TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–13

DOI: 10.1177/03611981231164386
journals.sagepub.com/home/trr
SAGE

Bayesian Network for Motorcycle Crash Severity Analysis

Subasish Das[1], Valerie Vierkant[2], Juan Cruz Gonzalez[2], Boniphace Kutela[3], and Abbas Sheykhfard[4]

Abstract

Given the lack of protective structural barriers and advanced restraints, motorcyclists are vulnerable road users. In 2020 in the United States, motorcycle-involved fatalities occurred 28 times more frequently per vehicle mile traveled than passenger car occupant fatalities, causing 5,579 motorcycle-related fatalities—the highest number of motorcyclists killed since 1975. By identifying patterns and relationships between key contributing factors, strategies for reducing motorcycle crashes can be developed. In addition to current efforts, additional research must be conducted using innovative avenues, with increased funding. Bayesian networks can better discover the relationships between potential speed compliance variables. This study used six years (2014 to 2019) of motorcycle crash data in Louisiana to determine the conditional probabilities of the influential factors. In addition to the high contribution of alcohol involvement, two-way undivided roadways, 35 to 44 year-old drivers involved in improper driving behaviors, and crash types are the underlying factors associated with a considerable increase in motorcycle crash severity. The findings of this study can also be used for decision making and strategy development for motorcycle safety.

Keywords

safety, motorcycles and mopeds, transportation safety management systems, motorcycle countermeasures, motorcycle safety gear

In 2020, motorcycle fatalities occurred 28 times more frequently per vehicle mile traveled than passenger car occupant fatalities in the United States. The U.S. had 5,579 motorcycle-related fatalities in 2020, the highest number since 1975 (*1*). Riding a motorcycle requires a high level of physical ability and coordination. The conventional method for analyzing crash severity has been to establish relationships between driver and traffic characteristics, environment and road conditions, and crash occurrence. Most models constructed using this methodology suffer from a reliance on general assumptions and aggregate measures. In addition, identifying contributing components based on observational data involves a vast array of relationships given the assumptions evaluated during various modeling procedures.

The purpose of Bayesian network (BN) analysis is to better uncover the interconnections between probable motorcycle crash severity factors. BN is an effective, directed graphical model built from Bayes' rule on posterior simulation. Its nodes and linkages depict random variables, their relationships, and conditional probability distributions for each variable's state. The Bayes rule is used to forecast the occurrence of a future event given the occurrence of a previous event (*2*). BN gives an intuitive picture of the interactions between components and the effects of modifying a variable on the dependent variable. Because it is a visual depiction, a causal map provides a greater knowledge of the system than standard analysis tools. The training data set used by the BN's learning algorithm is derived from actual events. The final learned BN can be used for scenario analysis of policy.

[1]Ingram School of Engineering, Texas State University, San Marcos, TX
[2]Texas A&M Transportation Institute, San Antonio, TX
[3]Texas A&M Transportation Institute, Houston, TX
[4]Department of Civil Engineering, Babol Noshirvani University of Technology, Babol, Mazandaran, Iran

Corresponding Author:
Subasish Das, subasish@txstate.edu

Although many studies have examined motorcycle crashes and related injuries, the number of motorcycle crashes continues to increase each year. These crashes can be attributed to various interacting factors. Most of the previous studies have applied traditional regressions that do not express the interdependence between various factors at the same moment. Advanced methods, such as Bayesian networks, provide the flexibility of the interdependence between multiple factors to understand the contributing factors and the patterns of their association. Furthermore, it is relatively hard to assess counterfactual scenarios under traditional regression models (*3*). Therefore, the present paper intended to fill this knowledge gap by extensively assessing motorcycle injury severity while considering a wide range of influential factors. To assess the sensitivities of the variable attributes, counterfactual scenarios were developed. In this study, we examined Louisiana motorcycle crash data from 2014 to 2019 to determine the features of motorcycle-related contributing factors. These databases contain over six years of crash records for motorcycle riders, including roadway, rider, vehicle, and crash-specific information. In this study, probabilistic measures of contributing factor attributes were derived based on the BN results for different injury types.

The continuation of this work after the introduction is arranged as follows. The next section summarizes related work, followed by a description of the data sources. Next, BN analysis theory is briefly explained. Finally, the last section presents the results and discussion, as well as observations for future research.

Literature Review

In this section, the authors review motorcycle-related safety studies that have been undertaken so far. The paper begins with an overview of conventional modeling to identify the influence of different factors, followed by crash injury analysis-related studies, and other specific safety issues associated with motorcycle mobility.

Chen and Fan (*4*) created a multinomial logit (MNL) model to evaluate the severity of pedestrian–vehicle collisions in North Carolina, United States. Medina et al. (*5*) conducted a 39-segment highway review using correlation, ANOVA, and multiple regression analysis, as well as a study of motorcycle drivers. Kostyniuk et al. (*6*) investigated the trends and patterns of motorcycle collisions in Michigan. Similarly, Naumann et al. (*7*) defined their opinion on the relationship between roadway conditions and motorcycle safety by describing how motor vehicle occupant deaths in the United States reached their lowest point in 2008 whereas motorcycle fatalities reached an all-time high. Using Ohio crash data from 2003 to 2007, Eustace et al. (*8*) examined the likelihood of a motorcyclist being seriously injured in a collision and the relevant risk factors. Ryb et al. (*9*) focused on the influence of advanced age on outcomes for wounded individuals in a Maryland data set connecting hospital discharge records and police reports.

Given the absence of an occupant compartment on motorbikes, reducing the number of collisions is essential for reducing fatalities and crashes. Cheng et al. (*10*) created five models representing distinct correlations of meteorological variables on motorcycle collision injuries often observed in crash data and compared fitness and performance at four levels of severity using a comprehensive Bayesian formulation. The outcomes demonstrated that designs with parameter differences in series and severity provided a higher fit and more accurate prediction of collisions. Das et al. (*11*) examined five years (2010 to 2014) of at-fault motorcycle rider-involved crashes in Louisiana by applying deep learning to model a high-dimensional input into a lower-dimensional output. Waseem et al. (*12*) advocated lowering speed restrictions on routes with a higher ratio of motorcycles, differentiating bikes from heavy cars, removing roadside fixtures, and controlling/reducing the risky behavior of motorcyclists. Using motorcycle crash data from 2001 to 2008 in Iowa, Shaheed and Gkritza (*13*) explored variables affecting the severity of collisions resulting from single-vehicle motorcycle crashes using a latent class technique. Pour-Rouholamin et al. (*14*) investigated motorcycle crashes in North Carolina. A partial proportional odds (PPO) logistic regression model was developed to examine the influence of injury severity. Moreover, two other popular ordered-response models, proportional odds and non-proportional odds models, as well as one similar unordered-response model, a multinomial logit model, were also developed to evaluate their performances compared with the PPO model. Cheng et al. (*10*) examined the impact of weather conditions on motorcycle crash injuries at four different severity levels using San Francisco motorcycle crash injury data. Chen et al. (*15*) conducted a study to identify risk and protective factors for crash involvement of older motorcyclists. Vajari et al. (*16*) evaluated the contributing factors to motorcycle crash severity at intersections. The multinomial logit model was used to evaluate motorcycle crashes. The results of the model demonstrated that certain factors increased the probability of fatal injuries. Agyemang et al. (*17*) explored and compared factors that are associated with motorcycle crash injury outcomes in rural and urban areas of Ghana. Preliminary analysis of the crash data revealed that more of the rural area crashes occurred under dark and unlit roadway conditions, while urban areas recorded more intersection-related crashes. Tamakloe et al. (*18*) investigated the latent patterns and chains of factors that simultaneously contribute to the

Table 1. Motorcycle-Related Injury Counts by Year

Year	Fatal (K)	Severe (A)	Moderate (B)	Complaint (C)	No injury (O)	Yearly total
2014	87	138	661	627	420	1,933
2015	94	155	686	654	445	2,034
2016	98	136	628	584	461	1,907
2017	97	176	618	540	424	1,855
2018	87	138	540	512	368	1,645
2019	92	149	531	453	364	1,589
Severity total	555	892	3,664	3,370	2,482	10,963

injury severity sustained by motorcycle crash casualties at intersections under different traffic control conditions in developing countries. Truong et al. (*19*) examined self-reported fatigue-related crashes among motorcycle taxi drivers in Hanoi, Vietnam. Wali et al. (*20*) quantified how different "policy-sensitive" factors are associated with the risk of motorcycle injury crashes while controlling for rider-specific, psycho-physiological, and other observed/unobserved factors. The results of the best-fit random parameters logit model with heterogeneity-in-means show that riders with partial helmet coverage have a significantly lower risk of injury crash involvement.

