ARTICLE IN PRESS

Safety Science xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Safety Science

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

Sensitivity analysis of driver's behavior and psychophysical conditions

Susana García-Herrero[a,*], J.M. Gutiérrez[b], Sixto Herrera[c], Amin Azimian[a], M.A. Mariscal[a]

[a] Escuela Politécnica Superior, Universidad de Burgos, Spain
[b] Instituto de Física de Cantabria, Consejo Superior de Investigaciones científicas (CSIC), Spain
[c] Dep. of Applied Mathematics and Computer Sciences, Universidad de Cantabria, Spain

ARTICLE INFO

Keywords:
Traffic accident
Drugs
Alcohol
Speed
Distraction
Human error
Bayesian network
Bias identification

ABSTRACT

To reduce traffic accidents, an accurately estimated model is needed to capture the true relationships between the injury severity and risk factors. This study aims to propose a robust procedure to address the biases in police-reported accident data and subsequently to conduct sensitivity analyzes in order to estimate the variations in injury severity and distraction probability based on drivers' behaviors/characteristics and psychophysical conditions. The results show that: (i) the excess speed will likely increase the probability of serious/fatal injury for drivers of all age groups by 10%; (ii) distraction and driver' errors will likely increase the probability of serious/fatal injury in all drivers driving at a proper speed up to 1.5%; (iii) alcohol and drug consumption can significantly increase the probability of being distracted and making errors by 28.5% and 33.5% respectively; (iv) Alcohol consumption reduces the probability of driving at an appropriate speed in drivers under 25 by 40%. However, the results for drugs consumption are not as significant as the ones for alcohol consumption.

1. Introduction

Road safety is a vitally important topic nowadays, given the chilling figures of the consequences of traffic accidents. According to the World Health Organization, around 1.35 million people die each year as a result of traffic accidents and up to 50 million people suffer from non-fatal injuries around the world (WHO, 2018), many of these injuries causedisability and pose significant financial consequences to victims, their families and countries which are equivalent to 0.4–4.1% of gross domestic product (Wijnen et al., 2019). In Spain, as in other industrialized countries, traffic accidents are one of the main causes of death within the occupational accident category. According to Directorate-General for Traffic (DGT) which is responsible for collecting traffic accident data in Spain, there were 44,017 drivers involved in traffic accidents on interurban roadways in 2016, of which 720 drivers were killed, 2,752 drivers were injured and hospitalized and 22,861 were injured without hospitalization (DGT, 2016).

Spain has been generally concerned with alcohol and drug use while driving and many legislative efforts have been made to mitigate traffic accidents of this nature. For instance, Spain has strict legal limit for alcohol in blood which is 0.5 g/l. In addition, there is no allowable limit for illicit drugs. Nevertheless, alcohol has a major role in a high percentage of Spain's road accidents. According to the crash statistics provided by DGT, 68% of drivers involved in casualty accidents were tested for alcohol in 2016. As for drivers under influence, the percentage of positive alcohol tests on interurban roads increases with the injury severity, from 5% in uninjured drivers to 25% in fatally injured (DGT, 2016).

Given the importance of alcohol and drug use in road safety, many research efforts have attempted to investigate their impacts on accident severity. However, most of these studies relied on police-reported accident data and ignored the reporting biases. The biases in accident data occur because of underreporting effects, especially for non-severe injury severities. As a result, the crash dataset degrades to an outcome-based sample overrepresented by fatal or serious injury severities (Yamamoto et al., 2008). This leads to biased estimates and erroneous inferences on the impact of critical variables such as visibility condition, alcohol and drug use.

Therefore, despite the fact that the negative impacts of alcohol/drug on driving have been already proved, quantifying their influences through the use of real data is complicated (especially for the associated bias problems) (Gjerde et al., 2019) and is still an open area of research (Kwon et al., 2015). Hence, the main objectives of this work are to: (a) develop a data-driven accident statistical dependence model to understand how changes in the states of important variables such as speed violation, driver's distraction and errors, and alcohol/drugs use would explain the variation in the probability of fatal/serious injury while addressing the reporting biases, and (b) understand the extent to which

* Corresponding author.
E-mail address: susanagh@ubu.es (S. García-Herrero).

https://doi.org/10.1016/j.ssci.2019.104586
Received 29 March 2019; Received in revised form 12 December 2019; Accepted 20 December 2019
0925-7535/

Please cite this article as: Susana García-Herrero, et al., Safety Science, https://doi.org/10.1016/j.ssci.2019.104586

the alcohol/drug consumption contribute to the probability of being distracted and making errors while driving.

2. Literature review

Driving under the influence of drugs and/or alcohol is a critical risk factor that impairs driving skills (Behnood and Mannering, 2017; Bukova-Zideluna and Villerusa, 2016; de Oña et al., 2014; Goss et al., 2008; Khanjani et al., 2017; Valen et al., 2019) because of reduced sense of risk and reaction time, drowsiness, speeding and aggressive behavior behind the wheel (Madrid, 2010). Excess speed or aggressive driving is considered as one of the most influential driving behaviors in fatal accidents (WHO, 2017), as slight increases in speed considerably raise the risk of being in an accident and the severity of the injuries (Castillo-Manzano et al., 2019; Pešić et al., 2019).

2.1. Trends in alcohol-related accidents

Traffic accidents caused by alcohol or drugs consumption are more common in a specific circumstance or in specific groups of the population rather than in others. For example, Owen et al. (2019) explained that accidents caused by alcohol consumption are more frequent in drivers between 25 and 35 years old and at night between 3 pm and 4 pm. Also, Gómez-Talegón et al. (2012) reported that the consumption of alcohol and drugs while driving are more common in young men on urban roads and on weekend nights. Also, Papalimperi et al. (2019) reported that fatal alcohol-related traffic accidents are more likely to happen during weekends and the summer period than during weekdays. Wu and Zhang (2018) demonstrated that the risk of having severe injuries is higher for alcohol-impaired drivers when the drivers are 65 years old or older and when the drivers make left turns in intersections. Valen et al. (2019) reported that at least one of the risk factors speeding, non-use of a seatbelt/helmet, and driving without a valid license were present among most of the drug/alcohol-impaired drivers fatally injured.

With the aim of reducing the consumption and the consequences of alcohol and drugs in road traffic, different approaches have been proposed: revising the legislative limits of blood alcohol (Ferrari et al., 2018; Pešić et al., 2019), increasing the public education campaigns, improving rehabilitation programs, promoting research, improving data collection, etc. For example, Pešić et al. (2019) suggested that the blood alcohol concentration limit should be the same for professional and non-professional drivers, because different limits confuse drivers. On the other hand, Ferrari et al. (2018) reported that a blood alcohol concentration below the legal limits is not likely a risk factor for accident occurrence.

2.2. Role of visibility in traffic safety

Prior studies also assessed the impact of light condition on accident severity. For example, Behnood and Mannering (2017) reported that accident occurring in daylight are more likely to increase the risk of minor injury and less likely to cause no injury or severe injury to drivers. Moreover, they found that dark and lighted roadways may increase the probability of no injury for the drivers under the influence of alcohol/drug. In contrast, Wang and Zhang (2017) showed that accident severity is likely to be higher in rural roadways, major arterials, not at intersection locations, locations with curves, dark and unlighted roadways, dry roadway and streets with high speed limits.

2.3. Role of distraction in traffic safety

Another aspect that can be detrimental for drivers is distraction caused by secondary tasks which can divert driver's attention away from the activities needed for safe driving (Bowden et al., 2019; Neyens and Boyle, 2008; Papantoniou et al., 2019; Sundfør et al., 2019). Secondary tasks include interaction with in-vehicle information systems (Reyes and Lee, 2008; Strayer et al., 2016), talking with passengers, texting or calling (Aksjonov et al., 2019), using intelligent personal assistant (Strayer et al., 2017), etc. For example, Bowden et al. (2019) reported that one-minute distractions is likely to negatively impact driver's performance for 40 s post-distraction. Moreover, Bowden et al. (2019) Impairs driver's response time and increase the speed variations during 0–20 s post-distraction. Besides, other researchers (Donmez and Liu, 2015; Neyens and Boyle, 2008) demonstrated that teenage and older drivers (65+) are more likely to sustain severe when engaged in phone conversation. In general, review of existing literature indicate that previous research mostly attempted to measure the impact of distraction on driver's performance and only few of them (Donmez and Liu, 2015; Neyens and Boyle, 2008) have attempted to quantify the impact of driver's distraction/error on injury severity. It should be noted that although the impact of alcohol on driver's injury, and the impact of distraction on driver's performance were studied in past research. But the literature still lacks the quantification of the impact of alcohol/drug consumption on driver's distraction/errors. For more information regarding the factors affecting injury severity please refer to the existing driver injury severity studies listed in Table 1 in Appendix A.

2.4. Existing methods in accident severity modeling

The majority of previous research efforts that investigated the effects of alcohol/drug, distraction and driver's errors on accident severity highly relied on classical statistical methods such as Logistic Regression (Buendia et al., 2015; Koopmans et al., 2015); Ordered Probit (Chiou et al., 2013) among others. However, in the last decades, the development of new data mining and machine learning techniques, together with the availability of data and computing resources have allowed researchers to apply these techniques to traffic safety field. The works of Chang and Wang (2006) and Halim et al. (2016) contain a sample review of the bibliography on the different techniques used to analyze and predict traffic accidents, justifying the use of current data mining techniques, such as the classification and regression trees, genetic algorithms, artificial neural networks, principal component analysis and fuzzy logic.

