Using data mining techniques to predict the severity of bicycle crashes

Published Version:
Using data mining techniques to predict the severity of bicycle crashes / Prati, Gabriele; Pietrantoni, Luca; Fraboni, Federico. - In: ACCIDENT ANALYSIS AND PREVENTION. - ISSN 0001-4575. - ELETTRONICO. - 101:(2017), pp. 44-54. [10.1016/j.aap.2017.01.008]

Availability:
This version is available at: https://hdl.handle.net/11585/577592 since: 2019-05-29

Published:
DOI: http://doi.org/10.1016/j.aap.2017.01.008

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Prati, G., Pietrantoni, L., & Fraboni, F. (2017). Using data mining techniques to predict the severity of bicycle crashes. Accident Analysis & Prevention, 101, 44–54. https://doi.org/10.1016/j.aap.2017.01.008

The final published version is available online at: https://doi.org/10.1016/j.aap.2017.01.008

Using Data Mining Techniques to Predict the Severity of Bicycle Crashes

Abstract

To investigate the factors predicting severity of bicycle crashes in Italy, we used an observational study of official statistics. We applied two of the most widely applied data mining techniques, CHAID decision tree technique and Bayesian network analysis. We used data provided by the Italian National Institute of Statistics on road crashes that occurred on the Italian road network during the period ranging from 2011 to 2013. In the present study, the dataset contains information about road crashes occurred on the Italian road network during the period ranging from 2011 to 2013. We extracted 49,621 road accidents where at least one cyclist was injured or killed from the original database that comprised a total of 575,093 road accidents. CHAID decision tree technique was employed to establish the relationship between severity of bicycle crashes and factors related to crash characteristics (type of collision and opponent vehicle), infrastructure characteristics (type of carriageway, road type, road signage, pavement type, and type of road segment), cyclists (gender and age), and environmental factors (time of the day, day of the week, month, pavement condition, and weather). CHAID analysis revealed that the most important predictors were, in decreasing order of importance, road type (0.30), crash type (0.24), age of cyclist (0.19), road signage (0.08), gender of cyclist (0.07), type of opponent vehicle (0.05), month (0.04), and type of road segment (0.02). These eight most important predictors of the severity of bicycle crashes were included as predictors of the target (i.e., severity of bicycle crashes) in Bayesian network analysis. Bayesian network analysis identified crash type (0.31), road type (0.19), and type of opponent vehicle (0.18) as the most important predictors of severity of bicycle crashes.

Keywords: data mining, cycling, bicycle crash, injury, fatality, safety, decision tree

1. Introduction

It is recognized that the use of bicycle as a mode of transport is associated with environmental and societal benefits (de Nazelle et al. 2011, Xia et al. 2013, Macmillan et al. 2014) as well as health benefits (Kelly et al. 2014, Götschi et al. 2016). However, there are also societal costs of bicycle use, especially in terms of consequences of bicycle crashes.

In Europe, 8% of people choose bicycles as the most common mode of daily transport (European Commission 2014). Nevertheless, cyclists still represent one of the road user categories with the highest risk of injuries and fatalities. From 2004 to 2013, cyclists' fatalities decreased by 32%, but from 2010 this tendency has stagnated, with less than a 1% year-to-year reduction. Furthermore, 31% of the fatalities happen at junctions (European Commission 2015). Risks for non-fatal accidents are higher for cyclists than for car drivers (de Hartog et al. 2010).

Similar to European data, in Italy, 6% of the population indicates the bicycle as the most common mode of transport (European Commission 2014). In 2014, there were 18.055 bicycle accidents and 273 cyclists' fatalities recorded in Italy, leading to a 9% increase in comparison to 2013. In Italy, the mortality index (deaths every 100 accidents) for cyclists is 1.42, which is more than double compared to car users (ISTAT 2015).

Various contributing factors to bicycle crashes have been identified in literature. Accident analysis revealed that violation of traffic rules plays a key role in fatal crashes involving cyclists. Red-light violation is one typical violation behaviour among cyclists (Wu et al. 2012, Pai and Jou 2014). Other violations commonly associated with collision were riding against traffic, in a wrong-way, or coming from an unexpected side of the road (Atkinson and Hurst 1983, Ashbaugh et al. 1995, Kim and Li 1996, Wachtel and Lewiston 1996, Wessels 1996, Räsänen et al. 1998, Vandenbulcke et al. 2014, Hamann et al. 2015).

Although fall and collisions with non-motorized users may happen more frequently, collisions involving motor vehicles account for majority of the reported bicyclists' fatalities and serious injuries (Rosenkranz and Sheridan, Rowe et al. 1995, Nicaj et al. 2009, Chong et al. 2010, Sze et al. 2011). Exposure to traffic increases the risk of collision (Hagel et al. 2014, Chen 2015). Another possible reason might be blind spot conflicts (Wachtel and Lewiston 1996).

Different approaches have been employed to investigate these factors. One of these approaches is based on investigating factors that increase the severity of bicycle crashes. Usually, collision data are gathered from official sources (Klassen et al. 2014). Using this approach, factors contributing to the severity of bicycle crashes have been studied at various levels: crash characteristics (e.g., type of collision and opponent vehicle), infrastructure characteristics (e.g., type of carriageway, road type, road signage, and type of road segment), cyclists (e.g., gender and age), and environmental factors (e.g., time of the day, darkness, day of the week, and weather). In terms of infrastructure characteristics, crashes in straight sections have been found to be the most fatal (Klop and Khattak 1999, Bíl et al. 2010). Bicycle crashes occurring at signalized intersections were less severe than those elsewhere (Eluru et al. 2008).

Among the crash characteristics, the involvement of trucks and heavy vehicles in the bicycle crash was found to increase the severity of bicycle crashes compared to other types of vehicles (Kim et al. 2007, Moore et al. 2011, Yan et al. 2011). Moreover, head-on and angle collisions were found to increase the level of bicyclist injury severity (Kim et al. 2007, Bíl et al. 2010, Yan et al. 2011).

In terms of environmental characteristics, the consequences of bicycle crashes tend to be less severe if they occur at day-time under good visibility, whereas crashes occurring in night-time traffic in places without streetlights have the worst consequences for cyclists (Klop and

Khattak 1999, Kim et al. 2007, Eluru et al. 2008, Bíl et al. 2010, Yan et al. 2011). Cycling in the night (e.g., from midnight to 6 a.m.) has been found to increase the likelihood of fatal injury (Stone and Broughton 2003, Eluru et al. 2008). Also, inclement weather (Kim et al. 2007) and foggy weather (Klop and Khattak 1999) were associated with more severe consequences of bicycle crashes. Concerning factors related to cyclists, there is evidence that male cyclists are more likely to suffer a bicycle fatal injury than female cyclists and that older cyclists (e.g., above 55 or 65 years old) are the most vulnerable age group (Kim et al. 2007, Eluru et al. 2008, Bíl et al. 2010).

The analysis of the predictors of the severity of bicycle crashes has been conducted using different types of analysis such as the generalized linear model of logistic regression, binary logit model, multinomial logit model, and mixed logit model (Klassen et al. 2014). However, because of the mass of complicated data on road accidents, it is difficult to use regression models to investigate the predictors of the severity of bicycle crashes. Firstly, regression models rely on different and strong statistical assumptions such as no outliers, linearity in modelling the relationship (Harrell 2001, Cohen et al. 2003, Tabachnick and Fidell 2012), which are hardly to be valid for accident data (Chang and Wang 2006, Yan et al. 2010, de Oña et al. 2011). Secondly, interaction may occur in complex forms and its detection using cross-product terms may be a daunting task (Yan et al. 2010). Thirdly, regression models may not satisfactorily handle many discrete variables or variables with a high number of categories (Harrell 2001, Cohen et al. 2003, Tabachnick and Fidell 2012).

Data mining techniques refer to an analytic process aimed at exploring large amounts of data (also known as 'big data' in the popular press) in search of structures, commonalities, hidden patterns (or rules) among data (Hand et al. 2001, Pujari 2001, Han et al. 2012). Data mining techniques such as CHAID decision tree technique and Bayesian network analysis have the following advantages: (1) no problem with outliers, (2) no assumption on variable distributions is made and a priori probabilistic knowledge about the severity of bicycle crashes is not needed, (3) many discrete variables or variables with a high number of categories are more properly handled compared to regression models, and (4) it is possible to extract information from large amounts of data (Breiman et al. 1984, Friedman et al. 1997, Sutton 2005, Strobl et al. 2009). CHAID decision tree technique and Bayesian network analysis have been successfully applied to investigate the predictors of head injury for pedestrians and cyclists (Badea-Romero and Lenard 2013), train–vehicle crashes at passive highway-rail grade crossings (Yan et al. 2010), traffic injury severity (Chang and Wang 2006, Mujalli et al. 2016), traffic accident injury severity on rural highways (de Oña et al. 2011, de Oña et al. 2013), and driver injury severity in rear-end crashes (Chen et al. 2015). However, to our knowledge, no research has used both CHAID decision tree technique and Bayesian network analysis in the study of the severity of bicycle crashes.

1.1 Study objectives

The main aim of the present study was to identify factors and rules crucial to the occurrence of fatal bicycle crashes. Crash characteristics (type of collision and opponent vehicle), infrastructure characteristics (type of carriageway, road type, road signage, pavement type, and type of road segment), cyclists (gender and age), and environmental factors (time of the day, day of the week, month, pavement condition, and weather) were considered as predictors of bicycle injury severity.

