MOBILITY & VEHICLE MECHANICS

https://doi.org/10.24874/mvm.2023.49.01.04
UDC: 656.84;519.22

INVESTIGATION OF THE SINGLE VEHICLE ACCIDENTS SEVERITY BY USING A PROBABILISTIC APPROACH

Drosu Alin [1], Corneliu Cofaru [2], Mihaela Virginia Popescu [3]*

Received in August 2022 *Accepted in October 2022*

RESEARCH ARTICLE

ABSTRACT: Single vehicle accidents (SVAs) arouse the interest not only of researchers, but also of the European Commission and other international road safety bodies, since a third part from all road fatalities from Europe are caused by the single vehicle accidents. The aim of this paper is to assess the severity of such accidents by estimating the probability of the fatalities (P_{d1}) and major injuries (P_{d2}) generated by the single vehicle accidents and to identify the factors affecting those probabilities. As for this research, a complex 6 -year-accidents data base has been used and the accidents records have been aggregated on a daily basis. A binary multiple logistic regression has been developed for each type of severity (fatality and major injury) using 86 predictors related to the place of accidents, road category and feature and characteristic, the number and the width of the lanes, horizontal road markings, safety components of the road, road surface characteristics and adherence, weather and lighting conditions, vehicles mileage and drivers' sex. The logistic models have been tested on their statistical significance and their explanatory efficiency was discussed. A descriptive analysis has been conducted for both models in order to discuss the distribution of the probability values. P_{d2} model has a better explanatory power than P_{d1} and its overall percentage of the predictions is 96.10 %. It is also has a very good homogeneity since all its predictors have positive values. An interesting finding is that no other predictors related to weather or lighting conditions do significantly explain the probabilities of a SVA to be of fatality or to generate major injuries. These constraints are to be further researched since the daily level of data aggregation studied in this paper influence the "immediate effect" of random phenomena, as weather or lighting conditions. P_{d2} model has a good applicability in the identifying, prioritizing and treating of the black spots. The Romanian road authority could also use it in driver education and injury risk identification in order to mitigate the severity of the accidents.

KEY WORDS: *Safety, single vehicle accidents, logistic regression, probabilistic*

[1]*Drosu ,Alin, Transylvania University of Braşov, Institution Adress, Country, name.familz name @dddd.com, ORCID ID: (*Corresponding author)*
[2] *Corneliu Cofaru, Departamentul Autovehicule şi transporturi Facultatea de Inginerie mecanică, Str. Politehnicii nr.1, Brasov, Romania, ccornel@unitbv.ro, ORCID ID:-*
[3]*Mihaela Virginia Popescu, Departamentul de Lingvistică teoretică şi aplicată, Facultatea de Litere, B-dul Eroilor nr. 25, Braşov, Romania, Corp T, sala TP7, mvp@unitbv.ro, ORCID ID:-*

