TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–14

DOI: 10.1177/0361198118822817
journals.sagepub.com/home/trr
SAGE

Crash Risk Evaluation and Crash Severity Pattern Analysis for Different Types of Urban Junctions: Fault Tree Analysis and Association Rules Approaches

Peijie Wu[1], Xianghai Meng[1], Li Song[1], and Wenze Zuo[1]

Abstract

Urban junctions usually present significant safety concerns, and the majority of all crashes in urban areas occur in or near junctions. Factors contributing to crash severity at junctions have been explored, but crash risk levels and crash severity patterns of different junction types have hardly been investigated. In order to fill this gap, this study analyzed the safety performance of six junction types and the factors contributing to crash severity, in order to assist city transportation authorities to implement effective countermeasures. Fault tree analysis (FTA) was applied for the risk evaluation of urban junctions and association rules (AR) algorithm was employed for the crash severity pattern analysis based on data from the U.K. STATS19 database from 2012 to 2016. Overall, four types of urban junctions with high crash risk level and over 4,000 AR contributing to crash severity are identified in the present paper. The results show that: (a) roundabouts and mini-roundabouts have the lowest fatality and casualty rates while T-junctions or staggered junctions and crossroads have the highest crash risk levels; (b) FTA may produce inaccurate outcomes because of incorrect logic gates, but AR can generate real potential relationships between crash severity and risk factors; (c) crash severity patterns are quite complex and the interdependence between risk factors is different for each junction type; (d) risk factors such as male driver, no physical crossing facilities within 50 meters, and give way or uncontrolled junction are common in high-risk junctions at night.

Overall road accident rates have been falling in the U.K. in recent years due to the introduction of a national road safety strategy (*1*). To reduce traffic casualties further, transportation authorities maintain budgets for road safety improvements, and prioritize spending on locations with high crash risk, such as urban junctions. Previous studies have shown that urban junctions usually arouse significant safety concerns, and crashes in urban areas primarily occur in or near junctions (*2*). Yet different types of urban junctions are constituted of different geometric designs, sight distance conditions, and traffic conflict points or angles, leading to different crash rates and crash severity patterns. Accordingly, it is an essential task of traffic safety analysts to analyze the safety performance of junction types in terms of crash severity as well as the factors contributing to high crash risk. Through this task, more insights can be gained into the potential causes of crashes at urban junctions, and effective countermeasures can be taken for different junctions. It is noteworthy that the crash risk is considered to be the product of crash likelihood and crash severity in a risk matrix (RM), which is the core conception of the present study.

Fault tree analysis (FTA) is a well-established technique, which is applied broadly for evaluation of the dependability of a wide range of systems, such as those in the automotive, aerospace, and construction industries (*3–5*). The logical connections between faults and their causes are represented graphically in a fault tree, which shows vividly the process of fault propagation through the whole system. Recently, the basic components of conventional FTA have been transferred from "fault" to "risk," and FTA has been successfully introduced in risk management (*6*, *7*). Nevertheless, FTA still has a deficiency in the manual definition of logic gates between each event, especially in cases where the relationships

[1]School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China

Corresponding Author:
Address correspondence to Xianghai Meng: mengxianghai100@126.com

between risks are unknown at the outset. With the popularity of artificial intelligence algorithms, AR analysis offers a new perspective to explore the potential relationships between crash severity and risk factors and to avoid the bias caused by FTA.

The main aim of the present study is to investigate the crash risk levels of six types of urban junctions and to identify the significant risk factors that contribute to crash severity at high-risk junctions. FTA and AR were compared, and the shortcomings of conventional FTA in risk analysis revealed. The crash severity patterns of high-risk junctions were obtained by AR algorithm. A better understanding of the crash severity patterns of junctions can help not only to facilitate the introduction of suitable countermeasures but also to explore safe driving strategies in an autonomous driving environment. The present paper starts with a literature review, followed by data description, methodology, results and discussion. The paper concludes with an overall summary and recommendations for future research.

Literature Review

In the existing traffic safety literature, the studies that model crash severity and investigate the significant factors of crash severity at urban junctions can be classified into three categories: studies of (a) roundabouts, (b) T-junctions and crossroads, and (c) other models of crash severity.

The first group of studies aim to identify factors that contribute to the crash severity as well as the safety effectiveness of roundabouts. Montello (*8*) investigated the factors contributing to crashes at urban roundabouts in Italy: interdependence between factors was determined using AR algorithm and it was found that the deviation angle, markings, and signs have significant effects on the safety of roundabouts. Daniels et al. (*9*) examined factors contributing to crash severity at roundabouts in Flanders, Belgium. They concluded that severity of injuries increases with the increase of drivers' ages, and that crashes at night and in built-up areas are more severe. Gross et al. and Daniels et al. (*10*, *11*) studied the improvement in safety performance of converting signalized intersections to roundabouts. The results of their studies suggest that roundabouts can significantly reduce both the number and severity of crashes.

The second group of studies, modeling the severity of crashes at T-junctions or crossroads, link injury severity with various risk factors in urban areas. It has been found that motorcyclists are more vulnerable to injuries than other vehicle drivers at T-junctions, and motorists were more likely to violate the right-of-way rules on non-built-up roads and in diminished light conditions (*12*). Nitsche et al. (*13*) identified the critical pre-crash scenarios at T-junctions and crossroads in an automated driving system. They also adopted AR to reveal common crash characteristics. Failure to give way, inappropriate maneuvers, and high speed limits are found in the crash data to be the main precipitating factors. Non-signalized intersections usually have greater crash severity compared with signalized junctions. The crash injury severity at three- and four-legged non-signalized intersections in Florida was analyzed by Haleem and Abdel-Aty (*2*) using multiple methods. That study suggested that traffic volume, number of left-turn movements, and young drivers are strongly associated with fatalities and casualties at non-signalized intersections. Thirdly, other literature has primarily focused on diverse models of crash severity, which include: ordered probit model (*14*), ordered logit model (*15*), mixed logit model (*16*), and single contributory factors, such as age of driver and lighting conditions (*17*, *18*).

The present study contributes in the following four ways. Firstly, it investigates the crash risk levels of six junction types that have seldom been studied and the interdependence between risk factors of crash severity. Secondly, it compares FTA and AR, and points out the shortcomings of FTA in risk management. Inaccurate results of FTA may be generated due to incorrect logic gates when the connection between each risk is initially unknown. Thirdly, this study confirms the advantages of AR analysis, which explores the underlying association between risk factors and their different combination forms in original datasets. Lastly, results yielded from this study provide more insight into the crash severity patterns of high-risk junctions at night, which can help traffic engineers make more effective traffic safety improvements within a limited budget.

