

# INVESTIGATION OF FACTORS THAT HAVE AFFECTED THE OUTCOMES OF ROAD TRAFFIC ACCIDENTS ON LITHUANIAN ROADS

---

TERESĖ LEONAVIČIENĖ<sup>1</sup>, SAUGIRDAS PUKALSKAS<sup>2</sup>,  
VIDMANTAS PUMPUTIS<sup>2</sup>, ERIKA KULEŠIENĖ<sup>3</sup>,  
VIDAS ŽURAUŠIS<sup>2\*</sup>

<sup>1</sup>*Dept of Mathematical Modelling, Vilnius Gediminas Technical University,  
Vilnius, Lithuania*

<sup>2</sup>*Dept of Automobile Engineering, Vilnius Gediminas Technical University,  
Vilnius, Lithuania*

<sup>3</sup>*BPN LT, Vilnius, Lithuania*

Received 3 January 2020; accepted 12 May 2020

**Abstract.** The purpose of this paper is to analyse the possibility for predicting the outcome of a road traffic accident concerning the traffic environment, personal traits of the traffic participant and the vehicle, i.e. aiming to answer the question whether specific values of the factors analysed to increase the likelihood of a fatal accident. The logistic regression model that allows identifying the relationship between the dependent and independent variables were used in the research. Other methods for describing and analysing categorical variables were also used alongside the logistic regression. When

---

\* Corresponding author. E-mail: [vidas.zuraulis@vgtu.lt](mailto:vidas.zuraulis@vgtu.lt)

Teresė LEONAVIČIENĖ (ORCID ID 0000-0003-3715-7303)

Saugirdas PUKALSKAS (ORCID ID 0000-0002-5051-1840)

Vidmantas PUMPUTIS (ORCID ID 0000-0002-2072-7755)

Erika KULEŠIENĖ (ORCID ID 0000-0002-5849-8142)

Vidas ŽURAUŠIS (ORCID ID 0000-0002-6210-7642)

Copyright © 2020 The Author(s). Published by RTU Press

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

analysing the results, it was recognised that the odds ratio above 1 shows a higher likelihood for a representative of the category in question to be involved in a fatal accident compared to a representative of the base category. Odds ratios of likelihoods for calculation of the road traffic accident types show that the likelihood of a fatal accident is statistically significant affected by rollovers or driving into obstacles, compared to vehicular collisions. When summarising the results, it was stated that most of the factors researched have an impact on the outcome of a road traffic accident. The influence of some factors has a higher probability of resulting in a fatal accident as compared to other factors.

**Keywords:** accident data, fatal accident, logistic regression, road safety, road traffic accident.

## Introduction

The prevention of road traffic accidents (hereinafter in the text – accident) around the world receives a lot of attention and funding. This area is primarily addressed because of unchanging accident statistics. As reported by the World Health Organisation (2013), approximately 1.24 million people die every year on the world roads, and another 20 to 50 million sustain injuries as a result of the accident. These injuries and deaths have an immeasurable impact on the affected families, whose lives are often changed irrevocably by these tragedies, and on the communities in which these people lived and worked (World Health Organization, 2013). It has been estimated that annual losses from accidents amount to about 3.8 billion euro in Lithuania (Pukalskas, Pečeliūnas, Sadauskas, Kilikevičienė, & Bogdevičius, 2015). Understandably, the massive losses incurred due to accidents on the global economy are difficult to quantify. In addition, to the prevention of accidents, the most relevant are the issues related to the analysis of accidents, i.e. investigation of individuals involved in traffic accidents and identification of factors most often affecting their consequences. Analysis of accumulated information has the potential of revealing solutions to problems, both technical and social. Such analysis is relevant throughout the world, and Lithuania is no exception in improving its accident statistics.

Issues related to accident statistics are often dealt with in scientific publications dedicated to transport and accident analysis. Finnish researchers, comparing the accident rates of the country and neighbouring Sweden, distinguished between urban and highway zones (Peltola & Luoma, 2017). Although accident rates in Swedish cities are less because of advanced urban planning, it is acknowledged that head-on collisions and single-vehicle accidents are the two most dangerous accident types. However, unlike in Finland, Sweden has developed an extensive network of centre-barrier roads. Following the model of Sweden, it is recommended that

data from the road register is used to analyse the safety of the road environment. The process of fatal accidents data collection has been identified as the sole limitation to obtain detailed accident analysis and to determine its causes. However, accident analysis does let be the identification of an apparent decline in the numbers of accident rates after the introduction of stringent road safety management measures in the Slovak Republic in 2009 (Brazinova & Majdan, 2016). Another identified measure related to the more accurate collection of accident data is also analysed in research conducted by Greek scientists Yannis, Papadimitriou, Chaziris, & Broughton (2014). Linking and matching methodology between police and hospitals accidents data of some European countries indicated the areas on which authorities focus their efforts (Yannis, Papadimitriou, Chaziris, & Broughton, 2014). The merging of hospital records and police records improves the collection of accident statistics and injury misclassification. While the Denning & Jennissen (2016) of their research emphasise that off-road crashes were not included in any of the national police reporting systems, all-terrain vehicle fatal accidents are the object of researchers from Iowa City, United States. The relevance of such research is related to the much higher popularity of all-terrain vehicles (special purpose vehicles) in the US than in Europe, which leads to more frequent accidents caused by these. It is noted that from 1998 to 2007 the rate of increase in a fatal accident on paved roads was nearly twice that of unpaved roads. As all-terrain vehicle design is unsuitable for paved roads (high centre of gravity, limited-slip rear differential, and all-terrain tyres) manufactures, safety agencies and stakeholders are under pressure to convey the message about the higher risk of this type of vehicle on paved roads.