Chawla et al. (*21*) examined motorcycle crash risk factors by employing data recently made available from the Federal Highway Administration Motorcycle Crash Causation Study (MCCS). Wali et al. (*20*) analyzed data from 321 motorcycle injury crashes from a comprehensive U.S. DOT FHWA's MCCS. Das et al. (*22*) investigated the matched case-control study of the MCCS containing a very wide list of crash-contributing factors associated with motorcycle crash occurrences.

The literature review shows that many factors interplay during the occurrence of a motorcycle crash. Limited studies have attempted to assess the interaction between multiple factors and counterfactual scenarios using motorcycle crash data. The current study aims to mitigate the research gap by extensively assessing motorcycle injury severity while considering a wide range of influential factors.

Methodology

Data Integration

To meet the research aims, the Louisiana Department of Transportation and Development (DOTD) provided six years' worth of traffic crash data (2014 to 2019). This database contains crash-related data from multiple relational databases. Using three primary data sets (crash, vehicle, and road inventory), the required data set was compiled. The vehicle database includes vehicle type. Only information on crashes involving motorcycles was examined for this study. Later, the crash identification numbers were integrated with the crash and DOTD (roadway inventory information) databases to produce comprehensive data. Table 1 lists motorcycle-related injury counts by year.

After deleting irrelevant information, a more precise database based on important contributing elements was constructed. The variable section method used previous study findings. According to the police-reported KABCO scale (where K = death, A = incapacitating injury or severe, B = non-incapacitating injury or moderate, C potential injury or complaint, and O = no injury), 4% of motorcyclists involved in collisions were killed, compared with 0.45% of all crash participants. In Louisiana, over 65% of motorcycle crashes occur during the day. Spring and fall are the worst times to ride a motorcycle, as these two seasons account for approximately 57% of all motorcycle crashes. Among collision categories, the proportion of single-motorcycle collisions is the highest. A higher proportion of segment-related collisions include motorcycles.

Bayesian Network

There are numerous data-driven techniques, including naive Bayesian networks (NBN), augmented Bayesian networks (ABN), and tree-augmented naive Bayes networks (TAN). Among these, TAN learning creates qualitative BNs that reflect risk influential factors' (RIFs) interactive relationships, which aids in generating insights on essential human variables that contribute to various sorts of crashes. Moreover, Friedman et al. (*23*) noted that TAN outperforms naive Bayes while retaining the computational simplicity and resilience that define naive Bayes. TAN has been demonstrated to be more competitive and precise than other data-driven network design methods (*23*).

Bayesian networks have numerous advantages over machine learning models or statistical models. BNs provide the flexibility of the interdependence between multiple factors to understand the contributing factors and the patterns of their association. For motorcycle crashes, there are multiple factors that interplay during a crash

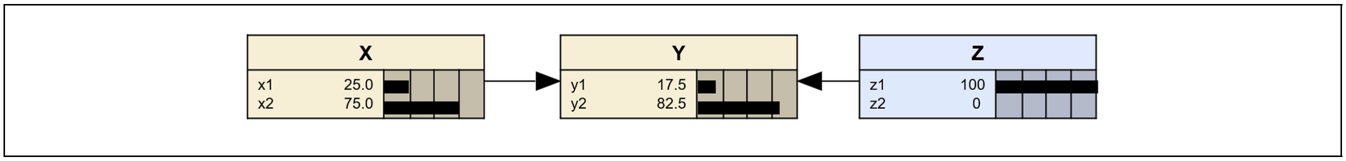

Figure 1. Schematic influence diagram.

occurrence. Thus, traditional regression models require a more complex modeling framework to explain the interaction between the underlying factors. On the other hand, machine learning algorithms have limited interpretability of the contributing factors. Furthermore, it is relatively hard to assess counterfactual scenarios under traditional regression models (*3*, *24*).

A BN determines a joint probability distribution over a set of random variables U, which is an annotated directed acyclic graph (DAG). Consider, $U = \{A_1, \ldots A_n,\ C\}$ where n stands for the number of RIFs, the variables $A_1, \ldots A_n$ are the RIFs and C is the class variable (for example, motorcycle rider injury types). Consider a graph structure where the class variable is the root, that is, $\prod C = \varnothing$ ($\prod C$ denotes the set of parents of C in U) and each RIF has the class variable as its unique parent, that is, $\prod A_i = \{C\}$ for $1 \leqslant i \leqslant n$. A BN defines a unique joint probability distribution over U given by:

$$P(A_1, \ldots A_n,\ C) = P(C).\ \prod_{i=1}^{n} P(A_i|C) \tag{1}$$

The DAG on $\{A_1, \ldots A_n\}$ is a tree if $\prod A_i$ contains only one parent for all A_i, except for one variable without parents (referred as the root). If function π can define a tree over $A_1, \ldots A_n$, if there is exactly one i such that $\pi(i) = 0$ (i.e., the root of the tree), and there is no sequence $i_1, \ldots i_k$ such that $\pi(ij) = i_{j+1}$ for $i \leqslant j < k$ and $\pi(i_k) = i_1$ (i.e., no cycles), such a function defines a tree network where $\prod A_i = \{C, \ldots A_{\pi(i)}\}$ if $\pi(i)>0$ and $\prod A_i = \{C\}$ if $\pi(i) > 0$, and $\prod A_i = \{\mathrm{C}\}$ if $\pi(i) = 0$.

Learning a TAN structure is an optimization problem. The solution function can be defined as:

$$I_p\left(A_i, A_j|C\right) = \sum\nolimits_{a_{ii}, a_{ji}, c_i} P\left(a_{ii}, a_{ji}, c_i\right) \log \frac{P\left(a_{ii}, a_{ji}|c_i\right)}{P(a_{ii}|c_i)P\left(a_{ji}|c_i\right)} \tag{2}$$

where

I_p represents the conditional mutual information,
a_{ii} is the i^{th} state of RIF A_i, a_{ji} is the i^{th} state of RIF, and
c_i is the i^{th} state of "crash type."

The optimization problem (to learn a TAN structure) is to find a tree defining function π over $A_{1,} \ldots A_n$ such that the log likelihood is maximized.

An influence diagram, also known as a Bayesian belief net or a decision diagram, is a graphical representation of factors that contribute to a conclusion or uncertainty (*25*, *26*). The influence diagram computes the likelihood of the outcome given all of the model's factors. Conditional probabilities establish a relationship between the inputs and the result. Decisions can also be incorporated into the influence diagram so that the decision maker can comprehend how each alternative influences the probability of a result (*27*, *28*). Using the Netica software, Figure 1 illustrates an influence diagram in which Y is an uncertain outcome with two possible outcomes y1 and y2 (with probabilities of 17.5% and 82.5%, respectively), X is an uncertain factor with two possible states x1 and x2 (with probabilities of 25% and 75%, respectively), and Z is a decision with two alternatives z1 and z2. The model's arrows indicate that the uncertainty of Y is conditionally reliant on the uncertainty of X and the choice Z. The graphical representation of decision node Z indicates that the alternative z1 is chosen.