2. Method

2.1 Road transport in Italy

Based on data from the National Institute of Statistics, the total population of Italy in 2016 is 60,665,552. According to Eurostat, in Italy in 2013, the total length of motorways was 6,751 kilometres (based on the last available data from 2013), whereas the total length of other roads was 249,288 kilometres. The total length of state, provincial and communal roads was 19,920, 154,948, and 74,420 kilometres, respectively. The motor vehicles movement on national territory was 51,293 million vehicles-kilometres. The estimated passenger road transport on national territory was 770,590 million of passenger-kilometres.

2.2 Data

The data used in this study were provided by the Italian National Institute of Statistics (ISTAT). The ISTAT gathers data about all road crashes collected by public institution. The data are collected through a broad collaboration among different public institutions: ISTAT, Italian Automobile Club, the Italian Ministry of Transport and Infrastructure, different National Police organizations, and local Municipalities.

Table 1

Descriptive Statistics of Crash Data

Variable	Count	%	Variable	Count	%	Variable	Count	%
Month			Type of carriageway			Pavement Condition		
January	2444	4.9	One-way carriageway	10786	21.7	Dry	45079	90.8
February	2163	4.4	Two-way carriageway	34953	70.4	Wet	4178	8.4
March	3427	6.9	Two carriageway	3138	6.3	Slippery	234	0.5
April	3899	7.9	Two carriageways or more	744	1.5	Frozen	87	0.2
May	5322	10.7	Road Type			Snowy	43	0.1
June	5562	11.2	Urban regional	533	1.1	Road Signage		
July	5588	11.3	Urban municipal	39327	79.3	Absent	4171	8.4
August	4901	9.9	Urban provincial	3505	7.1	Vertical	3265	6.6
September	5247	10.6	Urban national	1035	2.1	Horizontal	3988	8.0
October	4680	9.4	Rural municipal	934	1.9	Vertical and horizontal	38197	77.0
November	3609	7.3	Rural provincial	2972	6.0	Weather		
December	2779	5.6	Rural national	796	1.6	Clear	44072	88.8
Day			Other roads	181	0.4	Foggy	267	0.5
Monday	7034	14.2	Rural regional	338	0.7	Rainy	2381	4.8
Tuesday	8194	16.5	Pavement type			Hail	10	0.0
Wednesday	7813	15.7	Paved street	49173	99.1	Snow	74	0.1
Thursday	8156	16.4	Uneven paved street	318	0.6	Strong wind	49	0.1
Friday	7830	15.8	Non paved street	130	0.3	Other	2768	5.6

Variable	Count	%	Variable	Count	%	Variable	Count	%
Saturday	6494	13.1	Road segment			Type of Collision		
Sunday	4100	8.3	Intersection	11031	22.2	Head-on collision	3201	6.5
Severity			Roundabout	4424	8.9	Angle collision	25082	50.5
Injury	48798	98.3	Signalized intersection	7391	14.9	Sideswipe collision	9611	19.4
Fatality	823	1.7	Intersection with traffic lights or policeman	2646	5.3	Hit parked or stationary vehicle	2721	5.5
Cyclist age			Unsignalized intersection	1200	2.4	Hit pedestrian	257	0.5
0 - 14	3142	6.3	Grade crossing	26	0.1	Rear-end collision	3920	7.9
15 - 24	5919	11.9	Straight road	20350	41.0	Hit stopped vehicle	344	0.7
24 - 44	14550	29.3	Curve	1998	4.0	Hit obstacle	778	1.6
45 - 54	7974	16.1	Bump or bottleneck	152	0.3	Run-off-the-road	1912	3.9
55 - 64	6236	12.6	Slope	301	0.6	Sudden Braking	104	0.2
65 and older	11504	23.2	Tunnel with street light	69	0.1	Falling from the vehicle	1691	3.4
Not specified	296	0.6	Tunnel without street light	33	0.1	Opponent Vehicle		
Cyclist gender			Time of the day			Car	35246	71.0
Male	33912	68.3	6 a.m. to 6 p.m.	40676	82.0	Bus	365	0.7
Female	15709	31.7	6 p.m. to midnight	7881	15.9	Truck	3050	6.1
			Midnight to 6 a.m.	898	1.8	Powered two wheelers	2952	5.9
			Not specified	166	0.3	Other vehicles	945	1.9
						Multiple vehicles	910	1.8
						No opponent vehicles	6153	12.4

In the present study, the dataset contains information about road crashes occurred on the Italian road network during the period ranging from 2011 to 2013. At the time of the study, 2013 was the most recent available ISTAT data. In 2010 (Law L. 29/7/2010 n. 120) a new national traffic law was approved, with minor changes involving also bicycle use. Therefore, to have a trade-off between the need to have a large sample size and the need to control for change in road regulation, we chose a three-year period ranging from 2011 to 2013.

The ISTAT database does not include a distinction between different levels of injuries, thus making a distinction only between road crashes resulting in injuries or fatalities (within 30 days). As shown in Table 1, the database was rearranged and 15 categorical variables were selected: (1) month of the year, (2) day of the week, (3) time of the day (4) cyclist age, (5) cyclist gender, (6) road type, (7) accident location, (8) road pavement type, (9) road pavement condition, (10) type of junction, (11) road signage, (12) weather condition, (13) type of collision, (14) type of opponent vehicle and (15) outcome of the crash. Regarding the road type category, the administrative classification of the Italian Road Code classifies roads as highways, national roads, regional roads, provincial roads, or municipal roads (Maggiora 2005). Each type of road is built, owned and maintained by different organizations. Highways and national roads are owned by the central government and maintained by the national roads agency (ANAS) or by contractors. Typically, the responsibility for municipal roads, provincial roads, and regional roads rests with each respective level of government (e.g., the local government is responsible for municipal roads). Furthermore, the Italian Road Code categorises roads crossing urban communities with less than 10,000 inhabitants as urban national, urban regional, and urban provincial roads, respectively. Regarding the type of collision, the ISTAT database provides a classification in 12 categories.

2.3 Statistical Analysis

We analysed data about road crashes occurred on the Italian road network from 2011 to 2013 using CHAID decision tree technique and Bayesian network analysis. We divided the dataset into training data (70%) and test dataset (30 %). In the present study, the CHi-squared Automatic Interaction Detection (CHAID) and Bayes network techniques were employed using IBM SPSS Modeler version 18. The CHAID is a decision tree algorithm that allows splitting into more than two subgroups. In the present study, we employed exhaustive CHAID because of its superior ability to examine all possible splits. For the purpose of cross-validation, the dataset was split into two parts: a training dataset and a test dataset. Specifically, the total data was split into 70% for training and 30% for the test data. The training dataset was used to estimate the model parameters and build the model, while the test dataset was used to test the model for its applicability to independent data and to determine model's ability to generalize. Given the intrinsic imbalanced nature of the data, we altered the misclassification penalty using cost matrix manipulation (McCormick et al. 2013). Specifically, we chose a misclassification cost ratio of 100:1 to force CHAID to identify the fatal injury cases correctly more often (Roumani et al. 2013).

The CHAID was also used to reduce the set of variables because Bayes network work best with a small set of predictors. The Bayes network analysis is based on Bayesian probability theory. To calculate a posterior distribution for variables of interest, Bayesian probability employs prior distributions of each variable and joint distributions. In the present study, we used the tree augmented naïve Bayesian because it models interactions (i.e., it allows each predictor to depend on one other predictor). To reduce the impact of the intrinsic imbalanced nature of the accidents data on Bayesian network analysis, we carried out simple random oversampling

(Mujalli et al. 2016). We refer to IBM (2016) for a detailed description of the algorithm used in CHAID Bayesian network analysis.

3. Results

The original database comprised a total of 575,093 road accidents, from which we extracted the 49,621 road accidents where at least one cyclist was injured or killed. Of these roads accidents involving at least one injured or killed cyclist, the number of bicycle fatalities was 823 (1.7%).

3.1 CHAID Decision Tree Technique

The CHAID decision tree technique belongs to a group of rule-based classifiers, and orders the rules in a tree structure. The percentage of records having the particular value for the outcome variable, given values for the input variables represents the confidence (accuracy) of the produced rules. Using CHAID decision tree technique, the overall classification accuracy of the training set and testing set was 98%. In addition, the area under the curve (a goodness of fit measure for the classifier) of the training set and testing set was 0.83 and 0.81, respectively. That indicates quite accurate classification with no overfitting.

The relative importance of the input variables in the model is indicated by the length of the bars and their corresponding values in Figure 1. Predictor importance was determined by calculating the decrease in variance of the outcome variable (i.e., severity of bicycle crashes) due to each predictor, through a sensitivity analysis. The values of predictor importance are relative, and the sum of the values for all predictors on the display is 1.0. We refer to IBM (2016) for a detailed description of the algorithm used here. The x-axis shows the predictors while the y-axis shows the predictor importance score for each variable. According to Figure 1, road type (0.30), crash type (0.24), and age of cyclist (0.19) were the most important predictors in determining the severity of bicycle crashes. However, predictor importance scores are not revelatory of the reasoning behind their predictions. To get a deeper insight into the predictions of CHAID, we should explore the decision tree.