Coming back to the Swedish road safety policy, interest groups were identified, with some of them advocating higher speeds, promoting mobility and the economy, while others were prioritising road safety improvement and speed reduction (Svensson, Summerton, & Hrelja, 2014). The compromise is partly achieved by maintaining speed limits on regional and local roads to the competence of municipal governments. Besides, road safety policy this way is brought closer to the regions by avoiding state-only regulation.

As the accident rate in urban areas is raising all kind of accidents and vehicles is considered in detail analysis regardless of its small statistical sample (Russo & Comi, 2017). The limited amount of data requires specific mobility and safety simulation models been implemented because sustainable urban mobility without injuries and fatal accidents is the crucial target for the future transport network. A specific case study of the city of Zagreb was done to find out the relations between external factors and accident blackspots involving pedestrians (Čosić, Šimunović, &

Jakovljević, 2019). Accident data collecting and its statistical analysis were carried out before the determination of influential factors of pedestrian-vehicle accidents. The external factors related to road infrastructure, speed limits and traffic volume showed the association with accident rate. At the same time, factors of year and day period also meteorological conditions did not show significant relation. However, the visibility of drivers in rural roads is significantly affected by day period and conditions as it was experimentally tested using an eye-tracking camera in real traffic conditions (Madleňák, Hošťáková, Madleňáková, Drozdziel, & Török, 2018). During night driving the number of fixations on traffic signs was 38.50% less comparing to daylight conditions. In addition, visual smog has an even higher negative effect (81.70%), and that is significant issues ensuring road safety. In the other hand, motorways witness road events causing even more serious results. Frequently road catastrophes take place, with the participation of a few vehicles, confirming the statistics that they are caused by the speeding, the fatigue of drivers, the insufficient distance between vehicles, or incorrect change of traffic lane (Drozdziel & Wrona, 2018).

Comparing to the European Union (EU) context, the number of accidents in Lithuania is still very high (by deaths – position 22 in the EU). After a significant decrease in 2008–2010, the numbers of the road accidents and injuries for next decade are quite similar or even deteriorating, however keeping the decrease in the fatal accident (Lithuanian Road Administration..., 2019). With the increase in the number of vehicles and their managers, nonregistered accidents have become part of our lives. Where they are the only type of accidents, there is no need to analyse the problems of accidents. However, there are also many reported accidents resulting in injuries to people. A significant number of accidents are particularly painful as they involve human death or serious injuries. The amount of research summarising long-term observations of the road safety situation in Lithuania is low. The available studies only analyse the road safety issues of both the vehicles, the road and traffic participants (Antov & Smirnovs, 2016; Bureika, Gaidamauskas, Kupinas, Bogdevičius, & Steišūnas, 2017; Bureika, Žuraulis, & Sadauskas, 2012; Gailienė & Laurinavičius, 2017; Žuraulis, Nagurnas, Pečeliūnas, Pumputis, & Skačkauskas, 2018). Two hundred forty-six non-professional Lithuanian drivers participated in the study about psychological factors to risky driving, and it is one of the few studies of this type from middle-income Eastern European countries (Šeibokaitė, Endriulaitienė, Sullman, Markšaitytė, & Žardeckaitė-Matulaitienė, 2017). A report of the cross-sectional survey showed that cognitive processing of emotions complicates the decision-making process while driving, leading to an increase in

errors and lapses. This study encourages the testing of the fitness of drivers to drive in terms of emotional readiness. As questionnaire surveys are often used for accident issues, the log-linear model showed the best fit for such variables as gender, fault of drivers and time in Ankara city (Olmuş & Erbaş, 2012). Additionally, the survey showed that these variables are significantly associated with accident severity. Unfortunately, the variable as the gender of drivers do not directly relate to accidents and much more comprehensive analysis has to be done.

By reducing the number of fatal accidents, it is essential to summarise the information about accidents, attempting to find specific prevailing trends, to reveal the causes and to anticipate potential risk factors. Once the main hazards have been identified, and methodologies have been developed, targeted work is being done with road users to reduce the impact of the human factor:

- to improve their training and examination programmes,
- to integrate road safety topics into the school general education course,
- to draw attention in social advertising.

Since road users are identified as the main perpetrators of accidents, who inevitably tend to make mistakes, therefore, the likelihood and consequences of human errors can be reduced by improving the road infrastructure. The Vision Zero programme successfully implemented in Sweden is based on such a plan (Belin, Tillgren, & Vedung, 2012).

This work analyses the possibility for predicting the outcome of the accident concerning the traffic environment, personal traits of the traffic participant and the vehicle, i.e. aiming to answer the question whether specific values of the analysed factors (variables) increase the likelihood of a fatal accident.

## **1. Research methodology**

Many studies have analysed various data mining technologies to get the most useful objectives from accident data (Gupta, Solanki, & Singh, 2017; Li, Shrestha, & Hu, 2017; Shokohyar, Taati, & Zolfaghari, 2017). Data clustering technique and association rule mining were used to get valuable information from heterogeneous accident data (Kumar & Toshniwal, 2015).

Determination of cause is the foremost possible and final task for accident data mining where computational intelligence algorithms are applied (Xi, Gao, Niu, Ding, & Ning, 2013). Here the importance of each value with the association rules method is analysed for selected accident data layers.