ISTRAŽIVANJE JAČINE SUDARA POJEDINAČNOG VOZILA KORIŠĆENJEM PRINCIPA VEROVATNIĆEA

REZIME: Pojedinačne saobraćajne nezgode (SVA) izazivaju interesovanje ne samo istraživača, već i Evropske komisije i drugih međunarodnih tela za bezbednost saobraćaja na putevima, jer trećina svih poginulih na putevima u Evropi su su stradali u pojedinačnim saobraćajnim nezgodama. Cilj ovog rada je da proceni težinu ovakvih nezgoda procenom verovatnoće smrtnih slučajeva (Pd1) i teških povreda (Pd2) izazvanih pojedinačnim saobraćajnim nezgodama i da identifikuje faktore koji utiču na te verovatnoće. Što se tiče ovog istraživanja, korišćena je kompleksna šestogodišnja baza podataka o nezgodama i evidencija o nezgodama na dnevnoj bazi. Razvijena je binarna višestruka logistička regresija za svaku vrstu težine (smrt i teške povrede) koristeći 86 prediktora koji se odnose na mesto nezgode, kategoriju i profil i karakteristike puta, broj i širinu traka, horizontalne oznake na putu, bezbednosne komponente puta, karakteristike površine puta i prijanjanje, vremenski i svetlosni uslovi, kilometraža vozila i pol vozača. Logističkim modelima je ispitana statistička značajnost i diskutovana je njihova efikasnosti. Urađena je deskriptivna analiza za oba modela kako bi se diskutovalo o distribuciji vrednosti verovatnoće. Model P_{d2} ima bolju eksplanatornu sposobnost od P_{d1} i njegov ukupan procenat predviđanja je 96,10 %. Takođe, ima veoma dobru homogenost pošto svi njegovi prediktori imaju pozitivne vrednosti. Interesantan nalaz je da nijedan drugi prediktor koji se odnosi na vremenske uslove ili uslove osvetljenja ne objašnjavaju značajno verovatnoću da SVA bude sa smrtnim ishodom. Ova ograničenja treba dalje istraživati pošto dnevni nivo agregacije podataka koji se proučava u ovom radu utiče na „neposredan efekat“ nasumičnih pojava, kao što su: vremenski uslovi ili uslovi osvetljenja. P_{d2} model ima dobru primenljivost u identifikaciji, određivanju prioriteta i lečenju crnih tačaka. Rumunska uprava za puteve takođe bi mogla da ga koristi u obrazovanju vozača i identifikaciji rizika od povreda kako bi se ublažila težina nesreća.

KLJUČNE REČI: *Bezbednost, nezgoda pojedinačnog vozila, logistička regresija, verovatnoća*

INVESTIGATION OF THE SINGLE VEHICLE ACCIDENTS SEVERITY BY USING A PROBABILISTIC APPROACH

Drosu Alin, Corneliu Cofaru, Mihaela Virginia Popescu

INTRODUCTION

There are many definitions or classifications of the single vehicle accidents. While the American National Safety Council defines the single vehicle accidents (SVA) as collisions that involve pedestrians, fixed objects and wild and domestic animals, in Europe this type of accidents is represented by those collisions where the pedestrians or cyclists are not involved and the driver, rider and/or passengers are killed or seriously injured.

The single vehicle accidents are of interest not only for researchers, but also for the European Commission and other international road safety bodies, since a third part from all road fatalities from Europe are caused by the SVA [1]. At the European level, there is a lack of information in depth collision data focused on SVA and therefore more studies are required in order to find the factors that affect this type of collisions. In Romania, around 20% - 25% of the total number of fatalities is generated by SVA [2] and the Romanian authorities consider SVA as a major problem since the injury index is higher compared with other type of collisions (1.2 – 1.3 for SVA and 1.1 – 1.2 for non SVA). According to the Romanian authorities, the drivers' mood and their actions has a great impact in SVAs causation and injury severity. This fact does not confirm the findings of other international studies which show that the road environment highly impacts are the main factors causing SVA. Therefore, this paper will be focused more on researching the road and environmental factors that have a greater impact on the single vehicle accidents occurring on the Romanian territory.

The SVA are considered to have a higher severity than the multi vehicle accidents [3]. At the same time, the highest fatality percent of the road accidents is generated by the SVA [4]. While in the multi-vehicle accidents the severity of the crashes depends on how the drivers react when they are involved in such an event, in the case of SVA this is more influenced by the road environment and safety devices of the roadway. The hourly traffic volumes do not have a significant influence on the SVA's causational mechanism [5] and are negatively correlated with this type of collisions [6]. Even if the human error is one of the main factors that greatly influence the occurring of SVA, the environmental conditions significantly influence both accidents' probability and victims' severity [7]. The lighting conditions belong to those environmental characteristics which have a great impact on the probability of being involved in a SVA and on the severity of the victims and effects generated in this type of collisions [8]. Moreover, both probability and victims' severity are influenced by the weather conditions from the accident time when the drivers make errors or adopt dangerous decisions in their driving style, due to a poor visibility [9].