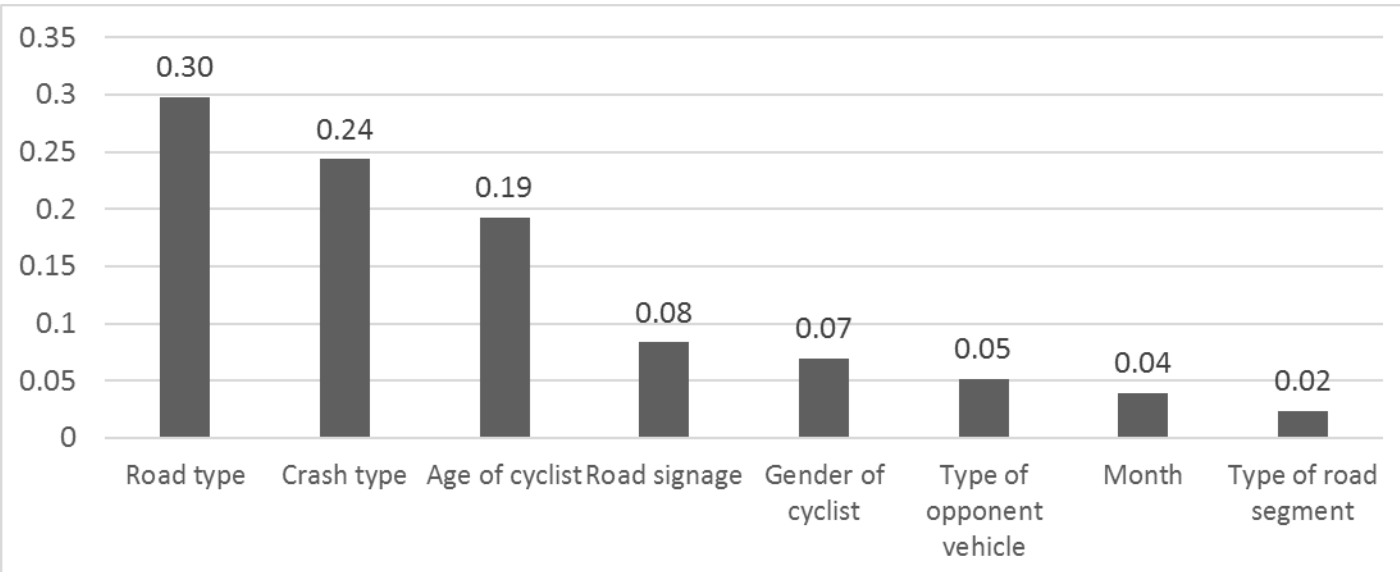

Figure 1. Predictor importance scores.

As it was explained earlier, CHAID is a classification method for building a decision tree. A decision tree split a data set into subgroups on the basis of the relationships between input variables (i.e., predictors of the severity of bicycle crashes) and the outcome variable (i.e., severity of bicycle crashes). At each tree node, the data is recursively split into two or more distinct groups by the values of an input variable, resulting in subgroups, which are then split again into smaller subgroups, and so on. To identify optimal splits, the CHAID employs the Chi-square independence test. The crosstabulations between each of the input variables and the outcome are examined and tested using a chi-square independence test. The CHAID selects the most significant input variable. If an input variable has more than two categories, the CHAID compares these categories, and those with no differences in the outcome are merged together. Therefore, the CHAID provides the details in the form of a decision tree model that classifies bicycle crashes resulting in non-fatal injury or fatal injury using a series of if–then–else rules. By using this type of decision tree model, researchers can understand the data structure or the combinations of variables that result in the highest (or lowest) risk for a condition of interest.

Figure 2 displays the final tree structure the severity of bicycle crashes. All bicycle crashes resulting in non-fatal injury or fatal injury were divided into 31 subgroups from root node to leaf nodes through different branches. The percentage of bicycle fatal crash varied from 0 to 11%. The tree structure involves eight splitting variables, including road type, road section type, cyclists' age, cyclists' gender, crash type, opponent vehicle, month, and road sign. The first optimal split in node 0 was according to road type, which classified bicycle crashes into four groups: if road type is urban regional, urban provincial or urban national, the tree predicts 2.37% of fatality crash; if road type is urban municipal, the percentage of fatality crash was 0.97%; if road type is rural municipal, the tree predicted 4.31% of fatality crash; and if road type is rural provincial, rural regional, or rural national, the percentage of fatality crash was 5.97%.

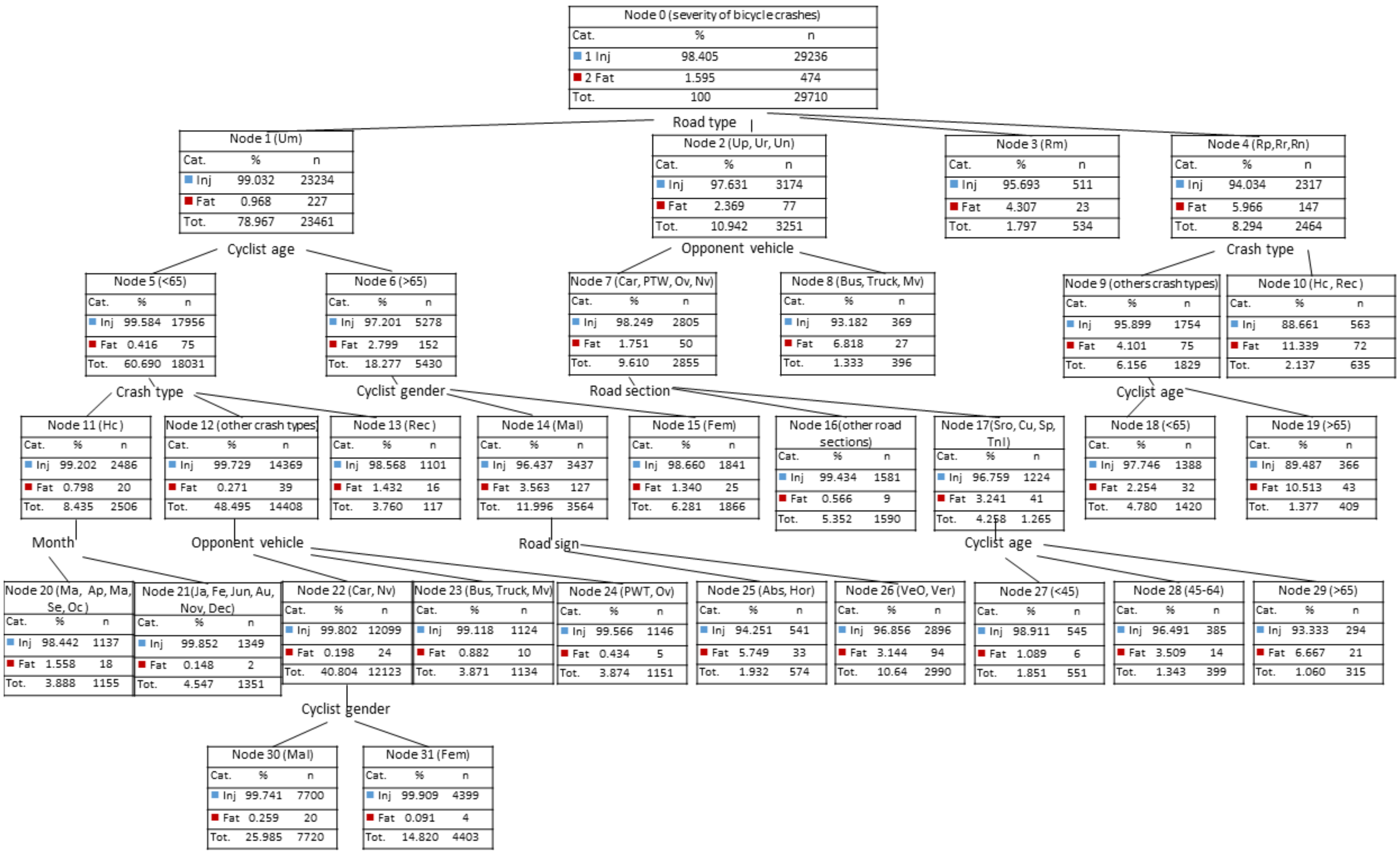

Figure 2. *Decision Tree.*

Note. Ja = January; Fe = February; Ma = March; Ap = April; Ma = May; Jun = June; Jul = July; Au = August; Se = September; Oc = October; Nov = November; Dec = December; <45 = age less than 45 years; 45-64 = age between 65 and 64 years; <65 = age less than 65 years; >65 = age 65 and older; Inj = Injury; Fat = Fatality; Mal = Male; Fem = Female; Ur = Urban regional; Um = Urban municipal ; Up = Urban provincial; Un = Urban national; Rm = Rural municipal; Rp = Rural provincial; Rn = Rural national; Oro = Other road; Rr = Rural regional; Int = Intersection; Rou = Roundabout; Sint = Signalized intersection; Itl = Intersection with traffic lights or policeman; Nsi = Non signalized intersection; Gc = Grade crossing; Sro = Straight road; Cu = Curve; Bob = Bump or bottleneck; Sp = Slope; Twl = Tunnel with street light; Tnl = Tunnel without street light; Abs = Absent; Ver = Vertical; Hor = Horizontal; VeO = Vertical and horizontal; Hc = Head-on collision; Ac = Angle collision; Sc = Sideswipe collision; Rec = Rear-end collision; Hp = Hit pedestrian; Hpsv = Hit parked or stationary vehicle; Hsv = Hit stopped vehicle; Ho = Hit obstacle in carriageway; Rr = Run-off-the-road; Sb = Sudden Braking; Ffv = Falling from the vehicle; Car = Car; Bus = Bus; Truck = Truck; Powered two wheelers = PTW; Ov = Other vehicles; Mv = Multiple vehicles; Nv = No opponent vehicles.