Exploring big data set of accidents as from China, data cleaning steps are necessary, so invalid data is deleted, and dirty data is removed (Chen, 2017). After that, the most accurate prediction models between Fisher Linear Discriminant, decision tree and random forest are analysed.

The publication presents the results of research obtained in collaboration with the Ministry of Transport and Communications of the Republic of Lithuania. Accidents involving deaths and injuries of people on the roads of the Republic of Lithuania from 2010 to 2015 are described and analysed in the research. The data set was compiled from records of 30 853 accidents (injuries and deaths). During this research, the focus was only on the perpetrators of the accident and the entire data set of 19 831 records was used. Only 18 175 records from this data set were recognised as suitable for the research due to insufficient data (no information provided about the perpetrators) in other records. Information on accidents is presented in various ways: general characteristics of the event, information on persons and vehicles involved in the accident. General event characteristics include accident number, date, time of the day, place (municipality, road number, type of road surface), number of vehicles and persons involved in the accident, number of injured and killed, meteorological conditions. Personal information is provided regarding the gender, age, driving experience, nature of injuries, sobriety, use of seat belt, airbag deployment. Vehicles are characterised by indicating the type, make, year of manufacture, country of registration and nature of the accident. Two additional variables are used in this research to distinguish accidents. One of these indicates the year in which the accident occurred and the other provides information on fatalities resulting from the accident. According to such classification, the total number of variables describing each accident is 29. However, some variables have been excluded because the number of values entered was below 5.00%, or they were assessed as not affecting the outcome of the accidents (for example, accident number or date). Thus, a total of 22 variables were used in the research.

Each of the variables analysed in the research draws from several to dozens of different values. For example, the values acquired for the variable Time of the accident are a day, dusk, dark, and the number of values of the variable indicating the Municipality of an accident is the highest – 62. Understandably, such a large number of different values is inappropriate for the research. Therefore, at the beginning of the research, the values obtained by many variables were merged into groups. Groups were formed without loss of information, and the probability of occurrence of the values of each group statistically significant differs from zero. For example, the variable indicating the municipality of the accident was reorganised by merging municipalities

into districts. After such grouping of acquired values into variables, the variable, which specifies the municipality of the accident has only 10 different values out of 62. The values acquired by the variables are discussed in more detail in the fourth section.

The logistic regression model, which allows identifying the relationship between the dependent and independent variables, was used in the research. Other methods for describing and analysing categorical variables were also used alongside the logistic regression in the research. The second section discusses in detail the logistic regression model and its application in evaluating the ratio of probabilities. The third section describes the set of research data, evaluating the influence of various factors on the possible outcome of the accident by using logistic regression tools. The fourth section outlines the result of the research and discusses the possibilities for further research.

## 2. Logistic regression model

Logistic regression is one of the most popular analysis methods for categorical data. The general principles in the logistic regression are the same as in the linear regression. In the logistic regression, the relationship between a response variable ( $Y$ ) and one or more explanatory variables ( $X$ ) is described, i.e. there is one dependent and one or more independent variables. The obtained method best fits the data. In the logistic regression, the dependent variable is dichotomous (the values of such variable are 0 – non-fatal accident and 1 – fatal accident). In the case when a dependent variable is influenced by more than one independent variable, multiple regression is assigned. In the general case, the independent variables are on different measurement scales. The general form of the logistic model (Eqs (1)–(2)) is the conditional probability (Hosmer, Lemeshow, & Sturdivant, 2013):

$$p(Y = 1 | X) = \frac{\exp(z(X))}{1 + \exp(z(X))}, \quad (1)$$

$$z(X) = a + b_1x_1 + b_2x_2 + \dots + b_kx_k, \quad (2)$$

where  $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$  – vector of  $k$  independent variables and  $a, b_1, b_2, \dots, b_k$  –

coefficients of the model. The coefficients  $b_1, b_2, \dots, b_k$  define the odds ratio (OR) for the dichotomous variables (Eq. (3)). An odds ratio is a

measure of association. For the dichotomous variable  $x_i$  coded 0 and 1, the relationship between OR and the regression coefficient is (Hosmer, Lemeshow, & Sturdivant, 2013):

$$\text{OR} = e^{b_i}. \quad (3)$$

Odds ratio approximates how much more likely (or unlikely) it is for the outcome to be present among those with  $x_i = 1$  than among those with  $x_i = 0$ . The influence of other variables (Eq. (4)) is not analysed, i. e. the model with

$$z(\mathbf{X}) = a + b_i x_i \quad (4)$$

is analysed. If  $b_i > 0$ , then  $\text{OR} > 1$  and if  $b_i < 0$ , then  $\text{OR} < 1$ .

In the logistic regression (Eq. (5)) for estimating the model parameters  $a, b_1, b_2, \dots, b_k$  the maximum likelihood method is used, and the maximum problem for the likelihood function (Eq. (6)) with probability (Eq. (7)) is solved:

$$z(\mathbf{X}) = a + b_1 x_1 + b_2 x_2 + \dots + b_k x_k, \quad (5)$$

$$L(a, b_1, b_2, \dots, b_k) = \prod_{Y=1} p \prod_{Y=0} (1-p), \quad (6)$$

$$p = \frac{\exp(z(\mathbf{X}))}{1 + \exp(z(\mathbf{X}))}. \quad (7)$$

### 3. Data analysis

This section contains an overview of the research data and the logistic regression model described in the second section used to determine the impact of each variable on the outcome of the accident.

