compared to the percentage of non-RwD crashes during weekdays. An Ohio study aiming at exploring driver characteristics related to run-off-road crashes found a higher percentage of those crashes occurred during weekends compared to non-run-off-road crashes (Alruwaished, 2014). Lack of visibility is considered a serious factor in RwD crashes, as among all lighting conditions, RwD crashes during darkness in the absence of streetlight are overrepresented (43.7 percent) in comparison to non-RwD crashes (13.2 percent). The percentage of RwD crashes due to the presence of animals on R2L highways appears to be higher with regards to non-RwD crashes. One insurance company

Table 2
Percentage distribution by selected variable class.

Category	RwD	%	Non-RwD	%	Category	RwD	%	Non-RwD	%
Crash and Environment Characteristics					**Vertical alignment**				
Time									
12 am–6 am	19,761	22.3	1,663	4.9	Level	76,530	86.2	31,254	91.5
6 am–12 pm	20,348	22.9	10,528	30.8	Dip-hump	190	0.2	74	0.2
12 pm-6 pm	23,465	26.4	16,301	47.7	Hillcrest	1,694	1.9	615	1.8
6 pm–12 am	25,216	28.4	5,666	16.6	Level-elevated	3,141	3.5	841	2.5
Day of week					On grade	7,020	7.9	1,283	3.8
Weekday	60,024	67.6	26,887	78.7	Other	215	0.2	91	0.3
Weekend	28,766	32.4	7,271	21.3	**Posted speed limit**				
Lighting condition					25 mph or less	834	0.9	1,433	4.2
Daylight	42,350	47.7	26,721	78.2	30 to 35 mph	2,299	2.6	2,747	8.0
Dark (no streetlight)	38,830	43.7	4,519	13.2	40 to 45 mph	10,004	11.3	6,631	19.4
Dark (streetlight)	2,316	2.6	1,037	3.0	50 to 55 mph	72,770	82.0	20,561	60.2
Dusk/dawn	3,258	3.7	1,160	3.4	60 mph or more	516	0.6	256	0.7
Other	74	0.1	28	0.1	unknown	2,367	2.7	2,530	7.4
Unknown	1,962	2.2	693	2.0	***Driver and Vehicle Characteristics***				
					Driver age				
Surface condition					15-19y	12,089	13.6	5,444	15.9
Dry	82,888	93.4	32,180	94.2	20–24y	14,764	16.6	4,997	14.6
Wet	5,578	6.3	1,860	5.4	25–34y	20,411	23.0	7,020	20.6
Icy	324	0.4	118	0.3	35–44y	14,419	16.2	4,822	14.1
Weather					45–54y	11,754	13.2	4,355	12.7
Clear	59,258	66.7	25,537	74.8	55–64y	7,007	7.9	3,065	9.0
Cloudy	15,099	17.0	5,190	15.2	65–74y	2,936	3.3	1,801	5.3
Rain	11,560	13.0	2,842	8.3	≥75y	1,530	1.7	1,359	4.0
Snow/sleet/hail	480	0.5	40	0.1	Unknown	3,880	4.4	1,295	3.8
Other	2,082	2.3	515	1.5	**Driver gender**				
Unknown	311	0.4	34	0.1	Male	55,971	63.0	20,987	61.4
Roadway condition					Female	29,002	32.7	12,021	35.2
No abnormalities	73,671	83.0	32,349	94.7	Unknown	3,817	4.3	1,150	3.4
Animal	9,669	10.9	160	0.5	**Driver license state**				
Bumps	223	0.3	97	0.3	Louisiana	77,116	86.9	29,995	87.8
Construction	533	0.6	520	1.5	Out of state	6,451	7.3	2,431	7.1
Holes/deep ruts	151	0.2	38	0.1	Unknown	5,223	5.9	1,732	5.1
Loose material	213	0.2	36	0.1	**Driver distraction**				
Previous crash	64	0.1	126	0.4	Not distracted	58,617	66.0	21,015	61.5
Shoulder abnormality	338	0.4	90	0.3	Cellphone	1,935	2.2	498	1.5
Water	1,676	1.9	247	0.7	Other electronic device(s)	408	0.5	161	0.5
Other	2,252	2.5	495	1.4	Inside vehicle	4,303	4.8	1,974	5.8
Roadway Characteristics					Outside vehicle	2,509	2.8	1,522	4.5
AADT									
400 or less	3,275	3.7	417	1.2	Unknown	21,018	23.7	8,988	26.3
401–1,000	11,636	13.1	1,708	5.0	**Alcohol**				
1,001–5,000	55,681	62.7	17,409	51.0	Yes	11,782	13.3	1,333	3.9
5,001–10,000	15,026	16.9	10,575	31.0	No	77,008	86.7	32,825	96.1
10,001–20,000	3,101	3.5	3,900	11.4	**Drugs**				
>20,000	71	0.1	149	0.4	Yes	3,303	3.7	358	1.0
Lane width					No	85,487	96.3	33,800	99.0
Less than 10 ft.	344	0.4	112	0.3	**Passenger presence**				
10 ft. to less than 11 ft.	20,273	22.8	5,429	15.9	Yes	21,473	24.2	8,936	26.2
11 ft. to less than 12 ft.	32,610	36.7	11,440	33.5	No	67,317	75.8	25,222	73.8
12 ft. and greater	35,563	40.1	17,177	50.3	**Vehicle type**				
Shoulder width					Car/SUV/van	51,033	57.5	18,206	53.3
2 ft. or less	14,712	16.6	4,245	12.4	Pickup truck	29,048	32.7	11,753	34.4
Greater than 2 ft. to 4 ft.	29,737	33.5	8,722	25.5	Bus	87	0.1	127	0.4
Greater than 4 ft. to 6 ft.	17,727	20.0	5,893	17.3	Large truck	4,093	4.6	2,060	6.0
Greater than 6 ft.	26,614	30.0	15,298	44.8	Motorcycle	2,097	2.4	290	0.8
Curve radius					Others	2,432	2.7	1,722	5.0
500 ft. or less	3,793	4.3	1,042	3.1	**Vehicle year**				
501–1,000 ft.	6,889	7.8	1,034	3.0	Before 2012	80,700	90.9	30,471	89.2
1,001–2,500 ft.	12,561	14.1	3,053	8.9	2012 or later	6,329	7.1	2,583	7.6
2,501–5,000 ft.	6,652	7.5	2,255	6.6	Unknown	1,761	2.0	1,104	3.2
5,001–10,000 ft.	4,558	5.1	1,939	5.7					
>10,000 ft.	3,029	3.4	1,549	4.5					
Tangent	51,308	57.8	23,286	68.2					

identifies Louisiana as a moderate risk state in terms of insurance claims due to collisions with animals (1 in 169 claims) during 2018–2019 (StateFarm, 2019).

- *Roadway Characteristics:* A higher percentage of RwD crashes occur on R2L with narrow lanes, narrow shoulders, moderate traffic volume (1001–5000 vpd), sharper curves in comparison to non-RwD crashes. Higher posted speed limit (50–55 mph) also appears to be prevalent with RwD crashes.
- *Driver and Vehicle Characteristics:* Drivers, aging from 25 to 54 years, show slightly higher representation in RwD crashes compared to non-RwD crashes. RwD crashes are overrepresented for alcohol than non-RwD crashes, 13.3 percent vs. 3.9 percent. Similar is found for drug involved crashes. Distraction due to cell phone usage is slightly higher in RwD crashes.

5. Results and discussions

5.1. Best logit model selection

In the automated process of final model development by minimizing AIC while including and excluding variables step by step, only the 'surface condition' variable was excluded. The model with all the initially selected variables is 110,991.1, which is reduced to 110,987.8 by excluding the 'surface condition' variable. The difference between keeping and excluding the variable in terms of AIC appears to be minimal; however, both 'surface condition' categories ('wet', 'icy' in reference to 'dry') were found insignificant if 'surface condition' was considered in the model. Performing a simple likelihood ratio (LR) test between 'model with surface condition' and 'model without surface condition' shows the inclusion of the 'surface condition' makes an insignificant change to the model ($LRstat = 0.634$ and $p - value = 0.728$). It is important to mention that 'weather condition' and 'surface condition' often represent an equivalent condition, for example – weather condition = clear and surface condition = dry, or rainy weather and wet surface. The researchers let the analytic model to minimize collinearity by discarding variables.