In the second level of the tree, the group including urban regional, urban provincial, and urban national road type led to another split based on type of opponent vehicle. If the opponent vehicle is a bus, a truck, or multiple vehicles are involved, the percentage of fatality crash was 6.82%, whereas if the opponent vehicle is a car, a powered two-wheeler or there was not opponent vehicle, the percentage of fatality crash was 1.75%. In the third level of the tree, for this group of opponent vehicle (i.e., cars, powered two wheelers or no opponent vehicle), type of road section segmented the data into two subgroups: in case of straight, curved, or steep road or tunnel without lighting, the percentage of bicycle crash was 3.24%, whereas in the other types of road section (e.g., intersection, roundabout, tunnel with lighting), the percentage of fatal crash was 0.57%. In the fourth level of the tree, the age of cyclists segmented the data concerning straight, curved, or steep road or tunnel without lighting into three subgroups. If the age of the cyclist was 65 years or higher, the percentage of fatal crash was 6.67%. The percentage of fatal crash decreased to 3.51% among cyclist aged between 45 and 64 years and to 1.09% among cyclist younger than 44 years.

In the second level of the tree, crash type segmented the group of rural provincial, rural regional, or rural national road type into two groups. In case of head-on or rear-end collisions, the percentage of fatal crash was 11.34%, whereas in the other types of crash, the percentage of fatality crash dropped to 4.10%. In the third level, the age of cyclists split the other type of crash category (i.e., excluding head-on or rear-end crash type) into two groups. If cyclists were 65 years old or older, the percentage of fatal crash was 10.51%, whereas this percentage among the other cyclists was 2.25%.

In the second level of the tree, the age of cyclists segmented the data regarding urban municipal in two groups. If the age of cyclist was 65 years or higher, the percentage of fatal crash was 2.80%, whereas if the age was lower than 65 years the percentage of fatal crashes decreased to 0.42%. In the third level, the gender of cyclists led to the split in cyclists aged 65 years or older. The percentage of fatal crash was 1.34% among female cyclists and 3.56% among male cyclists. In the fourth level of the tree, type of road sign led to another split among the group of male cyclists: if road sign was missing or there were only road markings, the percentage of fatal crash was 5.75%, whereas it dropped to 3.14% if a road sign was present. In the third level, crash type led to the split in the groups of cyclists aged less than 65 years. The percentage of fatal crash was 1.43% in case of rear-end crash type, 0.80% in case of head-on, fall from the vehicle, skid, and run-off-the-road, and 0.27% in the other types of crash. In the fourth level, head-on, fall from the vehicle, skid, and run-off-the-road crash types were segmented in two groups according to the month of the year. The percentage of fatality crash was 1.16% in March, April, May, September, and October, whereas was 0.15% in the other months of the year. In the fourth level, the 'other' type of crash (i.e., excluding rear-end, head-on, fall from the vehicle, skid, and run-off-the-road) was split into three groups according to the type of opponent vehicle. This finding indicates that the involvement of a bus, a truck or multiple vehicles had a higher percentage (0.88%) of fatal crash compared to crashes involving no vehicle or a car (0.20%) or PWT or other vehicles (0.43%). In the fifth level, the gender of cyclist led to the split in the group of car or no opponent vehicle. In case of male cyclists, the percentage of fatality crash was 0.26%, whereas in case of female cyclists the same percentage dropped to 0.09%.

3.2 Bayesian Network Analysis

The eight predictors of the severity of bicycle crashes that were selected using CHAID algorithm (see Figure 1) were included as predictors of the target (i.e., severity of bicycle crashes) in Bayesian network analysis. The accuracy of the Bayesian network model is 79% for both the training set and the test set, which is a good value. The area under the curve of both the training set and testing set was 0.86. A Bayesian network is a probabilistic graphical modelling technique that shows variables (referred to as nodes) in an acyclic graph. The acyclic graph represents the probabilistic, or conditional, independencies between the nodes described through the links in the network (also known as arcs). In other words, a Bayesian network model consists of the directed acyclic graph with nodes and a set of directed edges together with a conditional probability table for each node given values of its parent nodes. Figure 3 displays the resulting network graph of nodes that shows the association between the target and its predictors. In the tree augmented naïve Bayesian, each predictor (i.e., characteristics of bicycle crashes) has the target variable (i.e., severity of bicycle crashes) as a parent and can have one other predictor as a parent. The network was consistent of nine nodes, one for the target and one for each predictor. The relationship between the predictors is also displayed. The graphical model highlights the predictor importance (i.e., the relative importance of each predictor in estimating the model): the darkness indicates the closeness of the relationship to severity of bicycle crashes. The darkest coloured predictors, and, thus, the most important predictors of severity of bicycle crashes were crash type (0.31), road type (0.19), and type of opponent vehicle (0.18). As these three predictors were identified as the key determinants of severity of bicycle crashes, the three related relationships will be further discussed. As it was explained earlier, the Bayesian network model provides a conditional probability table for each related node. The Bayesian network model computes the joint probability distribution as a product of conditional probabilities for all nodes, given the values of each node's parents. Each column of the conditional probability table corresponds to a value of the predictor while each row corresponds to a combination of values of the target and parent predictor variables.

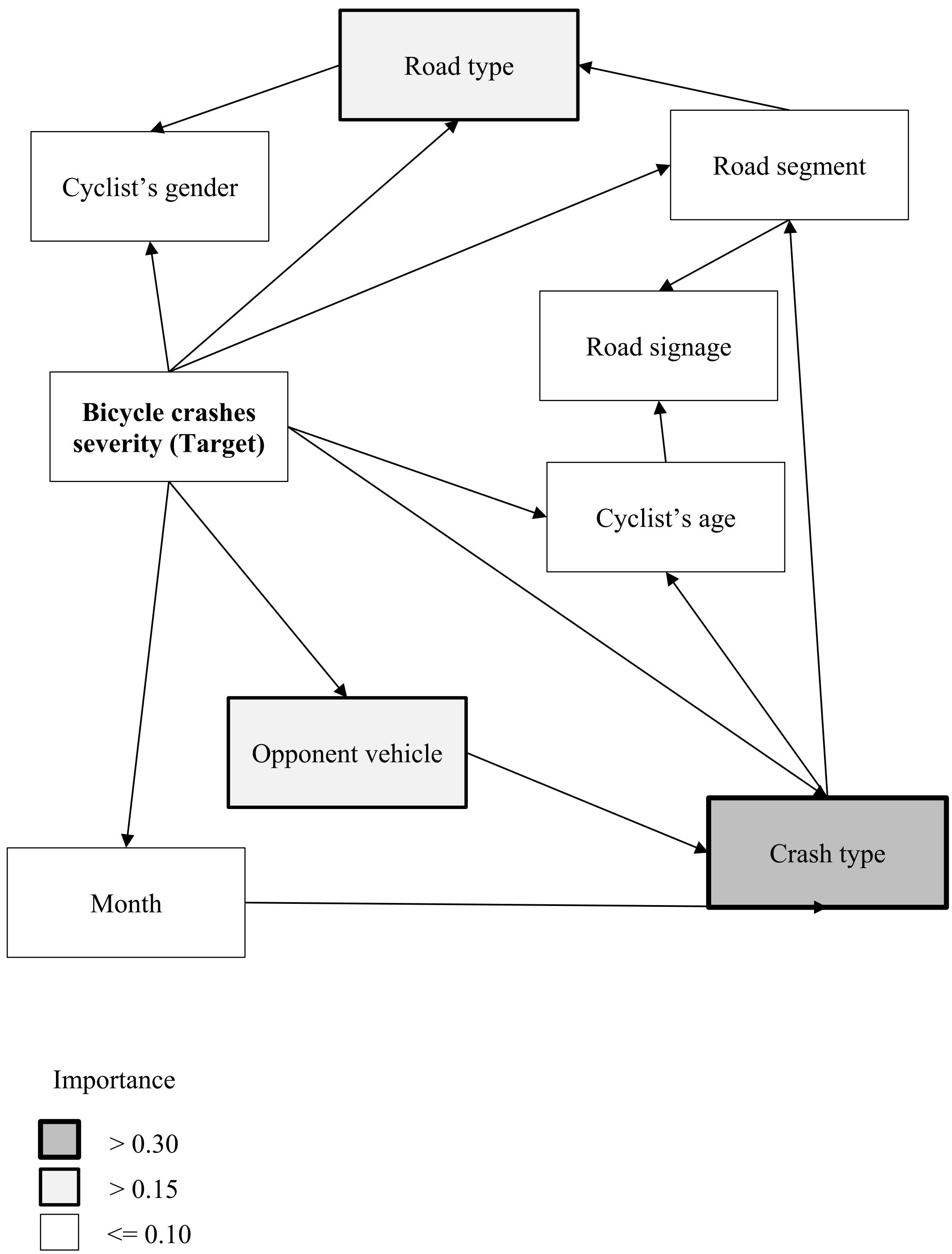

Figure 3. The Bayesian network model and predictor importance.

Table A1 (see Appendix A) summarizes the conditional probability for each values of crash type across all combination of values of target and month (i.e., its parents). The conditional probabilities of crash type suggest that fatality crashes were less probable than injuries crashes following angle crashes with another vehicle, especially in the period between February and December. In the same period, fatality crashes were more likely than injuries crashes following rear-end collisions.

Table A2 (see Appendix A) displays the conditional probabilities of road type taking into consideration the influence of road segment. Compared to injuries crashes, fatalities crashes were less likely in urban provincial road, especially at non-signalized intersection, straight road, and tunnel with street light. However, fatality crashes were more likely than injuries crashes in urban provincial road inside tunnel without street light.

Table A3 (see Appendix A) shows the conditional probabilities of type of opponent vehicle considering the influence of crash type. Fatality crashes were more likely than injuries crashes in collisions involving trucks following angle or sideswipe collisions and collisions involving multiple vehicles where a stopped vehicle was hit. Moreover, fatality crashes were more likely than injuries crashes in collisions involving a car following three types of collisions: angle, sideswipe, and hit stopped vehicle.