As mentioned in Section 1, accident information is presented with an indication of the general characteristics, the persons and the vehicles involved in the accident. Of particular interest in this research is the prediction of the values of a variable describing the outcome of the accident, in light of the values of other variables. The variable that acquires two values determines the outcome of the accident has two values: 0 – non-fatal accident, 1 – fatal accident.

Many research variables are of category type, acquiring many different meanings. It is necessary to rate the significance of each category, taking into account the probabilities of acquiring values of that category, to reduce the number of categories. In this research, the categories of variables with probabilities 0.05 were considered



**Table 1. Variables and their categories**

<b>Item No</b>	<b>Variable</b>	<b>Category codes or variable values</b>
1.	Type of the road traffic accident	1 – collision 2 – collision with a two-wheeled vehicle 3 – driving into an obstacle 4 – rollover 5 – other road traffic accidents
2.	Number of the participants on the road traffic accident	1 – no more than two participants 2 – more than two participants
3.	Number of the injured persons on the road traffic accident	0 – none 1 – 1 2 – above 1
4.	Number of vehicles involved in the road traffic accident	1 – 1 2 – 2 3 – above 2
5.	District of the road traffic accident	1 – Alytus 2 – Kaunas 3 – Klaipėda 4 – Marijampolė 5 – Panevėžys 6 – Šiauliai 7 – Tauragė 8 – Telšiai 9 – Utena 10 – Vilnius
6.	Road surface indicator in the place of accident 1	1 – asphalt concrete or cement concrete, 2 – other surfaces
7.	Road surface indicator in the place of accident 2	1 – dry road surface 2 – wet 3 – icy 4 – snowy 5 – contaminated road surface
8.	Time of the road traffic accident	1 – day 2 – dusk (twilight) 3 – dark
9.	Weather conditions	1 – sunny 2 – cloudy or raining 3 – others
10.	Horizontal curve of the road	1 – straight road section 2 – other
11.	Gender of the participant (perpetrator) in the road traffic accident	1 – female 2 – male
12.	Seat belt	1 – fastened, 2 – unfastened or unfitted

Table 1.

Item No	Variable	Category codes or variable values
13.	Airbag deployment	1 – deployed 2 – undeployed 3 – not installed
14.	Condition of injury or depth in the road traffic accident	1 – killed or died from injuries 2 – injured 3 – not injured
15.	The driving experience of a vehicle driver	Experience, in years. This variable has been redesigned during the research by forming three groups: 1 – driving experience up to 7 years 2 – driving experience from 7 to 18 years 3 – 18 years and more
16.	The sobriety of a person driving a vehicle	1 – sober 2 – drunk or intoxicated with narcotic substances 3 – refused check
17.	Driver age	Age, in years. The variable was reorganised during the research by forming three groups: 1 – drivers below 29 years old 2 – drivers from 29 to 45 years old 3 – 45 years and older
18.	Country of vehicle registration	1 – Lithuania 2 – other countries
19.	Vehicle type	1 – a motorised two-wheeled vehicle 2 – car 3 – truck 4 – other
20.	Vehicle make	1 – AUDI 2 – BMW 3 – FORD 4 – MERCEDES BENZ 5 – OPEL 6 – RENAULT 7 – TOYOTA 8 – VOLKSWAGEN 9 – other
21.	The primary impact of the vehicle	0 – no impact 1 – front 2 – rear 3 – lateral
22.	Outcome of the road traffic accident	0 – no deaths 1 – at least 1 death

insignificant and were combined with other similar categories by meaning, or a new category was created. After such data rearrangement, the number of different categories of variables ranges between 2 and 10 (Table 1).

First of all, all variables were briefly described. Then the application of the logistic regression model for the assessment of the ratio of the likelihood was evaluated.

The analysed data, as already mentioned, describes various aspects related to the accident. In this dataset, 7.38% are fatal accidents. Among the accidents that occurred between 2010 and 2015, the most common type of accidents was vehicle collisions, which account for 48.83% of all accidents. The accident usually involved 1 injury (75.10% of accidents). Most accidents were recorded in Kaunas, Klaipėda and Vilnius districts, including the cities themselves. The dominant type of road surface was asphalt or cement concrete. Route surface was usually dry during accidents (62.50%). Accidents dominate traffic dataset during daylight (55.50%). Most often, males were perpetrators of road accidents. Conforming to the research data, males were perpetrators of 76.90% fatal accidents.

The research of relations between the variable describing the outcome of the accident and other variables listed in Table 1 lead to make a cross-tabulation table. By applying the  $\chi^2$  criterion, it was found that all variables listed in Table 1, except for variables:

- number of the participants on the road traffic accident,
- number of injured persons on the road traffic accident,
- road surface indicator in the place of accident 2,
- the sobriety of a vehicle driver and
- driver age affects the outcome of the accident.

Although the impact of the listed variables on the outcome of the accident based on  $\chi^2$  criterion is questionable, they are included in the initial logistic regression model. It was decided after analysing the relationships between the category variables of this research. The relationships between multiple variables are shown in Figures 1 and 2. The presented association plots (Zeileis, Meyer, & Hornik, 2007) shows relations between the category variables gender of the perpetrator on the road traffic accident and outcome of the road traffic accident (Figure 1), number of the participants on the road traffic accident and outcome of the road traffic accident (Figure 2). The association plots indicated where data deviate from the expected values and the effects of such deviation.