Data

The open-access data, named the U.K. National Road Accident Database (STATS19), applied in the present study are provided by the U.K. Department for Transport (*19*). STATS19 refers to the national road crash database of casualties reported by the police on any road in Great Britain. In STATS19, crash severity falls into three levels: fatal, serious, and slight injuries (for detailed definitions of the three crash severity levels, see (*12*)). The database consists of three files compiled annually: accident file, vehicle file, and casualty file. In this study, only accident and vehicle files were connected according to the label "accident_index." The accident file contains the time and date of accident occurrence, weather, light conditions, etc. The vehicle file records vehicle and driver details, including driver's age and gender, vehicle maneuvers, etc.

This study focused on road crashes between two cars (excluding motorcycle, taxi, minibus, bus, and van) at six

Table 1. Summary of STATS19 data, 2012–2016

	Scenario 1: Daytime					Scenario 2: Nighttime				
Junction type	2012	2013	2014	2015	2016	2012	2013	2014	2015	2016
Roundabouts	1334	2329	2419	2432	2455	523	850	767	745	691
Mini-roundabouts	259	412	472	471	464	105	200	185	186	141
T- or staggered junctions	4663	7979	8470	8752	5816	1518	2498	2550	2338	2122
Crossroads	1963	2881	2981	3176	2605	989	1317	1431	1332	1248
More than four arms	267	273	164	240	184	87	92	69	88	87
Other junctions	333	459	500	609	590	87	163	157	160	188
Total	65,952				22,914					

types of urban junctions from 2012 to 2016. Over 80,000 car crashes that took place at urban junctions were extracted from the database, and the total crash data of every junction type during the five years were utilized for the application of FTA and AR, as shown in Table 1.

Methodology

Fault Tree Analysis

FTA was first developed by H.A. Watson and M.A. Mearns at Bell Labs for the analysis of a ballistic missile in the 1960s. The classic handbook on FTA was published by Vesely et al. in 1981 (*20*). FTA serves as a systematic and well-understood method for the estimation of safety and reliability of a complex system, both qualitatively and quantitatively. It is a deductive analysis tool that proceeds graphically from the occurrence of an undesired event to the identification of the root causes of the event (*21*). Generally, a fault tree, virtually a directed acyclic graph, consists of two important components: events and gates. An event is an occurrence within the system, typically the failure of a subsystem down to an individual component, and may be categorized as top event (TE), intermediate event (IE), basic event, and undeveloped event. Gates represent how failures are transferred via the system and how they affect the TE. There are four commonly used gates: AND gates, OR gates, k/N gates and INHIBT gates (*20*). A minimal cut set (MCS) is defined as the minimal set of basic events which jointly are capable of making the TE fail, which gives important information about the vulnerabilities of a system (*22*).

Let the output event of logic gates, designated by X_0, and the input events of logic gates be X_1 to X_N, where N denotes the number of inputs. When the connecting gate is either AND or OR, the probability of the output event, denoted as $P(X_0)$, is given by:

$$P(X_0) = \begin{cases} \prod_{i=1}^{N} P(X_i), \text{for AND gate} \\ 1 - \prod_{i=1}^{N} \{1 - P(X_i)\}, \text{for OR gate} \end{cases} \tag{1}$$

In the present study, conventional FTA was applied in the calculation of the probability of TEs (i.e., fatal crash, serious crash, and slight crash) and the analysis of the MCSs of each crash severity. FTA also permits the theoretical relation between the crash severity (TE), the risk categories (IEs) and the risk factors (basic events) to be clarified on the basis of AND/OR logic, as shown in Figure 1 (assuming that one crash is caused by human, vehicle, road, and environment risk factors together).

Association Rules

AR is one of the most popular data mining techniques, having been first introduced in 1993 for discovering buying patterns (*23*). In recent years, the AR method in data mining has been successfully applied to uncover potential patterns or rules in a variety of fields, such as, market basket analysis (*24*), medical science (*25*), and traffic safety (*8*).

AR analysis is the method of effectively identifying sets of items that occur together in a given event. It is based on the relative frequency of the number of times the sets of items occur alone and jointly in a database. In comparison with FTA, AR is characterized by short operation time in big datasets and exclusive dependence on datasets rather than the subjective evaluation of human researchers. The ARs are formed as "$A \rightarrow B$," where A denotes the antecedent and B denotes the consequent. It is worth mentioning that these rules should be interpreted as associations (i.e., potential risk factors that may cause a crash) between sets of items (i.e., crash data) rather than a direct causation. Three measures for selecting rules in AR analysis are support, confidence, and lift. Support is the percentage of specified cases in the data that contain both A and B. Confidence is the percentage of cases containing A which also contain B. Lift is the ratio of confidence to the percentage of cases containing B. The equations used to calculate these measures are as follows:

$$Support(A \rightarrow B) = P(A \cup B) \tag{2}$$

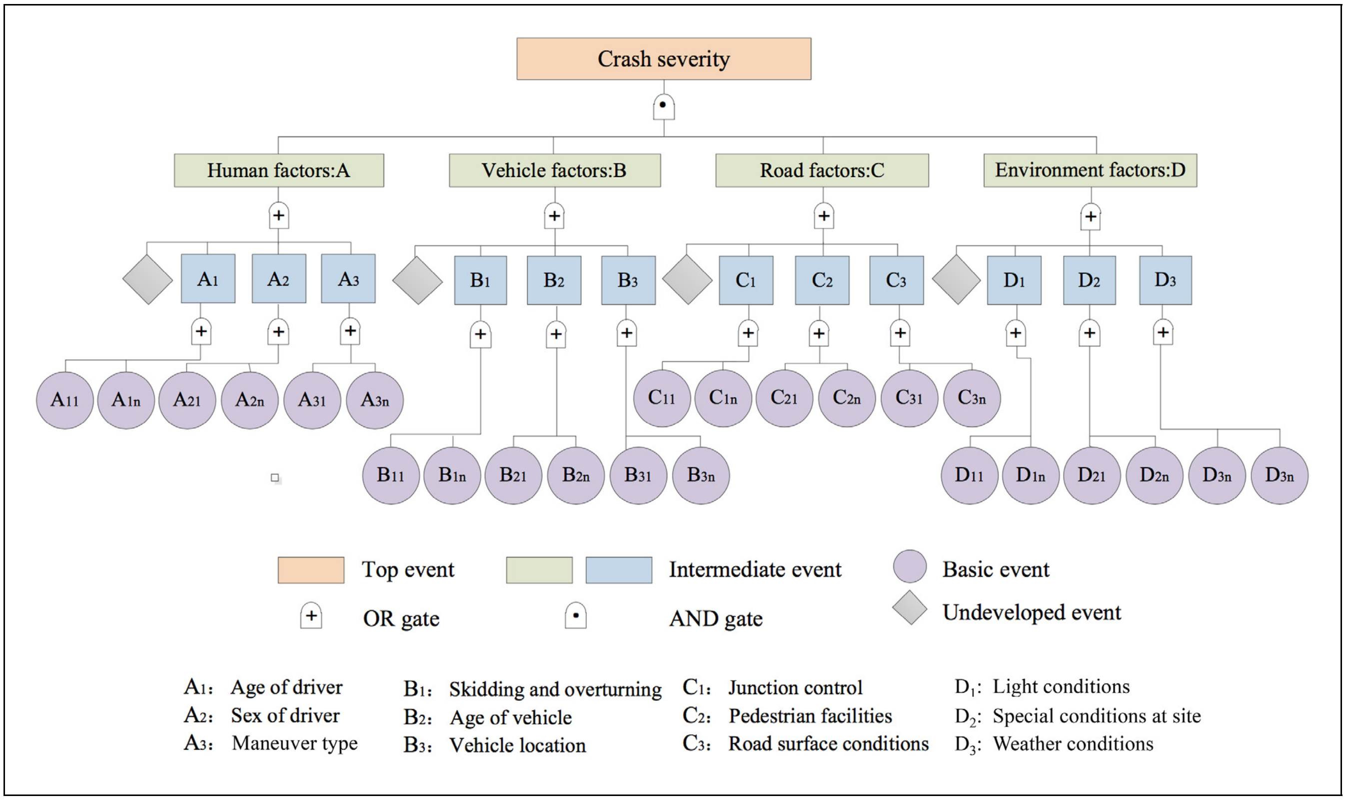

Figure 1. Fault tree structure for crash severity at urban junctions (for details of basic events, see Table 2).

$$Confidence(A \rightarrow B) = P(B|A) \quad (3)$$

$$Lift(A \rightarrow B) = \frac{P(B|A)}{P(B)} = \frac{Confidence(A \rightarrow B)}{P(B)} \quad (4)$$

where $P(\cdot)$ is the probability or percentage of cases.

Apriori is a well-known AR algorithm which finds the most frequent items and extends them to larger and larger item sets, provided the number of appearances is larger than the minimum support value (threshold) (*13*). Thus, in this study, the Apriori algorithm was employed for the analysis of crash severity AR, and the crash data were transferred to Boolean type (i.e., 1 or 0) before loading basic data into SPSS Clementine software (*26*). To identify strong associations, threshold values for support (S), confidence (C), and lift (L) were set as follows: $S \geqslant 30\%$, $C \geqslant 60\%$, and $L \geqslant 1$. Rules with antecedent valued as 0 were eliminated in the present study since they are meaningless for the final results.

Risk Matrix

RM is a table with several categories of likelihood (or frequency) for its rows and several categories of severity (or consequence) for its columns (*27*). By mapping "likelihood" and "severity" ratings to the corresponding risk priority levels, the RM is conducive to assessing risks and setting priorities to address potential hazards. It has been used extensively in many engineering projects. In addition, not all traffic crashes can be eliminated at once, and not all conceivable countermeasures are economically feasible. Given this, by using RM, the safety performance of different road infrastructure (i.e., road segments, tunnels, intersections) can be simply and effectively evaluated, and high-risk locations can be recognized by building a standard 3×3 crash risk matrix, as shown in Figure 2.

As shown in Figure 2, the crash risk increases with higher crash severity and probability of crash. For instance, when the crash severity is fatal and the probability of the crash is higher than 0.6, the corresponding crash risk will become the highest class, called "critical." The RM was constructed in the present study to evaluate the safety of six junction types, and the junctions with relatively high risk levels (i.e., medium, high and critical) should be addressed by city transportation authorities.

Results and Discussion

Crash Risk Evaluation

Studies have already investigated different factors associated with car crash severity in urban areas. From the

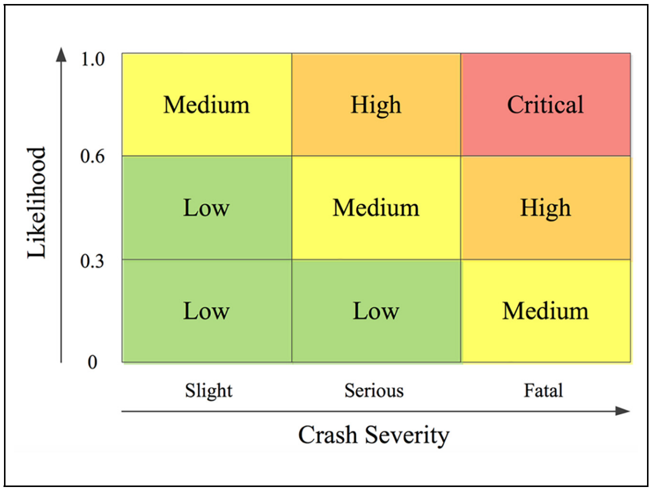

Figure 2. Crash risk matrix (green, yellow, orange, and red squares indicate low, medium, high, and critical risk levels).

literature review, 12 crash risks and 63 crash risk factors were identified and grouped into four risk categories, as listed in Table 2.

The FTA method was employed to calculate the total probability of the TE (i.e., fatal, serious and slight crash), and the RM was built to evaluate the safety performance of different types of urban junctions in two scenarios (i.e., daytime and nighttime), as shown in Table 3. The profiles of TEs' probability and crash risk levels for different urban junction types are listed in Table 3. FTA was only employed in datasets with a sample size greater than 200. When the sample size of data was less than 200, the probability of TEs was calculated by dividing the number of crashes in one level of severity by the number of crashes in the three levels of severity at one junction type.