Results and Discussions

This study developed the BN model in several steps. First, using Netica 6.04, a directed acyclic graph (DAG) representing the model was built (*29*). Netica is a practical, user-friendly software designed specifically for BN applications with a functional user interface. The networks are used to conduct many types of inference using the most advanced and efficient methods. Netica's expectation maximization (EM) technique (*30*) was then used to fit this network to many input-output combinations from the MC runs using the model. The process of learning yields estimates of the model parameters of the BN, also known as conditional probability distributions. Figure 2 depicts the layout of the final BN network, while the conditional probability tables of each node reflect the node's states as belief bars.

Figure 2 shows the initial data of motorcycle collisions modeled through a BN, centered around injury probabilities representing a real-world scenario. This figure shows proportions of different factors for each variable. The network shows that some of the attributes have higher likelihood in each variable category such as two-way–no physical separation streets (60.5%), between the hours of 1 pm and 6 pm (45.0%), on a straight-level

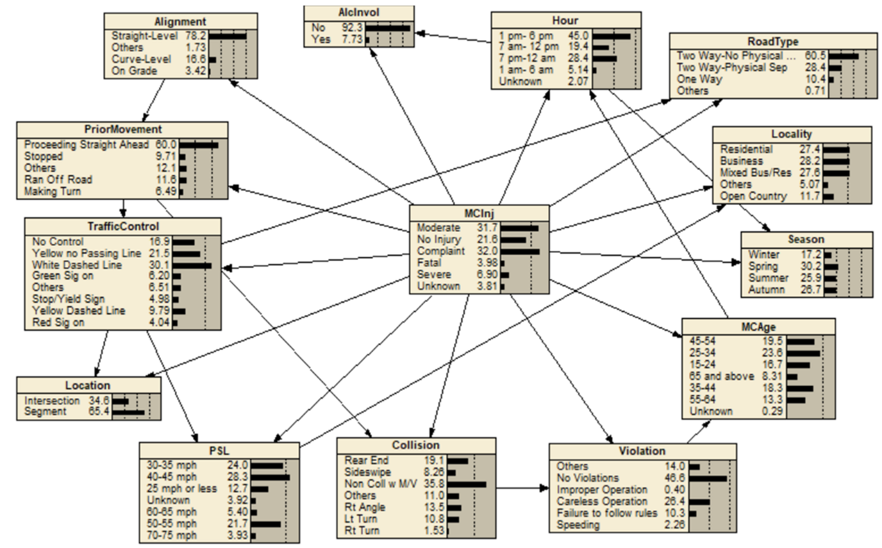

Figure 2. Bayesian network on the final data.
Note: PSL = posted speed limit; MC = motorcycle.

Table 2. Sensitivity of the Node "Motorcycle Injury"

Variable	Mutual information	% Entropy reduction
MCInj	2.160	100
MCAge	0.122	5.66
PriorMovement	0.073	3.37
Collision	0.060	2.78
TrafficControl	0.047	2.18
Violation	0.046	2.12
AlcInvol	0.039	1.79
PSL	0.035	1.61
Locality	0.012	0.56
Alignment	0.012	0.55
Hour	0.011	0.53
RoadType	0.008	0.36
Season	0.003	0.12
Location	0.002	0.09

Note: PSL = posted speed limit; MC = motorcycle.

alignment (78.2%), at a segment (65.4%), and with a posted speed limit (PSL) of 40 to 45 mph (28.3%) or 30 to 35 mph (24.0%). Fatal and severe injury crashes represent 11% of all crashes (which is proportionately high when compared with fatal all vehicle crash proportions). Alcohol (AlcInvol) was associated with 7.73% of the crashes. The most common locality of the crashes was almost evenly split between residential (27.4%), business (28.2%), and mixed bus/res (27.6%). Spring (30.2%) showed a slightly higher likelihood of motorcycle crashes compared with other seasons: fall (26.7%), Summer (25.9%), and winter (17.2%). The most common rider age group (MCAge) was 25 to 34 (23.6%), followed by 45 to 54 (19.5%) and 35 to 44 (18.3%). No violations were noted in 46.6% of the crashes; however, careless operation represented around one-fourth of these crashes (26.4% of the crashes). Non-collision with a motor vehicle was the most common collision type (35.8%), followed by rear-end collisions (19.1%). The most common prior movement was proceeding straight ahead (60.0%). White dashed lines were the most common traffic control device used (30.1%), followed by yellow no passing lines (21.5%) and no controls at all (16.9%). Some of these findings are in line with other safety and operation related studies (*31–34*).

Table 2 displays the results of the sensitivity analysis performed on the nodes by the software Netica. Entropy reduction appears in the third column of the table. Entropy is used in the theory of information to represent the disorder of a data set. The greater the entropy, the greater the data set's uncertainty. Table 2's decrease of entropy reveals that certain factors, including rider age, prior movement, collision kind, and traffic control type, are more sensitive.

Tables 3 and 4 provide conditional probability scores for two variable groups: (1) collision type, prior movement, and rider injury; and (2) rider age, rider injury, and violation type. For example, ran-off-road moderate crash injuries, which are non-collision with another vehicle or single-motorcycle crashes, show a probability score of 0.94, which indicates that a ran-off-road single-motorcycle crash has a likelihood of 0.94 to be involved in a moderate injury crash.

Table 3. Conditional Probability Table for Rider Injury, Prior Movement, and Collision Type

Rider injury (MCInj)	Prior movement before crash (PriorMovement)	Collision type						
		Rear end	Side swipe	Non collision with M/V	Others	Right angle	Left turn	Right turn
Fatal	Proceeding straight ahead	0.138	0.038	0.236	0.079	0.261	0.248	0
Fatal	Stopped	0.333	0.083	0.25	0	0.083	0.25	0
Fatal	Others	0.081	0.101	0.343	0.313	0.071	0.091	0
Fatal	Ran off road	0	0.008	0.969	0.024	0	0	0
Fatal	Making turn	0.4	0	0	0	0	0.6	0
Severe	Proceeding straight ahead	0.117	0.054	0.231	0.118	0.266	0.201	0.013
Severe	Stopped	0.333	0.061	0.182	0.242	0.061	0.091	0.03
Severe	Others	0.067	0.133	0.415	0.193	0.081	0.104	0.007
Severe	Ran off road	0.01	0.01	0.948	0.031	0	0	0
Severe	Making turn	0.15	0.1	0.35	0.1	0.15	0.125	0.025
Moderate	Proceeding straight ahead	0.147	0.074	0.338	0.073	0.209	0.145	0.014
Moderate	Stopped	0.433	0.04	0.317	0.06	0.056	0.075	0.02
Moderate	Others	0.095	0.148	0.434	0.139	0.065	0.104	0.016
Moderate	Ran off road	0.003	0.02	0.94	0.026	0.005	0.005	0
Moderate	Making turn	0.116	0.04	0.466	0.076	0.08	0.151	0.072
Complaint	Proceeding straight ahead	0.176	0.066	0.308	0.132	0.172	0.133	0.014
Complaint	Stopped	0.629	0.04	0.159	0.094	0.036	0.036	0.007
Complaint	Others	0.123	0.174	0.385	0.148	0.082	0.063	0.024
Complaint	Ran off road	0.003	0.009	0.936	0.041	0.005	0.005	0
Complaint	Making turn	0.147	0.079	0.385	0.138	0.053	0.153	0.044
No injury	Proceeding straight ahead	0.275	0.124	0.224	0.109	0.145	0.108	0.015
No injury	Stopped	0.69	0.066	0.039	0.133	0.04	0.024	0.008
No injury	Others	0.15	0.255	0.221	0.162	0.115	0.075	0.022
No injury	Ran off road	0.019	0.01	0.903	0.053	0.005	0.01	0
No injury	Making turn	0.159	0.107	0.207	0.119	0.107	0.2	0.1
Unknown	Proceeding straight ahead	0.188	0.188	0.2	0.165	0.206	0.047	0.006
Unknown	Stopped	0.345	0.138	0.069	0.276	0.138	0.034	0
Unknown	Others	0.179	0.221	0.111	0.371	0.091	0.023	0.003
Unknown	Ran off road	0	0.045	0.727	0.227	0	0	0
Unknown	Making turn	0	0.375	0.125	0	0.125	0.125	0.25

Note: MC = motorcycle; M/V = motor vehicle.