Among new data-mining techniques, the Bayesian Networks have been increasingly applied to the traffic accident studies (De Oña et al., 2013; de Oña et al., 2011; García-Herrero et al., 2016; Gregoriades and Mouskos, 2013; Liang and Lee, 2014; Mujalli and De ONa, 2011; Sun and Sun, 2015). Castro and Kim (2016) developed different accident severity models based on Bayesian networks, decision trees and artificial neural networks. Comparison of their models indicated that the Bayesian Networks outperformed other models in terms of accuracy. This finding confirms that the Bayesian Network would be an ideal method to evaluate the severity of traffic accidents, analyze their causes and risks and predict the likelihood of serious and fatal traffic accidents (Zong et al., 2013). Additionally, due to flexibility of Bayesian Networks, they could be used in combination with other statistical methods to analyze traffic accidents (Chen et al., 2015; Gregoriades et al., 2012).

Despite the vast availability of literature on accident severity analysis, most studies util1ized the frequentist approaches like logit models in different forms and very few studies have employed the Bayesian Networks in this context (see Table 1 in Appendix A). Bayesian Networks provides probabilistic presentation of the interactions and gives better estimation of risk and uncertainties compared to the frequentist models that only account for the expected values (Uusitalo, 2007). Furthermore, the sensitivity analysis in Bayesian Networks allows to measure the variation in a target variable in relation to other variables.

2.5. Research contributions

In general, the major contributions of this paper are as follows:

Table 1
Literature review on accident severity analysis.

	Paper	Objective	Key findings	Drivers' behaviors/ characteristics	Road characteristics	Vehicle characteristics	Environmental conditions	Psychophysical conditions	Distraction/ Error	Bias identification
Bayesian Networks	(De Oña et al., 2013)	To analyze traffic accident injury severity on rural highways using Latent class clusters and Bayesian network	The combined use of both methods (LCC and BNs) provide new information and insights on the main causes of accident severity that could be useful for road safety analysts.	X	X		X			No
	(de Oña et al., 2011)	To model accident severity using Bayesian network	Accident type, driver age, lighting and number of injuries are associated with a fatal or seriously injured accident	X	X	X	X			No
	(García-Herrero et al., 2016)	To identify the variables that most influence the injuries sustained by motorcycle and moped riders.	Influential factors were speeding, safety device infractions. The injuries were worse for riders younger than 24 and older than 48. Men were also more susceptible to being fatally or seriously injured.	X	X		X	X		No
	(Mujalli and De ONa, 2011)	To determine if it is possible to maintain or improve the performance of a model that is used to predict the injury severity of a traffic accident based on BNs reducing the number of variables considered in the analysis.	It is possible to reduce the number of variables used to model traffic accidents injury severity through BNs without reducing the performance of the model.	X	X		X			No
Other Methods	(Behnood and Mannering, 2017)	To identify factors affecting the injury severities of unimpaired, alcohol-impaired, and drug-impaired drivers	Unimpaired drivers are more responsive to variations in lighting, adverse weather, and road conditions. Age and gender were found to be important determinants of injury severity, but the effects varied significantly across all drivers, particularly among alcohol impaired drivers.	X	X		X	X		No
	(Neyens and Boyle, 2008)	To predict the likelihood of a severe injury for teenage drivers and their passenger using a national crash database	Teenage drivers have an increased likelihood of more severe injuries if distracted by a cell phone or by passengers than if the source of distraction was related to in-vehicle devices	X					X	No
		To compare the risk of road traffic injury in	Among child pedestrians and young drivers, males present	X		X				No

(continued on next page)

Table 1 (*continued*)

Paper	Objective	Key findings	Drivers' behaviors/ characteristics	Road characteristics	Vehicle characteristics	Environmental conditions	Psychophysical conditions	Distraction/ Error	Bias identification
(Santamariña-Rubio et al., 2014)	men and women, by age, mode of transport and injury severity, using the time people spend travelling as the exposure measure, in Catalonia for the period 2004–2008.	higher risk of slight and severe injury, and in the oldest groups women present higher risk. The death rate is always higher in men. There exists interaction between sex and age in road traffic injury risk. That is, injury risk is higher among men in some age groups, and among women in other groups, but these age groups vary depending on mode of these age groups vary depending on mode of							
(Bedard et al., 2002)	To identify the independent contribution of driver, crash, and vehicle characteristics to fatal injuries sustained by drivers.	Seatbelt use, reducing speed, and reducing the number and severity of driver-side impacts may prevent fatalities. Female gender (OR = 1.54) and blood alcohol concentration > 0.30(OR = 3.16) were also associated with higher fatality odd.	X		X		X		No
(Kashani and Mohaymany, 2011)	To identify the factors influencing crash injury severity on the roads in Iran using classification and regression trees (CART)	Improper overtaking and not using a seatbelt are the most important factors affecting the severity of injuries.	X	X		X			No
(Waller et al., 2003)	To examine alcohol's role in injuries, considering other important factors using regression analysis	Best predictors of injury severity were vehicle crush (TAD), safety belt use, and their interaction, and age. Also, alcohol was found to increase injury severity score	X				X		No
(Lemp et al., 2011)	To examine the factors affecting crash severity for persons involved in heavy duty truck crashes	The likelihood of fatalities and severe injury is estimated to rise with the number of trailers but fall with the truck length and gross vehicle weight rating (GVWR).	X	X	X	X			No
(Yasmin and Eluru, 2013)	To assess the injury severity of drivers of passenger vehicles	The likelihood of sustaining possible, on-incapacitating and incapacitating/fatal injuries is higher for crashes on both the medium and higher speed limit roads compared to the crashes on lower speed limit roads.	X	X	X	X	X		No
(Morgan and	To assess the effects that age, gender, and other factors have on	For all females and older males, the likelihood of severe injuries increased when crashes occurred	X	X	X	X			No

(*continued on next page*)

Table 1 (*continued*)

Paper	Objective	Key findings	Drivers' behaviors/ characteristics	Road characteristics	Vehicle characteristics	Environmental conditions	Psychophysical conditions	Distraction/ Error	Bias identification
Mannering, 2011)	crash severities by considering single-vehicle crashes that occurred on dry, wet, and snow/ice-covered roadway surfaces.	on wet or snow/ice surfaces–but for male drivers under45 years of age, the probability of severe injuries decreased on wet and snow/ice surfaces – relative to dry-surface crashes.							
(Kim et al., 2013)	To model driver-injury severity in single-vehicle crashes	Male driver, drunk driving, unsafe speed, older driver (65+) driving an older vehicle, and darkness without street lights increased the probability of fatal injury.	X		X	X			No
(Xie et al., 2012)	To analyze injury severities involving single-vehicle crashes on rural roads.	Driver age, DUI, seat belt usage, lighting condition and speed were associated to driver injury severity levels. Vehicle age and surface condition had no significant impact on driver injury severity.	X	X	X	X			No
(Eluru et al., 2012)	To identify the different factors that influence injury severity of highway vehicle occupants, in particular drivers, involved in a vehicle-train collision at highway-railway grade crossings.	The key factors influencing injury severity include driver age, time of the accident, presence of snow and/or rain, vehicle role in the crash and motorist action prior to the crash	X	X	X	X			No
(Paleti et al., 2010)	To examining the influence of aggressive driving behavior on driver injury severity in traffic crashes	Young drivers, drivers who are not wearing seat belt, under the influence of alcohol, not having a valid license, and driving a pick-up are found to be most likely to behave aggressively	X	X	X	X			No
(Feng et al., 2016)	To investigate the underlying risk factors of fatal bus accident severity to different types of driver	Some variables such as driver's age, driver's gender, risky behaviors and restraint system have similar impact on different type of drivers.	X	X	X	X			No
(Zou et al., 2017)	To investigate the differences between single-vehicle and multi-vehicle truck crashes	There exists a substantial difference between factors influencing single-vehicle and multi-vehicle truck crash severity. Also, Heterogeneity does exist in the truck weight, and it behaves differently in single-vehicle and multi-vehicle truck crashes.		X	X				No

(*continued on next page*)

Table 1 (*continued*)

Paper	Objective	Key findings	Drivers' behaviors/ characteristics	Road characteristics	Vehicle characteristics	Environmental conditions	Psychophysical conditions	Distraction/ Error	Bias identification
(Haghighi et al., 2018)	To examine the impact of a wide range of roadway geometric features on the severity outcomes of crashes occurred on rural two-lane highways.	There was a significant variation in severity outcomes of crashes occurred across segments which verifies the presence of hierarchical structure. Lower risk of severe crashes was found to be associated with the presence of 10-ft lane and/or narrow shoulders, lower roadside hazard rate, higher driveway density, longer barrier length, and shorter barrier offset.		X					No
(Chen et al., 2016)	To Investigate driver injury severity patterns in rollover crashes using support vector machine models	Comfortable driving environment conditions, driver alcohol or drug involvement, seatbelt use, number of travel lanes, driver demographic features, maximum vehicle damages in crashes, crash time, and crash location are significantly associated with driver incapacitating injuries and fatalities.	X		X	X	X		No
(Li et al., 2017)	To identify the association between one vehicle's attributes, as well as roadway engineering, environmental and crash characteristics, and the injury severity of occupants in the partnering vehicle in two-vehicle crashes	Inattentive driving, left-turn movement, heavy vehicle type of the at-fault vehicle, and angle and rear-end impact type increased probability of more severe injuries of not-at-fault vehicle occupants. Moreover, for at-fault vehicles, their probability of more severe injury was positively associated with inattentive driving and heavy vehicle type of the not-at-fault vehicle, and angle and approaching impact type.	X		X	X		X	No
(Naik et al., 2016)	To investigate the relationship between single-vehicle truck crash injury severity and detailed weather conditions.	Wind speed, rain, humidity, and air temperature were associated with single-vehicle truck crash injury severity	X	X		X			No
(Wang et al., 2019)	To simultaneously estimate the four common intersection crash consequence metrics, driver error, crash type, vehicle damage and injury severity by accounting for potential correlations due to common observed and unobserved factors	Driver's age, driving under the influence of drugs and alcohol, intersection geometry and control types, and adverse weather and light conditions are the most critical factors contributing to severe crash consequences. Moreover, driving under the influence of drugs or alcohol increases the possibility of reckless driving errors.		X		X	X	X	No