4. Discussion

Results from both CHAID decision tree technique and Bayesian network analysis revealed that crash type and road type were the most important predictors of the severity of bicycle accidents. According to CHAID decision tree technique, rear-end collisions increased the severity of bicycle accidents in urban municipal road and, especially, in rural provincial, rural regional, and rural national roads. In these types of rural roads, more than one out of ten bicycle injury collisions results in bicycle fatality. Bayesian network analysis showed that rear-end collisions were the most dangerous types of collisions, while angle crashes were the less dangerous. Rear-end collisions often imply an impact on cyclists who may not expect a crash with an oncoming motor vehicle and, therefore, are not ready to prevent the damages of the collision. However, the findings of Bayesian network analysis also showed that fatality crashes were more likely than injuries crashes in angle collisions involving a truck or a car. The most likely explanation for this apparent discrepancy is that, as in previous research (Yan et al. 2011), among patterns of types of crash, angle collisions occurred most frequently and, therefore, may involve different types of vehicle other than cars and trucks. Thus, when considering all the types of vehicle involved in angle collisions with bicycles, they may not be considered particularly dangerous. However, consistent with previous research (Moore et al. 2011, Yan et al. 2011), angle collisions involving cars or trucks significantly increased the level of bicyclist injury severity.

The results show that — in line with the literature (Macpherson et al. 2004, Amoros et al. 2011, Moore et al. 2011, Boufous et al. 2012) — the severity of bicycle crashes is different between rural and urban roads. There are clear differences (e.g., speed limits and actual speed, traffic flow, road design, lack of appropriate cycling infrastructure) between rural and urban roads which may impact the severity of bicycle crashes. The present study adds to literature by finding that the relationship between road type and severity of bicycle crashes is much more complex than the distinction between rural and urban roads. Urban regional, urban provincial or urban national roads cross small urban centres (urban communities with population less than 10,000 inhabitants). In these segments of urban roads, the speed of drivers of motorized vehicles is generally high (Montella et al. 2012). Motorized vehicle speed is one of the factors that increase the probability of a bicyclist suffering a fatal injury in a crash because of the increased kinetic energy and greater impact (Kim et al. 2007, Moore et al. 2011). Thus, a likely explanation is that over-speeding in urban areas is more probable in urban regional, urban provincial, or urban national than municipal roads. Indeed, a previous study on powered two-wheeler crashes in Italy revealed that crash severity is substantially lower in municipal roads than other urban roads (Montella et al. 2012).

Consistent with past research (McCarthy and Gilbert 1996, Kim et al. 2007, Yan et al. 2011), it was found in the current study that in a crash event where a large vehicle (i.e., truck or bus) was the opponent vehicle, the likelihood of fatality crash increases. The present study adds to literature by finding that this increased risk is not similar across urban regional, urban provincial, or urban national roads. Since over-speeding may be a problem in these types of urban roads (Montella et al. 2012), the bicycles' tendency to be in blind spots and bicycle's poor conspicuity may be exacerbated. In addition to being visible (i.e., to be usefully seen by satisfying geometric and optical requirements), a bicycle must also be conspicuous, that is, being able to attract the driver's attention (Langham and Moberly 2003). In rural roads, the increased crash severity does not seem to differ between large vehicles and other motorized vehicles because the increased speed limits are enough to increase the severity of collisions with every motorized vehicle.

In addition, this study goes beyond the existing literature by showing that not only the involvement of large vehicles increases the severity of bicycle crashes, but also the involvement of multiple vehicles. This is not surprising: when more vehicles are involved, multiple impacts are more likely (Tay and Rifaat 2007). However, the involvement of multiple vehicles has received little attention in the literature on bicycle safety though its occurrence is not rare. We note the in the present study, the involvement of truck was about 6% of the accidents and the involvement of multiple vehicles was about 2% the accidents. Moreover, a study on safety performance of roundabouts revealed that bicyclists were involved in 35% of the multiple vehicle-crashes (Daniels et al. 2010). The impact of multiple-vehicles accidents on bicycle safety may be an area for future research.

In line with several studies (Haileyesus et al. 2007, Bíl et al. 2010, Amoros et al. 2011), male cyclists were more likely to sustain a fatal injury than female cyclists. This variation may be explained by differences in bicycling exposure, risk-taking behaviours, and helmet use. Male cyclists have a greater exposure rate and case fatality rate than female cyclists (Li and Baker 1996).Compared to female cyclists, male cyclists have a higher tendency towards disregarding potential risks and committing traffic violations, including non-compliant roadway-crossing, disobeying the traffic signal at signalized intersections (Bernhoft and Carstensen 2008, Deffenbacher 2008, Yan et al. 2011, Johnson et al. 2013). Helmet use is lower in male cyclists than female cyclists (Harlos et al. 1999) and risk compensation has been observed only among male cyclists as helmeted male bicyclists tended to ride faster than non-helmeted ones (Messiah et al. 2012). In the present study, we have found that gender differences in severity of bicycle crashes are marked in urban municipal road, while in the other types of road, there is no evidence of gender differences. This finding suggests that gender differences in severity of bicycle crashes are context-specific. This could explain why evidence concerning gender differences in severity of bicycle crashes has been inconsistent, with some studies reporting no difference in this regard (Hoffman et al. 2010, Heesch et al. 2011). With regard to the fact that male cyclists were more likely to sustain a fatal injury than female cyclists in urban municipal roads, we argue that road type or (urban/rural) environment is a potentially important situational variable. Compared to rural environments, several factors are more characteristics of urban environments: congestion, rush-hour traffic, crowding, time-pressured commutes, more intersections and traffic lights (Deffenbacher 2008). These characteristics of urban environment may provide more chances to commit traffic violations and risk-taking behaviours which are more likely among male cyclists than female cyclists (Bernhoft and Carstensen 2008, Deffenbacher 2008, Yan et al. 2011, Johnson et al. 2013). Indeed, there is evidence that rural drivers are less likely to commit traffic violations than urban drivers (Zhang et al. 2013). This explanation should be examined in future research.

Injury severity increased among cyclists aged 65 and over compared to the youngest age group. This result is in line with the literature showing that injury severity increases with age (Eilert-Petersson and Schelp 1997, Rodgers 1997, Ekman et al. 2001, Stone and Broughton 2003, Kim et al. 2007, Eluru et al. 2008, Bíl et al. 2010, Yan et al. 2011, Boufous et al. 2012, Schepers 2012, Rivara et al. 2015). Physical fragility (susceptibility to injury) and, to a lesser extent, crash over-involvement due to of unsafe driving are likely to explain the excess death rates among older drivers per vehicle-mile of travel (Li et al. 2003, Anstey et al. 2005, Schepers 2012). Susceptibility to injury due to fragility of older cyclists seems to be one possible explanation for the increased likelihood of sustaining a fatal injury since the protection of cyclists is more worrisome than the protection of vehicle occupants. In the present study, in municipal roads (the less dangerous among all the types of road probably because of the low operating speed of motorized vehicles), the percentage of fatal injury is 0.42% among people aged less than 65 years, whereas is 2.80% among people aged 65 years and over. This finding seems to support the hypothesis of physical fragility: even a slight mishap can have serious consequences. Another possible explanation could be linked to risk factors associated with older age. As indicated in earlier studies (Eluru et al. 2008, Rivara et al. 2015) older individuals tend to have higher perception and reaction times which contribute to their higher injury risk propensity when cycling. Furthermore, Maring and van Schagen (1990) pointed out that even though age by itself was not the causal factor, older age was strongly associated with relevant variables such as less perceptual-motor speed and cognitive deterioration. Another risk factor for older cyclists, as it has been found for older drivers (Ball et al. 1993, Caird et al. 2005) could be their propensity to lower attentive states during the riding task.

In our study, the absence of road markings seems to increase the crash severity in older male cyclists in urban roads. Marked centre and edge lines provide a visual reference to guide motorists in the driving task, but potentially for cyclists as well. Schepers and den Brinker (2011) found that the characteristics of the visual design play a role in crashes where cyclists collide with a kerb, bollard or road narrowing, or ride onto the verge. They recommended a minimal level of guidance (e.g., edge markings) and conspicuity of obstacles (e.g., bollards).

In the present study, we found an increase in crash severity during spring (March, April, May) and the beginning of autumn (September and October), compared to other period of the year. As previously suggested in literature, the season and weather conditions have an influence on bicycle crashes (Liu et al. 1995, Kaplan and Giacomo Prato 2013). We believe that the unpredictability of the weather conditions in those specific months plays a key role in increasing the severity of crashes. As a matter of fact, the weather conditions are more variable in spring and autumn and thus road users could find themselves forced to drive or ride in adverse weather conditions without expecting it. Sudden bad weather could entail a more slippery road pavement and less conspicuity by the road users.

4.1 Limitation of the Study

Several limitations of this study also deserve comment. Although ISTAT collects the most complete data of road accidents in Italy, similar to other countries, some crashes and some important variables that may affect bicycle safety may be unavailable. We believe that the main limitation of the study is the limitation of the data available. For instance, vehicle speed prior to impact plays an important role in increasing the probability of fatal injury (Kim et al. 2007). Since the vehicle speed prior to impact was not collected, in-depth accident studies should enhance our understanding of the factors predicting the severity of bicycle crashes. Another important variable not included in the ISTAT database is the traffic flow condition. As a matter of fact, it is reasonable to argue that crashes in low traffic conditions could entail different risk factors and knowing the traffic conditions at the very moment of the crash could give more insight on the weight of different predictors and outcomes. Another flaw in the ISTAT database is that the classification of roads does not corresponds entirely to the functional classification present in the Italian Highway Code. In addition, when those data are available in crash databases, future studies are recommended to expand and update the extent of the current research. Finally, the predictors were based on previous theoretical and empirical work. Although the establishment of temporal ordering is essential for making firm causal interpretations, it is not sufficient. Some unobserved “third” variables may better explain the observed relations.