Table 4. Conditional Probability Table for Rider Injury, Rider Age, and Violation Type

Rider injury (MCInj)	Violation	Rider age						
		15–24	25–34	35–44	45–54	55–64	65 and above	Unknown
Fatal	Others	0.128	0.329	0.174	0.181	0.121	0.067	0
Fatal	No violations	0.067	0.174	0.221	0.235	0.208	0.094	0
Fatal	Improper operation	0	0	1	0	0	0	0
Fatal	Careless operation	0.166	0.225	0.199	0.225	0.119	0.066	0
Fatal	Failure to follow rules	0.17	0.191	0.085	0.319	0.17	0.021	0.043
Fatal	Speeding	0.125	0.484	0.188	0.125	0.078	0	0
Severe	Others	0.188	0.228	0.163	0.252	0.094	0.064	0.01
Severe	No violations	0.123	0.223	0.233	0.218	0.165	0.038	0
Severe	Improper operation	0	0.333	0.667	0	0	0	0
Severe	Careless operation	0.222	0.267	0.173	0.202	0.111	0.021	0.004
Severe	Failure to follow rules	0.227	0.173	0.2	0.16	0.16	0.067	0.013
Severe	Speeding	0.2	0.42	0.18	0.2	0	0	0
Moderate	Others	0.214	0.269	0.164	0.192	0.114	0.039	0.008
Moderate	No violations	0.136	0.24	0.224	0.215	0.147	0.038	0
Moderate	Improper operation	0.154	0.231	0.308	0.077	0.077	0.154	0
Moderate	Careless operation	0.207	0.262	0.162	0.199	0.126	0.039	0.005
Moderate	Failure to follow rules	0.189	0.216	0.172	0.191	0.156	0.066	0.011

(continued)

Table 4. (continued)

Rider injury (MCInj)	Violation	Rider age 15–24	25–34	35–44	45–54	55–64	65 and above	Unknown
Moderate	Speeding	0.33	0.33	0.138	0.073	0.064	0.064	0
Complaint	Others	0.175	0.273	0.183	0.177	0.131	0.055	0.007
Complaint	No violations	0.147	0.245	0.199	0.214	0.156	0.038	0
Complaint	Improper operation	0.133	0.267	0.133	0.2	0.133	0.133	0
Complaint	Careless operation	0.221	0.257	0.174	0.177	0.115	0.05	0.005
Complaint	Failure to follow rules	0.214	0.222	0.156	0.205	0.145	0.046	0.011
Complaint	Speeding	0.218	0.309	0.273	0.073	0.109	0.018	0
No injury	Others	0.161	0.24	0.187	0.155	0.096	0.158	0.003
No injury	No violations	0.13	0.223	0.204	0.236	0.166	0.041	0
No injury	Improper operation	0.111	0.222	0.167	0.222	0.222	0.056	0
No injury	Careless operation	0.249	0.222	0.147	0.176	0.108	0.097	0
No injury	Failure to follow rules	0.171	0.204	0.182	0.193	0.149	0.096	0.004
No injury	Speeding	0.294	0.235	0.147	0.147	0.029	0.147	0
Unknown	Others	0.009	0.009	0.026	0.009	0.013	0.931	0.004
Unknown	No violations	0.019	0.045	0.038	0.025	0.013	0.86	0
Unknown	Improper operation	0	0	0	0	0	1	0
Unknown	Careless operation	0.037	0.037	0.037	0	0	0.89	0
Unknown	Failure to follow rules	0.019	0.019	0.019	0	0	0.943	0
Unknown	Speeding	0.167	0	0	0	0	0.833	0

Note: MC = motorcycle.

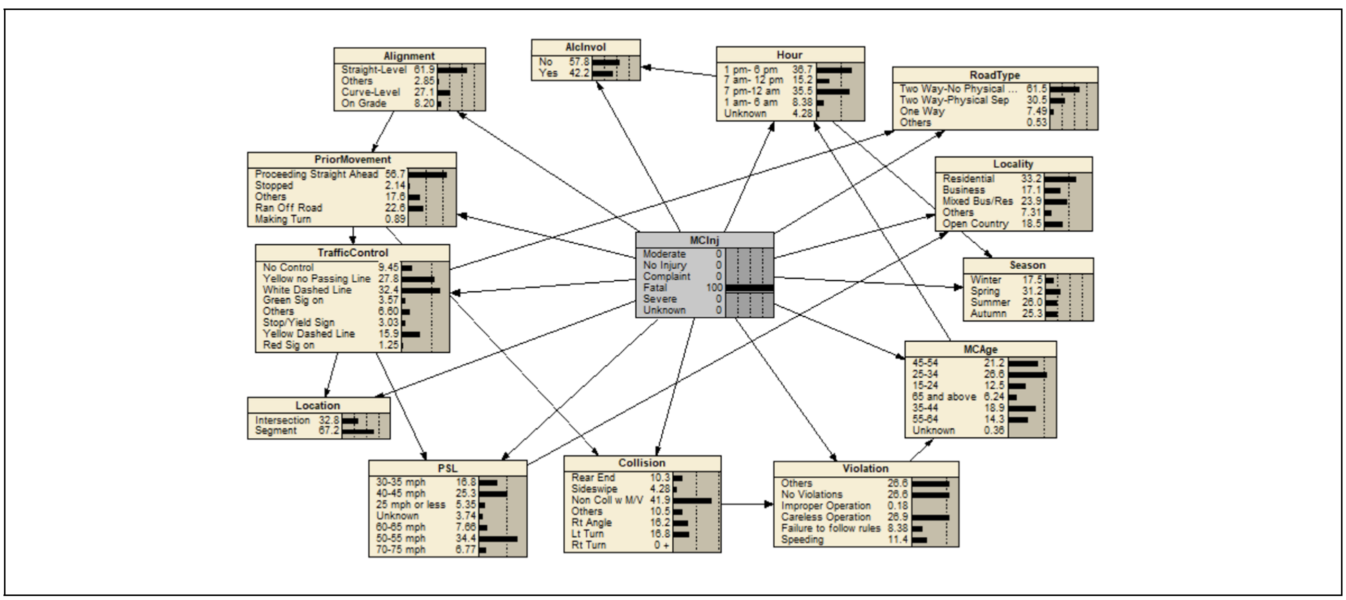

Figure 3. Counterfactual Bayesian network considering all motorcycle injuries are fatal.
Note: PSL = posted speed limit; MC = motorcycle.