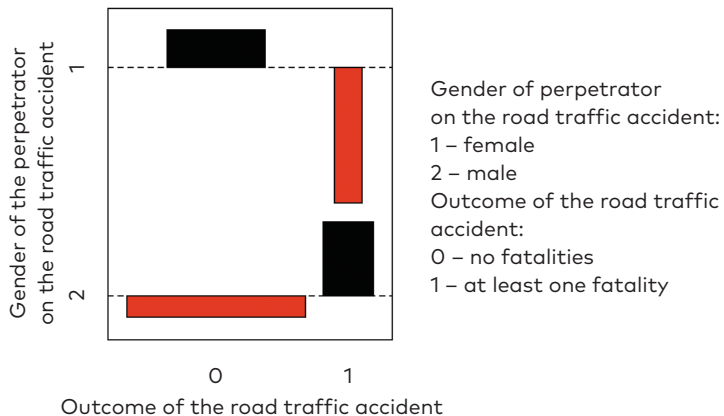
In the case, when the box is above one of the horizontal dotted lines, it is considered that the observed count is more than the expected count. In the case, when the box is below one of the horizontal dotted lines, it is considered that the observed count is below the expected count.

Figure 1 shows that the variable outcome of the road traffic accident depends on the variable gender of the participant in the road traffic accident. In the case of non-fatal accidents, the value observed in the female group was higher than predicted (the black box above the 1<sup>st</sup> dotted line). The value in the male group is higher than that predicted for a fatal accident (the black box above the 2<sup>nd</sup> dotted line). Red boxes below the 1<sup>st</sup> and 2<sup>nd</sup> lines correspond to the cases, where the observed values were less than the predicted values.

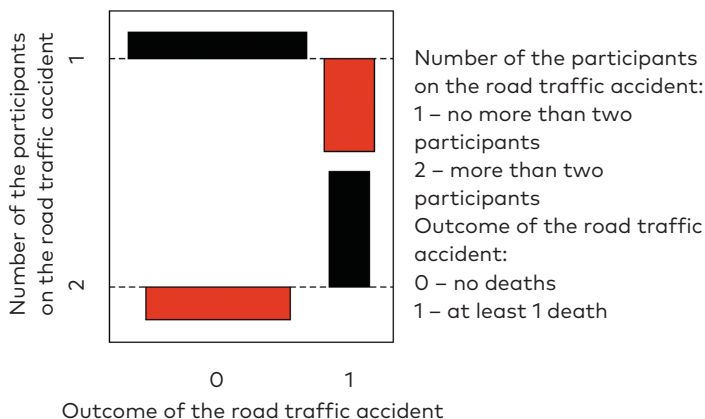
When the number of the participants on the road traffic accident is below 2, the number of observed values is higher than the values predicted for non-fatal accidents (the black box above the 1<sup>st</sup> dotted line in Figure 2). When the number of the participants on the road traffic accident is no less than 2, the number of observed values is higher than the values predicted for a fatal accident (the black box above the 2<sup>nd</sup> dotted line). Red boxes below the 1<sup>st</sup> and 2<sup>nd</sup> lines correspond to the cases, where the observed values were less than the predicted values. Similarly, the relationships between other category variables and the variable outcome of the road traffic accident were investigated. However, Figures 1 and 2 do not answer the question of how strongly the values of each category variable affect the result of the variable outcome of the road traffic accident.

The logistic regression model described in Section 2 was used to reveal the relationships among the variables. The relationships among the variables were evaluated by analysing the relationship of probabilities.

Because the variables in the logistic regression model (both dependent and independent) are categorical, the independent variables were first transformed by introducing dummy (design) variables.



**Figure 1.** Association plot for categorical variables based on the gender



**Figure 2.** Association plot for categorical variables based on the number of the participants

Tables 2 and 3 present the schemes for rearrangement of two variables conversion and pseudo-entry schemes (number of the participants on the road traffic accident and road surface indicator in the place of accident 2) and introduction of dummy variables. Similar actions were performed with all research variables.

**Table 2.** Rearrangement of variables

Number of the participants on the road traffic accident	Dummy variable	
	Number of the participants on the road traffic accident	
	1	2
1 – no more than two participants	0	0
2 – more than two participants	0	1

**Table 3.** Introduction of dummy variables

Road surface indicator in the place of accident 2	Dummy variable				
	Road surface indicator in the place of accident				
	1	2	3	4	5
Dry road surface	0	0	0	0	0
Wet	0	1	0	0	0
Icy	0	0	1	0	0
Snowy	0	0	0	1	0
Contaminated road surface	0	0	0	0	1

The logistic regression model described in Section 2 by Eqs (1)–(7) is modified using dummy variables. Assume that the  $j^{\text{th}}$  categorical variable has  $s$  independent levels. Then the number of the dummy variable is  $(s-1)$ , and the logistic regression model is (Hosmer, Lemeshow, & Sturdivant, 2013):

$$p = \frac{\exp(z(\mathbf{X}))}{1 + \exp(z(\mathbf{X}))}, \quad (8)$$

$$z(\mathbf{X}) = a + b_1x_1 + \sum_{l=1}^{s-1} b_{jl}D_{jl} + \dots + b_kx_k. \quad (9)$$