(*continued on next page*)

Table 1 (*continued*)

Paper	Objective	Key findings	Drivers' behaviors/ characteristics	Road characteristics	Vehicle characteristics	Environmental conditions	Psychophysical conditions	Distraction/ Error	Bias identification
(Osman et al., 2018)	To analyze passenger-car crash injury severity in different work zone configurations	Partial control of access, roadways classified as rural, crashes during evening times, crashes during weekends, and curved roadways are key factors that increase the likelihood of severe outcomes.		X		X			No
(Sundfør et al., 2019)	To conduct a comprehensive mapping of the types of inattention that contribute to fatal road crashes.	Failure to check for information in blind spots or behind other sight obstructions is a typical form of inattention. Distraction by use of mobile phones contributed to between two and four percent of all fatal crashes, while other sources of distraction, within or outside of the vehicle, contributed to about ten percent.						X	No
(Yu et al., 2019)	To investigate highway single-vehicle crashes and the effects of significant contributing factors to driver injury severity	Urban indicator and principle road type were found to be random parameters and have significant heterogeneity in the mean as a function of male indicator and driver's age indicators.	X	X	X	X			No
(Wu and Zhang, 2018)	To identify and quantify the impacts of alcohol/non-alcohol-influenced driver's behavior and demographic features as well as geometric and environmental characteristics on driver's injury severities around intersections.	The probability of having severe injuries is higher for non-alcohol-influenced drivers when the drivers are 65 years old or older.	X	X	X	X	X		No
(Wang and Zhang, 2017)	To identify and quantify the impacts of several key roadway and environmental factors to the traffic crash severities	Higher crash severity are associated with rural roadways, major arterials, not at intersection locations, locations with curves, during nighttime when it is dark without street light, dry roadway conditions, and high speed limits.		X		X			No

Table 2
Database. Records in the database for the states of each variable.

Variables	Number of samples	% of Cases	Comments
Type of Vehicle			
Car	107,427	88,9%	Car
Van	10,768	8,9%	Van
Off-the-road vehicles	2636	2,2%	Off-the-road vehicles
Anomalies			
Yes	78,256	64,8%	With anomalies in the tires, blowout, address, brakes or other anomalies
None	853	0,7%	Without any anomalies in the car
Unknown	41,722	34,5%	Unspecified
Maneuver			
Normal driving	40,262	33,3%	Following straight path
Overtaking	1220	1,0%	Overtaking
Fast maneuver	877	0,7%	Fast maneuver to save obstacle, pedestrian or animal
Others	41,047	34,0%	Taking a curve, changing lanes, circulating reverse, crossing the road, entering the circulation, spinning, detained, braking, stopped or parked
Unknown	37,425	31,0%	Unspecified
Zone			
Inter-Urban	48,109	39,8%	Roads
Crossing road	1846	1,5%	Section of interurban road that runs through urban land
Urban	69,585	57,6%	Street
Motorway	1255	1,0%	Motorway or urban highway
Condition-Firme			
Good	97,242	80,5%	Dry and clean
Bad	17,675	14,6%	With mud or loose gravel, wet, very waterlogged or flooded, with ice snow or oil and others
Unknown	5914	4,9%	It is unknown or not specified
Meteo			
Good	84,776	70,2%	Clear day
Bad	17,149	14,2%	Cloudy, weak rain, strong rain, hailing, snowing
Unknown	18,906	15,6%	It is unknown or not specified

Table 3
Variable: Overall severity of accident. Labels and percentage of cases.

Overal accident severity state	Label	Percentage
1	Fatal or serious injury	7,46%
2	Minor injury	92,54%

1. Develop a Bayesian Network model while accounting for the biases in accident dataset.
2. Perform sensitivity analysis to determine the degree to which variation in fatal/serious injury probability is explained by driver's behavior/characteristics and psychophysical conditions.
3. Perform sensitivity analysis to determine how changes in alcohol/drug consumption would affect the probability of committing speed violations, driving distractedly and making errors while driving.

3. Data collection

3.1. Statistical questionnaire of traffic accidents with victims

The statistical questionnaire of traffic accidents with victims is a tool established by DGT (BOE, 1993, 2014) to gather information related to the traffic accident such as accident date, accident location, charactersitics of accident-involved vehicles and persons, road characteristics, accident type, accident victims, number of fatalities, trip purpose, drivers' actions, presumed violations made by the driver (administrative, speeding and others), psychophysical conditions of the drivers, etc. The statistical questionnaires are generally collected by the Civil Guard General Directorate or by a local police officers and are then recorded in the ARENA2 software upon approval of the National Registry of Traffic Accident Victims. The ARENA2 is a "*traffic accident information system that is part of the computer infrastructure of the DGT and is designed to gather, store, integrate, distribute and use accident data in Spain. It includes all of the information sources that exist* nationally and can be consulted and used as a source of data by all direct consumers of accident information" (Ramos, 2012).

3.2. Database

The database used in our research was provided by DGT. This database consists ofthree data tables related to the traffic accidents that occured in Spain in 2016. The first one contains general data on accidents consisting of a total of 102,362 records. The second includes data on people, consisting of a total of 174,679 records of drivers involved in accidents. Finally, the third table contains microdata on vehicles, containing a total of 179,295 records of vehicles involved in accidents. In order to relate the tables, two identifiers were used, ID_Accidente and ID_vehiculo, which assign to each accident and vehicle, respectively, a unique identification. These relationships are based on the structure of the accident victim database, in which each record in the accident table have one or more vehicles involved in an accident, and each accident-involved vehicle may have zero (p.e. a parked car) or one driver.

The database used in this study is based on the Table of Drivers involved in traffic accidents with victims in Spain in 2016 which includes a total of 174,679 records. First, the database was filtered to classify the drivers based on vehicle type; cars, vans and off-the-road vehicles. Second, a new variable called Overall Severity was extracted from the table of accidents which represents the injury severity for drivers, vehicle occupants and pedestrians. Overal severity is used as the target variable in this study and has two possible states:

(1) Fatal/Serious Injury (F/SI) if at least one of the vehicle occupants (drivers and passengers and pedestrians) was fatally or seriously injured (***F/SI***) in the accident, and
(2) Minor/No Injury (M/NI) if the vehicle occupants and pedestrians had no or minor injury

As a result, the final database contains 120,831 records for the drivers involved traffic accident. The variables included in the model, the states of each variable, the number of cases, its percentage and the explanation of each state are show in Table 2 in Appendix B. In addition, the percentage of records in the database for each state of the Overall Severity, with respect to the total, are shown in Table 3.

4. Methodology

This section first describes the overall research procedure and principles underlying the development of the proposed approach (Section 4.1). Then, it provides a summary of the Bayesian Network methodology and then explains how it is used to perform sensitivity analysis (Section 4.2).

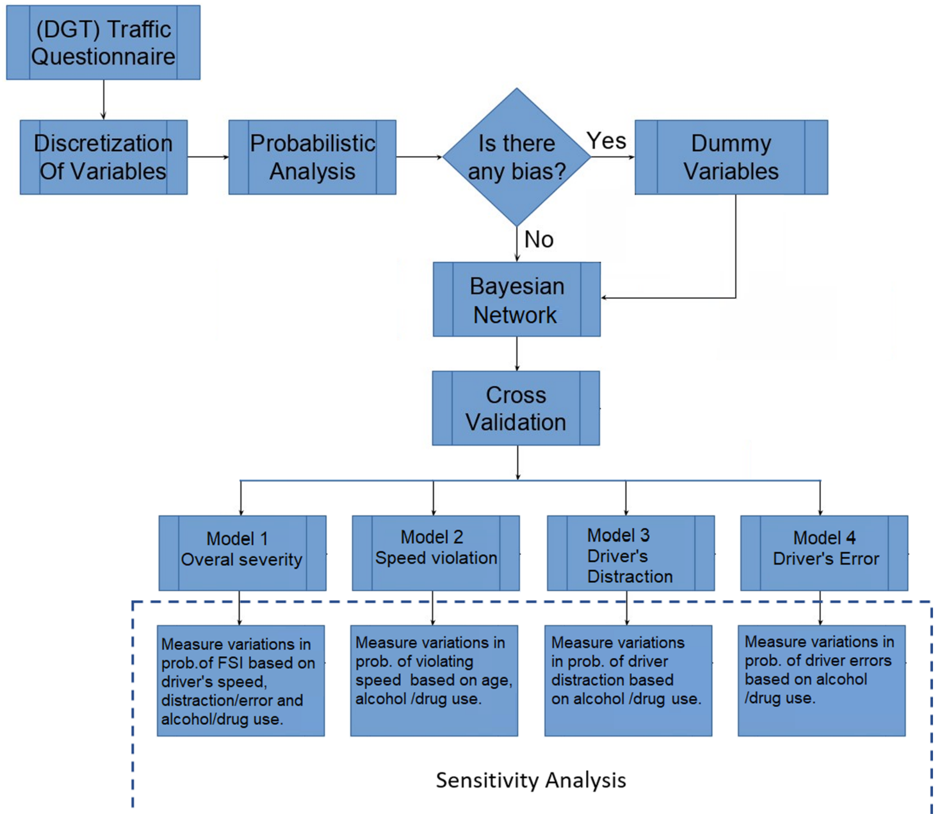

Fig. 1. Flowchart of overall methodology.