4.2 Conclusions and Recommendations

The issue of cyclist safety is crucial. In the present study, we employed CHAID decision tree technique and Bayesian network analysis to determine the predictors of the severity of bicycle crashes. According to the results of CHAID analysis, the most important predictors were, in decreasing order of importance, road type, crash type, age of cyclist, road signage, gender of cyclist, type of opponent vehicle, month, and type of road segment. These eight variables were included as predictors of the target (i.e., severity of bicycle crashes) in Bayesian network analysis. By applying Bayesian network on these eight predictors, crash type, road type, and type of opponent vehicle resulted as the most important predictors of severity of bicycle crashes.

These findings suggest the importance of divisions on rural roads (i.e., rural provincial, rural regional, or rural national), which can separate bicycles from motor vehicles maintain high operating speeds. A bikeway separated from motorized traffic is likely to reduce the possibility of bicyclists riding with high-speed traffic, and, thus, reduce the risk for leading to those most dangerous patterns of crashes (i.e., rear-end, head-on). A bikeway separated from motorized traffic could be effective at reducing the severity of crashes in rural roads where motor vehicles maintain high operating speeds and head-on and rear-end collisions are more fatal. It is interesting to note that older cyclists are more concerned about the absence of a bikeway separated from motorized traffic and tend to feel the presence of cycle paths most important for their comfort (Bernhoft and Carstensen 2008). In addition, an in-bicycle consumer-friendly vehicle detection system could warn motorists of the cyclist's presence by flashing lights and, at the same time, inform the cyclist about the speed and distance of approaching vehicles. Another recommended countermeasure for reducing the frequency of rear-end collisions is increasing rear conspicuity of bicycles or bicyclists. According to a systematic review (Kwan and Mapstone 2006), fluorescent materials in yellow, red, and orange colours improve detection and recognition of cyclists in the daytime. For night-time conspicuity, lamps, flashing lights, and retroreflective materials in red and yellow colours increase detection and recognition. Bicycle lights improve conspicuity and decrease the risk of an accident and they are assumed to decrease severity due to reduced reaction time and the ability to take evasive action for the vehicle driver involved in the accident (Kim et al. 2007). In several countries, it is therefore mandatory to use lights during night-time, including Italy. Evidence-based public campaigns and police enforcement can increase the willingness to use bicycle lights.

In municipal urban roads, a bikeway separated from motorized traffic could be less effective in reducing the severity of bicycle crashes given the low risk of fatal bicycle crashes. (Mulvaney et al. 2015). Integration as opposed to segregation, as expressed by the concept of urban shared spaces (Hamilton-Baillie 2008a, b, Biddulph 2012, Karndacharuk et al. 2014), could be the most promising approach to reduce the severity of bicycle crashes. In urban regional, urban provincial, or urban national roads, speed-reducing measurements, such as speed calming measures, speed bumps, and elevated bicycle crossings, could be effective countermeasures to mitigate the problem of excessive speed. A driving simulator experiment revealed that perceptual cues such as gateways (aimed at reducing the speed of vehicles entering in the urban area) and traffic calming devices (aimed at complementing the gateway effect inside the urban area) have proved to be effective in reducing speed in rural highway crossing a small urban community (Galante et al. 2010). A recent review of the literature recommends the use of 30 km (20 mph) speed restrictions in urban areas to effectively reduce bicycle crashes (Mulvaney et al. 2015).

Finally, given that large vehicles increased the severity of bicycle crashes, in-vehicle systems that detect and alert drivers of the cyclists' presence in traffic could be useful. Also, infrastructure-based detection and cooperative systems could be useful to improve detection of cyclists and may assist drivers in minimizing blind spots.

Acknowledgements

We thank Víctor Marín Puchades and Marco De Angelis for helping us to prepare the dataset.

Funding source

This work was supported by the European Commission under the Horizon 2020 Framework Programme of the European Union (2014-2020). Project XCYCLE contract number: 635975.

References

Amoros, E., Chiron, M., Thélot, B., Laumon, B., 2011. The injury epidemiology of cyclists based on a road trauma registry. BMC Public Health 11 (1), 1-12.

Anstey, K.J., Wood, J., Lord, S., Walker, J.G., 2005. Cognitive, sensory and physical factors enabling driving safety in older adults. Clin Psychol Rev 25 (1), 45-65.

Ashbaugh, S.J., Macknin, M.L., Vanderbrug Medendorp, S., 1995. The ohio bicycle injury study. Clinical Pediatrics 34 (5), 256-260.

Atkinson, J.E., Hurst, P.M., 1983. Collisions between cyclists and motorists in new zealand. Accident Analysis & Prevention 15 (2), 137-151.

Badea-Romero, A., Lenard, J., 2013. Source of head injury for pedestrians and pedal cyclists: Striking vehicle or road? Accident Analysis & Prevention 50, 1140-1150.

Ball, K., Owsley, C., Sloane, M.E., Roenker, D.L., Bruni, J.R., 1993. Visual attention problems as a predictor of vehicle crashes in older drivers. Investigative Ophthalmology & Visual Science 34 (11), 3110-3123.

Bernhoft, I.M., Carstensen, G., 2008. Preferences and behaviour of pedestrians and cyclists by age and gender. Transportation Research Part F: Traffic Psychology and Behaviour 11 (2), 83-95.

Biddulph, M., 2012. Radical streets? The impact of innovative street designs on liveability and activity in residential areas. Urban Design International 17 (3), 178-205.

Bíl, M., Bílová, M., Müller, I., 2010. Critical factors in fatal collisions of adult cyclists with automobiles. Accident Analysis & Prevention 42 (6), 1632-1636.

Boufous, S., De Rome, L., Senserrick, T., Ivers, R., 2012. Risk factors for severe injury in cyclists involved in traffic crashes in victoria, australia. Accident Analysis & Prevention 49, 404-409.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees Wadsworth International Group, Belmont, CA, USA.

Caird, J.K., Edwards, C.J., Creaser, J.I., Horrey, W.J., 2005. Older driver failures of attention at intersections: Using change blindness methods to assess turn decision accuracy. Human Factors: The Journal of the Human Factors and Ergonomics Society 47 (2), 235-249.

Chang, L.-Y., Wang, H.-W., 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. Accident Analysis & Prevention 38 (5), 1019-1027.

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. Accident Analysis & Prevention 80, 76-88.

Chen, P., 2015. Built environment factors in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach. Safety Science 79, 336-343.

Chong, S., Poulos, R., Olivier, J., Watson, W.L., Grzebieta, R., 2010. Relative injury severity among vulnerable non-motorised road users: Comparative analysis of injury arising from bicycle–motor vehicle and bicycle–pedestrian collisions. Accident Analysis & Prevention 42 (1), 290-296.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2003. Applied multiple regression/correlation analysis for the behavioral sciences Erlbaum, Mahwah, NJ.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accid Anal Prev 42 (2), 393-402.

De Hartog, J.J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? Environmental Health Perspectives 118 (8), 1109-1116.

De Nazelle, A., Nieuwenhuijsen, M.J., Anto, J.M., Brauer, M., Briggs, D., Braun-Fahrlander, C., Cavill, N., Cooper, A.R., Desqueyroux, H., Fruin, S., Hoek, G., Panis, L.I., Janssen, N., Jerrett, M., Joffe, M., Andersen, Z.J., Van Kempen, E., Kingham, S., Kubesch, N., Leyden, K.M., Marshall, J.D., Matamala, J., Mellios, G., Mendez, M., Nassif, H., Ogilvie, D., Peiro, R., Perez, K., Rabl, A., Ragettli, M., Rodriguez, D., Rojas, D., Ruiz, P., Sallis, J.F., Terwoert, J., Toussaint, J.F., Tuomisto, J., Zuurbier, M., Lebret, E., 2011. Improving health through policies that promote active travel: A review of evidence to support integrated health impact assessment. Environ Int 37 (4), 766-77.

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. Accident Analysis & Prevention 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. Accident Analysis & Prevention 43 (1), 402-411.

Deffenbacher, J.L., 2008. Anger, aggression, and risky behavior on the road: A preliminary study of urban and rural differences. Journal of Applied Social Psychology 38 (1), 22-36.

Eilert-Petersson, E., Schelp, L., 1997. An epidemiological study of bicycle-related injuries. Accident Analysis & Prevention 29 (3), 363-372.

Ekman, R., Welander, G., Svanstrom, L., Schelp, L., Santesson, P., 2001. Bicycle-related injuries among the elderly--a new epidemic? Public Health 115 (1), 38-43.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention 40 (3), 1033-1054.

European Commission, 2014. Quality of transport. Special eurobarometer 422a / wave eb82.2 – tns opinion & social.

European Commission, 2015. Traffic safety basic facts on main figures. In: Transport, D.G.F. ed. European Commission.

Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Machine Learning 29 (2), 131-163.

Galante, F., Mauriello, F., Montella, A., Pernetti, M., Aria, M., D'ambrosio, A., 2010. Traffic calming along rural highways crossing small urban communities: Driving simulator experiment. Accid Anal Prev 42 (6), 1585-94.

Götschi, T., Garrard, J., Giles-Corti, B., 2016. Cycling as a part of daily life: A review of health perspectives. Transport Reviews 36 (1), 45-71.