After the introduction of dummy variables, a logistic regression model (Eqs (8)–(9)) with all the research variables was made. The hypothesis (Eqs (10)–(11)) whether the values of all the coefficients included in the model are equal to zero was tested, i.e.:

$$H_0 : b_1 = b_2 = \dots = b_k = 0, \quad (10)$$

$$H' : \exists i, b_i \neq 0. \quad (11)$$

Likelihood ratio test statistic is 6909.4 and  $p$ -value below 0.05, and that means that not all coefficients in the model are equal to zero. The model correctly predicts outcomes for 99.80% of non-fatal accidents and 67.90% of fatal accidents. The total percentage of correct forecasts is 97.00%. Therefore, such a model reflects the available data well (above 50.00% predicts fatal and non-fatal accidents), but is insufficient. However, it is more relevant to the other issue, i.e. the impact of each variable on the outcome of the accident. The effects of the identified factors on the outcome of the road traffic accident were studied using the ORs (Eq. (3)). Few dummy variables described each factor in the logistic regression model, and now one of these variables must be as a reference (Table 4).

Table 4 does not include variables with the impact on the outcome variable is statistically insignificant ( $\chi^2$  test and likelihood ratio test  $p$ -values above 0.05):

- number of the participants on the road traffic accident ( $p$ -value is equal to 0.14),
- number of injured persons on the road traffic accident ( $p$ -value is equal to 0.23),
- road surface indicator in the place of accident 2 ( $p$ -value is equal to 0.08),
- driver age ( $p$ -value is equal to 0.96).

Table 4. The effects of the identified factors on the outcome of the road traffic accident

Variable	Category	Odds ratio	confidence interval		p-value
			lower	upper	
Type of the road traffic accident (relative to collision)	collision with a two-wheeled vehicle	1.42	1.21	1.67	0.000
	driving on an obstacle	1.88	1.69	2.09	0.000
	rollover	1.93	1.63	2.28	0.000
	other road traffic accidents	1.04	0.86	1.25	0.700
Number of vehicles involved in the road traffic accident (relative to 2 vehicles)	1	1.74	1.57	1.93	0.000
	above 2	1.87	1.64	2.13	0.000
District of the road traffic accident (relative to Vilnius)	Kaunas	0.86	0.73	1.00	0.050
	Klaipėda	0.64	0.53	0.77	0.000
	Marijampolė	1.83	1.52	2.21	0.000
	Panevėžys	1.26	1.06	1.50	0.010
	Šiauliai	0.77	0.63	0.94	0.010
	Tauragė	1.46	1.12	1.89	0.005
	Telšiai	1.44	1.15	1.80	0.001
	Utena	2.27	1.82	2.84	0.000
Road surface indicator in the place of accident 1 (relative to asphalt/cement concrete)	Alytus	1.12	0.88	1.42	0.360
	another surface	1.46	1.24	1.72	0.000
Day/night time of the road traffic accident (relative to day)	dusk (twilight)	1.67	1.14	1.97	0.000
	dark	2.27	2.06	2.51	0.000
Meteorological conditions (relative to sunny)	cloudy or raining	1.17	1.06	1.29	0.001
	others	1.86	1.56	2.23	0.000
Horizontal curve of the road (relative to the straight section of the road)	others	1.55	1.38	1.74	0.000
Gender of the participant in the road traffic accident (relative to a female)	male	2.35	2.04	2.70	0.000
Seat belt (relative to fastened)	unfastened or not equipped	2.21	1.91	2.55	0.000
Airbag deployment (relative to deployed)	did not deploy	1.55	1.37	1.76	0.000
	not installed	1.71	1.52	1.93	0.000

Table 4.

Variable	Category	Odds ratio	confidence interval		p-value
			lower	upper	
Condition of injuries and deaths on the road traffic accident (relative to killed or died from injuries)	injured	0.06	0.05	0.07	0.000
	not injured	0.11	0.10	0.13	0.000
The driving experience of a vehicle driver (relative to below 7 years)	from 7 to 18 years	1.12	0.66	1.26	0.070
	no less than 18 years old	1.18	1.05	1.32	0.004
The sobriety of a vehicle driver (relative to sober)	drunk or intoxicated with narcotic substances	1.34	1.14	1.57	0.000
	refused check	0.51	0.07	3.79	0.510
Country of vehicle registration (relative to Lithuania)	other countries	3.15	2.56	3.87	0.000
Vehicle type (relative to car)	motor two-wheeled vehicle	1.06	0.86	1.30	0.570
	truck	3.33	2.88	3.86	0.000
	other type	1.28	1.09	1.50	0.000
Vehicle name (relative to Volkswagen)	Audi	1.11	0.94	1.32	0.230
	BMW	1.02	0.81	1.29	0.850
	Ford	0.65	0.49	0.86	0.000
	Mercedes Benz	1.47	1.18	1.84	0.000
	Opel	0.80	0.64	1.01	0.060
	Renault	0.99	0.79	1.23	0.900
	Toyota	0.64	0.47	0.86	0.000
	another name	1.05	0.92	1.20	0.480
Primary vehicle impact (relative to front)	rear	0.50	0.42	0.60	0.000
	lateral	0.78	0.69	0.89	0.000
	no impact	0.34	0.08	1.40	0.140

Analysis of the results of Table 4 shows that not all calculated OR values are statistically significant (not all  $p$ -values are below 0.05). However, as these are separate categories of statistically significant variables, there is no reason to exclude them from the analysis.

## 4. Discussion and results

This section discusses the results of the data analysis performed in this research, and based on the available data, identifies the variables that have the most significant impact on the outcome of the accident.