The examination of multicollinearity was performed through a conservative rule of thumb for $\mathrm{GVIF}^{\frac{1}{2\times \mathrm{Df}}}$ greater than 2. None of the variables in the final dataset had a value of greater than 2, irrespective of including 'surface condition'. This indicates multicollinearity is not an issue for the model with selected variables. The researchers decided not to include 'surface condition' in the final model. The GVIF values of the variables in the final model have been presented in Table 3.

Probabilities of an RwD crash predicted by the final model were compared against actual binary outcomes (RwD crash = 1, non-RwD crash = 0) through the estimates of sensitivity and specificity. This required a nonparametric approach (i.e. no distributional assumption) for estimating decision threshold comparing all possible values of sensitivity and specificity. The model performed considerably well with a high accuracy in terms of AUC over 81.9%.

5.2. Results of the logit model

The model output is presented in terms of the coefficients, their standard errors, the z-statistic in Table 4, and the resulting odds ratios have also been included in the same table. All the mathematical interpretations are RwD crashes on R2L highways for each selected variable class type in reference to preselected reference types. Despite having a very low frequency in Table 2, some variable classes might show high affinity to RwD crashes.

Key takeaways from the odds ratio interpretations are described below:

- *Time:* The higher than 1 odds-ratios during both 6 pm – 12 am and 12 am – 6 am indicate a higher likelihood of RwD crashes during nighttime. Visibility during nighttime has been documented as a key factor that distinguishes RwD crashes from non-RwD crashes (Hashemi and Archilla, 2017; Lord et al., 2011) and nighttime retro-reflectivity in signs and markings has been continually recognized as a common recommendation for preventing RwD crashes based on the prevalence of RwD crashes occurring during nighttime (Jalayer et al., 2015; Julian, 2013; McGee, 2018; Nambison and Hallmark, 2011).
- *Day of the week:* The phenomenon of higher likelihood during weekends is often connected to drivers' higher tendency of alcohol consumption and subsequently losing control of vehicles as alcohol-related crashes are known to occur more during weekends (NHTSA, 2017).
- *Lighting condition:* The absence of daylight increases the likelihood of RwD crashes with relatively reduced crash likelihood during dusk or dawn or dark with streetlight. Rural roadway streetlighting is primarily recommended at intersections, railroad crossings, bridges, and sharp curves (AASHTO, 2018), 'presence of lighting facility' is nonetheless expected to prevent the risk of RwD crashes.
- *Weather condition:* Difficult maneuvering of vehicles on wet or slippery pavement surface due to these adverse weather conditions possibly contributed to roadway departures. The results could significantly vary in other locations and could well be higher especially in the territories with very high intensity of snows and sleets. Variable speed limits have often been recommended in similar extreme weather conditions (Al-Kaisy et al., 2012; Katz et al., 2017).
- *Roadway condition:* In addition to colliding with animals, a driver gets surprised by the unexpectedly animal(s) running across the road, specifically in rural areas, and loses control of the vehicle causing a roadway departure. However, the risk could possibly be attributed to the population of wildlife near the R2L highways and speeding (Huijser et al., 2004), animal movements during nighttime (Khattak, 2000). Poor roadway travel lane conditions loose material, holes/deep ruts, bumps, shoulder abnormalities (edge drop-offs, holes, or ruts on the shoulder) also possess a considerable risk for vehicles

Table 3
Checking multicollinearity with generalized inflation variance factors.

Variables	Df	GVIF	$GVIF^{\frac{1}{2\times Df}}$	Variables	Df	GVIF	$GVIF^{\frac{1}{2\times Df}}$
Time	3	2.679	1.179	Posted speed limit	5	1.160	1.015
Day of week	1	1.034	1.017	Driver age	8	5.560	1.113
Lighting condition	5	2.764	1.107	Driver gender	2	5.810	1.553
Weather condition	5	1.160	1.015	Driver license state	2	2.234	1.223
Roadway condition	9	1.152	1.008	Driver distracted	5	1.173	1.016
AADT	5	1.218	1.020	Alcohol	1	1.099	1.048
Lane width	3	1.122	1.019	Drugs	1	1.033	1.016
Shoulder width	3	1.177	1.028	Passenger presence	1	1.049	1.024
Curve radius	6	1.061	1.005	Vehicle type	5	1.464	1.039
Vertical alignment	5	1.039	1.004	Vehicle year	2	2.787	1.292

Table 4
Logistic regression model results.

Variables	Estimate	S.E.	z-value	Odds ratio (OR)	Variables	Estimate	S.E.	z-value	Odds ratio (OR)
(Intercept)	0.427	0.03	14.207	1.533	**Vertical alignment (Ref: level)**				
Time (Ref: 6am-12 pm)					Level-elevated	0.472	0.045	10.443	1.604
12 am–6 am	0.972	0.038	25.317	2.643	On grade	0.615	0.035	17.612	1.85
12 pm–6 pm	−0.269	0.018	−15.045	0.764	**Posted speed limit (Ref: 50 to 55 mph)**				
6 pm–12 am	0.075	0.029	2.613	1.078	25 mph or less	−1.69	0.051	−33.011	0.184
Day of week (Ref: weekday)					30 to 35 mph	−1.343	0.034	−39.184	0.261
Weekend	0.301	0.018	17.166	1.351	40 to 45 mph	−0.776	0.021	−37.426	0.46
Lighting condition (Ref: daylight)					60 mph or more	−0.379	0.087	−4.356	0.684
Dark (no streetlight)	0.801	0.029	27.609	2.229	Unknown	−1.248	0.035	−35.914	0.287
Dark (streetlight)	0.177	0.049	3.613	1.194	**Driver age (Ref: 25-34y)**				
Dusk/dawn	0.228	0.04	5.647	1.256	15–19y	−0.135	0.025	−5.408	0.874
Unknown	0.228	0.055	4.133	1.256	55–64y	−0.075	0.03	−2.508	0.927
Weather (Ref: clear)					65–74y	−0.227	0.038	−5.923	0.797
Cloudy	0.249	0.02	12.146	1.282	≥75y	−0.479	0.047	−10.292	0.619
Rain	0.647	0.026	24.967	1.91	**Driver gender (Ref: male)**				
Snow/sleet/hail	1.737	0.173	10.067	5.681	Female	−0.051	0.018	−2.928	0.95
Unknown	0.697	0.202	3.459	2.008	Unknown	0.697	0.086	8.064	2.008
Roadway condition (Ref: no abnormalities)					**Driver license state (Ref: Louisiana)**				
Animal	2.844	0.082	34.875	17.187	Out of state	0.086	0.029	2.932	1.089
Construction	−0.168	0.071	−2.352	0.845	unknown	0.235	0.049	4.807	1.265
Holes/deep ruts	0.66	0.203	3.252	1.935	**Driver distraction (Ref: not distracted)**				
Loose material	1.118	0.199	5.611	3.06	Cellphone	0.424	0.057	7.389	1.527
Previous crash	−2.241	0.177	−12.66	0.106	Inside vehicle	0.081	0.032	2.515	1.084
Shoulder abnormality	0.685	0.133	5.156	1.983	Outside vehicle	−0.392	0.04	−9.838	0.676
Water	0.893	0.077	11.675	2.443	Unknown	−0.042	0.019	−2.268	0.959
Other	0.315	0.056	5.608	1.371	**Alcohol (Ref: no)**				
AADT (Ref: 1,001–5,000)					Yes	0.76	0.034	22.621	2.138
400 or less	0.737	0.058	12.81	2.09	**Drugs (Ref: no)**				
401–1,000	0.582	0.03	19.2	1.789	Yes	1.093	0.061	17.895	2.984
5,001–10,000	−0.596	0.018	−32.607	0.551	**Passenger presence (Ref: no)**				
10,001–20,000	−1.069	0.03	−35.641	0.343	Yes	−0.125	0.017	−7.162	0.883
>20,000	−1.364	0.161	−8.494	0.256	**Vehicle type (Ref: car/SUV/van)**				
Lane width (Ref: 12ft or greater)					Pickup truck	−0.224	0.018	−12.681	0.8
Less than 10ft.	0.368	0.124	2.96	1.445	Bus	−0.976	0.16	−6.081	0.377
10 ft. to less than 11 ft.	0.22	0.021	10.325	1.247	Large truck	−0.123	0.034	−3.588	0.884
11 ft. to less than 12 ft.	0.158	0.017	9.366	1.171	Motorcycle	0.984	0.069	14.271	2.675
Shoulder width (Ref: greater than 6ft)					Others	−0.471	0.039	−11.977	0.624
2 ft. or less	0.4	0.024	16.658	1.492	**Vehicle year (Ref: before 2012)**				
Greater than 2 ft. to 4 ft.	0.312	0.019	16.097	1.366	2012 or later	−0.191	0.029	−6.621	0.826
Greater than 4 ft. to 6 ft.	0.241	0.022	11.163	1.273	Unknown	−0.784	0.073	−10.717	0.457
Curve radius (Ref: tangent)									
500 ft. or less	0.545	0.042	13.081	1.725	**Summary statistics**				
501–1,000 ft.	0.957	0.039	24.849	2.604	AIC	110,988			
1,001–2,500 ft.	0.539	0.025	21.938	1.715	Log-likelihood	−55,415.89			
2,501–5,000 ft.	0.328	0.029	11.189	1.388	McFadden's R^2	0.237			
5,001–10,000 ft.	0.159	0.033	4.883	1.173					