The presence of the hard shoulders and the lanes' width strongly influence the severity of a SVA [10]. Therefore, knowing the causational factors of the single vehicle accidents not only mitigate the accident and injury risk, but also will help the road safety authorities to implement appropriate counter measures.

1. DATA PREPARATION

In this paper, a complex accident database was used and all the road crashes occurred between 01.01.2013 – 31.12.2018 were included. According to the SVA's European definition, the collisions with one vehicle have been filtered, regardless the category (personal vehicle, light or heavy commercial vehicle, moped or motorcycle). Those accident records which involved pedestrians or cyclists were removed and all the data were aggregated on a daily basis. Finally, the dataset used for this research has 2.190 rows (corresponding to those 6 years included in this analyze).

2. METHODOLOGY

The severity of a road accident is generally shown by the severity of the victims. Therefore, using the Binary Multiple Logistic Regression (BMLR), two types of probabilities are assessed for predictive purposes by using SPSS: the probability of a SVA to be fatal or not (Pd1) and the probability for a SVA's victims to have or not to have major injuries (Pd2). The road accidents' severity assessed by using BMLR techniques is often found in the literature, as it is shown in other researches [11], [12].

The response variable of a BMLR is a probability P that always takes values between 0 and 1 and its mathematic model can be written as follows:

$$P = \frac{e^y}{1+e^y}, \tag{1}$$

where,

- y represents the odds of an accident to be labelled as a fatal/not fatal or with major injuries/no major injuries and take the values between zero and infinite.

It is also defined as the probability "p" for a SVA to be fatal or to have major injuries divided by the probability "q = 1-p" of a SVA to be not fatal or to have no major injuries. Using the natural logarithms, we can write the following relation:

$$\ln(y) = \ln\left(\frac{p}{q}\right) = \ln\left(\frac{p}{1-p}\right), \tag{2}$$

Mathematically, the relation of y (odds) can be written as:

$$y = \exp\left(b_0 + b_1 * \mathrm{X}_1 + b_2 * \mathrm{X}_2 + ... + b_k * \mathrm{X}_k + ... + b_n * X_n\right), \tag{3}$$

where,

- b_0, b_1, b_2, … b_k *are the logit coefficients related to the regression's coefficients and* b_0 *is* the intercept.
- X_1, X_2, …,X_k, …, X_n are the predictors which will be included in the analyses and k = 1, 2,…,n represents the number of these predictors.

In this paper, the logit coefficients b_k of the predictors will be assessed using the odds. In order to do this assessment, an OR odds ratio will be calculated and will show to which extend the odds will be changed for a SVA to be fatal/non-fatal or with major injuries/no major injuries, when a predictor changes its value by own unit, when the other predictors remain constant. The equation for OR odds ratio is the following:

$$QR(\text{odds ratio}) = e^{b_k}, \tag{4}$$

Two kinds of statistical tests will be used in order to check the significance of the logistic models, as SPSS outputs: G^2 chi-square likelihood ratio test and Hosmer and Lemeshow's

Goodness of Fit Test. G^2 test will be used in order to check whether all predictors' logit coefficients, except the intercept, have null values. Its formula is the following:

$$G^2 = (-2\text{LLnull}) - (\text{-2LLmodel}). \qquad (5)$$

2LLnull shows the errors associated with the model that has the intercept b_o as the only parameter and 2LLmodel shows the errors associated with the model that includes both the intercept b_o and predictors.

In case there are no major differences between the observed and estimated data, then the model is considered a good one. This is shown by the Hosmer and Lemeshow's Goodness of Fit Test's value which should not be statistically of significance ($p>0.05$).