4.1. Overall procedure

The Bayesian network proposed in this study represents the impacts of drivers' characteristics (e.g., age and gender), behavioral factors (e.g., speed violations, distractions and errors) and psychophysical conditions (e.g., alcohol/drugs use) on accident severities. Furthermore, the model can be used to analyze the differences among the likelihood of violating speed limit, driving distractedly or making errors as a function of alcohol and drug use. As shown in Fig. 1, once the accident data collected from the DGT Traffic questionnaire and the variables of interest were identified and discretized which is discussed in Section 3, they will be analyzed to identify possible biases. In order to detect the biased variables, for each variable, the difference between the percentage of severe injuries for drivers with known states (e.g., positive or negative alcohol test result) and the percentage of severe injuries for the drivers with unknown state (e.g., unknown alcohol test result) will be estimated. In this sense, if the difference is significant, then this implies that the variable is biased, and a dummy variable is needed to be defined for the variable in order to isolate homogeneous samples. In the next, the Bayesian network will be learned from all variables including the dummy variables to estimate the conditional probability for fatal/injury severity. The resulting model will be evaluated using 10-fold cross-validation approach to assess its prediction accuracy. Later on, four models/classifiers will be considered from the Bayesian Network to conduct sensitivity analysis. The results from the first model will be used to estimate the variations in the probability of fatal or serious injurie as a function of speeds, drivers' errors and distraction, gender and age. The second model will be used to estimate the variations in the probabilities of driver error as a function of alcohol and drugs. The third model will determine the variations in the probability of speed violation as a function of age, alcohol and drugs, and the last model will determine the variations in the probability of speed violation as a function of the age of the driver and alcohol and drugs uses.

4.2. Bayesian networks

Bayesian networks (BNs) are probabilistic graphical models based on directed acyclic graphs (DAG) which combine probability and graph theories to efficiently learn the joint probability distribution (JPD) of a multivariate problem involving discrete variables. As a result, Bayesian networks explicitly represent our knowledge of the given problem in probabilistic terms through the DAG and the joint probability distribution (JPD) of the variables that comprise it (Castillo, Gutiérrez, and Hadi, 1997):

$$p(x) = p(x_1, \ldots, x_n) \tag{1}$$

where x_i corresponds to a realization of the aleatory variable X_i. The particular JPD structure for a given problem is obtained by a factorization (using the Bayes rule) as a set of conditional probability functions, which are obtained from the dependence/independence structure among the variables reflected in the DAG. This allows to factorize the JPD using the product of several conditional probabilities, as follows:

$$p(x_1, x_2, x_3, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \vee \pi_i) \tag{2}$$

where π_i is the set of parents of node X_i in the graph. The independencies in the graph are thus immediately translated into the probabilistic model in a very practical manner. In this study, this methodology was used to build probabilistic models reflecting the significant relationships (probabilistic dependences/independences) between the driver's behavior considered (as given by the factors described previously) and the overall severity of the accident, which let us analyze the severity of traffic accidents as a function of the driver's behavior. There are currently many programs that can be used to solve this problem efficiently, such as Netica Software, Hugin Investigator, Genie, Matlab, R or Microsoft with MSBNx sofware. In this study, The Bayes Net (https://github.com/bayesnet/bnt) toolbox for Matlab (Matlab, 2014) have been proposed to perform the analysis.

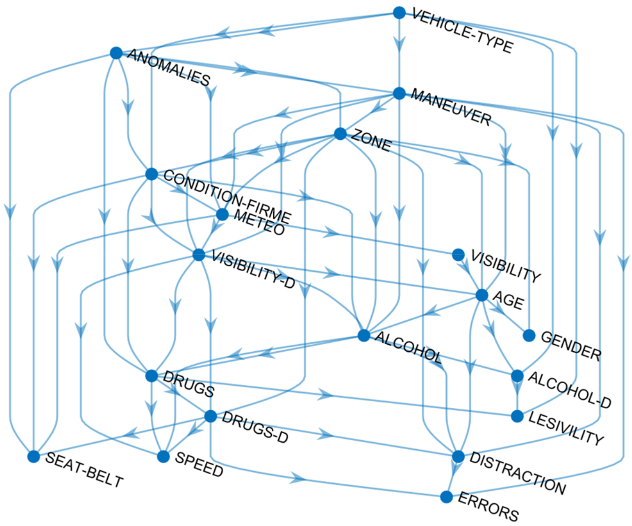

Fig. 2. Directed acyclic graph corresponding to the Bayesian.

Moreover, it is noteworthy to mention that, once the target variable is identified, the Bayesian Network can be interpreted as a Bayesian Classifier by means of the JPD and, then, it can be evaluated in order to indentify the skill of our model and to avoid model over-fitting. To this aim, a 10-fold cross-validation was developed creating a random partition of the database in ten subsamples, so using the 90% to train and the remaining 10% to evaluate, and repeating this procedure ten times, one for each subsample. The evaluation of each model was done using several parameters. First, the area under the ROC (Receiver Operating Characteristic) curve (AUC), which is a standard score for probabilistic and binary classifiers that varies from 0.5 (random guess) to 1 (perfect performance), was considered as a measure of the overall accuracy of the model. Secondly, the sensitivity and specificity were considered to identify if the model presents some bias to one of the categories. Both indices are defined as:

$$Sensitivity = TP/P \wedge Specifity = TN/N \tag{3}$$

where TP and TN are the number of predicted true positives and negatives respectively. Whereas, P and N are the number of observed positives and negatives respectively. On the one hand, note that the output of the Bayesian Network is a probability, so, in order to make the binary validation, the observed “a-priori” probability of the category has been considered as threshold to define the occurrence from the obtained probability. On the other hand, note that for each category of the target variable we have a binary classifier (occurring or not occurring) and, as a result, a value for each of the defined parameters. This is mainly relevant for those variables with several categories (e.g. Distractions). Finally, once the model has been evaluated and its predictability tested, the 100% of the database has been considered to train the model used for the sensitivity analysis.

5. Results and discussion

This section presents the results of the sensitivity analysis conducted on the Bayesian network. Section 5.1 represents the outputs from the Bayesian Network, which is developed by introducing dummy variables for “Visibility”, “Alcohol” and “Drugs” in order to address the biases in accident dataset. The performance of Bayesian Network has been evaluated using 10-fold cross-validation method. Section 5.2 summarizes the a-priori probabilities which represent the probabilities of fatal/serious injury in relation to the variable and their states in the Bayesian Network. Section 5.3 represents four classifiers/models obtained from the Bayesian Network in order to perform the sensitivity analysis. As sensitivity analysis in Bayesian Network is defined based on the relationship between the network nodes/variables, conditional probabilities associated with each variable have been estimated. The sensitivity of changes in the probability of Fatal/Serious injury as a result of changes in the state of speed violations, distraction/errors, and alcohol and drug use have been explained in Sections 5.3.1–5.3.3 respectively. Finally, the sensitivity of changes in the probability of driver’s distraction/error and speed violations as a result of alcohol and drug use have been summarized and discussed in Sections 5.3.4 and 5.3.5 respectively.

5.1. Bias identification and model estimation

As the procedure in which the questionnaires are filled out by the corresponding authorities introduces some biases in the analysis (e.g., the police officers complete the accident questionnaire more exhaustively for serious accidents when reporting the visibility conditions and carrying out the alcohol and drug tests on the drivers involved in these accidents), an extensive analysis of the “a-priori” probabilities of serious injury was performed for each variable in the accident. As the knowledge of the variables, “Visibility”, “Alcohol” and “Drugs” can dramatically modify the probability of serious injury, three dummy variables corresponding to these variables have been introduced in the models to reflect the known/unknown state of these variables. These “dummy variables” isolate homogeneous subsamples and generate valid model and unbiased parameter estimate. Fig. 2 represents the DAG obtained from the learnt Bayesian Network after accommodating the dummy variables in the model. DAG is obtained by applying the score-based greedy learning algorithm proposed by Buntine (1991) with a regularization term to penalize the model's complexity. Later, the parameters given by the DAG are obtained using maximum likelihood estimation. It should be noted that, in this study, the analysis was performed using Matlab (Matlab, 2014), in particular the last stable version (R2018b), the toolbox Bayes Net (Murphy, 2001; Toolbox, 2001) and MeteoLab toolbox (Gutiérrez et al., 2004).

As discussed earlier, a 10-fold cross-validation approach was considered to evaluate the results from the learnt Bayesian Network shown in Table 4. First, as could be expectable in the case of binary variables the accuracy given by the AUC, is approximately similar for both categories while the Sensitivity and Specificity exchange their values in both variables. Moreover, the results demonstrate that 85% of the traffic accidents with fatal/severe injuries are predicted by the model correctly. This value decreases to 70% for the accidents with minor/no injuries, slightly overestimating the cases of severe injury in spite of the initial probabilities of both categories (see Table 3). As a result, the developed model has shown a good performance and is able to smooth the imbalance effects between both sample categories.