to force into an RwD crash. Water on roadway describes a condition with a measurable amount of standing or running water located on the roadway that might have contributed to the crash and also presents a high relative risk to RwD crashes.

- *AADT:* Odds of having an RwD crash turned out to be lower compared to non-RwD crash with higher AADT. This result was also supported by Zhu et al. (2010) study – where RwD crashes in the analysis of R2L fatal crash data from four states (Alabama, Georgia, Mississippi, South Carolina) were found to be negatively associated with AADT as a continuous variable.
- *Lane width and shoulder width:* Narrower the R2L lane width and shoulder width the higher the odds of RwD crashes. Similar findings were obtained for lane width (Hallmark et al., 2011), shoulder width (Kweon and Lim, 2019), both separately (Lord et al., 2011) and as a combination (Stein and Neuman, 2007). Significant safety improvements due to lane width increase (Dell'Acqua and Russo, 2011; Hauer, 2007) and shoulder width increase (Gitelman et al., 2019; Peng et al., 2012) on R2L have been found.
- *Horizontal and vertical alignment:* The highest odds of RwD crashes were on R2L highways with a radius of 501–1,000 ft. with respect to tangents. A possible explanation could be other factors associated with RwD on a curve – approach tangent length, speed limit, available sight, radii of adjacent curves, vertical grade, etc. and any geometric properties that influence drivers speed choice (Donnell et al., 2019). In terms of vertical alignment, the odds of an RwD crash was found to be significant on level-elevated and was on-grade R2L segment, where and on-grade segment indicates roadway going up or down a hill or bridge approach. Countermeasures such as in-lane curve warning pavement markings (Donnell et al., 2019), rumble strips (Galgamuwa and Dissanayake, 2019; Torbic, 2009) have been proven effective in reducing run-off road crashes on R2L.
- *Speed limit:* Interestingly, the highest odds for a RwD crash was on R2L segments with a speed limit of 50–55 mph among the known five speed limit groups. Previous studies have dichotomized speed limits and found strong associations of RwD crashes on roadways greater than 35 mph (Hashemi and Archilla, 2017) and greater than 60 mph (Cejun and Subramanian, 2009).
- *Driver age and gender:* Generally, younger drivers are often known to be involved with aggressive driving behaviors such as speeding, alcohol impairment, etc., which possibly contribute to higher odds of RwD crashes. In this analysis, drivers aged 25–34 years were found more prone to RwD crashes on R2L highways than other age groups. The estimated odds ratio for males to be involved in an RwD crash with respect to females is slightly higher, which could be attributable to the fact that vehicle miles driven by male drivers are higher than vehicle miles driven by female drivers.
- *Driver license state:* Roadway familiarity of the driver is an influential factor to driver's driving performance and relative involvement could be different for different crash types (Intini et al., 2019, 2018). In our study on R2L, out-of-state drivers have 9% higher odds of having an RwD crash compared to in-state drivers. This odds ratio has been estimated to be up to two times in two separate studies (Harootunian et al., 2014, 2013).
- *Distraction:* Distracted driving is among the most predominant driving behaviors that are often characterized along with RwD crashes, as a majority of ROR crashes or near-crash events have been attributed to different forms of distraction (Shane et al., 2009). The estimated odds of distraction due to any cellphone use (talking, manipulating, navigating, etc.) with RwD crashes was the highest in our analysis, followed by other in-vehicle distractions (e.g. eating, smoking, reading, etc.), whereas distractions that occurred due to outside sources were found to be 1.5 times more associated with non-RwD crashes.
- *Alcohol and drugs:* Intoxication with alcohol or drugs may diminish cognitive abilities such as driver's concentration, reaction time, and consequently the ability to control vehicle maneuvering for a certain period of time – posing a high risk of an RwD crash. Cejun and Subramanian (2009) estimated higher SVROR fatal crash risk due to alcohol and Romano and Pollini (2014) estimated high odds of being fatally injured in general due to being intoxicated simultaneously by alcohol with drugs. Road departure safety has also been improved by alcohol and drug education and enforcement (FHWA, 2017b).
- *Passenger presence:* Passenger presence and their behavior and interaction with drivers have been conventionally interpreted to have an adverse effect on driving, leading to an RwD crash. However, studies have also suggested the presence of passengers could influence all drivers to be more cautious (Orsi et al., 2013; Vollrath et al., 2002). Older drivers with passengers were found to be safer during nighttime (Hing et al., 2003), and younger drivers were also reported to be in reduced risk due to assistance with guidance for directions (McDonald and Sommers, 2016). However, in this study, the RwD crash likelihood on R2L highways with the presence of passenger(s) was found as lower than the absence of passengers.
- *Vehicle type:* Among all the vehicle groups, only motorcycles were found to have more association with the RwD crash than passenger car/van/SUV, whereas all other vehicle groups showed a lower likelihood of RwD crashes. Interestingly enough, according to the Ohio study, motorcycles were more associated with basically no-injury ROR crashes, with a higher likelihood of injury ROR crashes on roadways with a posted speed limit under 40 mph (Alruwaished, 2014).
- *Vehicle year:* Recommended by U.S. government's revamped 5-star safety ratings program in 2012 (NHTSA, 2012), newer vehicle models have been equipped and pre-installed with lane departure warning and lane keeping assistance which have been abundantly proven to have positive effects on preventing RwD crashes, e.g. Cicchino (2018) and Sternlund et al. (2017). In our analysis, vehicles of 2012 or later (newer models) were estimated to have lower odds of having RwD crashes than vehicles manufactured the year before 2012.

5.3. Results of association rule mining

5.3.1. Top contributing factors from logit model

As the estimated odds ratios listed in Table 4 present high likelihood (odds ratio > 1) of a number of characteristics with statistical significance (p-value < 0.05) in comparison to perceived normal and most frequent scenarios, the predominant contributing factors were identified by ordered descending odds ratios for RwD crashes. Any 'unknown' factors were discarded while choosing from the numerical hierarchy of odds ratios ordered ascendingly. Inverse odds ratios for reference variable class were also considered if the odds ratio was significant. From the logit model with the original crash dataset, the hierarchically selected ten contributing factors are presented in Fig. 2.