The explanatory power of the estimated logistic models is given by two indicators which are similarly to the R^2 coefficient that is found in the linear regression: Cox & Snell R Square and Nagelkerke R Square. The classification table generated by SPSS will be used in order to check the correctness of the predictions. This shows how high the determination coefficient R^2 for establishing the proportions is. In the other words, R^2 will show the degree of compatibility between the observed and estimated values.

The logit coefficients will be tested in order to see if they significantly influence the values of the dependent variables P_{d1} or P_{d2}. The interpretation and the testing of these coefficients will be done by using the Wald test which tests the null hypothesis when a coefficient has the value of zero. The interpretation of the coefficients will be conducted by using these factors. Their positive values show an increase of the odds for a SVA to be fatal or to generate major injuries, in this case being a direct relationship between the predictor and the correspondent odds. Contrary, the negative values show a decrease of those odds.

In order to interpret the logit coefficients by using the percent, the following formula will be used:

$$Exp(B)\% = 100 * \left(e^{b_k} - 1\right), \qquad (6)$$

3. RESULTS AND DISCUSSIONS

All the predictor variables that are used in this paper are investigated as risk factors related to the road, road environment, vehicle, driver and accident itself, as it is shown in table 1.

Table 1. Risk factors' category and the predictors used in the Binary Logistic Regressions.

Risk factors related to:	Category of the variable	Number of response variables included in this set	Variable's type
Road	Road type	6	Categorical
	Road feature	8	Categorical
	Road feature's configuration	13	Categorical
	Road inclination	4	Categorical
	Number of traffic lanes	8	Continuous
	Lanes' width	4	Categorical
	Roadway condition	5	Categorical
Road environment	Built up area/Outside built up areas	2	Categorical
	Road markings	4	Categorical
	Hard shoulders characteristics	3	Categorical

	Safety devices on the road	9	Categorical
	Roadway adherence	9	Categorical
	Lighting conditions	10	Categorical
	Weather conditions	7	Categorical
Vehicle	Odometer's displaying	7	Categorical
Driver	Male/female	2	Categorical
Crash	Accidents' causes	33	Categorical

As it was specified above at chapter 2, the data were aggregated on daily basis. Due to this aggregation level, all the categorical variables have been automatically transformed in continuous ones and there is no longer need for them to be transformed to dummy variables as it is in the logistic regressions that use this type of predictors. For example, the "Road type" category has 6 response variables that are categorical ones and each of them can take only the values of 0 or 1: street, national road, inter-county road, inter-village roads, highway and other roads. In case of an analysis made on an individual basis, it would have been necessary to transform those 6 variables in 5 dummy variables (k-1 variables). In our case (daily aggregation level), the values of a variable is given by the cumulating the 0 and 1 values of each of the variables counted that precise day. When redundant variables are found by SPSS, meaning that some variables are collinear with others, these are automatically removed by SPSS. Therefore, even if the number of response variables are 134 (Table 1), the total number of predictors taken into analyses are 86 as the figure 1 shows.

3.1 Estimation of P_{d1}

In order to estimate the P_{d1}, a binary multiple logistic regression is developed in SPSS and its primary outputs are shown in Figure 1:

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	413.559	86	.000
	Block	413.559	86	.000
	Model	413.559	86	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	2551.885[a]	.172	.232

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	7.579	8	.476

Figure. 1. Primary outputs for P_{d1} model consisting in G2 test, values of the Cox & Snell R Square and Nagelkerke R Square and Hosmer and Lemeshow test.

As it can be seen in figure 1, P_{d1} model is statistically significant since G2(86) = 413.559, $p<0.001$. This is also shown by the Hosmer and Lemeshow's Goodness of Fit test which is not significant ($p=0.476>0.05$), demonstrating an influence of the predictors on Pd1 variance. The Cox & Snell R^2 value is 0.172 and Nagelkerke R^2 value is 0.232, showing that all the predictors have an overall contribution between 17.20 % and 23.20 % to the variation of P_{d1}.