Counterfactual Scenarios

Netica software uses a counting-learning technique to conduct the parameter learning of conditional probability tables (CPT) from examples based on the TAN model. After the CPTs have been constructed and retrieved, the posterior probabilities for each variable can be computed. The statistical examination of the likelihood of variables provides intriguing preliminary findings that offer useful insights into safety precautions and crash prevention. Figure 3 shows the BN for motorcycle crashes by considering the counterfactual scenario of all motorcycle injuries as fatal crashes. In this scenario,

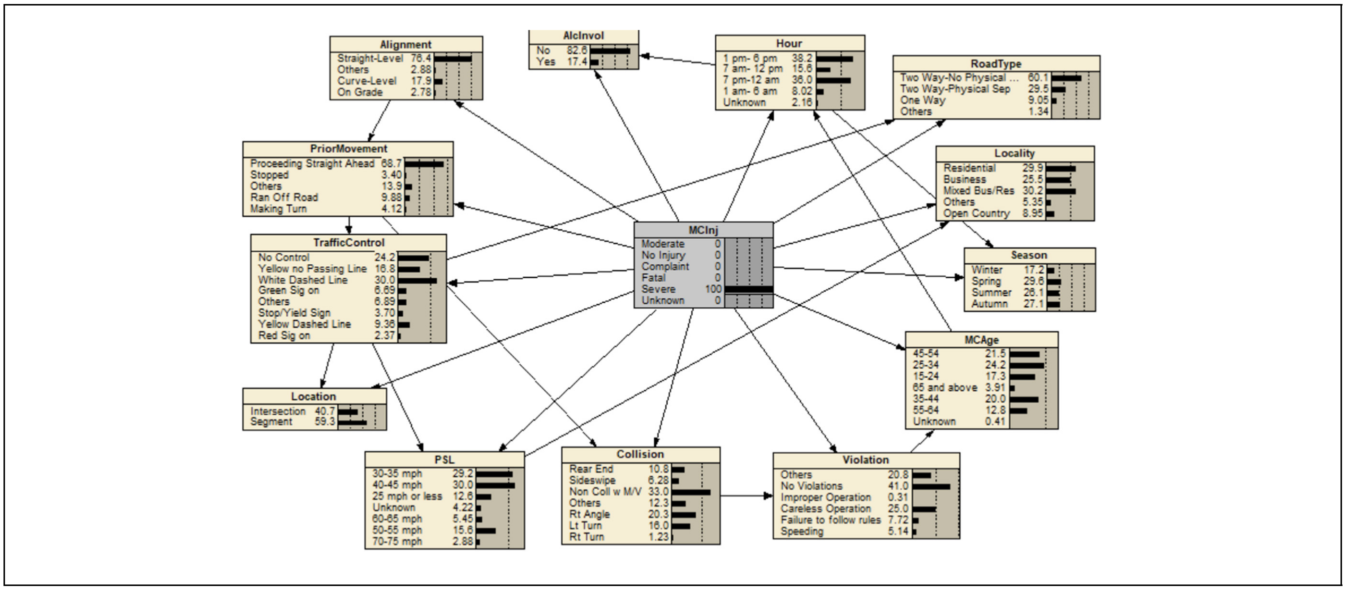

Figure 4. Counterfactual Bayesian network considering all motorcycle injuries are severe.
Note: PSL = posted speed limit; MC = motorcycle.

alcohol involvement increased to 42.2% (5.46 times the final data BN network proportions). The proportion of crashes where no violations occurred decreased from 46.6% to 26.6% (a 43% reduction). For PSL, the most common speed limit for the fatal crash data was 50 to 55 mph, compared with 40 to 45 mph in the full data. In the fatal model scenario, number of locations, with speed limit 50 to 55 mph had an occurrence of 34.4% compared with 21.7% in the real-world data scenario (1.59 times the full data). For locality, the crash occurring in a business area decreased to 17.1% (a reduction of 39%). The use of a yellow dashed line as a traffic control device increased in the fatal model to 15.9% (1.6 times higher than the full data), whereas no control decreased to 9.45% (a reduction of 44%). Ran-off-road crashes became more prevalent in the fatal crash model, increasing to 22.6% (1.94 higher odds). Curve level made up a greater proportion of the alignment in the fatal injury scenario, increasing to 27.1% (1.63 higher odds). Hour, road type, location, season, and age did not have major changes compared with the real-world scenario.

Figure 4 shows the BN for motorcycle crashes by considering the counterfactual scenario of all motorcycle injuries as severe injury crashes. Alcohol involvement in this scenario increased to 17.4% compared with 7.73% of the full data (an increase of 2.25 times the full data). This is a notably smaller increase in alcohol involvement compared with the fatal crash scenario, where the alcohol involvement increased by 5.46 times the full data. In collision type, rear end decreased from 19.1% in the full data to 10.8% in the severe injury data (a decrease of 43%). Right angle and left turn crashes were both more prevalent in the severe injury model, increasing by 1.5 and 1.48 times the full data, respectively. No traffic control device being used became more common in the severe injury model, increasing to 24.2% (an increase of 1.43 times the full data). This contrasts with the fatal injury model, where instances in which no traffic control devices were used decreased compared with the real-world model. A prior movement of being stopped became less common in the severe injury model, decreasing to 3.40% (a decrease of 82%). Hour, road type, alignment, locality, season, age, and violations had little change from the full data.

Figure 5 represents the BN for motorcycle crashes by considering the counterfactual scenario of all motorcycle injuries as moderate injury crashes. In the following network, when all crashes are listed as moderate, there were no major changes between moderate and real-world scenarios. Some of the minor changes include the straight-level alignment increasing from 60% to 65.5% (1.09 times more than the real-world scenario), non-collisions with motor vehicles increasing from 35.8% to 43.5% (1.22 times more than original data), and finally a minor increase from 60% to 65.5% (1.09 times more than the full data) of motorcycles proceeding straight ahead. The change between traffic control, location, PSL, collision, violation, age, season, alcohol, type of road, and hour were negligible. Moderate scenarios represent 31.7% of the real-world crash scenarios, explaining the reason why these two BN changes are small.

Figure 6 shows the BN for motorcycle crashes by considering the counterfactual scenario of all motorcycle injuries as complaint injury crashes. There was a minor

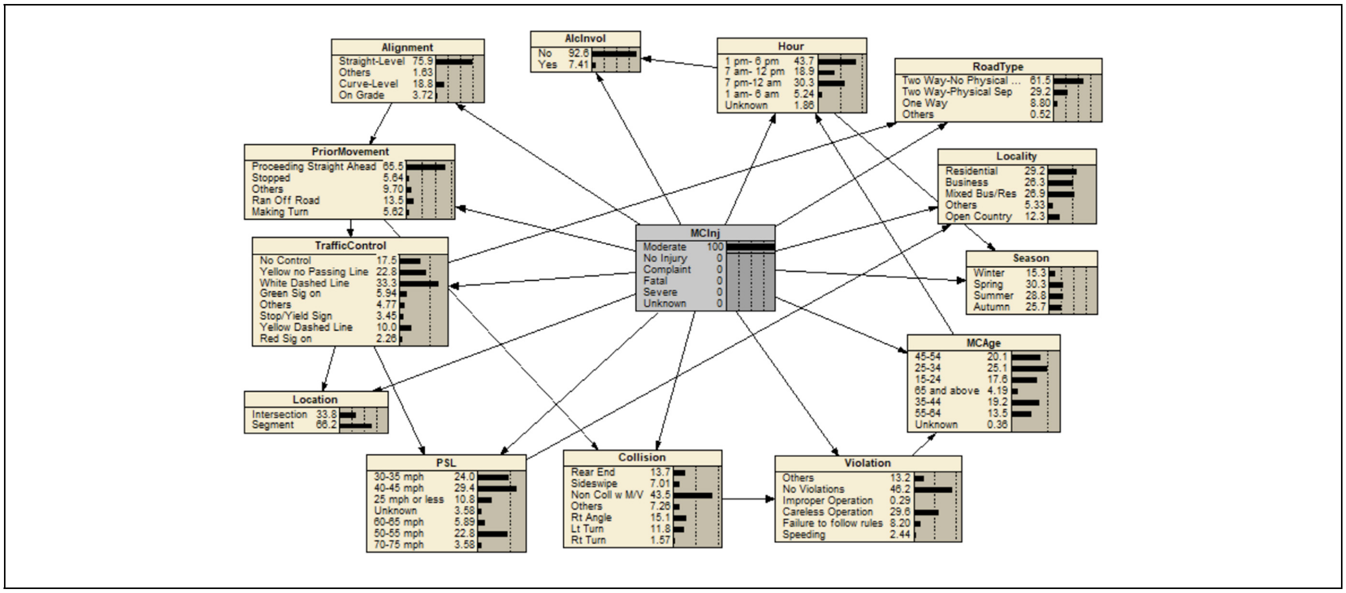

Figure 5. Counterfactual Bayesian network considering all motorcycle injuries are moderate.
Note: PSL = posted speed limit; MC = motorcycle.