The odds of a roadway departure due to animal presence was found to be very high (17.187) compared to normal roadway conditions. No studies were found to have directly estimated the risk of roadway departure due to presence or collision with animals on roadway. Snow/sleet/hail had the highest odds (5.681) to be associated with RwD crashes on R2L highways despite having a very low frequency (Table 2). Other factors related to drivers (drugs, alcohol), roadway characteristics (speed limit of 50–55 mph and AADT of 1,001–5,000 vehicles per day, curve radius – 501–1,000 ft.), crash environment (time: 12 am – 6 am, darkness without streetlight) appear to have relatively high odds ratios. It is important to understand how multiple factors with higher likelihood are associatively involved with RwD crash scenarios.

5.3.2. Apriori rule results

The results were grouped and presented in three different tables according to the categorization of ten selected contributing factors in the consequents – crash environment (Table 5), roadway characteristics (Table 6), vehicle with driver at-fault in crashes (Table 7). A serial number was provided to the rules for identification and explanation. For example, rule 1 (in Table 4) is {lighting_condition = dark_(no_street_light)} ⇒ {roadway_condition = animal} [S = 0.08, C = 0.18, L = 1.689]. It indicates that the likelihood of a crash in presence of animals on the roadway during dark with no lighting is likely to happen 1.689 times than all dark with no lighting crashes. For each itemset groups, the rules are sorted based on the lift measures.

With the addition of new itemsets, more crash scenarios are created – of which a large number often involves most frequent and normal conditions. The interpretability of these rules relies researcher's requirement based on the selection of threshold of support, confidence, and lift values. Among the top ten contributing factors used as consequents, the ones with a high percentage in the RwD dataset (e.g. speed limit = 50 to 55 mph, AADT = 1,001–5,000, etc.) ended up generating rules with other most frequent items. Based on the results of rule mining, several key observations on *RwD crashes on R2L roads in Louisiana* can be made are –

- The most noteworthy association of RwD crashes due to animals on the roadway is the absence of streetlight in darkness (rule 1, rule 6–15).
- The weather condition 'snow/sleet/hail' have higher odds of RwD crashes in the logit model; however, due to their small share in the original dataset (Table 2), rules for generating associative factors required lower values of minimum support and minimum confidence. Factors present RwD crashes during snow/sleet/hail are the most frequent and normal condition in general weekday, daylight, no driver distraction, no alcohol, AADT of 1,001–5,000 vehicles per day (vpd), etc. Variable speed limits might be useful in improving safety during inclement weather (Al-Kaisy et al., 2012; Katz et al., 2017).
- Male drivers have been significantly linked with RwD crashes in the logit model, as previously discussed. The ARM analysis furthermore revealed that RwD crashes during 12 am to 6 am can be strongly associated with male drivers (rule 32), especially during the absence of streetlight (rule 36) and conjointly with no presence of passengers (rule 41).
- RwD crashes in absence of streetlight are also likely to occur during 6 pm – 12 am (rule 46) and also on R2L roads with a speed limit between 50–55 mph (rule 47).

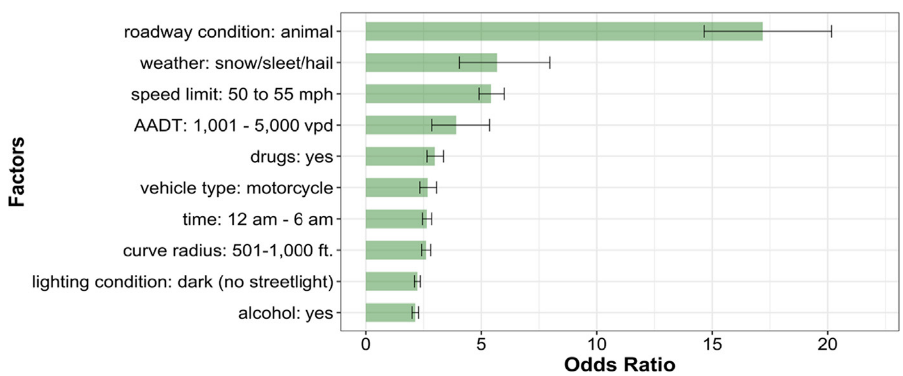

Fig. 2. Selected 10 contributing factors ordered by odds ratios.

Table 5
Top 5 of 2-, 3-, and 4-itemset rules of associated factors of RwD crash environment contributing factors.

	S/N	Antecedent(s)	*s*	*c*	*l*	Count
Consequent: {roadway_condition = animal}, s ≥ 0.05, c ≥ 0.05, l > 1						
2-itemset (total rules = 11)	1	{lighting_condition = dark_(no_street_light)}	0.080	0.184	1.689	7,143
	2	{driver_distraction = no}	0.100	0.151	1.386	8,847
	3	{curve_radius = tangent}	0.075	0.130	1.191	6,656
	4	{vehicle_type = car_van_SUV}	0.072	0.125	1.147	6,373
	5	{alcohol = no}	0.108	0.124	1.141	9,567
3-itemset (total rules = 79)	6	{driver_distraction = no, lighting_condition = dark_ (no_street_light)}	0.074	0.249	2.288	6,606
	7	{alcohol = no, lighting_condition = dark_(no_street_light)}	0.079	0.227	2.080	7,057
	8	{curve_radius = tangent, lighting_condition = dark_ (no_street_light)}	0.056	0.217	1.994	4,984
	9	{day = weekday, lighting_condition = dark_(no_street_light)}	0.056	0.205	1.887	5,016
	10	{lighting_condition = dark_(no_street_light), vehicle_type = car_van_SUV}	0.053	0.203	1.866	4,677
4-itemset (total rules = 249)	11	{alcohol = no, driver_distraction = no, lighting_condition = dark_(no_street_light)}	0.074	0.302	2.771	6,541
	12	{curve_radius = tangent, driver_distraction = no, lighting_condition = dark_(no_street_light)}	0.052	0.288	2.649	4,627
	13	{driver_distraction = no, lighting_condition = dark_ (no_street_light), weather_condition = clear}	0.057	0.277	2.544	5,026
	14	{day = weekday, driver_distraction = no, lighting_condition = dark_(no_street_light)}	0.052	0.273	2.508	4,653
	15	{driver_distraction = no, lighting_condition = dark_ (no_street_light), speed_limit = 50_to_55mph}	0.069	0.264	2.421	6,106
Consequent: {weather_condition = snow_sleet_hail}, s ≥ 0.0025, c ≥ 0.005, l > 1						
2-itemset (total rules = 12)	16	{day = weekday}	0.005	0.007	1.331	432
	17	{lighting_condition = daylight}	0.003	0.007	1.223	280
	18	{driver_distraction = no}	0.004	0.007	1.212	384
	19	{alcohol = no}	0.005	0.006	1.117	465
	20	{aadt = 1,001–5,000}	0.004	0.006	1.100	331
3-itemset (total rules = 66)	21	{day = weekday, driver_distraction = no}	0.004	0.009	1.607	349
	22	{day = weekday, lighting_condition = daylight}	0.003	0.008	1.570	260
	23	{curve_radius = tangent, day = weekday}	0.003	0.008	1.461	276
	24	{aadt = 1,001–5,000, day = weekday}	0.003	0.008	1.451	294
	25	{alcohol = no, day = weekday}	0.005	0.008	1.442	420
4-itemset (total rules = 167)	26	{aadt = 1,001–5,000, day = weekday, driver_distraction = no}	0.003	0.010	1.823	246
	27	{curve_radius = tangent, day = weekday, driver_distraction = no}	0.003	0.010	1.761	226
	28	{alcohol = no, day = weekday, driver_distraction = no}	0.004	0.009	1.737	342
	29	{day = weekday, driver_distraction = no, passenger_presence = no}	0.003	0.009	1.658	275
	30	{day = weekday, driver_distraction = no, drugs = no}	0.004	0.009	1.655	348
Consequent: {time = 12am-6am}, s ≥ 0.1, c ≥ 0.05, l > 1						
2-itemset (total rules = 9)	31	{lighting_condition = dark_(no_street_light)}	0.189	0.432	1.943	16,793
	32	{driver_gender = male}	0.154	0.244	1.096	13,651
	33	{passenger_presence = no}	0.177	0.234	1.051	15,752
	34	{weather_condition = clear}	0.152	0.228	1.026	13,527
	35	{vertical_alignment = level}	0.197	0.228	1.025	17,456
3-itemset (total rules = 53)	36	{driver_gender = male, lighting_condition = dark_ (no_street_light)}	0.132	0.464	2.086	11,691
	37	{lighting_condition = dark_(no_street_light), passenger_presence = no}	0.151	0.457	2.055	13,407
	38	{lighting_condition = dark_(no_street_light), roadway_condition = no_abnormalities}	0.150	0.454	2.042	13,322
	39	{lighting_condition = dark_(no_street_light), vehicle_year = before_2012}	0.173	0.437	1.965	15,400
	40	{lighting_condition = dark_(no_street_light), vertical_alignment = level}	0.167	0.435	1.953	14,827
4-itemset (total rules = 135)	41	{driver_gender = male, lighting_condition = dark_ (no_street_light), passenger_presence = no}	0.107	0.484	2.175	9,474
	42	{driver_gender = male, lighting_condition = dark_ (no_street_light), roadway_condition = no_abnormalities}	0.104	0.475	2.134	9,256
	43	{lighting_condition = dark_(no_street_light),	0.120	0.470	2.112	10,679