Looking at the classification table from figure 2, we can notice that the overall percentage of the predictions is 58.60 %.

Classification Table[a,b]

Observed			Predicted		
			P_{d1}		Percentage Correct
			0	1	
Step 0	P_{d1}	0	0	905	.0
		1	0	1281	100.0
	Overall Percentage				58.6

a. Constant is included in the model.

b. The cut value is .500

Figure 2. Classification table of the P_{d1} predictions.

After conducting the Wald test, only 14 predictors are of significance, as it is shown in table 2.

Table 2. The significant predictors and their logit coefficients B, Wald test of the individual significance, the exponential values of the coefficients (odds values) and their percentages.

Predictors	b_k	S.E.	Wald	df	Sig.	$Exp(b_k)$	OR=$Exp(b_k)$ %
Outside built-up areas (OBA)	1.527	0.674	5.134	1	0.023	4.605	360.50
Built up areas (BUA)	1.445	0.675	4.591	1	0.032	4.244	324.40
Curve (CRV)	-0.724	0.342	4.496	1	0.034	0.485	-51.50
Intersection (INT)	-0.878	0.367	5.722	1	0.017	0.416	-58.40
Railway crossing (RWC)	-0.922	0.391	5.551	1	0.018	0.398	-60.20
LN = 1	-0.365	0.157	5.395	1	0.020	0.694	-30.60%
LN = 2	-0.198	0.082	5.844	1	0.016	0.820	-18.00%
LN = 3	-0.180	0.073	6.110	1	0.013	0.836	-16.40
Paved hard shoulder (PHS)	0.084	0.040	4.431	1	0.035	1.087	8.70
Safety barriers (STB)	0.668	0.300	4.976	1	0.026	1.951	95.10%
Safety slides (STS)	0.484	0.189	6.525	1	0.011	1.622	62.20%
No road markings (NRM)	0.298	0.131	5.179	1	0.023	1.347	34.70%
Odometer info:							
25000 km – 50000 km (OD25-50)	0.215	0.094	5.217	1	0.022	1.239	23.90%
Female (FML)	-0.167	0.040	17.759	1	0.000	0.846	-15.40%
b_0	-1.419	0.133	113.956	1	0.000	0.242	-

When the number of SVA occurring outside of built-up areas increases by one unit, the odds of them being fatal increase by 360.50 % [OR(OBA) = 360.50 %, Wald(1) = 5.134, $p<0.05$].

Similarly, for those SVA that occur inside the built-up areas the odds increase by 324.40 % [OR(BUA) = 324.40 %, Wald(1) = 4.591, $p<0.05$], showing that a similar fatality risks is assigned for both types of travels (outside and inside built-up areas), although the travel speed is notably higher outside built-up areas than inside of cities or villages.

Regarding the road's features, the SVAs' number that occurs in curves, intersections and railway crossings significantly explain the variation of P_{d1}. There is an indirect relationship between each of these road's features and the variation of P_{d1}, as it is shown by the negative values of the odds: [OR(CRV) = -51.50 %, Wald(1) = 4.496, $p<0.05$], [OR(INT) = -58.40 %, Wald(1) = 5.722, $p<0.05$] and [OR(RWC) = -60.20 %, Wald(1) = 5.551, $p<0.05$]. In other words, any increase of the SVAs' number that occurs in this type of road's features will lead to a decrease of the probability of SVA of being fatal. The indirect relationship between the P_{d1}'s variation and the SVAs' number that occur in curves is an interesting finding of this paper, since it contradicts other papers in which the curves are considered as risk factors having a high significance in explaining the variation of the severity's increase of a SVA that occurs in these roads' features [4], [13], [14], [15]. When the drivers have to approach a curve, an intersection or a railway crossing, they become more attentive and decrease the travel speed. Apart from this, as it can be seen in the table 2, each SVA that occur on a railway crossing leads to a decrease by 60.20 % of the odds for that collision to be a fatal one. The number of those SVA occurring on roads with 1, 2 or 3 lanes (LN = 1, LN = 2 or LN = 3) significantly explain the variation of P_{d1}, in an indirect relationship, as it is shown by the negative values of the logit coefficients.