The results of the RM suggest that the average probability of all three crash severities is much lower in the daytime than in the nighttime, and the average probability of crashes in the nighttime is 4.12 times that in the daytime. This result may be attributed to the higher speed of vehicles and poor light conditions at night (*35*). Roundabouts and mini-roundabouts have the lowest probability of all crashes compared with the other five junction types, which is consistent with the studies of Gross et al. (*10*). In brief, the ranking of high crash risk level for urban junction types (top four highest risk levels) is: T- or staggered junction, crossroads, junction with more than four arms, and other junctions. The relatively high crash risk of T- and staggered junctions and crossroads is presumably attributed to poor sight distance conditions and more conflict points (*35*).

Comparison between FTA and AR

Based on the risk factors identified in Table 2, FTA and AR were employed to explore the contributory risk factors of four junction types with high risk level (i.e., T- or staggered junction, crossroads, junctions with more than four arms, and other junctions) at night. High crash risk factors of each category with the top five MCSs' probabilities in FTA and the top five support (%) in AR were selected for the comparison. The results of FTA and AR are quite different from each other, though there are some common high risk factors, as shown in Table 4.

As the logic gates of the FTA were different from the real risk factors combinations in the original datasets, the high risk factors identified by FTA were all four-factor combinations in each of the MCSs, and only one factor with the highest frequency in each of the risk categories can be selected. In comparison with FTA, AR has the advantage of dependence on datasets, which avoids incorrect subjective understanding from human researchers. It is surprising to find that the high risk factors identified by AR were fairly complicated. For instance, there were one-factor, two-factor, three-factor, and four-factor combinations in the AR with high degrees of support, confidence, and lift.

The disadvantage of FTA can be observed from the obvious differences between results of FTA and AR. In conventional FTA, the basic event is normally a "fault" in the system, and the relationship between each basic event must be known at the outset. However, when FTA is applied in risk management, the basic event is a "risk" rather than a "fault," and the real relationship between each of the risks is mostly unknown when building the fault tree structure. Therefore, the logic gates of FTA can be "noisy," meaning they have a chance of failure (*22*). For instance, the crash risk of male driver, the crash risk combination of male driver and give way junction, or other combinations of crash risk factors are all likely to cause a serious crash in a T- or staggered junction, which is difficult to be expressed in the logic gates of FTA.

Crash Severity Pattern Analysis

To generate AR among the crash severity and contributory risk factors in high-risk junctions at night, the Apriori algorithm was used in the present study. The association algorithm identified over 4,000 rules with support greater than 30%, confidence greater than 60%, and lift greater than one. Among these rules, only the top ten support values in each junction type were selected for the crash severity pattern. The result of the selected AR was the crash severity, thus providing statistical evidence that different crash severities of various junction types are dependent on different contributory risk factors (*8*).

Table 2. Identification of Potential Risks of Crash Severity at Urban Junctions

Crash risk categories	Crash risks	Crash risk factors	Sources
A: Human factors	A1: Age of driver	A11: Young driver (age 16–20) A12: Middle-aged driver (age 56–65) A13: Old driver (age over 66)	Paleti et al. (*28*), Daniels et al. (*9*)
	A2: Sex of driver	A21: Male A22: Female	Fountas et al. (*14*), Bogue et al. (*15*), Jiang et al. (29)
	A3: Maneuver type	A31: Reversing A32: Parked	Haleem and Abdel-Aty (2)
		A33: Waiting to go – held up A34: Slowing or stopping A35: Moving off A36: U-turn A37: Turning left A38: Waiting to turn left A39: Turning right A40: Waiting to turn right A41: Changing lane to left A42: Changing lane to right A43: Overtaking moving vehicle – offside A44: Overtaking static vehicle – offside A45: Overtaking – nearside A46: Going ahead left-hand bend A47: Going ahead right-hand bend A48: Going ahead other	Sobhani et al. (*30*), Daniels et al. (*9*)
B: Vehicle factors	B1: Skidding and overturning	B11: Skidded B12: Skidded and overturned B13: Jackknifed B14: Jackknifed and overturned B15: Overturned	Castro et al. (*31*), Jiang et al. (*29*)
	B2: Age of vehicle	B21: Vehicle age (5–9 years) B22: Vehicle age (over 10 years)	Chen et al. (*32*)
	B3: Vehicle location	B31: Approaching junction or waiting/ parked at junction approach B32: Cleared junction or waiting/ parked at junction exit B33: Leaving roundabout B34: Entering roundabout B35: Mid-junction, on roundabout or on main road	Chen et al. (*32*), Jiang et al. (*29*)
C: Road factors	C1: Junction control	C11: Authorized person C12: Auto traffic signal C13: Stop sign C14: Give way or uncontrolled	Gross et al. (*10*)
	C2: Pedestrian facilities	C21: No physical crossing facilities within 50 meters C22: Zebra crossing C23: Pelican, puffin, toucan or similar non-junction pedestrian light crossing C24: Pedestrian phase at traffic signal junction	Ravishankar and Nair (*33*)
	C3: Road surface Conditions	C31: Wet or damp C32: Snow C33: Frost or ice C34: Flood over 3 cm deep	Montella (*34*) Fountas et al. (*14*)
D: Environment factors	D1: Light conditions	D11: Nighttime – lights unlit D12: Nighttime – no lighting D13: Nighttime – lighting unknown	Sasidharan and Donnell (*35*)

(continued)

Table 2. (continued)

Crash risk categories	Crash risks	Crash risk factors	Sources
	D2: Special conditions at site	D21: Auto traffic signal out D22: Auto signal part defective D23: Road sign or marking defective or obscured D24: Roadworks D25: Road surface defective D26: Oil or diesel D27: Mud	Montella (*34*)
	D3: Weather conditions	D31: Raining no high winds D32: Snowing no high winds D33: Fine and high winds D34: Raining and high winds D35: Snowing and high winds D36: Fog or mist	Jiang et al. (*29*)

Table 3. Crash Risk Evaluation of Urban Junctions in U.K. between 2012 and 2016

		Scenario 1: Daytime		Scenario 2: Nighttime	
Junction type	Crash severity	Probability	Risk level	Probability	Risk level
Roundabouts	Fatal	0.001	Low	0.000	Low
	Serious	0.086	Low	0.220	Low
	Slight	0.096	Low	0.217	Low
Mini-roundabouts	Fatal	0.001	Low	0.002	Low
	Serious	0.075	Low	0.032	Low
	Slight	0.118	Low	0.239	Low
T- or staggered junctions	Fatal	0.002	Low	0.030	Low
	Serious	0.084	Low	0.759	**High**
	Slight	0.095	Low	0.629	**Medium**
Crossroads	Fatal	0.001	Low	0.003	Low
	Serious	0.076	Low	0.806	**High**
	Slight	0.108	Low	0.338	Low
More than four arms	Fatal	0.014	Low	0.000	Low
	Serious	0.116	Low	0.057	Low
	Slight	0.128	Low	0.778	**Medium**
Other junctions	Fatal	0.001	Low	0.000	Low
	Serious	0.072	Low	0.062	Low
	Slight	0.110	Low	0.760	**Medium**
Average probability of crashes		0.066	-	0.272	-

Note: Bold font indicates that the crash risk level of junction type is relatively high.