When analysing the results of Table 4, that the OR is above 1 shows the higher OR for a representative of the category in question to be involved in a fatal accident compared to a representative of the base category. Ratios of likelihoods for calculation of the accident type show that the OR of a fatal accident is statistically significant affected by rollovers or driving into an obstacle, compared to vehicle collisions. Regarding the impact of the number of vehicles involved in the road traffic accident on the outcome of the road traffic accident, the OR of a fatal accident involving more than 2 vehicles (OR = 1.87) is higher than in the case of 2 vehicles. It is also noted that the OR of a fatal accident is higher even if there is only one vehicle in the accident (OR is equal to 1.74). The odds ratio value above 1 with the  $p$ -value below 0.05 is obtained analysing the districts (according to Vilnius district) and observed that the risk of the fatal accident is higher in the districts of Marijampolė, Panevėžys, Tauragė, Telšiai, and Utena. The outcome of the road traffic accident is also strongly influenced by the time of the day. Nighttime or twilight traffic conditions are more complicated compared to the daytime and is illustrated by the OR values, which are 2.27 and 1.67, respectively. Meteorological conditions of fog, snow, strong wind and other, has more impact on fatal accident compared to sunny weather (OR is equal to 1.86). On slopes or horizontal curves of the road, the OR increases to 1.55 when compared to a straight section of the road. The ratio of values with the maximum likelihood also include the OR calculated for males compared to females (the likelihood of a fatal accident is above 2 times higher for male drivers). Non-fastened or not-equipped seat belts compared to fastened seat belts increase the likelihood of fatal accident (OR is equal to 2.21). Airbags unfitted or not deployed in vehicles also have a statistically significant effect on the outcome of the road traffic accident compared to airbag deployment. Compared to sober drivers, drivers intoxicated with alcohol or narcotic substances have a higher probability of being involved in a fatal accident (OR is equal to 1.34). It is noticeable higher probability value ratios for vehicles registered in other countries (the OR above 3.0 when compared to vehicles registered in Lithuania). A probability ratio higher than 3.0 was also pertained for trucks compared to passenger cars.

## Conclusions

When summarising the results obtained during the particular statistical analysis, it is stated that most of the research factors have an impact on the outcome of the road traffic accident. The influence of some factors is higher, i.e. they increase the likelihood of involvement in a fatal accident more, as compared to others.

1. Regarding the impact of the number of vehicles involved in the road traffic accident on the outcome of the road traffic accident, the probability of a fatal accident involving more than 2 vehicles (OR = 1.87) is higher than in the case of 2 vehicles. It is also noted that the OR of a fatal accident is higher even if there is only one vehicle in the accident (OR is equal to 1.74).
2. The outcome of the road traffic accident is also strongly influenced by the time of the day. Nighttime or twilight traffic conditions are more complicated compared to the daytime and is illustrated by the odds ratio values, which are 2.27 and 1.67, respectively.
3. Meteorological conditions of fog, snow, strong wind and other, has more impact on fatal accident compared to sunny weather (odds ratio is equal to 1.86).
4. The ratio of values with the maximum probability also include the odds ratio calculated for males compared to females (the likelihood of a fatal accident is above 2 times higher for male drivers).
5. Unfastened or not-equipped seat belts compared to fastened seat belts increase the likelihood of fatal accident (odds ratio is equal to 2.21).
6. Compared to sober drivers, drivers intoxicated with alcohol or narcotic substances have a high probability of being involved in a fatal accident (odds ratio is equal to 1.34).
7. The logistic regression model presented in this research correctly predicts about 67.90% of a fatal accident, and it can be used as an auxiliary tool to identify the factors affecting the outcome of the road traffic accident concerning more on the impact of each factor separately.

## REFERENCES

- Antov, D., & Smirnovs, J. (2016). 78 Improving road safety in the Baltic states – role of strategies. *Injury Prevention*, 22(Suppl 2), A30.1-A30. <https://doi.org/10.1136/injuryprev-2016-042156.78>
- Belin, M. Å., Tillgren, P., & Vedung, E. (2012). Vision Zero – a road safety policy innovation. *International Journal of Injury Control & Safety Promotion*, 19(2), 171-179. <https://doi.org/10.1080/17457300.2011.635213>
- Brazinova, A., & Majdan, M. (2016). Road traffic mortality in the Slovak Republic in 1996–2014. *Traffic Injury Prevention*, 17(7), 692-698. <https://doi.org/10.1080/15389588.2016.1143095>