Each single vehicle accident that occur on those roads where safety barriers are installed increases by 95.10 % the probability of that SVA to be a fatal one [OR(STB) = 95.10 %, Wald(1) = 4.976, $p<0.05$]. On the roads where safety slides are installed, each SVA increase by 62.20 % the probability of that collision to generate fatalities [OR(STS) = 62.20 %, Wald(1)= 6.525, $p<0.05$]. The safety barriers are usually installed on a those roads that have a high risk of collisions and where the speed is inappropriate to the roadway configuration. When the number of the SVA occurring on those roads where the horizontal marking is missing increases with one unit, this leads to an increase by 34.70 % of the odds that SVA to be a fatal one [OR(NRM) = 34.70 %, Wald(1) = 5.179, $p<0.05$] and this confirms the previous research findings that considered the longitudinal road marking as a risk factor [16]. Usually, the roadway where the horizontal marking is missing can be found on the areas where road works are developing and where the drivers pay more attention simultaneously with a speed diminishing.

Another interesting finding of this paper concerns the number of female drivers involved in single vehicle accidents. This number significantly explains the variation of Pd1 in an indirect relationship and each SVA in which a female driver was involved leads to a decrease by 15.40 % of the odds of that SVA to be a fatal one [OR(FML) = -15.40 %, Wald(1) = 17.759, $p<0.001$]. This statement is in conflict with the conclusions found on [17] in which the female drivers have increased odds to be involved in SVA on the mountain roads and the severity of the accidents caused by the female drivers is higher comparing with the male drivers [15]. The very low p value of this predictor ($p<0.001$) shows that this predictor has the highest explanatory power from all the variables.

Using the predictors presented in table 2, which significantly explain the P_{d1}'s variation, the logistic equation can be written with the formula (3) and the estimated values of P_{d1} will be calculated using formula (1).

A descriptive analysis conducted on the P_{d1} distribution shows a positive asymmetry tendency of the P_{d1} distribution curve as the skewness value (0.73) is a positive one, as it is shown in the table 3.

Table 3: Statistical parameters of the P_{d1} distribution curve.

Pd1 Descriptive Statistics		
N	Valid	2190
	Missing	1
Mean		.585937
Median		.573255
Mode		.0000
Std. Deviation		.2044395
Skewness		.073
Std. Error of Skewness		.052
Kurtosis		-.914
Std. Error of Kurtosis		.105

As it is presented in figure 3, since the skewness value is a positive one, there are many higher P_{d1} values in the detriment of the smaller P_{d1} values. The mean and median have close values and this shows the tendency of the P_{d1} distribution curve towards a normal distribution. This is also confirmed by research [17] that states that any distribution which has the skewness value between [-0.80, 0.80] can be treated as a normal distribution.