5.2. A-priori probabilities

After having built the Bayesian Network, a-priori probabilities for each node have been estimated. A-priori probabilities represent the probability of fatal and/or serious injury (F/SI) in relation to all variables including “visibility”, “alcohol” and “drug”. As shown in Table 5, poor visibility was likely to increase the probability of F/SI by 10.4% (24.7%–14.3%). In addition, alcohol and drug consumption were likely to increase the probability of F/SI by 3.3% (14.8%–11.5%) and 45.9% (58.2%–12.3%) respectively. This implies that drug consumption is likely to have more adverse impact on driver’s accident severity than other variables.

Table 4
Validation parameters for the 10-fold cross-validation.

Overall severity	AUC	Sensitivity	Specificity
M/NI	0,839	0,700	0,850
F/SI	0,839	0,850	0,700

Table 5
Initial probabilities of drivers based on the states of each variable.

Variable			P(F/SI)		
Vehicle-Type	Car	Van	Off road		
	0,074	0,075	0,082		
	Normal	Overtaking	Fast maneuver	Other Unknown	
Maneuver	0,094	0,148	0,104	0,081 0,043	
	Inter-Urban	Crossing road	Urban	Motorway	
Zone	0,104	0,085	0,054	0,037	
	Good	Bad			Unknown
Condition-Firme	0,074	0,084			0,050
	Good	Bad			Unknown
Meteo	0,077	0,089			0,049
	Good	Bad			Unknown
Visibility	**0,143**	**0,247**			**0,027**
	< 25	25–65	> 65		Unknown
Age	0,074	0,072	0,109		0,037
	Male	Women			Unknown
Gender	0,076	0,074			0,030
	Non	Yes			Unknown
Alcohol	**0,115**	**0,148**			**0.058**
	Non	Yes			Unknown
Drugs	**0,123**	**0,582**			**0.068**
	Yes	Non			Unknown
Seat-Belt	0,079	0,070			0,061
	Appro.	Inappro.	Excessive		Unknow
Speed	0,079	0,129	0,273		0,063
	Non	Yes			Unknown
Distraction	0,084	0,095			0,070
	Non	Yes			Unknown
Errors	0,084	0,092			0,063

Note: Values with bold letters correspond to the suspicious variable for which a “dummy” variable has been introduced in the model.

Table 7
Sensitivity analysis of the probability of a fatal/serious injury in a traffic accident based on speed, gender and age.

Age	Speed violations	Male	Female	Unknown
< 25	Appropriate speed	0,069	0,064*	0,061
	Inappropriate Speed	0,143*	0,133*	0,119
	Excessive speed	0,244*	0,229	0,219
	Unknown	0,061	0,063*	0,042
25–65	Appropriate speed	0,077	0,073	0,068
	Inappropriate Speed	0,127*	0,122*	0,112
	Excessive speed	0,296*	0,285*	0,273
	Unknown	0,060*	0,065*	0,044
> 65	Appropriate speed	0,127*	0,127*	0,110
	Inappropriate Speed	0,128*	0,128	0,108
	Excessive speed	0,302*	0,301	0,272
	Unknown	0,088*	0,084	0,066
Unknown	Appropriate speed	0,058	0,060	0,060
	Inappropriate Speed	0,115	0,117	0,108
	Excessive speed	0,000*	0,000*	0,000*
	Unknown	0,041*	0,042*	0,025*

Note: Values highlighted with an asterisk (*) reflect significant differences at a significance level of 95%.

5.3. Sensitivity analysis

Based on the learnt Bayesian Network which includes the Joint Probability Distribution of all the variables, four classifiers/models have been obtained based on four target variables (see Table 6). Model 1 considers all the variables to assess their impacts on overall accident severity which is a target variable. While Model 2, Model 3 and Model 4 have been considered to further analyze the impacts of drug and alcohol consumption on speed violations, driver’s distractions and drivef́s errors, respectively. To assess the reliability of these models, the validation scores were estimated and were summarized in Tables 6. With the exception of the excessive speed all estimated AUCs range between 0.79 and 0.90 reflecting the accuracy of the four classifiers obtained from the learnt Bayesian Network. Also in most cases a higher equilibrium between the Sensitivity and Specificity is obtained with the values around the 80%. The main differences are obtained for the Unknown category which switch the behavior of both the negative and positive cases.

The influence of the driver’s behavior on the severity of an accident was determined by conducting a sensitivity analysis with Model 1 of the Bayesian network. This has been done by: First, estimating the probability of the accident’s severity based on speed violations (differentiating between excess speed and inappropriate speed), gender and age (see Section 5.3.1). Second, estimating the probability of accident severity based on the driver’s behaviors (differentiating between distractions and errors), gender and age (see Section 5.3.2). Third, estimating the probability of overall accident severity based on the driver’s psychophysical conditions (drug and alcohol consumption), gender and age (see Section 5.3.3). Furthermore, to measure the variations in the probability of driving distractedly/making error, and committing speed violations as a result of alcohol/drug use, similar sensitivity analysis have been conducted which are explained in Sections 5.3.4 and 5.3.5, respectively. In all cases, a 95% confidence interval has been considered to evaluate the statistical significance of the changes (the hypothesis test of difference between proportions/probabilities).

5.3.1. Overall accident severity based on drivers’ speed violations

Table 7 shows the results of the first sensitivity analysis for Model 1, which gives the estimated probabilities for a F/SI in a traffic accident based on the driving speed and the driver’s gender and age. From the results it can be seen that a male aged under 25 driving at an appropriate speed at the time of the accident would have a probability of a F/SI of 6.9%. In the case of inappropriate speed, the F/SI probability rises to 14.3%. And finally, if the driver is exceeding the speed limit, the probability of a F/SI would be 24.4%. Additionally, the increased probability of a fatal or serious injury in people driving at an inappropriate speed are different in males and females (24,4% for young male against 22,9% for young women) which is consistent with findings of Lawpoolsri et al. (2007) that reported young male drivers tend to

Table 6
AUC, sensitivity and specificity obtained for the four models considered and the different states of the target variables.

Model/category	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.
1-Overall Severity	M/NI			F/SI								
	0,86	0,78	0,80	0,86	0,80	0,78						
	Appro.			Inappro.			Excessive			Unknown		
2-Speed Violation	0,89	0,96	0,69	0,81	0,84	0,66	0,64	0,79	0,74	0,90	0,71	0,99
	Non			Yes						Unknown		
3-Distraction	0,92	0,88	0,81	0,79	0,76	0,79				0,92	0,76	0,94
	Non			Yes						Unknown		
4-Driver's Errors	0,91	0,80	0,82	0,85	0,89	0,65				0,95	0,82	0,92

Table 8
Analysis of the probability of death or serious injury in a traffic accident based on driver distraction, gender and age.

Age	Distraction	Male	Female	Unknown
< 25	No	0,082	0,074	0,065
	Yes	0,096	0,088	0,081
	Unknown	0,070	0,070	0,046
25–65	No	0,082*	0,077	0,068
	Yes	0,092*	0,087	0,082
	Unknown	0,067*	0,070*	0,050
> 65	No	0,127*	0,128*	0,105
	Yes	0,134*	0,133	0,109
	Unknown	0,101*	0,097*	0,078
Unknown	No	0,054	0,057	0,052
	Yes	0,059	0,060	0,057
	Unknown	0,042*	0,043*	0,025*

Note: Values highlighted with an asterisk (*) reflect significant differences at a significance level of 95%.

engage in more risky driving behaviors compared to young women.

5.3.2. Overall accident severity based on drivers' distractions and errors

Table 8 represents the results from the sensitivity analysis corresponding to Model 1 and it contains probabilities of a F/SI in an accident based on whether the driver was distracted behind the wheel. From the results, a non-distracted male drivers aged between 25 and 65 has a F/SI probability of 8.2%. While, this value rises to 9.2% for distracted ones. Moreover, in every case, by age and gender, the probability of F/SI under distracted condition would be higher than that under non-distracted condition which are consistent with the findings of Gong and Fan (2017) and Choudhary and Velaga (2019). Table 9 shows that drivers' errors slightly increase the risk of F/SI significantly in young and middle aged male drivers. Also, it can be seen that the difference between the probabilities of F/SI when driving with and without errors is around 1% in male drivers under 25. However, such difference is zero in drivers over 65 years.

5.3.3. Overall accident severity based on drug and alcohol consumption

Tables 10 and 11 show the results from the sensitivity analysis corresponding to Model 1 and it gives the probability of a fatal or serious injury in a traffic accident based on the driver's age, gender and psychophysical conditions. In other words, they present the results by age and gender when driving normally or under the influence of alcohol and drugs (see Tables 10 and 11). The results imply that driving under the effects of drugs or alcohol, drastically increases the severity of an accident in men and women and in all age groups which are consistent with those in previous studies (Chen et al., 2015; Robertson et al., 2017).

Table 9
Analysis of the probability of death or serious injury in a traffic accident based on driver errors, gender and age.

Age	Error	Male	Female	Unknown
< 25	No	0,084	0,077	0,065
	Yes	0,094*	0,086	0,072
	Unknown	0,061*	0,063*	0,044
25–65	No	0,082*	0,078	0,067
	Yes	0,089*	0,084*	0,078
	Unknown	0,060*	0,064*	0,047
> 65	No	0,127*	0,129*	0,104
	Yes	0,127*	0,126*	0,106
	Unknown	0,089*	0,086	0,071
Unknown	No	0,051	0,053	0,045*
	Yes	0,058	0,058	0,053
	Unknown	0,039*	0,040*	0,024*

Notes: Values highlighted with an asterisk (*) reflect significant differences at a significance level of 95%.

Table 10
Probability of a F/SI in a traffic accident based on age, gender and alcohol consumption.