- Bureika, G., Gaidamauskas, E., Kupinas, J., Bogdevičius, M., & Steišūnas, S. (2017). Modelling the assessment of traffic risk at level crossings of Lithuanian railways. *Transport*, 32(3), 282-290.  
<https://doi.org/10.3846/16484142.2016.1244114>
- Bureika, G., Žuraulis, V., & Sadauskas, V. (2012, October). Research on automobile technical state impact on road traffic accident level in the country. In *Transport Means-2012: Proceedings of the 16th International Conference* (pp. 69-72).
- Chen, C. (2017). Analysis and forecast of traffic accident big data. In *ITM Web of Conferences* (Vol. 12, p. 04029). EDP Sciences.  
<https://doi.org/10.1051/itmconf/20171204029>
- Ćosić, M., Šimunović, L., & Jakovljević, M. (2019). Relationships between external factors and pedestrian accident blackspots – a case study of the city of Zagreb. *Promet-Traffic & Transportation*, 31(3), 329-340.  
<https://doi.org/10.7307/ptt.v31i3.3119>
- Denning, G. M., & Jennissen, C. A. (2016). All-terrain vehicle fatalities on paved roads, unpaved roads, and off-road: evidence for informed roadway safety warnings and legislation. *Traffic Injury Prevention*, 17(4), 406-412.  
<https://doi.org/10.1080/15389588.2015.1057280>
- Drozdziel, P., & Wrona, R. (2018, April). Legal and utility problems of accidents on express roads and motorways. In *2018 XI International Science-Technical Conference Automotive Safety* (pp. 1-5). IEEE.  
<https://doi.org/10.1109/AUTOSAFE.2018.8373315>
- Gailienė, I., & Laurinavičius, A. (2017). The need and benefit of slab track: case of Lithuania. *Gradevinar*, 69(5), 387-396.  
<https://doi.org/10.14256/JCE.1776.2016>
- Gupta, M., Solanki, V. K., & Singh, V. K. (2017). Analysis of datamining technique for traffic accident severity problem: a review. In *Proceedings of the Second International Conference on Research in Intelligent & Computing in Engineering* (pp. 197-199). <https://doi.org/10.15439/2017R121>
- Hosmer, Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- Kumar, S., & Toshniwal, D. (2015). A data mining framework to analyse road accident data. *Journal of Big Data*, 2(1), 26.  
<https://doi.org/10.1186/s40537-015-0035-y>
- Li, L., Shrestha, S., & Hu, G. (2017, June). Analysis of road traffic fatal accidents using data mining techniques. In *2017 IEEE 15th International Conference on Software Engineering Research, Management & Applications (SERA)* (pp. 363-370). <https://doi.org/10.1109/SERA.2017.7965753>
- Lithuanian Road Administration under the Ministry of Transport and Communications of the Republic of Lithuania (2019). *Statistics of fatal and injury road accidents in Lithuania, 2015–2018*. (in Lithuanian)
- Madleňák, R., Hošťáková, D., Madleňáková, L., Drozdziel, P., & Török, A. (2018). The analysis of the traffic signs visibility during night driving. *Advances in Science & Technology Research Journal*, 12(2).  
<https://doi.org/10.12913/22998624/92103>

- Olmuş, H., & Erbaş, S. (2012). Analysis of traffic accidents caused by drivers by using Log-linear models. *Promet – Traffic & Transportation*, 24(6), 495-504. <https://doi.org/10.7307/ptt.v24i6.1201>
- Peltola, H., & Luoma, J. (2017). Comparison of road safety in Finland and Sweden. *European Transport Research Review*, 9(1), 3. <https://doi.org/10.1007/s12544-016-0220-x>
- Pukalskas, S., Pečeliūnas, R., Sadauskas, V., Kilikevičienė, K., & Bogdevičius, M. (2015). The methodology for calculation of road accident costs. *Transport*, 30(1), 33-42. <https://doi.org/10.3846/16484142.2015.1020871>
- Russo, F., & Comi, A. (2017). From the analysis of European accident data to safety assessment for planning: the role of good vehicles in urban area. *European Transport Research Review*, 9(1), 9. <https://doi.org/10.1007/s12544-017-0225-0>
- Shokohyar, S., Taati, E., & Zolfaghari, S. (2017). The Effect of Drivers' Demographic Characteristics on Road Accidents in Different Seasons Using Data Mining. *Promet - Traffic & Transportation*, 29(6), 555-567. <https://doi.org/10.7307/ptt.v29i6.2342>
- Svensson, T., Summerton, J., & Hrelja, R. (2014). The politics of speed—local and regional actors' views on speed limits, traffic safety and mobility in Sweden. *European Transport Research Review*, 6(1), 43-50. <https://doi.org/10.1007/s12544-013-0109-x>
- Šeibokaitė, L., Endriulaitienė, A., Sullman, M. J., Markšaitytė, R., & Žardeckaitė-Matulaitienė, K. (2017). Difficulties in emotion regulation and risky driving among Lithuanian drivers. *Traffic Injury Prevention*, 18(7), 688-693. <https://doi.org/10.1080/15389588.2017.1315109>
- World Health Organization (2013). *Global status report on road safety 2013: supporting a decade of action: summary* (No. WHO. NMH. VIP 13.01). World Health Organization.
- Xi, J., Gao, Z., Niu, S., Ding, T., & Ning, G. (2013). A hybrid algorithm of traffic accident data mining on cause analysis. *Mathematical Problems in Engineering*, 2013. <https://doi.org/10.1155/2013/302627>
- Yannis, G., Papadimitriou, E., Chaziris, A., & Broughton, J. (2014). Modeling road accident injury under-reporting in Europe. *European Transport Research Review*, 6(4), 425-438. <https://doi.org/10.1007/s12544-014-0142-4>
- Zeileis, A., Meyer, D., & Hornik, K. (2007). Residual-based shadings for visualising (conditional) independence. *Journal of Computational & Graphical Statistics*, 16(3), 507-525. <https://doi.org/10.1198/106186007X237856>
- Žuraulis, V., Nagurnas, S., Pečeliūnas, R., Pumputis, V., & Skačkauskas, P. (2018). The analysis of drivers' reaction time using cell phone in the case of vehicle stabilisation task. *International Journal of Occupational Medicine & Environmental Health*, 31(5), 633. <https://doi.org/10.13075/ijomeh.1896.01264>