(*continued on next page*)

Table 5 (*continued*)

	S/N	Antecedent(s)	*s*	*c*	*l*	Count
		passenger_presence = no, roadway_condition = no_abnormalities}				
	44	{driver_gender = male, lighting_condition = dark_(no_street_light), speed_limit = 50_to_55mph}	0.114	0.467	2.098	10,125
	45	{driver_gender = male, lighting_condition = dark_(no_street_light), vertical_alignment = level}	0.116	0.466	2.096	10,285
Consequent: {lighting_condition = dark_(no_street_light)}, s ≥ 0.2, c ≥ 0.05, l > 1						
2-itemset (total rules = 8)	46	{time = 6 pm-12am}	0.219	0.769	1.759	19,401
	47	{speed_limit = 50_to_55mph}	0.376	0.459	1.050	33,400
	48	{driver_distraction = no}	0.299	0.452	1.034	26,513
	49	{vehicle_type = car_van_SUV}	0.259	0.451	1.031	23,016
	50	{driver_gender = male}	0.284	0.450	1.029	25,185
3-itemset (total rules = 42)	51	{drugs = no, time = 6 pm-12am}	0.210	0.772	1.765	18,639
	52	{driver_gender = male, speed_limit = 50_to_55mph}	0.244	0.473	1.081	21,680
	53	{speed_limit = 50_to_55mph, weather_condition = clear}	0.259	0.473	1.080	23,016
	54	{speed_limit = 50_to_55mph, vehicle_type = car_van_SUV}	0.224	0.472	1.080	19,879
	55	{driver_distraction = no, speed_limit = 50_to_55mph}	0.261	0.472	1.078	23,160
4-itemset (total rules = 78)	56	{speed_limit = 50_to_55mph, vertical_alignment = level, weather_condition = clear}	0.231	0.481	1.100	20,543
	57	{driver_distraction = no, speed_limit = 50_to_55mph, vertical_alignment = level}	0.230	0.481	1.100	20,384
	58	{driver_gender = male, speed_limit = 50_to_55mph, vertical_alignment = level}	0.214	0.481	1.099	19,018
	59	{driver_gender = male, driver_license_state = Louisiana, speed_limit = 50_to_55mph}	0.219	0.476	1.089	19,430
	60	{drugs = no, speed_limit = 50_to_55mph, weather_condition = clear}	0.248	0.473	1.082	22,025

Table 6
Top 5 of 2-, 3-, and 4-itemset rules of associated factors of RwD crash roadway contributing factors.

	S/N	Antecedent(s)	*s*	*c*	*l*	Count
	Consequent: {speed_limit = 50_to_55mph}, s ≥ 0.4, c ≥ 0.05, l > 1					
2-itemset (total rules = 9)	61	{curve_radius = tangent}	0.487	0.843	1.028	43,234
	62	{driver_distraction = no}	0.553	0.838	1.022	49,110
	63	{driver_license_state = Louisiana}	0.717	0.825	1.007	63,654
	64	{aadt = 1,001–5,000}	0.517	0.825	1.006	45,930
	65	{vehicle_type = car_van_SUV}	0.474	0.824	1.006	42,072
3-itemset (total rules = 46)	66	{curve_radius = tangent, driver_license_state = Louisiana}	0.427	0.848	1.035	37,932
	67	{curve_radius = tangent, vertical_alignment = level}	0.433	0.846	1.032	38,412
	68	{curve_radius = tangent, vehicle_year = before_2012}	0.442	0.845	1.031	39,279
	69	{curve_radius = tangent, drugs = no}	0.469	0.842	1.027	41,623
	70	{alcohol = no, curve_radius = tangent}	0.425	0.841	1.026	37,772
4-itemset (total rules = 63)	71	{curve_radius = tangent, driver_license_state = Louisiana, drugs = no}	0.410	0.847	1.034	36,442
	72	{curve_radius = tangent, drugs = no, vertical_alignment = level}	0.416	0.845	1.031	36,959
	73	{curve_radius = tangent, drugs = no, vehicle_year = before_2012}	0.425	0.844	1.030	37,780
	74	{alcohol = no, curve_radius = tangent, drugs = no}	0.417	0.841	1.026	37,004
	75	{alcohol = no, driver_distraction = no, driver_license_state = Louisiana}	0.439	0.840	1.025	38,961
	Consequent: {aadt = 1,001–5,000}, s ≥ 0.3, c ≥ 0.05, l > 1					
2-itemset (total rules = 9)	76	{vehicle_type = car_van_SUV}	0.363	0.631	1.007	32,219
	77	{speed_limit = 50_to_55mph}	0.517	0.631	1.006	45,930
	78	{vehicle_year = before_2012}	0.572	0.629	1.003	50,784
	79	{lighting_condition = daylight}	0.300	0.629	1.003	26,650
	80	{alcohol = no}	0.545	0.628	1.001	48,363
3-itemset (total rules = 38)	81	{speed_limit = 50_to_55mph, vehicle_type = car_van_SUV}	0.301	0.635	1.013	26,730
	82	{speed_limit = 50_to_55mph, vehicle_year = before_2012}	0.473	0.634	1.010	42,017
	83	{vehicle_type = car_van_SUV, vehicle_year = before_2012}	0.332	0.633	1.010	29,489
	84	{alcohol = no, speed_limit = 50_to_55mph}	0.448	0.632	1.008	39,819
	85	{passenger_presence = no, speed_limit = 50_to_55mph}	0.392	0.632	1.008	34,810

Table 6 (*continued*)