Figure 3 shows the one-factor, two-factor, three-factor, and four-factor rules, along with their degree of support (the values of confidence and lift are all greater than one, so they are omitted from the figure). According to this result, we confirm that AR allows a better assessment of the interdependences among the crash risk factors than FTA. For each type of junction, one color of circle represents a high risk factor and each overlapping area means that several factors cause the crash jointly. Crash severity pattern can be seen as different combinations of crash risk factors, which includes one-factor, two-factor and multiple-factor rules. These combinations are in different proportions to the total crash number of each level crash severity at a particular junction type.

The three most frequent risk factors in all high risk level junctions at night are: {Sex_of_driver = Male}, {Pedestrian_facilities = No physical crossing facilities within 50 meters}, and {Junction_control = Give way or uncontrolled}, as shown in Figure 3. This suggests that these risk factors are strongly correlated with the occurrence of crash casualties at urban junctions. Consistent with previous research (*14*), this result shows that male drivers are more likely to cause serious crashes than female drivers, because the majority of vehicle drivers

Table 4. Contributory Risk Factors of High Crash Risk Urban Junctions at Night Identified by FTA and AR

Junction type	Crash severity	FTA: High risk factors (MCSs)	AR: High risk factors (Rules, only antecedent)	Common high risk factors	Different high risk factors
T- or staggered junction	Serious	• Male & Mid-junction, on roundabout or on main road & Give way or uncontrolled & Nighttime-lights lit • Male & Mid-junction, on roundabout or on main road & No physical crossing facilities within 50 meters & Nighttime-lights lit • Going ahead other & Mid-junction, on roundabout or on main road & Give way or uncontrolled & Nighttime-lights lit • Going ahead other & Mid-junction, on roundabout or on main road & No physical crossing facilities within 50 meters & Nighttime-lights lit • Male & Vehicle age (over 10 years) & Give way or uncontrolled & Nighttime-lights lit	• Young driver (age 16–20) • Give way or uncontrolled • Give way or uncontrolled & Young driver (age 16–20) • No physical crossing facilities within 50 meters • No physical crossing facilities within 50 meters & Young driver (age 16–20)	• Give way or uncontrolled • No physical crossing facilities within 50 meters	• Young driver • Male • Mid-junction, on roundabout or on main road • Nighttime-lights lit • Going ahead other • Vehicle age (over 10 years)
	Slight	• Male & Mid-junction, on roundabout or on main road & Give way or uncontrolled & Nighttime-lights lit • Going ahead other & Mid-junction, on roundabout or on main road & Give way or uncontrolled & Nighttime-lights lit • Male & Mid-junction, on roundabout or on main road & No physical crossing facilities within 50 meters & Nighttime-lights lit • Male & Vehicle age (over 10 years) & Give way or uncontrolled & Nighttime-lights lit • Going ahead other & Mid-junction, on roundabout or on main road & No physical crossing facilities within 50 meters & Nighttime-lights lit	• Give way or uncontrolled • No physical crossing facilities within 50 meters • No physical crossing facilities within 50 meters & Give way or uncontrolled • Male • Male & Give way or uncontrolled	• Give way or uncontrolled • No physical crossing facilities within 50 meters • Give way or uncontrolled • Male	• Mid-junction, on roundabout or on main road • Nighttime-lights lit • Going ahead other • Vehicle age (over 10 years)

(continued)

Table 4. (continued)

Junction type	Crash severity	FTA: High risk factors (MCSs)	AR: High risk factors (Rules, only antecedent)	Common high risk factors	Different high risk factors
Crossroads	Serious	• Male & Mid-junction, on roundabout or on main road & Auto traffic signal & Nighttime-lights lit • Going ahead other & Mid-junction, on roundabout or on main road & Auto traffic signal & Nighttime-lights lit • Male & Mid-junction, on roundabout or on main road & No physical crossing facilities within 50 meters & Nighttime-lights lit • Going ahead other & Mid-junction, on roundabout or on main road & No physical crossing facilities within 50 meters & Nighttime-lights lit • Male & Vehicle age (over 10 years) & Auto traffic signal & Nighttime-light lit	• Male • Mid-junction, on roundabout or on main road • Going ahead other • Auto traffic signal • No physical crossing facilities within 50 meters	• Male • Mid-junction, on roundabout or on main road • Going ahead other • Auto traffic signal • No physical crossing facilities within 50 meters	• Nighttime-lights lit • Vehicle age (over 10 years)
More than four arms	Slight	• Male & Mid-junction, on roundabout or on main road & Auto traffic signal & Nighttime-lights lit • Going ahead other & Mid-junction, on roundabout or on main road & Auto traffic signal & Nighttime-lights lit • Male & Mid-junction, on roundabout or on main road & No physical crossing facilities within 50 meters & Nighttime-lights lit • Going ahead other & Mid-junction, on roundabout or on main road & No physical crossing facilities within 50 meters & Nighttime-lights lit • Male & Vehicle age (over 10 years) & Auto traffic signal & Nighttime-lights lit	• Auto traffic signal • Male • Male & Auto traffic signal • Mid-junction, on roundabout or on main road • Wet or damp	• Auto traffic signal • Male • Mid-junction, on roundabout or on main road	• Nighttime-lights lit • Going ahead other • No physical crossing facilities within 50 meters • Vehicle age (over 10 years)

(continued)

Table 4. (continued)

Junction type	Crash severity	FTA High risk factors (MCSs)	AR High risk factors (Rules, only antecedent)	Common high risk factors	Different high risk factors
Other junctions	Slight	• Male & Approaching junction or waiting/parked at junction & Give way or uncontrolled & Nighttime-lights lit • Male & Approaching junction or waiting/parked at junction & No physical crossing facilities within 50 meters & Nighttime-lights lit • Going ahead other & Approaching junction or waiting/parked at junction & Give way or uncontrolled & Nighttime-lights lit • Going ahead other & Approaching junction or waiting/parked at junction & No physical crossing facilities within 50 meters & Nighttime-lights lit • Male & Vehicle age (over 10 years) & Give way or uncontrolled & Nighttime-lights lit	• No physical crossing facilities within 50 meters • Give way or uncontrolled • Male • Give way or uncontrolled & No physical crossing facilities within 50 meters • Approaching junction or waiting/parked at junction & No physical crossing facilities within 50 meters	• No physical crossing facilities within 50 meters • Male • Approaching junction or waiting/parked at junction	• Give way or uncontrolled • Nighttime-lights lit • Going ahead other • Vehicle age (over 10 years)