Age	Alcohol	Male	Female	Unknown
< 25	No	0,112	0,099*	0,143
	Yes	0,163*	0,145	0,125
25–65	No	0,114	0,104*	0,122
	Yes	0,147*	0,137	0,124
> 65	No	0,165*	0,169*	0,135
	Yes	0,183*	0,183	0,142

Note: Values highlighted with an asterisk (*) reflect significant differences at a significance level of 95%.

Table 11
Probability of a F/SI in a traffic accident based on age, gender and drugs consumption.

Age	Drug	Male	Female	Unknown
< 25	No	0,123	0,115	0,000
	Yes	0,569*	0,579*	0,000
25–65	No	0,121*	0,118*	0,086
	Yes	0,570*	0,574*	0,550
> 65	No	0,186*	0,180	0,000
	Yes	0,678	0,686	0,000

Note: Values highlighted with an asterisk (*) reflect significant differences at a significance level of 95%.

Note that, in both cases we have considered the dummy variables corresponding to the drug or alcohol consumption to obtain a homogeneous and unbiased sample. In addition, the significance of the changes in the probabilities is referred to the probability of the subsample filtered by the corresponding dummy variable. In general, and in agreement with previous studies, the probabilities of a F/SI in a traffic accident are higher for men than women. The results indicate that the probability of a F/SI does not always rise with age. For instance, the consequences of alcohol for drivers under the 25 are worse than other groups, but in the case of drugs use there are not differences in all age groups.

Additionally, it can be seen that a young female driver without alcohol influence has a F/SI likelihood of 9.9%. In contrast, the corresponding probability under alcohol influence is 14.5% (4.6% higher). For a male driver in the same age range, the alcohol consumption is likely to increase the F/SI probability up to 5.1%. Referring to Table 11, the differences between the estimated probability of a F/SI for an individual with and without drug influence across all age ranges are relatively high. For example, young male drivers under 25 with drug influence has an accident rate of 56.9% which is 44.6% higher than that in the same group of male drivers without drug influence.

5.3.4. Alcohol and drugs impact on driving distraction and errors

To further investigate the impacts of psychophysical conditions on distraction and errors sensitivity analysis has been performed based on Model 3 and Model 4. The evidence variables used in Model 3 were "Alcohol-D" and "Alcohol". The variable "Alcohol-D" was used to eliminate the uncertainty generated by the sample bias. Hence the following results correspond to the drivers who were subjected to the alcohol test and the estimated values in the table show the relative probabilities. The same procedure was followed to estimate the relative probabilities for drug/distraction. Table 12 shows represents the probability that a driver being distracted under distracted or None-distracted condition. The quantification of these probabilities yields clear results, such as, on the one hand, how alcohol consumption raises the likelihood of distracting from 15,2% to 43,7%. On the other hand, the drugs impact is different and there is no direct effect between drug use and driving distractions.

Table 13 indicates the relationships between alcohol and drug

Table 12
Probability of driver distractions based on psychophysical conditions.

	Distraction None	Distraction Yes
Alcohol none	84,8%	15,2%
Alcohol yes	56,3%	43,7%
Drugs none	81,9%	18,1%
Drugs yes	84,1%	15,9%

Table 13
Probability of driver errors based on age and psychophysical conditions.

	Errors none	Errors yes
Alcohol none	62,3%	37,7%
Alcohol yes	28,8%	71,2%
Drugs none	57,6%	42,4%
Drugs yes	61,0%	39,0%

consumption and drivers' errors at the wheel based on Model 4. There is a direct relationship between the consumption of alcohol and the probability that the driver has an error while driving, for example, from the results, the probability of committing an error with and without alcohol influence are 71.2% and 37.7% in respectively. However, when it comes to the drug consumption, there are no difference in the probability of making or not making mistakes during driving (42,4% vs 39,0%).

5.3.5. Alcohol and drugs impact on speed violations

To analyze the effect that drug and alcohol consumption have on speed violations, a Bayesian network was developed for Model 2. The sensitivity analysis was carried out by taking as the study variable the speed violations. The evidence variables were the driver's psychophysical conditions and drivers' age. The results, presented in Table 14, show the relative probabilities of driving at an adequate, inappropriate or excessive speed depending on the age range and the whether or not the driver is under alcohol influence. To carry out a sensitivity analysis, the artificial variable "Alcohol-D" has been taken into account which corresponds to the records of those drivers who were subjected to the alcohol test.

The results indicate that alcohol consumption significantly reduces the likelihood of driving at an appropriate speed in each age range which is consistent with previous studies (Bogstrand et al., 2015; Phillips and Brewer, 2011; Stübig et al., 2012). For example, the probability of driving at an appropriate speed in young drivers not under the influence of alcohol is 78.2%. In contrast, when driving under the influence of alcohol, the likelihood of driving at an appropriate speed drops to 38,1%. The alcohol consumption is likely to increase the risk of speed violations in all age groups, especially in drivers under 25. As shown in Table 14 the alcohol consumption has increased the probability of driving at excessive speed from 2.3% to 13.8% in young drivers.

On the other hand, Table 15 shows the probability of moving at appropriate speeds depending on age and drug consumption. As in the case of alcohol, to carry out this sensitivity analysis, the variable "Drugs-D" was used to consider only the drivers subjected to the drug consumption test. We see in this case, the absence of a relationship between both factors (drug consumption and variation in driving speed). That is, the probabilities of driving at adequate speeds are similar regardless of drug use by drivers.

Table 14
Probability of driving at various speeds based on age and alcohol consumption.

Age	Alcohol	Appropriate speed	Inappropriate speed	Excessive speed
< 25	None	78,2%	19,5%	2,3%
	Yes	38,1%	48,1%	13,8%
25–65	None	90,3%	9,0%	0,7%
	Yes	61,2%	32,5%	6,3%
> 65	None	93,3%	6,4%	0,3%
	Yes	74,5%	22,9%	2,5%

Table 15
Probability of driving at various speeds based on age and drug consumption.

Age	Drugs	Appropriate speed	Inappropriate speed	Excessive speed
< 25	None	67,9%	26,1%	6,1%
	Yes	68,4%	26,2%	5,4%
25–65	None	85,9%	11,8%	2,3%
	Yes	86,2%	11,8%	2,1%
> 65	None	92,7%	6,5%	0,8%
	Yes	92,6%	6,7%	0,7%

6. Conclusions

The main goal of this work is to define a robust process to address possible inhomogeneities in the sample which could result in biased parameter estimates by introducing dummy variables for the suspicious variables. First, three variables including Visibility, Drugs and Alcohol have been found to be suspicious as the police officers might overstate these them when there are not serious injuries in the traffic accident. As a result, three dummy variables (Visibility-D, Drugs-D and Alcohol-D) have been and were used in the sensitivity analysis to filter homogeneous subsamples.

Second, the Bayesian Networks have been trained and evaluated using a 10-fold cross-validation process. The resulting AUCs fall between 0.77 and 0.85 which represents a good model performance. By taking the advantage of the Bayesian Network's properties which includes the DAG and the JPD, four different models have been developed in order to conduct the sensitivity analysis. From the models, it was found that, on average, the excess speed will likely increase the probability of serious/fatal injury for drivers of all age groups by 10%. Additionally, there is not a significant difference between the probability of fatal/serious injury in males and females for all age groups and speed states. Also, distraction and driver' errors will likely increase the probability of serious/fatal injury in all drivers up to 1.5% even when they are driving at a proper speed. This implies that distracted drivers may have lower chance to detect hazard on roadway and therefore, they may not be able to reduce the impact speed with other vehicle(s) or object. As distracted driving is prevalent among young drivers, education and computer-based training program are effective tool to improve their knowledge and safety awareness (Kumfer et al., 2017; Rodwell et al., 2018).

Additionally, it was found that driving under the effects of drugs/alcohol will likely increase the probability of fatal/serious injury up to 5%. This is consistent with the findings of previous studies (Chen et al., 2015; Robertson et al., 2017). Besides, alcohol and drug consumption can significantly increase the probability of being distracted and making errors by 28.5% (from 15.2% to 43.7%) and 33.5% (from 37,7% to 71,2%) respectively. Finally, alcohol consumption significantly reduces the probability of driving at an appropriate speed. This reduction is relatively high for drivers under 25. That is the probability of driving at an appropriate speed for young None-alcohol involved drivers is %78. However, this value drops to 38.1% when it comes to young alcohol involved drivers. The results for drugs consumption are not as significant as the ones for alcohol consumption.

As found in this study, alcohol use poses significant threat to the health and safety of drivers, especially drivers under 25, by impairing their attention and performance. Some of the effective preventive actions that can be used by transportation authorities and decision makers in order to reduce alcohol-related accidents and/or the consumption of alcohol while driving are (a) lower the legal blood alcohol limits (b) obligate the car manufacturers to equip the vehicles with ignition interlock devices to prevent drunk driving, and (d) Run enforcement campaigns targeting drink and drug driving.

7. Limitations and future research

Our study presents some limitations. First, it is noteworthy to mention that the data collection system for traffic accidents in Spain has been modified since 2013, however this was not applied to all regions equally. For this reason, the database for years 2014 and 2015 are not homogeneous. Therefore, it is not possible to carry out a cross-sectional study and our research has been carried out based on 2016 traffic accident data. In addition, the drug test results in our database are recorded as dichotomous outcomes (positive/negative). Therefore, it is impossible to analyze the influence of different levels of drugs consumption.

In terms of future research, it would be valuable to study the change in the habits of drug users and alcohol consumption across the time in order to assess the influence of alcohol consumption campaigns, legislation changes, new drugs in the market, etc. on the consequences of traffic accidents. it is recommended to adopt new data collection methods to measure different levels of alcohol and drug consumption. Such information would help law enforcement agencies to promote accident reduction strategies.