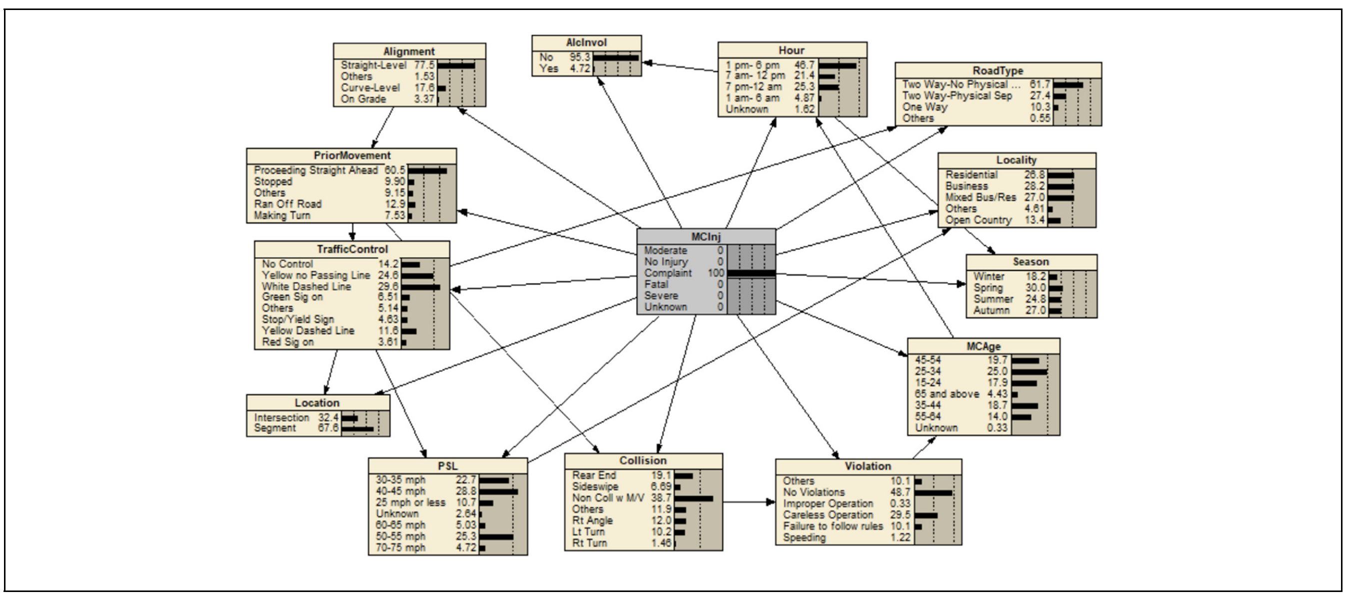

Figure 6. Counterfactual Bayesian network considering all motorcycle injuries are complaint.
Note: PSL = posted speed limit; MC = motorcycle.

change from 7.73% to 4.72% (a decrease of 39%) for alcohol involvement. Changes in all other aspects, including alignment, hour, location, season, type of road, and so forth, were considered negligible. The following scenario represents 32% of the full crash data scenario, so little change can be expected in different aspects of this scenario and the moderate scenario. On the other hand, in scenarios such as fatal, severe, and unknown, we can expect more changes because these scenarios represent a small percentage of the full crash data scenarios.

Figure 7 shows the BN for motorcycle crashes by considering the counterfactual scenario of all motorcycle injuries as no injury crashes. In this scenario, alcohol involvement decreased to 3.98% (a reduction of 48.51%). The rear-end collision type increased to 31.5% for the no injury model (1.65 times increase from the full data). For traffic control devices, stop/yield signs increased to 8.53 (1.71 times the full data), and yellow dashed line decreased to 6.75% (a decrease of 31%). For prior movement, stopped increased to 19.6% (2.02 times

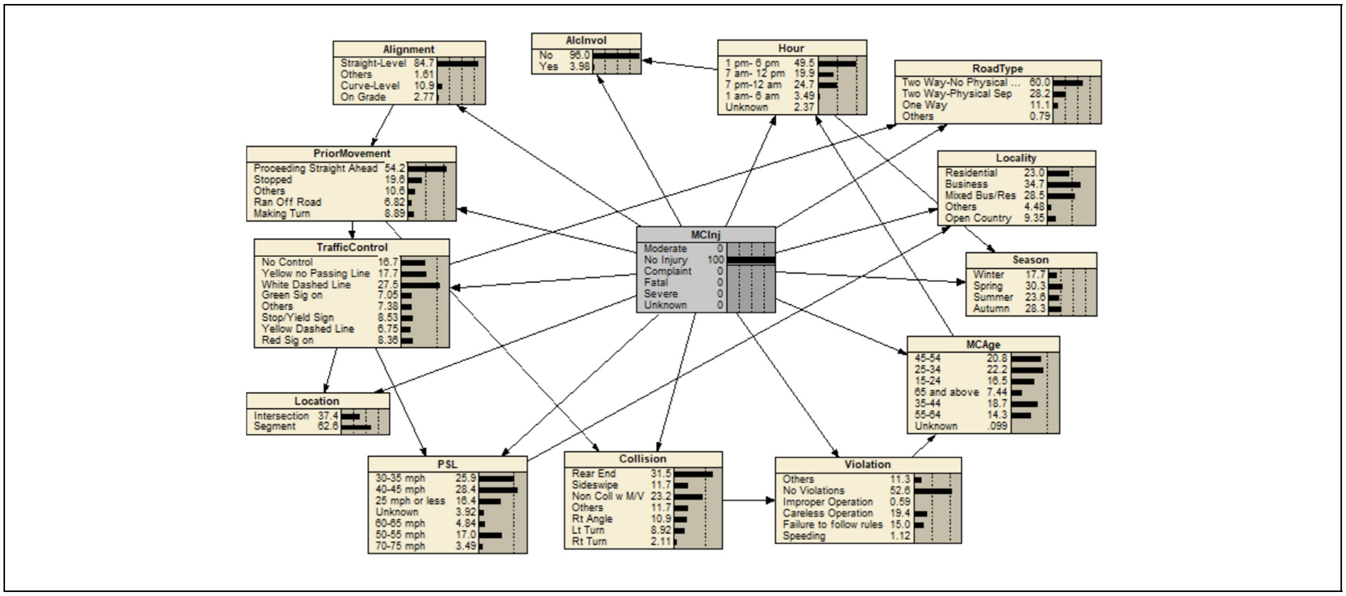

Figure 7. Counterfactual Bayesian network considering all motorcycle injuries are no injury.
Note: PSL = posted speed limit; MC = motorcycle.