	S/N	Antecedent(s)	*s*	*c*	*l*	Count
4-itemset (total rules = 75)	86	{alcohol = no, speed_limit = 50_to_55mph, vehicle_year = before_2012}	0.408	0.635	1.012	36,220
	87	{passenger_presence = no, speed_limit = 50_to_55mph, vehicle_year = before_2012}	0.356	0.634	1.011	31,653
	88	{roadway_condition = no_abnormalities, speed_limit = 50_to_55mph, vehicle_year = before_2012}	0.393	0.634	1.010	34,928
	89	{drugs = no, speed_limit = 50_to_55mph, vehicle_year = before_2012}	0.455	0.634	1.010	40,367
	90	{driver_gender = male, speed_limit = 50_to_55mph, vehicle_year = before_2012}	0.304	0.633	1.010	26,966
	Consequent: {curve_radius = 501_to_1000ft.}, s ≥ 0.04, c ≥ 0.05, l > 1					
2-itemset (total rules = 6)	91	{roadway_condition = no_abnormalities}	0.071	0.085	1.097	6,270
	92	{aadt = 1,001–5,000}	0.051	0.081	1.049	4,533
	93	{passenger_presence = no}	0.061	0.080	1.032	5,390
	94	{vehicle_year = before_2012}	0.072	0.079	1.017	6,366
	95	{driver_license_state = Louisiana}	0.068	0.078	1.005	6,014
3-itemset (total rules = 27)	96	{aadt = 1,001–5,000, roadway_condition = no_abnormalities}	0.047	0.089	1.153	4,135
	97	{passenger_presence = no, roadway_condition = no_abnormalities}	0.056	0.087	1.123	4,931
	98	{roadway_condition = no_abnormalities, vehicle_year = before_2012}	0.065	0.086	1.111	5,808
	99	{driver_license_state = Louisiana, roadway_condition = no_abnormalities}	0.061	0.086	1.108	5,441
	100	{drugs = no, roadway_condition = no_abnormalities}	0.068	0.085	1.098	6,001
4-itemset (total rules = 48)	101	{aadt = 1, 001–5, 000, roadway_condition = no_abnormalities, vehicle_year = before_2012}	0.043	0.090	1.166	3,834
	102	{aadt = 1, 001–5, 000, driver_license_state = Louisiana, roadway_condition = no_abnormalities}	0.040	0.090	1.162	3,577
	103	{aadt = 1, 001–5, 000, drugs = no, roadway_condition = no_abnormalities}	0.044	0.089	1.152	3,950
	104	{driver_license_state = Louisiana, passenger_presence = no, roadway_condition = no_abnormalities}	0.048	0.088	1.137	4,267
	105	{passenger_presence = no, roadway_condition = no_abnormalities, vehicle_year = before_2012}	0.051	0.088	1.135	4,523

Table 7
Top 5 of 2-, 3-, and 4-itemset rules of associated factors of RwD crash vehicle and driver contributing factors.

	S/N	Antecedent(s)	*s*	*c*	*l*	Count
	Consequent: {drugs = yes}, s ≥ 0.02, c ≥ 0.03, l > 1					
2-itemset (total rules = 10)	106	{alcohol = yes}	0.021	0.155	4.162	1,824
	107	{roadway_condition = no_abnormalities}	0.036	0.044	1.180	3,235
	108	{weather_condition = clear}	0.029	0.043	1.159	2,555
	109	{driver_gender = male}	0.025	0.040	1.087	2,264
	110	{driver_license_state = Louisiana}	0.034	0.040	1.062	3,047
3-itemset (total rules = 40)	111	{alcohol = yes, roadway_condition = no_abnormalities}	0.020	0.156	4.196	1,784
	112	{roadway_condition = no_abnormalities, weather_condition = clear}	0.028	0.050	1.348	2,510
	113	{driver_license_state = Louisiana, roadway_condition = no_abnormalities}	0.034	0.047	1.268	2,986
	114	{driver_gender = male, roadway_condition = no_abnormalities}	0.025	0.047	1.268	2,221
	115	{roadway_condition = no_abnormalities, vehicle_type = car_van_SUV}	0.022	0.047	1.262	1,953
4-itemset (total rules = 55)	116	{driver_license_state = Louisiana, roadway_condition = no_abnormalities, weather_condition = clear}	0.026	0.054	1.454	2,322
	117	{roadway_condition = no_abnormalities, speed_limit = 50_to_55mph, weather_condition = clear}	0.024	0.052	1.408	2,138
	118	{roadway_condition = no_abnormalities, vertical_alignment = level, weather_condition = clear}	0.025	0.052	1.385	2,249
	119	{roadway_condition = no_abnormalities, vehicle_year = before_2012, weather_condition = clear}	0.026	0.051	1.380	2,340
	120	{passenger_presence = no, roadway_condition = no_abnormalities, weather_condition = clear}	0.022	0.051	1.368	1,964
	Consequent: {vehicle_type = motorcycle}, s ≥ 0.02, c ≥ 0.03, l > 1					
2-itemset (total rules = 11)	121	{time = 12 pm-6 pm}	0.011	0.042	1.759	975
	122	{day = weekend}	0.012	0.036	1.523	1,035
	123	{driver_gender = male}	0.022	0.035	1.471	1,944
	124	{lighting_condition = daylight}	0.016	0.035	1.465	1,465
	125	{weather_condition = clear}	0.020	0.030	1.256	1,758

(*continued on next page*)

Table 7 (*continued*)

	S/N	Antecedent(s)	*s*	*c*	*l*	Count
3-itemset (total rules = 71)	126	{driver_gender = male, lighting_condition = daylight}	0.015	0.052	2.192	1,339
	127	{day = weekend, driver_gender = male}	0.011	0.050	2.120	937
	128	{lighting_condition = daylight, weather_condition = clear}	0.014	0.045	1.888	1,241
	129	{lighting_condition = daylight, time = 12 pm-6 pm}	0.011	0.044	1.846	947
	130	{driver_gender = male, weather_condition = clear}	0.018	0.043	1.825	1,631
4-itemset (total rules = 237)	131	{driver_gender = male, lighting_condition = daylight, weather_condition = clear}	0.013	0.065	2.759	1,134
	132	{driver_gender = male, lighting_condition = daylight, passenger_presence = no}	0.013	0.057	2.420	1,156
	133	{driver_distraction = no, driver_gender = male, lighting_condition = daylight}	0.011	0.056	2.376	977
	134	{driver_gender = male, driver_license_state = Louisiana, lighting_condition = daylight}	0.014	0.053	2.249	1,210
	135	{alcohol = no, driver_gender = male, lighting_condition = daylight}	0.014	0.053	2.238	1,249
	Consequent: {alcohol = yes}, $s \geq 0.06$, $c \geq 0.05$, $l > 1$					
2-itemset (total rules = 10)	136	{lighting_condition = dark_(no_street_light)}	0.086	0.198	1.489	7,674
	137	{day = weekend}	0.063	0.196	1.475	5,629
	138	{driver_gender = male}	0.105	0.166	1.254	9,314
	139	{roadway_condition = no_abnormalities}	0.129	0.155	1.169	11,429
	140	{weather_condition = clear}	0.099	0.149	1.120	8,809
3-itemset (total rules = 53)	141	{lighting_condition = dark_(no_street_light), roadway_condition = no_abnormalities}	0.083	0.253	1.906	7,413
	142	{driver_gender = male, lighting_condition = dark_(no_street_light)}	0.069	0.243	1.830	6,115
	143	{day = weekend, roadway_condition = no_abnormalities}	0.061	0.223	1.684	5,460
	144	{lighting_condition = dark_(no_street_light), weather_condition = clear}	0.064	0.215	1.617	5,720
	145	{lighting_condition = dark_(no_street_light), vehicle_year = before_2012}	0.082	0.207	1.559	7,285
4-itemset (total rules = 122)	146	{driver_gender = male, lighting_condition = dark_(no_street_light), roadway_condition = no_abnormalities}	0.066	0.303	2.281	5,897
	147	{lighting_condition = dark_(no_street_light), roadway_condition = no_abnormalities, weather_condition = clear}	0.062	0.273	2.054	5,547
	148	{driver_license_state = Louisiana, lighting_condition = dark_(no_street_light), roadway_condition = no_abnormalities}	0.074	0.264	1.993	6,553
	149	{lighting_condition = dark_(no_street_light), roadway_condition = no_abnormalities, vehicle_year = before_2012}	0.079	0.263	1.979	7,040
	150	{lighting_condition = dark_(no_street_light), roadway_condition = no_abnormalities, speed_limit = 50_to_55}	0.073	0.258	1.947	6,456

- The majority of the rules include the most frequent itemsets in the dataset. R2L roads with a speed limit 50–55 mph on a tangent section (rule 67) are common combinations of road characteristics – no driver distraction, no alcohol, no drugs, tangent section, no passengers, clear weather, Louisiana drivers, weekday, no roadway abnormalities, car/van/SUV, etc.
- Motorcycle crashes have higher odds to occur than car/van/SUV according to the logit model, however, association rules show that they are likely to occur at weekends (rule 121) or during 12 pm to 6 pm (rule 122), and understandably involve male drivers (rule 123).
- Association of some itemsets are quite evident – for example, in rule 129, the combination of time 12 pm-6 pm and daylight, lighting condition has presumably added no new knowledge.
- Drivers involved with drug intoxication are also likely to be under the influence of alcohol, the opposite of which is not necessarily prevalent according to the top five rules – as no top rules can be found that associate alcohol involvement include drugs (rule 136–150).
- Drugs and alcohol related RwD crashes also showed association with male drivers (rule 109 and rule 138). Alcohol related RwD crashes showed a strong affinity to occur in absence of streetlight (rule 146–150). Education and enforcement are useful in reducing alcohol impaired crashes (FHWA, 2017b).