Hagel, B.E., Romanow, N.T.R., Morgunov, N., Embree, T., Couperthwaite, A.B., Voaklander, D., Rowe, B.H., 2014. The relationship between visibility aid use and motor vehicle related injuries among bicyclists presenting to emergency departments. Accident Analysis & Prevention 65, 85-96.

Haileyesus, T., Annest, J.L., Dellinger, A.M., 2007. Cyclists injured while sharing the road with motor vehicles. Inj Prev 13.

Hamann, J.C., Peek-Asa, C., Lynch, C.F., Ramirez, M., Hanley, P., 2015. Epidemiology and spatial examination of bicycle-motor vehicle crashes in iowa, 2001–2011. Journal of Transport & Health 2 (2), 178-188.

Hamilton-Baillie, B., 2008a. Shared space: Reconciling people, places and traffic. Built Environment 34 (2), 161-181.

Hamilton-Baillie, B., 2008b. Towards shared space. Urban Design International 13 (2), 130-138.

Han, J., Pei, J., Kamber, M., 2012. Data mining: Concepts and techniques Elsevier, Waltham, USA.

Hand, D.J., Mannila, H., Smyth, P., 2001. Principles of data mining MIT press, Massachusetts, USA.

Harlos, S., Warda, L., Buchan, N., Klassen, T.P., Koop, V.L., Moffatt, M.E.K., 1999. Urban and rural patterns of bicycle helmet use: Factors predicting usage. Injury Prevention 5 (3), 183-188.

Harrell, F., 2001. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis Springer, New York.

Heesch, K.C., Garrard, J., Sahlqvist, S., 2011. Incidence, severity and correlates of bicycling injuries in a sample of cyclists in queensland, australia. Accid Anal Prev 43.

Hoffman, M.R., Lambert, W.E., Peck, E.G., Mayberry, J.C., 2010. Bicycle commuter injury prevention: It is time to focus on the environment. J Trauma 69 (5), 1112-7; discussion 1117-9.

IBM, 2016. IBM SPSS modeler 18.0 algorithms guide. IBM Corporation, North Castle Drive, Armonk, NY.

Istat, 2015. I.Stat is the warehouse of statistics currently produced by the italian national institute of statistics.

Johnson, M., Charlton, J., Oxley, J., Newstead, S., 2013. Why do cyclists infringe at red lights? An investigation of australian cyclists' reasons for red light infringement. Accid Anal Prev 50, 840-7.

Kaplan, S., Giacomo Prato, C., 2013. Cyclist-motorist crash patterns in denmark: A latent class clustering approach. Traffic Inj Prev 14 (7), 725-33.

Karndacharuk, A., Wilson, D.J., Dunn, R., 2014. A review of the evolution of shared (street) space concepts in urban environments. Transport Reviews 34 (2), 190-220.

Kelly, P., Kahlmeier, S., Gotschi, T., Orsini, N., Richards, J., Roberts, N., Scarborough, P., Foster, C., 2014. Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship. Int J Behav Nutr Phys Act 11, 132.

Kim, J.-K., Kim, S., Ulfarsson, G.F., Porrello, L.A., 2007. Bicyclist injury severities in bicycle–motor vehicle accidents. Accident Analysis & Prevention 39 (2), 238-251.

Kim, K., Li, L., 1996. Modeling fault among bicyclists and drivers involved in collisions in hawaii, 1986-1991. Transportation Research Record: Journal of the Transportation Research Board 1538, 75-80.

Klassen, J., El-Basyouny, K., Islam, M.T., 2014. Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: A city of edmonton case study. Safety Science 62, 295-304.

Klop, J., Khattak, A., 1999. Factors influencing bicycle crash severity on two-lane, undivided roadways in north carolina. Transportation Research Record: Journal of the Transportation Research Board 1674, 78-85.

Kwan, I., Mapstone, J., 2006. Interventions for increasing pedestrian and cyclist visibility for the prevention of death and injuries. Cochrane Database Syst Rev (4), Cd003438.

Langham, M., Moberly, N., 2003. Pedestrian conspicuity research: A review. Ergonomics 46 (4), 345-363.

Li, G., Baker, S.P., 1996. Exploring the male-female discrepancy in death rates from bicycling injury: The decomposition method. Accid Anal Prev 28 (4), 537-40.

Li, G., Braver, E.R., Chen, L.H., 2003. Fragility versus excessive crash involvement as determinants of high death rates per vehicle-mile of travel among older drivers. Accid Anal Prev 35 (2), 227-35.

Liu, X., Shen, D., Huang, J., 1995. Analysis of bicycle accidents and recommended countermeasures in beijing, china. Transportation research record (1487), 75-83.

Macmillan, A., Connor, J., Witten, K., Kearns, R., Rees, D., Woodward, A., 2014. The societal costs and benefits of commuter bicycling: Simulating the effects of specific policies using system dynamics modeling. Environ Health Perspect 122 (4), 335-44.

Macpherson, A.K., To, T.M., Parkin, P.C., Moldofsky, B., Wright, J.G., Chipman, M.L., Macarthur, C., 2004. Urban/rural variation in children’s bicycle-related injuries. Accident Analysis & Prevention 36 (4), 649-654.

Maggiora, E., 2005. Le strade comunali e provinciali: Regime giuridico, classificazione, uso, circolazione, polizia, responsabilità: Problemi e casi pratici [municipal and provincial roads: Jurisdiction, classification, usage, traffic, police, responsibilities. Problems and practical cases] Giuffrè, Milano.

Maring, W., Van Schagen, I., 1990. Age dependence of attitudes and knowledge in cyclists. Accident Analysis & Prevention 22 (2), 127-136.

Mccarthy, M., Gilbert, K., 1996. Cyclist road deaths in london 1985–1992: Drivers, vehicles, manoeuvres and injuries. Accident Analysis & Prevention 28 (2), 275-279.

Mccormick, K., Abbott, D., Brown, M.S., Khabaza, T., Mutchler, S.R., 2013. IBM SPSS modeler cookbook Packt Publishing, Birmingham, UK.

Messiah, A., Constant, A., Contrand, B., Felonneau, M.L., Lagarde, E., 2012. Risk compensation: A male phenomenon? Results from a controlled intervention trial promoting helmet use among cyclists. Am J Public Health 102 Suppl 2, S204-6.

Montella, A., Aria, M., D’ambrosio, A., Mauriello, F., 2012. Analysis of powered two-wheeler crashes in italy by classification trees and rules discovery. Accident Analysis & Prevention 49, 58-72.

Moore, D.N., Schneider Iv, W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. Accident Analysis & Prevention 43 (3), 621-630.

Mujalli, R.O., López, G., Garach, L., 2016. Bayes classifiers for imbalanced traffic accidents datasets. Accident Analysis & Prevention 88, 37-51.

Mulvaney, C.A., Smith, S., Watson, M.C., Parkin, J., Coupland, C., Miller, P., Kendrick, D., Mcclintock, H., 2015. Cycling infrastructure for reducing cycling injuries in cyclists. Cochrane Database of Systematic Reviews (12).

Nicaj, L., Stayton, C., Mandel-Ricci, J., Mccarthy, P., Grasso, K., Woloch, D., Kerker, B., 2009. Bicyclist fatalities in new york city: 1996–2005. Traffic Injury Prevention 10 (2), 157-161.

Pai, C.-W., Jou, R.-C., 2014. Cyclists' red-light running behaviours: An examination of risk-taking, opportunistic, and law-obeying behaviours. Accident Analysis & Prevention 62, 191-198.

Pujari, A.K., 2001. Data mining techniques Universities press, Hiderguda, India.

Räsänen, M., Summala, H., Pasanen, E., 1998. The safety effect of sight obstacles and road-markings at bicycle crossings. Traffic engineering & control 39 (2), 98-102.

Rivara, F.P., Thompson, D.C., Thompson, R.S., 2015. Epidemiology of bicycle injuries and risk factors for serious injury. Injury Prevention 21 (1), 47-51.

Rodgers, G.B., 1997. Factors associated with the crash risk of adult bicyclists. Journal of Safety Research 28 (4), 233-241.

Rosenkranz, K.M., Sheridan, R.L., Trauma to adult bicyclists: A growing problem in the urban environment. Injury 34 (11), 825-829.

Roumani, Y.F., May, J.H., Strum, D.P., Vargas, L.G., 2013. Classifying highly imbalanced icu data. Health Care Management Science 16 (2), 119-128.

Rowe, B.H., Rowe, A.M., Bota, G.W., 1995. Bicyclist and environmental factors associated with fatal bicycle-related trauma in ontario. Canadian Medical Association Journal 152 (1), 45-53.

Schepers, P., 2012. Does more cycling also reduce the risk of single-bicycle crashes? Injury Prevention 18 (4), 240-245.

Schepers, P., Den Brinker, B., 2011. What do cyclists need to see to avoid single-bicycle crashes? Ergonomics 54 (4), 315-27.

Stone, M., Broughton, J., 2003. Getting off your bike: Cycling accidents in great britain in 1990–1999. Accident Analysis & Prevention 35 (4), 549-556.

Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods 14 (4), 323-348.

Sutton, C.D., 2005. Classification and regression trees, bagging, and boosting. In: C.R. Rao, E.J.W., Solka, J.L. eds. Handbook of statistics. Elsevier, pp. 303-329.

Sze, N.N., Tsui, K.L., So, F.L., Wong, S.C., 2011. Bicycle-related crashes in hong kong: Is it possible to reduce mortality and severe injury in the metropolitan area? Hong Kong Journal of Emergency Medicine 18 (3), 136.