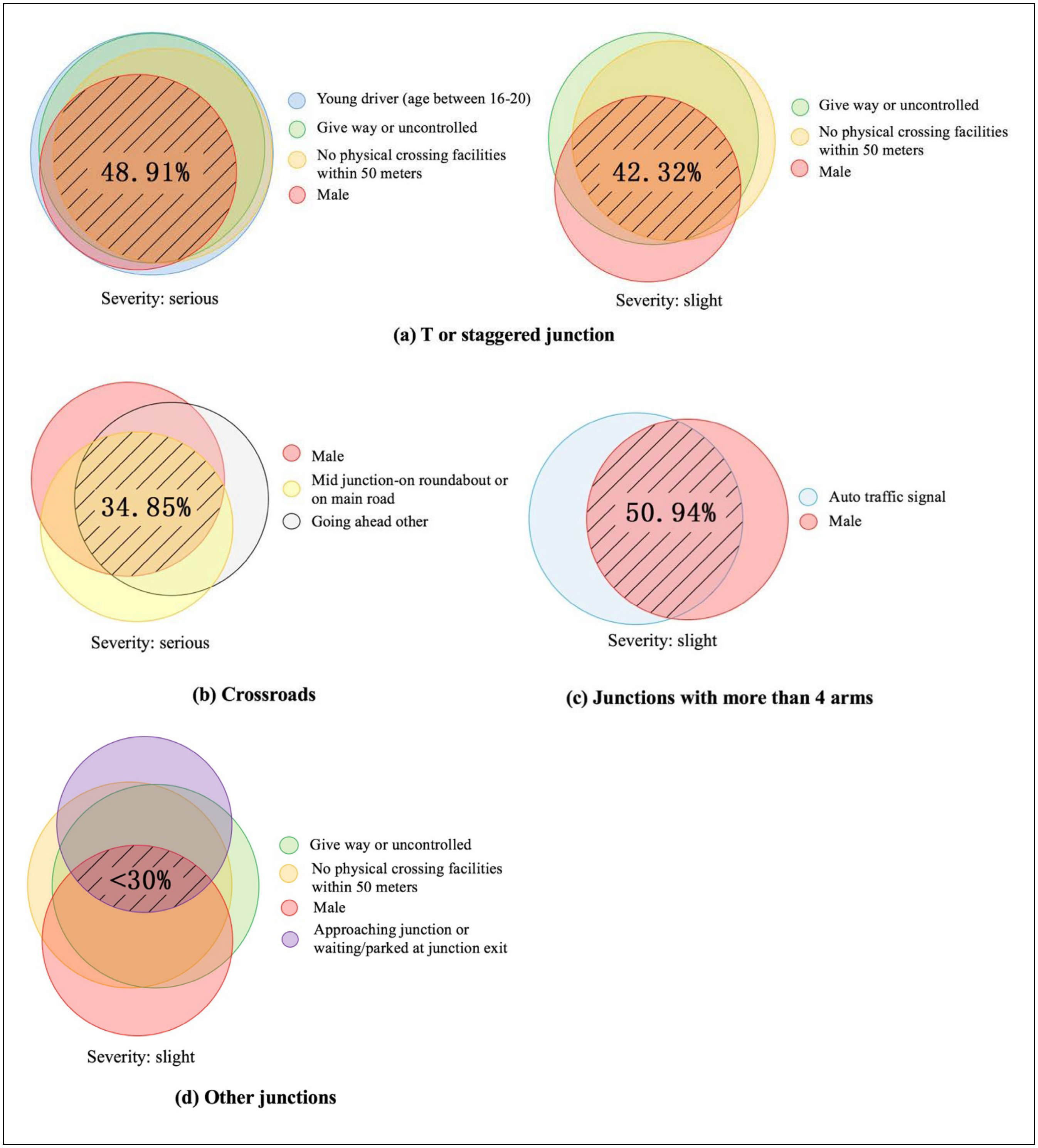

Figure 3. Crash severity pattern of urban junction types with high risk level (the area of circles denotes the degree of support).

are men and they are more likely to engage in dangerous driving behaviors, such as overtaking, excess speed, and drink driving. When there are no physical crossing facilities in junctions, drivers tend to go at higher speed without taking special notice of crossing pedestrians. In addition, if the junction control is "give way" or it is uncontrolled, then there will be more conflict points and chaotic traffic than with traffic signal control method (*2*).

It is worth noting that young driver (aged between 16 and 20) is the major risk factor for serious crashes in T- or staggered junctions at night, with degree of support even reaching 100%. This result can probably be attributed to aggressive or risky driving behaviors and lack of driving experience of young drivers (*28*). It is clear that {Sex_of_driver = Male}, {Vehicle_location = Mid junction}, and {Maneuver_type = Going ahead other} are the three critical risk factors of serious crashes at crossroads. Since the middle area of a crossroads is the location with biggest crash exposure, and overtaking behaviors will increase the possibility of vehicle collision, the combination of these three factors will cause severe crashes at crossroads at night (*29*).

Besides, the combination forms of risk factors in each junction type are quite complex and they are different from each other. It reminds us that different safety improvement measures should be taken in accordance with the types of urban junctions. Through the analysis of the crash severity patterns of high-risk junctions, traffic management authorities can have better awareness of potential crash risks in different urban junctions and make effective countermeasures (e.g., introduce traffic signal systems in uncontrolled junctions; prohibit overtaking in junctions) in high-risk junctions.