6. Conclusions

The aim of the study was to understand the dichotomy between RwD and non-RwD crashes based on the influential contributing factors and subsequently explore the association of key factors. The model developed in this study can be incorporated in transportation safety to identify and predict crashes that have high potential to result in RwD crashes. The interpretation of the RwD related crash, roadway and vehicle characteristics identified from the logit model subsequent discussions linking up with previous literature. A number of characteristics across all variables showed significant individual association with RwD crashes on R2L highways.

From the discussion of logit model results, it is understandable that an RwD crash could be instigated due to a simultaneous presence of two or more contributing factors creating a complicated crash scenario. While the impacts of individual independent factors can be assessed by the logit model, the 'ARM' can potentially be used as a convenient, effective, and intuitive tool for structuring the crash contributing factors to estimate an associative impact on the roadway departure likelihood.

Our study evidently shows that a number of meaningful interesting rules are extractable from roadway departure crash data using unsupervised data mining techniques like ARM. A number of rules in this study were generated to supplement the key contributing factors aiming to shed more light on RwD crashes. For example, we identified from the rules that animal presence coupled with poor visibility due to the absence of streetlights can be strongly associated with a higher likelihood of RwD crashes. Male drivers are involved in RwD crashes during the early morning (12 am to 6 am) especially during the absence of streetlight and with no presence of passengers. Male drivers are also associated with alcohol and drug related RwD crashes. The drivers in an RwD crash while being intoxicated by drugs are also likely to be under the influence of alcohol.

This study approach of using ARM alongside the logit model is expected to help highway safety specialists to draw upon their combined expertise from the analysis results to reach the goal of strategic reduction of RwD crashes through the appropriate selection of countermeasures. To put an instance into perspective – installation of streetlights on the long stretches of R2L sections may not be cost-effective; however, our knowledge from ARM results indicate that providing better nighttime visibility by clearing roadside verges and installing retroreflective warning road signs with explicit animal traffic information, can contribute to reducing the number of RwD crashes resulting from the presence of animals on roads. The current findings are limited to the likelihood of the factors that are associated with the RwD crashes. The proposed countermeasures are targeted towards the reduction of RwD crashes. However, the countermeasures are usually designed for overall (both RwD and non-RwD) crash reduction.

This study is not without limitations. Several other factors have not been considered mainly due to the unavailability of data. There is a lack of comprehensive database on the presence of pavement markings, centerline rumble strips, or shoulder rumble strips on R2L over the analysis period. The research team had to eliminate the possibility of using those characteristics, which could have an influential impact on preventing an RwD crash. Vehicle related information such as 'vehicle component failure', or 'movement due to collision avoidance' could also be helpful. Vehicle operating speed is known to play a very critical role in RwD crashes. However, due to its substantial underreporting, the posted speed limit was used as a substitute. For similar reasons, other important variables like skid number, curve direction, presence of pavement edge drop-off, and qualitative visibility condition could not be used in the analysis. Exploring further relationships among RwD crash factors from various perspectives of road users, crash environment, and roadway features using a non-parametric approach such as ARM explains a wide array of RwD crash scenarios and eventually could help researchers better understand the RwD crash mechanism.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was sponsored by the Louisiana Transportation Research Center (SIO: DOTLT1000295). The authors also would like to thank Mary Kathryn Sevin for her assistance in preparing the manuscript. The authors declare that the contents of this article have not been published previously. All the authors have contributed to the work described, read, and approved the contents for publication in this journal.

References

AASHTO, 2018. A Policy on Geometric Design of Highways and Streets, 7th Edition. American Association of State Highway and Transportation Officials, Washington DC.

AASHTO, 2008. Driving Down Lane-Departure Crashes: A National Priority, April 2008.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Al-Kaisy, A., Ewan, L., Veneziano, D., 2012. Evaluation of a Variable Speed. Salem.

Alruwaished, A.F., 2014. Characteristics of Drivers Who Cause Run-Off- Road-Crashes on Ohio Roadways. University of Dayton.

Avelar, R., Geedipally, S., Das, S., Wu, L., Kutela, B., Lord, D., Tsapakis, I., 2019. Evaluation of Roadside Treatments to Mitigate Roadway Departure Crashes. TxDOT Report.

Bayardo, R.J., Agrawal, R., 1999. Mining the most interesting rules. In: Proc. of the Fifth ACM SIGKDD Int'l Conf on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, United States, San Diego, pp. 145–154. https://doi.org/10.1145/312129.312219.

Cejun, L., Subramanian, R., 2009. Factors Related to Fatal Single-Vehicle Run-Off-Road Crashes. Washington, DC.

Cicchino, J.B., 2018. Effects of lane departure warning on police-reported crash rates. J. Safety Res. 66, 61–70. https://doi.org/10.1016/j.jsr.2018.05.006.

Das, S., Dutta, A., Jalayer, M., Bibeka, A., Wu, L., 2018. Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety. Int. J. Transp. Sci. Technol. 7, 114–123. https://doi.org/10.1016/j.ijtst.2018.02.001.

Das, S., Sun, X., Goel, S., Sun, M., Rahman, M.A., Dutta, A., 2020as. Flooding related Traffic Crashes: Findings from Association Rules. J. Transp. Saf. Secur.

Das, S., Sun, X., Sun, M., 2020b. Rule-based safety prediction models for rural two-lane run-off-road crashes. Int. J. Transp. Sci. Technol.

Dell'Acqua, G., Russo, F., 2011. Safety Performance Functions for Low-Volume Roads, in: 90th Annual Meeting of the Transportation Research Board. Washington DC.

Donnell, E., Porter, R.J., Li, L., Hamilton, I., Himes, S.C., Wood, J., 2019. Reducing Roadway Departure Crashes at Horizontal Curve Sections on Two-Lane Rural Highways. Washington DC.

Federal Highway Administration (FWHA), 2017a. Roadway Departure Safety [WWW Document]. URL https://safety.fhwa.dot.gov/roadway_dept/ (accessed 8.10.18).

Federal Highway Administration (FWHA), 2017b. Oregon Roadway Departure Safety Implementation Plan Update 2009-2015.

FHWA, 2019. Driving FoRRRwD - A Systemic Approach to Reducing Roadway Departures [WWW Document]. Calif. Local Tech. Assist. Progr. URL http://californialtap.org/pdf/EDC_5RuralRwDDeparture_CA_webinar_9_19_ALBIN_.pdf

Fox, J., 2016. Applied Regressions Analysis and Generalized Linear models. SAGE Publications.

Fox, J., Monette, G., 1992. Generalized collinearity diagnostics. J. Am. Stat. Assoc. 87, 178–183. https://doi.org/10.2307/2290467.

Galgamuwa, U., Dissanayake, S., 2019. Evaluation of the safety effectiveness of lane-departure countermeasures on two-lane undivided roadways using crash modification factors. J. Transp. Saf. Secur. 11, 377–397. https://doi.org/10.1080/19439962.2017.1419523.

Gitelman, V., Doveh, E., Carmel, R., Hakkert, S., 2019. The influence of shoulder characteristics on the safety level of two-lane roads: A case-study. Accid. Anal. Prev. 122, 108–118. https://doi.org/10.1016/j.aap.2018.10.003.