Tabachnick, B.G., Fidell, L.S., 2012. Using multivariate statistics Pearson, Boston, MA.

Tay, R., Rifaat, S.M., 2007. Factors contributing to the severity of intersection crashes. Journal of Advanced Transportation 41 (3), 245-265.

Vandenbulcke, G., Thomas, I., Int Panis, L., 2014. Predicting cycling accident risk in brussels: A spatial case–control approach. Accident Analysis & Prevention 62, 341-357.

Wachtel, A., Lewiston, D., 1996. Risk factors for bicycle-motor vehicle collisions at intersections. Journal of Safety Research 3 (27), 195.

Wessels, R., 1996. Bicycle collisions in washington state: A six-year perspective, 1988-1993. Transportation Research Record: Journal of the Transportation Research Board 1538, 81-90.

Wu, C., Yao, L., Zhang, K., 2012. The red-light running behavior of electric bike riders and cyclists at urban intersections in china: An observational study. Accident Analysis & Prevention 49, 186-192.

Xia, T., Zhang, Y., Crabb, S., Shah, P., 2013. Cobenefits of replacing car trips with alternative transportation: A review of evidence and methodological issues. Journal of Environmental and Public Health 2013, 14.

Yan, X., Ma, M., Huang, H., Abdel-Aty, M., Wu, C., 2011. Motor vehicle–bicycle crashes in beijing: Irregular maneuvers, crash patterns, and injury severity. Accident Analysis & Prevention 43 (5), 1751-1758.

Yan, X., Richards, S., Su, X., 2010. Using hierarchical tree-based regression model to predict train–vehicle crashes at passive highway-rail grade crossings. Accident Analysis & Prevention 42 (1), 64-74.

Zhang, G., Yau, K.K.W., Chen, G., 2013. Risk factors associated with traffic violations and accident severity in china. Accident Analysis & Prevention 59, 18-25.

Appendix A.

Table A1

Crash Type/Month Conditional Probabilities

Month	Severity of Bicycle Crashes	Run-off-the-road	Head-on collision	Sudden Braking	Angle collision	Falling from the vehicle	Sideswipe collision	Rear-end collision	Hit pedestrian	Hit parked or stationary vehicle	Hit stopped vehicle	Hit obstacle in carriageway
January	Fatality	0.06	0.09	0	0.51	0.05	0.13	0.13	0	0	0	0.02
February	Fatality	0.03	0.08	0	0.43	0.05	0.08	0.32	0	0.02	0	0
March	Fatality	0.08	0.11	0	0.41	0.09	0.09	0.19	0	0.03	0.01	0
April	Fatality	0.05	0.03	0	0.31	0.09	0.15	0.34	0	0	0.01	0
May	Fatality	0.03	0.15	0	0.4	0.06	0.08	0.19	0	0.02	0	0.06
June	Fatality	0.04	0.11	0	0.35	0.04	0.12	0.31	0	0.03	0	0.01
July	Fatality	0.05	0.15	0	0.35	0.07	0.15	0.24	0	0	0	0
August	Fatality	0.03	0.05	0	0.41	0.03	0.21	0.27	0	0.01	0	0
September	Fatality	0.05	0.15	0	0.42	0.04	0.14	0.19	0.01	0	0	0
October	Fatality	0.07	0.05	0	0.31	0.06	0.2	0.2	0.01	0.05	0	0.03
November	Fatality	0.01	0.11	0.02	0.31	0.06	0.21	0.26	0	0.02	0	0
December	Fatality	0	0.07	0	0.42	0.05	0.15	0.28	0	0.03	0	0
January	Injury	0.03	0.08	0	0.53	0.03	0.18	0.07	0	0.05	0.01	0.01
February	Injury	0.03	0.07	0	0.54	0.04	0.18	0.07	0	0.05	0.01	0.01

March	Injury	0.05	0.06	0	0.51	0.03	0.19	0.08	0	0.05	0.01	0.02
April	Injury	0.04	0.06	0	0.51	0.04	0.2	0.07	0.01	0.05	0.01	0.01
May	Injury	0.04	0.06	0	0.5	0.04	0.21	0.07	0.01	0.05	0.01	0.02
June	Injury	0.04	0.07	0	0.49	0.04	0.2	0.08	0.01	0.05	0	0.02
July	Injury	0.04	0.07	0	0.5	0.03	0.2	0.08	0	0.06	0.01	0.02
August	Injury	0.04	0.07	0	0.5	0.04	0.2	0.08	0.01	0.05	0	0.02
September	Injury	0.04	0.06	0	0.49	0.03	0.22	0.08	0	0.05	0.01	0.01
October	Injury	0.04	0.05	0	0.53	0.04	0.19	0.07	0.01	0.06	0.01	0.01
November	Injury	0.03	0.07	0	0.52	0.02	0.19	0.08	0	0.07	0.01	0.01
December	Injury	0.03	0.05	0	0.57	0.03	0.17	0.08	0	0.06	0	0.01

Table A2

Road Type/Road Segment Conditional Probabilities

Severity of Bicycle Crashes	Road segment	Urban municipal	Urban provincial	Urban national	Rural municipal	Rural provincial	Rural national	Other road	Rural regional	Urban regional
Fatality	Intersection	0.62	0.05	0.03	0.07	0.16	0.05	0	0.02	0
Fatality	Roundabout	0.54	0.08	0	0.02	0.16	0.12	0.02	0.05	0
Fatality	Signalized intersection	0.4	0.1	0.04	0.06	0.27	0.08	0	0.05	0.01
Fatality	Intersection with traffic lights or policeman	0.63	0.09	0.02	0	0.12	0.12	0	0	0.02
Fatality	Non signalized intersection	0.65	0.14	0	0.15	0.06	0	0	0	0
Fatality	Straight road	0.42	0.1	0.02	0.04	0.27	0.09	0.01	0.03	0.02
Fatality	Curve	0.32	0.19	0	0.1	0.21	0.1	0.02	0.07	0
Fatality	Bump or bottleneck	0.39	0.23	0	0.18	0.2	0	0	0	0
Fatality	Slope	0.54	0.12	0	0	0.09	0.1	0.14	0	0

Fatality	Tunnel with street light	0.33	0	0	0	0.33	0.35	0	0	0
Fatality	Tunnel	0.7	0	0	0	0	0	0	0	0.3
Injury	Intersection	0.84	0.07	0.02	0.01	0.03	0.01	0	0	0.01
Injury	Roundabout	0.75	0.11	0.03	0.01	0.08	0.01	0	0.01	0
Injury	Signalized intersection	0.79	0.07	0.02	0.02	0.06	0.02	0	0.01	0.01
Injury	Intersection with traffic lights or policeman	0.89	0.05	0.02	0.01	0.02	0.01	0	0	0.01
Injury	Non signalized intersection	0.81	0.1	0.02	0.02	0.03	0.01	0	0	0.01
Injury	Straight road	0.8	0.06	0.02	0.02	0.06	0.01	0	0.01	0.01
Injury	Curve	0.57	0.09	0.02	0.07	0.18	0.03	0.01	0.01	0.01
Injury	Bump or bottleneck	0.66	0.05	0.02	0.04	0.16	0.02	0.02	0	0.01
Injury	Slope	0.64	0.08	0.01	0.04	0.11	0.04	0.07	0	0.02
Injury	Tunnel with street light	0.68	0.03	0.03	0	0.09	0.12	0.03	0.03	0

Injury	Tunnel without street light	0.33	0	0	0	0.44	0.17	0	0	0.06

Table A3

Type of Opponent Vehicle/Crash Type Conditional Probabilities

Crash type	Crash severity	Car	Bus	Truck	PTW	Other vehicles	Multiple vehicles	No opponent
Head-on collision	Fatality	0.78	0	0.11	0.04	0	0.05	0.01
Angle collision	Fatality	0.67	0.01	0.17	0.06	0.04	0.05	0.01
Sideswipe collision	Fatality	0.42	0.08	0.36	0.03	0.06	0.04	0
Rear-end collision	Fatality	0.8	0.01	0.1	0.02	0.02	0.04	0.02
Hit pedestrian	Fatality	0	0	0	0	0	0	1
Hit parked or stationary vehicle	Fatality	0.59	0	0.12	0	0.08	0.2	0
Hit stopped vehicle	Fatality	0	0	0	0	0	0	1
Hit obstacle in carriageway	Fatality	0	0	0	0	0	0	1
Run-off-the-road	Fatality	0	0	0	0	0	0	1
Sudden Braking	Fatality	0	0	0	0	0	0	1
Falling from the vehicle	Fatality	0	0	0	0	0	0	1
Head-on collision	Injury	0.73	0.01	0.06	0.08	0.02	0.02	0.08
Angle collision	Injury	0.82	0.01	0.06	0.07	0.01	0.01	0.02
Sideswipe collision	Injury	0.75	0.02	0.09	0.07	0.03	0.02	0.02
Rear-end collision	Injury	0.73	0	0.07	0.08	0.04	0.05	0.03

Crash type	Crash severity	Car	Bus	Truck	PTW	Other vehicles	Multiple vehicles	No opponent
Hit pedestrian	Injury	0	0	0	0	0	0	1
Hit parked or stationary vehicle	Injury	0.83	0.01	0.09	0.01	0.02	0.04	0.01
Hit stopped vehicle	Injury	0	0	0	0	0	0	1
Hit obstacle in carriageway	Injury	0	0	0	0	0	0	1
Run-off-the-road	Injury	0	0	0	0	0	0	1
Sudden Braking	Injury	0	0	0	0	0	0	1
Falling from the vehicle	Injury	0	0	0	0	0	0	1