Conclusion

In the present study, FTA was applied to evaluate the crash risk level of six urban junction types (both in daytime and nighttime scenarios) in the U.K. Next, from data mining technology, AR analysis was employed to investigate the contributory factors and crash severity patterns of high-risk junctions. By comparing the high risk factors identified by FTA and AR, the shortcomings of conventional FTA were highlighted, as the wrong logic gates of FTA will generate some misleading and inaccurate results in risk management. Therefore, the use of AR was recommended for the analysis of the potential risk factors associated with crash severity. Finally, the Venn diagram of crash severity patterns provided more understanding of the interdependence between risk factors for researchers and traffic engineers. To sum up, the following conclusions can be drawn:

- The average probability of crashes in the nighttime is much higher than that in the daytime under the use of FTA to calculate the occurrence probability of TEs.
- Among six types of urban junctions, roundabouts and mini-roundabouts have the lowest fatalities and casualties. T- or staggered junctions, crossroads, junctions with more than four arms, and other junctions, have relatively high crash risk levels. Especially, T- or staggered junctions and crossroads have highest crash risk.
- AR has the advantage of exploring potential crash risk in original data, and avoiding the inaccurate results caused by logic gates in FTA. The high crash risks of each junction type are quite complicated, because one-factor, two-factor, and multiple-factor combinations make up different percentages of all crash data, and the relationship between risk factors is difficult to express by the logic gates of FTA.
- Urban junctions with high risk levels have different crash severity patterns. Risk factors including male driver, no physical crossing facilities within 50 meters, and give way or uncontrolled junction, are common in all high-risk junctions, and they will increase the crash severity in urban junctions. Other risk factors including young drivers (age between 16 and 20), mid-junction location and overtaking also should be addressed by traffic management administrations.

In conclusion, junctions are the high crash risk parts of a city road network, and different types of junctions have unique characteristics and special crash severity patterns. When the FTA method is applied in risk evaluation, incorrect logic gates may produce misunderstanding of results, especially when the relationship of each risk is unknown at the outset. Therefore, the AR algorithm is a good way to overcome the disadvantages of conventional FTA. It has been proven to generate more objective results of risk factors in the present study and has a huge potential to be effectively used in big datasets. In the near future, the application of AR in other safety research topics (e.g., real-time crash risk evaluation, safety issues of autonomous vehicles) is quite promising, and the determination of parameters of AR could be further investigated. Because the STATS19 dataset is lacking other important information about crashes, more studies are needed to emphasize more micro-level and detailed potential risk factors of junctions, including aggressive driving behavior, traffic volume, and vehicle speed. When it comes to the safety evaluation of urban junctions in developing countries, such as China and India, the first priority is to build a national crash database, as developed countries have done. Many minor crashes are usually unreported in developing countries, which affects the final results of any crash risk analysis. Therefore, the real crash risk level maybe higher than the analyzed crash risk level in developing countries. Unlike developed countries, there are multiple modes of transport on the road in developing countries, so safety issues of urban junctions in developing countries remains a field for further research.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (No. 71701055). The authors sincerely thank all the teachers and classmates who gave many valuable suggestions on this paper.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Peijie Wu, Xianghai Meng; data collection: Peijie Wu, Li Song; analysis and interpretation of results: Peijie Wu, Wenze Zuo. All authors reviewed the results and approved the final version of the manuscript.

References

1. *Reported Road Casualties in Great Britain: Main Results 2013*. Department for Transport. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/324580/rrcgb-main-results-2013.pdf. Accessed June 26, 2014.
2. Haleem, K., and M. Abdel-Aty. Examining Traffic Crash Injury Severity at Unsignalized Intersections. *Journal of Safety Research*, Vol. 41, No. 4, 2010, pp. 347–357. https://doi.org/10.1016/j.jsr.2010.04.006.
3. Kabir, S., T. Azad, M. Walker, and Y. Gheraibia. Reliability Analysis of Automated Pond Oxygen Management System. *Proc., 2015 18th International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, IEEE, New York, 2016, pp. 144–149. https://10.1109/ICCITechn.2015.7488058.
4. Walker, M., and Y. Papadopoulos. Qualitative Temporal Analysis: Towards a Full Implementation of the Fault Tree Handbook. *Control Engineering Practice*, Vol. 17, No. 10, 2009, pp. 1115–1125. https://doi.org/10.1016/j.conengprac.2008.10.003.
5. Chi, C. F., S. Z. Lin, and R. S. Dewi. Graphical Fault Tree Analysis for Fatal Falls in the Construction Industry. *Accident Analysis and Prevention*, Vol. 72, No. 3, 2014, pp. 359–369. https://doi.org/10.1016/j.aap.2014.07.019.
6. Hyun, K. C., S. Min, H. Choi, J. Park, and I. M. Lee. Risk Analysis Using Fault-Tree Analysis and Analytic Hierarchy Process Applicable to Shield TBM Tunnels. *Tunnelling and Underground Space Technology Incorporating Trenchless Technology Research*, Vol. 49, No. 1, 2015, pp. 121–129. https://doi.org/10.1016/j.tust.2015.04.007.
7. Tupper, L. L., M. Chowdhury, and J. Sharp. Tort Liability Risk Prioritization Through the Use of Fault Tree Analysis. Presented at 93th Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.
8. Montella, A. Identifying Crash Contributory Factors at Urban Roundabouts and Using Association Rules to Explore Their Relationships to Different Crash Types. *Accident Analysis and Prevention*, Vol. 43, No. 4, 2011, pp. 1451–1463. https://doi.org/10.1016/j.aap.2011.02.023.
9. Daniels, S., T. Brijs, E. Nuyts, and G. Wets. Externality of Risk and Crash Severity at Roundabouts. *Accident Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 1966–1973. https://doi.org/10.1016/j.aap.2010.06.001.
10. Gross, F., C. Lyon, B. Persaud, and R. Srinivasan. Safety Effectiveness of Converting Signalized Intersections to Roundabouts. *Accident Analysis and Prevention*, Vol. 50, No. 1, 2013, pp. 234–241. https://doi.org/10.1016/j.aap.2012.04.012.
11. Daniels, S., T. Brijs, E. Nuyts, and G. Wets. Explaining Variation in Safety Performance of Roundabouts. *Accident Analysis and Prevention*, Vol. 42, No. 2, 2010, pp. 393–402. https://doi.org/10.1016/j.aap.2009.08.019.
12. Pai, C. W. Motorcyclist Injury Severity in Angle Crashes at T-Junctions: Identifying Significant Factors and Analysing What Made Motorists Fail to Yield to Motorcycles. *Safety Science*, Vol. 47, No. 8, 2009, pp. 1097–1106. https://doi.org/10.1016/j.ssci.2008.12.007.
13. Nitsche, P., P. Thomas, R. Stuetz, and R. Welsh. Pre-crash Scenarios at Road Junctions: A Clustering Method for Car Crash Data. *Accident Analysis and Prevention*, Vol. 107, 2017, pp. 137–151. https://doi.org/10.1016/j.aap.2017.07.011.
14. Fountas, G., P. C. Anastasopoulos, and M. Abdel-Aty. Analysis of Accident Injury Severities Using a Correlated Random Parameters Ordered Probit Approach with Time Variant Covariates. *Analytic Methods in Accident Research*, Vol. 18, 2018, pp. 57–68. https://doi.org/10.1016/j.amar.2018.04.003.
15. Bogue, S., R. Paleti, and L. Balan. A Modified Rank Ordered Logit Model to Analyze Injury Severity of Occupants in Multivehicle Crashes. *Analytic Methods in Accident Research*, Vol. 14, 2017, pp. 22–40. https://doi.org/10.1016/j.amar.2017.03.001.
16. Uddin, M., and N. Huynh. Truck-involved Crashes Injury Severity Analysis for Different Lighting Conditions on Rural and Urban Roadways. *Accident Analysis and Prevention*, Vol. 108, 2017, pp. 44–55. https://doi.org/10.1016/j.aap.2017.08.009.
17. Osman, M., S. Mishra, and R. Paleti. Injury Severity Analysis of Commercially-Licensed Drivers in Single-Vehicle Crashes: Accounting for Unobserved Heterogeneity and Age Group Differences. *Accident Analysis and Prevention*, Vol. 118, 2018, pp. 289–300. https://doi.org/10.1016/j.aap.2018.05.004.
18. Gray, R. C., M. A. Quddus, and A. Evans. Injury Severity Analysis of Accidents Involving Young Male Drivers in Great Britain. *Journal of Safety Research*, Vol. 39, No. 5, 2008, pp. 483–495. https://doi.org/10.1016/j.jsr.2008.07.003.
19. *Road Safety Data*. Department for Transport. https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data. Accessed January 1, 2005.
20. Vesely, W. E., F. F. Goldberg, N. H. Roberts, and D. F. Haasl. *Fault Tree Handbook*, Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commision, 1981. https://doi.org/NUREG-0492.
21. Goodman, G. V. R. An Assessment of Coal Mine Escapeway Reliability Using Fault Tree Analysis. *Mining Science and Technology*, Vol. 7, No. 2, 1988, 205–215. https://doi.org/10.1016/S0167-9031(88)90610-X.
22. Ruijters, E., and M. Stoelinga. Fault Tree Analysis: A Survey of the State-of-the-Art in Modeling, Analysis and Tools. *Computer Science Review*, Vol. 15–16, No. 3, 2015. pp. 29–62. https://doi.org/10.1016/j.cosrev.2015.03.001.
23. Agrawal, R., T. Imielinski, and A. N. Swami. Mining Association Rules between Sets of Items in Large Databases.