Acknowledgments

We would like to thank the staff at DGT for cooperating in this study, and in particular for allowing us to use the road accident data, which made it possible for us to complete this work.

This study was supported by funds from the following projects:

“Modelo cuantitativo de Red Bayesiana con capacidad predictiva de la gravedad del accidente en función de los comportamientos y actuaciones de las personas” [Qualitative Bayesian Network Model for Predicting Accident Severity Based on Personal Behaviors and Actions]. Ref. SPIP2015-01852 supported by funds from DGT (Directorate-General for Traffic).

“Modelización mediante técnicas de machine learning de la influencia de las distracciones del conductor en la seguridad vial” [Modeling the influence of driver distractions on road safety using machine learning techniques]. Ref. BU300P18 supported by funds from FEDER (Fondo Europeo de Desarrollo Regional, Junta de Castilla y León).

Appendix A

See Table 1.

Appendix B

See Table 2.

References

Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., Herrmann, M., 2019. Detection and evaluation of driver distraction using machine learning and fuzzy logic. IEEE Trans. Intell. Transp. Syst. 20 (6), 2048–2059. https://doi.org/10.1109/tits.2018.2857222.

Bedard, M., Guyatt, G.H., Stones, M.J., Hirdes, J.P., 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accid. Anal. Prev. 34 (6), 717–727.

Behnood, A., Mannering, F., 2017. The effects of drug and alcohol consumption on driver injury severities in single-vehicle crashes. Traffic Inj. Prev. 18 (5), 456–462. https://doi.org/10.1080/15389588.2016.1262540.

BOE, 1993. Orden de 18 de febrero de 1993 por la que se modifica la estadística de accidentes de circulación. from https://www.boe.es/eli/es/o/1993/02/18/(2).

BOE, 2014. Orden INT/2223/2014, de 27 de octubre, por la que se regula la comunicación de la información al Registro Nacional de Víctimas de Accidentes de Tráfico. from https://www.boe.es/boe/dias/2014/11/29/pdfs/BOE-A-2014-12411.pdf.

Bogstrand, S.T., Larsson, M., Holtan, A., Staff, T., Vindenes, V., Gjerde, H., 2015. Associations between driving under the influence of alcohol or drugs, speeding and seatbelt use among fatally injured car drivers in Norway. Accid. Anal. Prev. 78, 14–19. https://doi.org/10.1016/j.aap.2014.12.025.

Bowden, V.K., Loft, S., Wilson, M.D., Howard, J., Visser, T.A.W., 2019. The long road home from distraction: Investigating the time-course of distraction recovery in driving. Accid. Anal. Prev. 124, 23–32. https://doi.org/10.1016/j.aap.2018.12.012.

Buendia, R., Candefjord, S., Fagerlind, H., Balint, A., Sjoqvist, B.A., 2015. On scene injury severity prediction (OSISP) algorithm for car occupants. Accid. Anal. Prev. 81, 211–217. https://doi.org/10.1016/j.aap.2015.04.032.

Bukova-Zideluna, A., Villerusa, A., 2016. 849 Prevalence of alcohol among car drivers in road accidents in Latvia: an overview from 2010 to 2014: BMJ Publishing Group Ltd.

Buntine, W., 1991. Theory refinement on Bayesian networks. In: Paper Presented at the Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence.

Castillo-Manzano, J.I., Castro-Nuno, M., Lopez-Valpuesta, L., Vassallo, F.V., 2019. The complex relationship between increases to speed limits and traffic fatalities: Evidence from a meta-analysis. Saf. Sci. 111, 287–297. https://doi.org/10.1016/j.ssci.2018.08.030.

Castillo, E., Gutiérrez, J.M., Hadi, A.S., 1997. Expert Systems and Probabilistic Network Models. Springer Verlag.

Castro, Y., Kim, Y.J., 2016. Data mining on road safety: Factor assessment on vehicle accidents using classification models. Int. J. Crashworthiness 21 (2), 104–111. https://doi.org/10.1080/13588265.2015.1122278.

Chang, L.Y., Wang, H.W., 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. Accid. Anal. Prev. 38 (5), 1019–1027. https://doi.org/10.1016/j.aap.2006.04.009.

Chen, C., Zhang, G., Qian, Z., Tarefder, R.A., Tian, Z., 2016. Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accid. Anal. Prev. 90, 128–139.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., Guan, H., 2015. A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. Accid. Anal. Prev. 80, 76–88. https://doi.org/10.1016/j.aap.2015.03.036.

Chiou, Y.-C., Hwang, C.-C., Chang, C.-C., Fu, C., 2013. Modeling two-vehicle crash severity by a bivariate generalized ordered probit approach. Accid. Anal. Prev. 51, 175–184. https://doi.org/10.1016/j.aap.2012.11.008.

Choudhary, P., Velaga, N.R., 2019. Effects of phone use on driving performance: A comparative analysis of young and professional drivers. Saf. Sci. 111, 179–187. https://doi.org/10.1016/j.ssci.2018.07.009.

de Oña, J., de Oña, R., Eboli, L., Forciniti, C., Mazzulla, G., 2014. How to identify the key factors that affect driver perception of accident risk. A comparison between Italian and Spanish driver behavior. Accid. Anal. Prev. 73, 225–235.

De Oña, J., López, G., Mujalli, R., Calvo, F.J., 2013. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. Accid. Anal. Prev. 51, 1–10.

de Oña, J., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accid. Anal. Prev. 43 (1), 402–411.

DGT, 2016. Main figures on road traffic accidents. Spain 2016. from http://www.dgt.es/Galerias/seguridad-vial/estadisticas-e-indicadores/publicaciones/principales-cifras-siniestralidad/2017-2799_Summary_Main_figures_on_road_safety_data_Spain_2016_ACCESIBLE.pdf.

Donmez, B., Liu, Z., 2015. Associations of distraction involvement and age with driver injury severities. J. Saf. Res. 52, 23–28.

Eluru, N., Bagheri, M., Miranda-Moreno, L.F., Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accid. Anal. Prev. 47, 119–127.

Feng, S., Li, Z., Ci, Y., Zhang, G., 2016. Risk factors affecting fatal bus accident severity: Their impact on different types of bus drivers. Accid. Anal. Prev. 86, 29–39.

Ferrari, D., Manca, M., Banfi, G., Locatelli, M., 2018. Alcohol and illicit drugs in drivers involved in road traffic crashes in the Milan area. A comparison with normal traffic reveals the possible inadequacy of current cut-off limits. Forensic Sci. Int. 282, 127–132.

García-Herrero, S., Mariscal, M., Beltrán, R., Gutiérrez, J., 2016. Variables influencing the severity of motorcycle and moped traffic accidents in Spain. Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016 (Glasgow, Scotland, 25-29 September 2016), 202.

Gjerde, H., Ramaekers, J., Mørland, J., 2019. Methodologies for establishing the relationship between alcohol/drug use and driving impairment-Differences between epidemiological, experimental, and real-case studies. Forensic Sci. Rev. 31 (2), 141–160.

Gómez-Talegón, T., Fierro, I., González-Luque, J.C., Colás, M., López-Rivadulla, M., Javier Álvarez, F., 2012. Prevalence of psychoactive substances, alcohol, illicit drugs, and medicines, in Spanish drivers: A roadside study. Forensic Sci. Int. 223 (1), 106–113. https://doi.org/10.1016/j.forsciint.2012.08.012.

Gong, L., Fan, W., 2017. Modeling single-vehicle run-off-road crash severity in rural areas: Accounting for unobserved heterogeneity and age difference. Accid. Anal. Prev. 101, 124–134. https://doi.org/10.1016/j.aap.2017.02.014.

Goss, C.W., Van Bramer, L.D., Gliner, J.A., Porter, T.R., Roberts, I.G., DiGuiseppi, C., 2008. Increased police patrols for preventing alcohol-impaired driving. Cochrane Database of Systematic Reviews 4.

Gregoriades, A., Mouskos, K., Michail, H., Katzis, K., 2012. Combining traffic simulation with Bayesian networks for improved quantification of accident risk index. In: Bruzzone, A., Longo, F., Kropf, P., Piera, M.A. (Eds.), Proceedings of the 2012 Summer Computer Simulation Conference, pp. 97–105.

Gregoriades, A., Mouskos, K.C., 2013. Black spots identification through a Bayesian Networks quantification of accident risk index. Transportation Res. Part C: Emerging Technol. 28, 28–43.