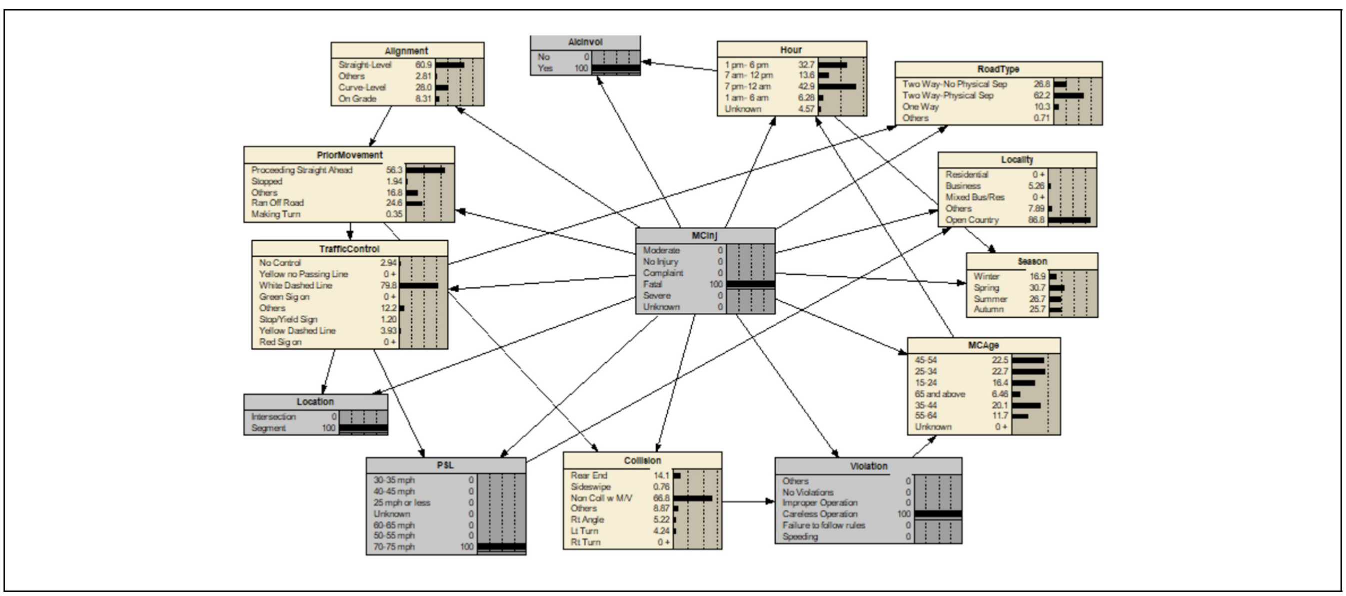

Figure 8. Special case.
Note: PSL = posted speed limit; MC = motorcycle.

the full data), and ran off road decreased to 6.82% (a decrease of 41%). For alignment, curve level decreased to 10.9 (a decrease of 34%). Hour, road type, season, age, violations, PSL, and location had only minor changes from the full data.

One of the main advantages of the BNs is the so-called joint strategies, by which two or more hypotheses may be integrated to determine the probability of another variable. Figure 8 shows the BN for motorcycle crashes if the severe injury crashes, alcohol involvement, segment location, a PSL of 70 to 75 mph, and careless operation violations were all 100% (referred to as the special case model). For hour, 7 pm to 12 pm became the most common time compared with 1 pm to 6 pm in the full data. In the special case model, 7 pm to 12 pm increased to 42.9% (1.5 times the full data) and 1 pm to 6 pm decreased to 32.7% (a decrease of 27%). For road type, two-way–physical separation became the most common type compared with two-way–no physical separation in the original model. Two-way–physical separation increased to 62.2% in the special case model (2.19 times the original model) and two-way–no physical separation decreased to 26.8% (a decrease of 56%). For locality, open country became the most common area, increasing to 86.8% (7.12 times the full data). Non collision with a motor vehicle remained the most common collision type but increased to 66.8% (an increase of 1.87 times the original model). White dashed line also remained the most common traffic control device but increased to 79.8% (2.65 times the original model). Yellow no passing line, green sign on, and red sign on decreased all the way to zero. For prior movement, stopped decreased to 1.94% (a decrease of 80%), and ran off road increased to 24.6% (an increase of 2.12 times the original model). For alignment, curve level increased to 28.0% (an increase of 1.67 times the original model). Age had only minor changes.

Conclusions

Motorcycle safety is a major safety concern. In 2020, the U.S. experienced the highest number of motorcycle fatalities since 1975. Although many studies have explored motorcycle crashes and crash-related injuries, there is still a need for advanced methods to understand the contributing factors and their association patterns. This paper aimed to fill this knowledge gap by conducting an extensive assessment of motorcycle injury severity with the inclusion of a wide range of influential factors using six years of Louisiana crash data (2007 to 2016) via BN analysis.

Counterfactual scenarios were developed to see the sensitivity of the variable attributes. For fatal crash scenarios, alcohol involvement, PSL (50 to 55 mph), off roadway crash, and curve level showed higher probability scores compared with the full data. Locations within business areas showed lower probability scores. For serious injury scenarios, alcohol involvement, rear-end crash, right angle crash, and left turn crash showed higher probabilities compared with the full data. Non-use of traffic control device also show higher probability scores. In moderate injury scenarios, straight-level alignment, non-collisions with motor vehicle, and proceeding ahead crashes showed minor increases in probability scores compared with the full data. Finally, for complaint crash scenarios, alcohol involvement showed a minor increase compared with the full data. The results provided a general overview of the important factors and the level of sensitivity of each of these variables. This study generated BN plots and conditional probability measures for all variables by considering counterfactual scenarios based on motorcycle crash injury types. Overall, this study developed a framework of BN analysis for examining the counterfactual scenarios based on crash injury types. The methodological framework can be reused in addressing other emerging traffic safety issues such as ran-off-road crashes, single-vehicle crashes, and nighttime crashes.

The government can target offenders who engage in risky riding behaviors with education programs across different platforms. Licensing procedures can include motorcycle safety related specific training courses. As the licensing system lacks specific safety focus related educational programs, individuals develop their skills through self-training. As part of the prerequisites for a motorcycle license, education programs can be significantly effective in the long run. To enhance the safety of motorcyclists, we also recommend road-specific adjustments. This trend may be explained by riders' speeding and unsafe maneuvers in the absence of traffic control systems. There can be some benefit in installing traffic signs on such roads, although more serious measures may be needed depending on the condition of the secondary street and the number of motorcycles on the street. It is possible to designate a separate exclusive path for motorcycle riders on secondary road segments with high motorcycle traffic.

Admittedly, there are certain limitations to this research. First, the data set is limited to six years. The current study did not include 2020. Future studies can use 2020 data to perform sensitivity testing on the impact of COVID-19. Second, the current analysis is limited to the defined variables. Some of the variables were later discarded because of missing values. A robust data set with complete information can provide additional insights. For future research, it would be beneficial to examine the influence of policy, enforcement, and regulations on the crash injury types of motorcycle crashes. Additionally, multi-source data can be used to mitigate the current data limitations.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Subasish Das; data collection: Subasish Das; analysis and interpretation of results: Subasish Das; draft manuscript preparation: Subasish Das, V. Vierkant, J. Gonzalez, Boniphace Kutela, and Abbas Sheykhfard. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Subasish Das https://orcid.org/0000-0002-1671-2753
Valerie Vierkant https://orcid.org/0000-0001-7457-497X
Boniphace Kutela https://orcid.org/0000-0002-5450-1623
Abbas Sheykhfard https://orcid.org/0000-0002-9536-3108