Proc., 1993 ACM SIGMOD International Conference on Management of Data, 1993. https://10.1145/170036.170072.
24. Valle, M. A., G. A. Ruz, and R. Morrás. Market Basket Analysis: Complementing Association Rules with Minimum Spanning Trees. *Expert Systems with Applications*, Vol. 97, 2018, pp. 146–162. https://doi.org/10.1016/j.eswa.2017.12.028.
25. Kim, J., C. Bang, H. Hwang, D. Kim, C. Park, and S. Park. IMA: Identifying Disease-Related Genes Using Mesh Terms and Association Rules. *Journal of Biomedical Informatics*, Vol. 76, 2017, pp. 110–123. https://doi.org/10.1016/j.jbi.2017.11.009.
26. *IBM SPSS Modeler 17.0 User's Guide*. IBM, Inc., Armonk, New York, 2015.
27. Cox, L. A. Jr. What's Wrong with Risk Matrices? *Risk Analysis*, Vol. 28, No. 2, 2008, pp. 497–512. https://10.1111/j.1539-6924.2008.01030.x.
28. Paleti, R., N. Eluru, and C. R. Bhat. Examining the Influence of Aggressive Driving Behavior on Driver Injury Severity in Traffic Crashes. *Accident Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 1839–1854. https://doi.org/10.1016/j.aap.2010.05.005.
29. Jiang, X., B. Huang, R. L. Zaretzki, S. Richards, X. Yan, and H. Zhang. Investigating the Influence of Curbs on Single-Vehicle Crash Injury Severity Utilizing Zero-Inflated Ordered Probit Models. *Accident Analysis and Prevention*, Vol. 57, No. 3, 2013, pp. 55–66. https://doi.org/10.1016/j.aap.2013.03.018.
30. Sobhani, A., W. Young, D. Logan, and S. Bahrololoom. A Kinetic Energy Model of Two-Vehicle Crash Injury Severity. *Accident Analysis and Prevention*, Vol. 43, No. 3, 2010, pp. 741–754. https://doi.org/10.1016/j.aap.2010.10.021.
31. Castro, M., R. Paleti, and C. R. Bhat. A Spatial Generalized Ordered Response Model to Examine Highway Crash Injury Severity. *Accident Analysis and Prevention*, Vol. 52, No. 4, 2013, pp. 188–203. https://doi.org/10.1016/j.aap.2012.12.009.
32. Chen, C., G. Zhang, J. Yang, J. C. Milton, and A. Alcántara. An Explanatory Analysis of Driver Injury Severity in Rear-End Crashes Using a Decision Table/Naïve Bayes Hybrid Classifier. *Accident Analysis and Prevention*, Vol. 90, 2016, pp. 95–107. https://doi.org/10.1016/j.aap.2016.02.002.
33. Ravishankar, K. V. R., and P. M. Nair. Pedestrian Risk Analysis at Uncontrolled Midblock and Unsignalised Intersections. *Journal of Traffic and Transportation Engineering (English Edition)*, Vol. 5, No. 2, 2018, pp. 137–147. https://doi.org/10.1016/j.jtte.2017.06.005.
34. Montella, A. Identifying Crash Contributory Factors at Urban Roundabouts and Using Association Rules to Explore their Relationships to Different Crash Types. *Accident Analysis and Prevention*, Vol. 43, No. 4, 2011, pp. 1451–1463. https://doi.org/10.1016/j.aap.2011.02.023.
35. Sasidharan, L., and E. T. Donnell. Propensity Scores-Potential Outcomes Framework to Incorporate Severity Probabilities in the Highway Safety Manual Crash Prediction Algorithm. *Accident Analysis and Prevention*, Vol. 71, No. 10, 2014, pp. 183–193. https://doi.org/10.1016/j.aap.2014.05.017.