Gutiérrez, J.M., Cano, R., Cofiño, A.S., Sordo, C.M., 2004. Redes probabilísticas y

neuronales en las ciencias atmosféricas: Ministerio de Medio Ambiente, Secretaría General Técnica.
Haghighi, N., Liu, X.C., Zhang, G., Porter, R.J., 2018. Impact of roadway geometric features on crash severity on rural two-lane highways. Accid. Anal. Prev. 111, 34–42.
Halim, Z., Kalsoom, R., Bashir, S., Abbas, G., 2016. Artificial intelligence techniques for driving safety and vehicle crash prediction. Artif. Intell. Rev. 46 (3), 351–387. https://doi.org/10.1007/s10462-016-9467-9.
Kashani, A.T., Mohaymany, A.S., 2011. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Saf. Sci. 49 (10), 1314–1320.
Khanjani, N., Mousavi, M., Dehghanian, A., Jahani, Y., Souri, H., 2017. The role of drug and alcohol use and the risk of motor vehicle crashes in Shiraz, Iran, 2014: A case–crossover study. Traffic Inj. Prev. 18 (6), 573–576.
Kim, J.-K., Ulfarsson, G.F., Kim, S., Shankar, V.N., 2013. Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. Accid. Anal. Prev. 50, 1073–1081.
Koopmans, J.M., Friedman, L., Kwon, S., Sheehan, K., 2015. Urban crash-related child pedestrian injury incidence and characteristics associated with injury severity. Accid. Anal. Prev. 77, 127–136. https://doi.org/10.1016/j.aap.2015.02.005.
Kumfer, W., Liu, H.C., Wu, D.Y., Wei, D.L., Sama, S., 2017. Development of a supplementary driver education tool for teenage drivers on rural roads. Saf. Sci. 98, 136–144. https://doi.org/10.1016/j.ssci.2017.05.014.
Kwon, O.H., Rhee, W., Yoon, Y., 2015. Application of classification algorithms for analysis of road safety risk factor dependencies. Accid. Anal. Prev. 75, 1–15.
Lawpoolsri, S., Li, J., Braver, E.R., 2007. Do speeding tickets reduce the likelihood of receiving subsequent speeding tickets? A longitudinal study of speeding violators in Maryland. Traffic Inj. Prev. 8 (1), 26–34.
Lemp, J.D., Kockelman, K.M., Unnikrishnan, A., 2011. Analysis of large truck crash severity using heteroskedastic ordered probit models. Accid. Anal. Prev. 43 (1), 370–380.
Li, L., Hasnine, M.S., Nurul Habib, K.M., Persaud, B., Shalaby, A., 2017. Investigating the interplay between the attributes of at-fault and not-at-fault drivers and the associated impacts on crash injury occurrence and severity level. J. Transportation Saf. Security 9 (4), 439–456.
Liang, Y., Lee, J.D., 2014. A hybrid Bayesian Network approach to detect driver cognitive distraction. Transportation Res. Part C: Emerging Technol. 38, 146–155.
Madrid, C.A., 2010. Guía para la Prevención de los Accidentes de Tráfico con Relación Laboral.
Matlab, 2014. Matlab 2014b. from http://es.mathworks.com/.
Morgan, A., Mannering, F.L., 2011. The effects of road-surface conditions, age, and gender on driver-injury severities. Accid. Anal. Prev. 43 (5), 1852–1863.
Mujalli, R.O., De ONa, J., 2011. A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks. J. Saf. Res. 42 (5), 317–326.
Murphy, K., 2001. The bayes net toolbox for matlab. Comput. Sci. Statistics 33 (2), 1024–1034.
Naik, B., Tung, L.-W., Zhao, S., Khattak, A.J., 2016. Weather impacts on single-vehicle truck crash injury severity. J. Saf. Res. 58, 57–65.
Neyens, D.M., Boyle, L.N., 2008. The influence of driver distraction on the severity of injuries sustained by teenage drivers and their passengers. Accid. Anal. Prev. 40 (1), 254–259.
Osman, M., Paleti, R., Mishra, S., 2018. Analysis of passenger-car crash injury severity in different work zone configurations. Accid. Anal. Prev. 111, 161–172.
Owen, R., Ursachi, G., Fosdick, T., Horodnic, A.V., 2019. Driving while impaired by alcohol: An analysis of drink-drivers involved in UK collisions. Traffic Inj. Prev. 1–7.
Paleti, R., Eluru, N., Bhat, C.R., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Accid. Anal. Prev. 42 (6), 1839–1854.
Papalimperi, A.H., Athanaselis, S.A., Mina, A.D., Papoutsis, I.I., Spiliopoulou, C.A., Papadodima, S.A., 2019. Incidence of fatalities of road traffic accidents associated with alcohol consumption and the use of psychoactive drugs: A 7-year survey (2011–2017). Exp. Ther. Med. 18 (3), 2299–2306.
Papantoniou, P., Yannis, G., Christofa, E., 2019. Which factors lead to driving errors? A structural equation model analysis through a driving simulator experiment. IATSS Res. 43 (1), 44–50.
Pešić, D., Antić, B., Smailović, E., Marković, N., 2019. Driving under the influence of alcohol and the effects of alcohol prohibition—Case study in Serbia. Traffic Inj. Prev. 1–5.
Phillips, D.P., Brewer, K.M., 2011. The relationship between serious injury and blood alcohol concentration (BAC) in fatal motor vehicle accidents: BAC = 0.01% is associated with significantly more dangerous accidents than BAC = 0.00%. Addiction 106 (9), 1614–1622. https://doi.org/10.1111/j.1360-0443.2011.03472.x.
Ramos, L.d.E., 2012. Sistema de Información de Accidentes de tráfico de la DGT. Retrieved 28/06/2015, from http://comunidad.medialab-prado.es/sites/default/files/sistema-de-Informacion-de-Accidentes-de-trafico-DGT.ppt.
Reyes, M.L., Lee, J.D., 2008. Effects of cognitive load presence and duration on driver eye movements and event detection performance. Transportation Res. Part F: Traffic Psychology Behaviour 11 (6), 391–402.
Robertson, R.D., Mainegra-Hing, M., Pashley, C.R., Brown, S.W., Vanlaar, W.G.M., 2017. Prevalence and trends of drugged driving in Canada. Accident; Anal. Prevention 99 (Pt A). https://doi.org/10.1016/j.aap.2016.12.008.
Rodwell, D., Hawkins, A., Haworth, N., Larue, G.S., Bates, L., Filtness, A., 2018. A mixed-methods study of driver education informed by the goals for driver education: Do young drivers and educators agree on what was taught? Saf. Sci. 108, 140–148.
Santamariña-Rubio, E., Pérez, K., Olabarria, M., Novoa, A.M., 2014. Gender differences in road traffic injury rate using time travelled as a measure of exposure. Accid. Anal. Prev. 65, 1–7.
Strayer, D.L., Cooper, J.M., Turrill, J., Coleman, J.R., Hopman, R.J., 2016. Talking to your car can drive you to distraction. Cognitive Res.: Principles Implications 1 (1), 16.
Strayer, D.L., Cooper, J.M., Turrill, J., Coleman, J.R., Hopman, R.J., 2017. The smartphone and the driver's cognitive workload: A comparison of Apple, Google, and Microsoft's intelligent personal assistants. Can. J. Exp. Psychology/Revue canadienne de psychologie expérimentale 71 (2), 93.
Stübig, T., Petri, M., Zeckey, C., Brand, S., Müller, C., Otte, D., Haasper, C., 2012. Alcohol intoxication in road traffic accidents leads to higher impact speed difference, higher ISS and MAIS, and higher preclinical mortality. Alcohol 46 (7), 681–686. https://doi.org/10.1016/j.alcohol.2012.07.002.
Sun, J., Sun, J., 2015. A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data. Transportation Res. Part C-Emerging Technol. 54, 176–186. https://doi.org/10.1016/j.trc.2015.03.006.
Sundfør, H.B., Sagberg, F., Høye, A., 2019. Inattention and distraction in fatal road crashes–Results from in-depth crash investigations in Norway. Accid. Anal. Prev. 125, 152–157.
Toolbox, f.M., 2001. From https://github.com/bayesnet/bnt.
Uusitalo, L., 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecol. Model. 203 (3–4), 312–318.
Valen, A., Bogstrand, S.T., Vindenes, V., Frost, J., Larsson, M., Holtan, A., Gjerde, H., 2019. Fatally injured drivers in Norway 2005–2015—Trends in substance use and crash characteristics. Traffic Inj. Prev. 20 (5), 460–466. https://doi.org/10.1080/15389588.2019.1616700.
Waller, P.F., Hill, E.M., Maio, R.F., Blow, F.C., 2003. Alcohol effects on motor vehicle crash injury. Alcohol.: Clin. Exp. Res. 27 (4), 695–703.
Wang, K., Bhowmik, T., Yasmin, S., Zhao, S., Eluru, N., Jackson, E., 2019. Multivariate copula temporal modeling of intersection crash consequence metrics: A joint estimation of injury severity, crash type, vehicle damage and driver error. Accid. Anal. Prev. 125, 188–197.
Wang, Y., Zhang, W., 2017. Analysis of roadway and environmental factors affecting traffic crash severities. Transp. Res. Procedia 25, 2119–2125.
WHO, 2017. Managing Speed.
WHO, 2018. Accidentes de tránsito. From https://www.who.int/es/news-room/fact-sheets/detail/road-traffic-injuries.
Wijnen, W., Weijermars, W., Schoeters, A., Van den Berghe, W., Bauer, R., Carnis, L., Martensen, H., 2019. An analysis of official road crash cost estimates in European countries. Saf. Sci. 113, 318–327.
Wu, Q., Zhang, G., 2018. Formulating alcohol-influenced driver's injury severities in intersection-related crashes. Transport 33 (1), 165–176.
Xie, Y., Zhao, K., Huynh, N., 2012. Analysis of driver injury severity in rural single-vehicle crashes. Accid. Anal. Prev. 47, 36–44.
Yamamoto, T., Hashiji, J., Shankar, V.N., 2008. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Accid. Anal. Prev. 40 (4), 1320–1329.
Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accid. Anal. Prev. 59, 506–521.
Yu, H., Li, Z., Zhang, G., Liu, P., 2019. A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence. Anal. Methods Accident Res 100110.
Zong, F., Xu, H., Zhang, H., 2013. Prediction for traffic accident severity: comparing the bayesian network and regression models. Math. Problems Eng. https://doi.org/10.1155/2013/475194.
Zou, W., Wang, X., Zhang, D., 2017. Truck crash severity in New York city: an investigation of the spatial and the time of day effects. Accid. Anal. Prev. 99, 249–261.