References

1. NHTSA. Traffic Safety Facts: Motorcycle. 2020 Data. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813306. Accessed July 8, 2022.
2. Koski, T., and J. M. Noble. *Bayesian Networks, An Introduction*. John Wiley & Sons, Ltd, Chichester, England, 2009.
3. Garces, G., A. Rakotondranaivo, and E. Bonjour. Improving Users' Product Acceptability: An Approach based on Bayesian Networks and A Simulated Annealing Algorithm. *International Journal of Production Research*, Vol. 54, No. 17, 2016, pp. 5151–5168. https://doi.org/10.1080/00207543.2016.1156183
4. Chen, Z., and W. (David) Fan. A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity in North Carolina. *International Journal of Transportation Science and Technology*, Vol. 8, No. 1, 2019, pp. 43–52. https://doi.org/10.1016/j.ijtst.2018.10.001.
5. Medina, A. M. F., and J. C. T. Soto. Roadway Factors Associated with Motorcycle Crashes, ITE Technical Conference and Exhibit, FL, April 4–6, 2011.
6. Kostyniuk, L. P., and A. D. Nation. Transportation Research Institute. Motorcycle Crash Trends in Michigan: 2001-2005, University of Michigan, Transportation Research Institute, Ann Arbor, MI, 2005.
7. Naumann, R. B., A. M. Dellinger, B. A. West, J. L. Annest, and G. W. Ryan. Motorcycle Injuries across the USA, 2001–2008: Injury Comparisons and Projections. *Injury Prevention*, Vol. 16, 2010, pp. A81–A82.
8. Eustace, D., V. K. Indupuru, and P. Hovey. Identification of Risk Factors Associated with Motorcycle-Related Fatalities Ohio. *Journal of Transportation Engineering*, Vol. 137, No. 7, 2011, pp. 474–480.
9. Ryb, G. E., P. C. Dischinger, J. A. Kufera, and T. J. Kerns. Trends in Motorcycle Injuries, Deaths and Case Fatality Rates in the U.S.A. Population. *Annals of Advances in Automotive Medicine*, Vol. 53, 2009, pp. 2.
10. Cheng, W., G. S. Gill, T. Sakrani, M. Dasu, and J. Zhou. Predicting Motorcycle Crash Injury Severity Using Weather Data and Alternative Bayesian Multivariate Crash Frequency Models. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 172–180. https://doi.org/10.1016/j.aap.2017.08.032.
11. Das, S., A. Dutta, K. Dixon, L. Minjares-Kyle, and G. Gillette. Using Deep Learning in Severity Analysis of At-Fault Motorcycle Rider Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672(34): 122–134.
12. Waseem, M., A. Ahmed, and T. U. Saeed. Factors Affecting Motorcyclists' Injury Severities: An Empirical Assessment Using Random Parameters Logit Model with Heterogeneity in Means and Variances. *Accident Analysis & Prevention*, Vol. 123, 2019, pp. 12–19. https://doi.org/10.1016/j.aap.2018.10.022.
13. Shaheed, M. S., and K. Gkritza. A Latent Class Analysis of Single-Vehicle Motorcycle Crash Severity Outcomes. *Analytic Methods in Accident Research*, Vol. 2, 2014, pp. 30–38. https://doi.org/10.1016/j.amar.2014.03.002.
14. Pour-Rouholamin, M., M. Jalayer, and H. Zhou. Modelling Single-Vehicle, Single-Rider Motorcycle Crash Injury Severity: An Ordinal Logistic Regression Approach. *International Journal of Urban Sciences*, Vol. 21, No. 3, 2017, pp. 344–363. https://doi.org/10.1080/12265934.2017.1311801.
15. Chen, S. -J., C. -Y. Chen, and M. -R. Lin. Risk Factors for Crash Involvement in Older Motorcycle Riders. *Accident Analysis & Prevention*, Vol. 111, 2018, pp. 109–114. https://doi.org/10.1016/j.aap.2017.11.006.
16. Vajari, M. A., K. Aghabayk, M. Sadeghian, and N. Shiwakoti. A Multinomial Logit Model of Motorcycle Crash Severity at Australian Intersections. *Journal of Safety Research*, Vol. 73, 2020, pp. 17–24. https://doi.org/10.1016/j.jsr.2020.02.008.
17. Agyemang, W., E. K. Adanu, and S. Jones. Understanding the Factors That Are Associated with Motorcycle Crash Severity in Rural and Urban Areas of Ghana. *Journal of Advanced Transportation*, Vol. 2021, 2021, 6336517. https://doi.org/10.1155/2021/6336517.
18. Tamakloe, R., S. Das, E. N. Aidoo, and D. Park. Factors Affecting Motorcycle Crash Casualty Severity at Signalized and Non-Signalized Intersections In Ghana: Insights From A Data Mining And Binary Logit Regression Approach. *Accident Analysis & Prevention*, Vol. 165, 2022, p. 106517. https://doi.org/10.1016/j.aap.2021.106517.
19. Truong, L. T., H. T. T. Nguyen, and R. Tay. A Random Parameter Logistic Model of Fatigue-Related Motorcycle Crash Involvement in Hanoi, Vietnam. *Accident Analysis & Prevention*, Vol. 144, 2020, 105627. https://doi.org/10.1016/j.aap.2020.105627.
20. Wali, B., A. J. Khattak, and A. J. Khattak. A Heterogeneity Based Case-Control Analysis of Motorcyclist's Injury Crashes: Evidence from Motorcycle Crash Causation Study. *Accident Analysis & Prevention*, Vol. 119, 2018, pp. 202–214. https://doi.org/10.1016/j.aap.2018.07.024.
21. Chawla, H., I. Karaca, and P. T. Savolainen. Contrasting Crash- and Non-Crash-Involved Riders: Analysis of Data from the Motorcycle Crash Causation Study. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2673, No. 7, 2019, pp. 122–131. https://doi.org/10.1177/0361198119851722.
22. Das, S., A. Dutta, and I. Tsapakis. Topic Models from Crash Narrative Reports of Motorcycle Crash Causation Study. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2675, No. 9, 2021, pp. 449–462. https://doi.org/10.1177/03611981211002523.
23. Friedman, N., D. Geiger, and M. Goldszmidt. Bayesian Network Classifiers. *Machine Learning*, Vol. 29, 1997, pp. 131–163.
24. Das, S. *Artificial Intelligence in Highway Safety*. CRC Press, Boca Raton, FL, 2022.
25. Howard, R., and J. Matheson. *Influence Diagrams. The Principles and Applications of Decision Analysis*. Strategic Decisions Group, Menlo Park, CA, 1984.

26. Pearl, J. *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Reasoning*. Morgan Kaufmann Publishers, Los Altos, 1988.
27. Shachter, R. Model Building with Belief Networks and Influence Diagrams. In *Advances in Decision Analysis: From Foundations to Applications* (W. Edwards, R. Miles, and D. Von Winterfeldt, eds.), Cambridge University Press, Cambridge, 2007, pp. 177–207.
28. Das, S., J. Gonzalez, J. Liu, and X. Kong. Encroachment-related Work zone Crash Injury Analysis using Data-driven Bayesian Network. Transportation Research Board Annual Meeting, Washington DC, IEEE, New York, 2023.
29. Norsys. (2020). Netica 6.04. https://www.norsys.com/netica.html. Accessed July 12, 2022.
30. Lauritzen, S. L. The EM Algorithm for Graphical Association Models with Missing Data. *Computational Statistics& Data Analysis*, Vol. 19, 1995, pp. 191–201.
31. Kutela, B., S. Das, and B. Dadashova. Mining Patterns of Autonomous Vehicle Crashes Involving Vulnerable Road Users to Understand the Associated Factors. *Accident Analysis & Prevention*, Vol. 165, 2022, 106473.
32. Das, S., B. Brimley, T. Lindheimer, and A. Pant. *Safety Impacts of Reduced Visibility in Inclement Weather*. Center for Advancing Transportation Leadership and Safety (ATLAS Center). UTC Center, Michigan, 2017.
33. Sheykhfard, A., F. Haghighi, G. Fountas, S. Das, and A. Khanpour. How do Driving Behavior and Attitudes toward Road Safety Vary between Developed and Developing Countries? Evidence from Iran and the Netherlands. *Journal of Safety Research*, 2023. https://doi.org/10.1016/j.jsr.2023.02.005.
34. Das, S., and I. Tsapakis. Interpretable Machine Learning Approach in Estimating Traffic Volume on Low-Volume Roadways. *International Journal of Transportation Science and Technology*, Vol. 9, No. 1, 2020, pp. 76–88.