Hahsler, M., Grun, B., Hornik, K., 2008. The arules package: mining association rules and frequent itemsets, version 0.6-6.

Hallmark, S., Boyle, L.N., Boyle Ng, L., Qiu, Y., 2011. Assessing the Likelihood of a Lane Departure Event Using Naturalistic Driving Study Data, in: 3rd International Conference on Road Safety and Simulation. Indianapolis, p. 18p.

Harootunian, K., Aultman-hall, L., Lee, B.H.Y., 2013. Assessing the relative crash fault of out-of-state drivers in Vermont. USA. J. Transp. Saf. Secur. 37–41. https://doi.org/10.1080/19439962.2013.856983.

Harootunian, K., Lee, B.H.Y., Aultman-hall, L., 2014. Odds of fault and factors for out-of-state drivers in crashes in four states of the USA. Accid. Anal. Prev. 72, 32–43. https://doi.org/10.1016/j.aap.2014.06.012.

Hashemi, M., Archilla, A.R., 2017. Exploratory Analysis of Roadway Departure Crashes Contributing Factors Based on Classification and Regression Trees, in: ITE Western District Annual Meeting, San Diego. San Diego, pp. 1–8.

Hashemi, M., Archilla, A.R., 2016. Potential Factors Affecting Roadway Departure Crashes in Oahu, Hawaii, in: ITE Western District Annual Meeting, Albuquerque. Albuquerque, p. 10.

Hauer, E., 2007. Lane Width and Safety.

Hing, J.Y.C., Stamatiadis, N., Aultman-Hall, L., 2003. Evaluating the impact of passengers on the safety of older drivers. J. Safety Res. 34, 343–351. https://doi.org/10.1016/j.jsr.2003.09.011.

Huijser, M.P., McGowen, P., Fuller, J., Hardy, A., Kociolek, A., Clevenger, A.P., Smith, D., Ament, R., 2004. Wildlife-Vehicle Collision Reduction Study: Report to Congress. Virginia.

Intini, P., Berloco, N., Colonna, P., Ranieri, V., Ryeng, E., 2018. Exploring the relationships between drivers' familiarity and two-lane rural road accidents. A multi-level study. Accid. Anal. Prev. 111, 280–296. https://doi.org/10.1016/j.aap.2017.11.013.

Intini, P., Colonna, P., Olaussen Ryeng, E., 2019. Route familiarity in road safety: A literature review and an identification proposal. Transp. Res. Part F Traffic Psychol. Behav. 62, 651–671. https://doi.org/10.1016/j.trf.2018.12.020.

Jalayer, M., Zhou, H., Satterfield, C., 2015. Overview of safety countermeasures for roadway departure crashes. ITE J. 86, 39–46.

Julian, F., 2013. Strategic Directions on Roadway Departure Crashes: Supporting the Decade of Action. In: Strategic Directions on Roadway Departure Crashes. Transportation Research Board, Washington, DC, pp. 3–4.

Katz, B., Ma, J., Rigdon, H., Sykes, K., Huang, Z., Raboy, K., 2017. Synthesis of Variable Speed Limit Signs, Fhwa-Hop-17-003. Washington DC.

Khattak, A.J., 2000. Human Fatalities in Animal-Related Highway Crashes 158–166.

Kweon, Y., Lim, I., 2019. Development of Safety Performance Functions for Network Screening of Roadway Departure Crashes in Virginia. Richmond.

LADOTD, 2018. Destination Zero Deaths: Louisiana Strategic Highway Safety Plan. Baton Rouge.

LADOTD, n.d. Louisiana Department of Transportation and Development Highway Crash List [WWW Document]. URL http://www8.dotd.la.gov/crash1/ (accessed 7.5.18).

Lord, D., Brewer, M.A., Fitzpatrick, K., Geedipally, S.R., Peng, Y., 2011. Analysis of Roadway Departure Crashes on Two-Lane Rural Roads in Texas 7.

McDonald, C.C., Sommers, M.S., 2016. "Good passengers and not good passengers:" adolescent drivers' perceptions about inattention and peer passengers. J. Pediatr. Nurs. 31, e375–e382. https://doi.org/10.1016/j.pedn.2016.07.006.

McGee, H.W., 2018. Practices for Preventing Roadway Departures, Practices for Preventing Roadway Departures. https://doi.org/10.17226/25165

Nambison, S., Hallmark, L.S., 2011. Lane Depature Safety Countermeasures: Strategic Action Plan for Iowa Department Of Transportation.

NHTSA, 2017. Traffic Safety Facts: 2017 Data - Alcohol-Impaired Driving. Washington, DC.

NHTSA, 2012. NHTSA Announces Model Year 2012 Vehicles to be Rated Under Government 5-Star Safety Ratings Program 2012 Model Year Vehicle Test List [WWW Document]. Off. Public Aff. U.S. Dep. Transp. URL https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjk8Yu7gKznAhVGQq0KHUlRCgAQFjAAegQIBRAB&url=https%3A%2F%2Fwww.nhtsa.gov%2Fstaticfiles%2Fcommunications%2Fpdf%2Fnhtsa1711.pdf&usg=AOvVaw3m7gf5w7ypvrP6-ueKwnPO (accessed 9.2.19).

Orsi, C., Marchetti, P., Montomoli, C., Morandi, A., 2013. Car crashes: The effect of passenger presence and other factors on driver outcome. Saf. Sci. 57, 35–43. https://doi.org/10.1016/j.ssci.2013.01.017.

Pande, A., Abdel-Aty, M., 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Saf. Sci. 47, 145–154. https://doi.org/10.1016/j.ssci.2007.12.001.

Peng, Y., Geedipally, S., Lord, D., 2012. Effect of roadside features on single-vehicle roadway departure crashes on rural two-lane roads, in: 91th Annual Meeting of the Transportation Research Board. Washington DC. https://doi.org/10.3141/2309-03

Rahman, M.A., Sun, X., Sun, M., Shan, D., 2020. Investigating Characteristics of Cellphone Distraction with Significance Tests and Association Rule Mining. IATSS Res. https://doi.org/10.1016/j.iatssr.2020.09.001.

R Development Core Team, 2019. R: A Language and Environment for Statistical Computing.

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., 2019. Support Functions and Datasets for Venables and Ripley's MASS.

Romano, E., Pollini, R.A., 2014. Patterns of drug use in fatal crashes. Natl. Inst. Heal. Addict. 108, 1428–1438. https://doi.org/10.1111/add.12180.

Shane, B.M., Jonathan, M.H., Sheila, G.K., Thomas, A.D., 2009. Contributing Factors to Run-Off-Road Crashes and Near-Crashes. Washington, DC.

StateFarm, 2019. Herd Mentality: Avoiding Animal Collisions [WWW Document]. StateFarm Newsroom.

Stein, W.J., Neuman, T.R., 2007. Mitigation Strategies for Design Exceptions. Washington DC.

Sternlund, S., Strandroth, J., Rizzi, M., Lie, A., Tingvall, C., 2017. The effectiveness of lane departure warning systems—A reduction in real-world passenger car injury crashes. Traffic Inj. Prev. 18, 225–229. https://doi.org/10.1080/15389588.2016.1230672.

Torbic, D.J., 2009. NCHRP Report 641 – Guidance for the Design and Application of Shoulder and Centerline Rumble Strips. Washington DC.

Vollrath, M., Meilinger, T., KrÜger, H.P., 2002. How the presence of passengers influences the risk of a collision with another vehicle. Accid. Anal. Prev. 34, 649–654. https://doi.org/10.1016/S0001-4575(01)00064-1.

Weng, J., Zhu, J.Z., Yan, X., Liu, Z., 2016. Investigation of work zone crash casualty patterns using association rules. Accid. Anal. Prev. 92, 43–52. https://doi.org/10.1016/j.aap.2016.03.017.

Zhu, H., Dixon, K.K., Washington, S., Jared, D.M., 2010. Predicting single-vehicle fatal crashes for two-lane rural highways in Southeastern United States. Transp. Res. Rec. 88–96. https://doi.org/10.3141/2147-11.