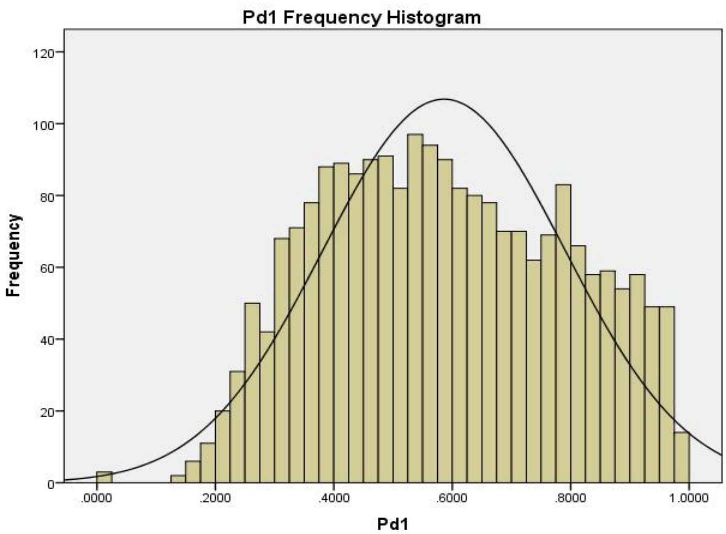

Figure 3. Frequency histogram of the Pd1 valuesAs is it also shown in the table 3 and figure 3, since the kurtosis value (-0.914) is a negative one, the estimated values of Pd1 are scattered and heterogeneous.

4.2 P_{d2} estimation

The primary outputs generated after running a binary logistic regression for P_{d2} model are shown in figure 4 in which we can see that P_{d2} model is also statistically significant, since $G^2(86) = 242.819$, $p<0.001$ and the Hosmer and Lemeshow's Goodness of Fit test is not of significance ($p= 0.983>0.05$).

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	242.819	86	.000
	Block	242.819	86	.000
	Model	242.819	86	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	482.256[a]	.105	.372

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	1.924	8	.983

Figure 4: Primary outputs for P_{d2} model consisting in G2 test, values of the Cox & Snell R Square si Nagelkerke R Square and Hosmer and Lemeshow test.

The following predictors explain the variance of P_{d2} in a proportion of 10.50 % - 37.20 % (as it shown by the Cox & Snell R^2 and Nagelkerke R^2 values): the number of daily accidents that occurred in simple "T" intersections, on those roads with unpaved shoulders, where the lane's width is over 3.75 m and in which cars having more than 100.000 km are involved, as it is shown in the table 5.

Table 5. The significant predictors and their logit coefficients B, Wald test of the individual significance, the exponential values of the coefficients (odds values) and their percentages.

Predictors	b_k	S.E.	Wald	df	Sig.	Exp(b_k)	OR=Exp(b_k) %
Simple "T" intersections (TIN)	1.314	0.567	5.363	1	0.021	3.721	272.10
Lane width >3.75m (LW3)	0.534	0.255	4.391	1	0.036	1.706	70.60
Unpaved hard shoulder (UHS)	0.276	0.139	3.961	1	0.047	1.318	31.80
Odometer info: 100,000 km – 150,000 km (OD100-150)	0.929	0.212	19.214	1	0	2.532	153.20
Odometer info: >150,000 km (OD>150)	0.329	0.143	5.276	1	0.022	1.389	38.90
b_0	-0.161	0.333	0.233	1	0.629	0.851	-

In case of P_{d2}, the overall percentage of the predictions is 96.10 %, as it is shown in the figure 6.

Classification Table[a,b]

Observed			Predicted		
			P_{d2}		Percentage Correct
			0	1	
Step 0	P_{d2}	0	0	86	.0
		1	0	2100	100.0
	Overall Percentage				96.1

a. Constant is included in the model.

b. The cut value is .500

Figure 6: Classification table of the Pd2 predictions.

The logistic equation can be now written using formula (3) and is as follows:

$$Y_{d2} = exp(-0.161 + 1.314*TIN + 0.534*LW3 + 0.276*UHS + \\ + 0.929*(OD100-150) + 0.329*(OD>150)) , \quad (7)$$

The estimated values of P_{d2} can be now calculated using formula (1).

The simple “T” intersections are one of the most dangerous road features, since each SVA taking place here increases the major injury probability by 272.10 % [OR(TIN) = 272.10 %, Wald(1) = 5.363, P<0.05]. Most of these intersections can be found on those roads that are situated outside of the built-up areas where the signalization is poor and the travel speeds are higher.

On the roads where the lane’s width is >3.75 m, each increase of the SVAs’ number with one unit will increase the odds with 70.60 %, this being in line with the findings [19] which states that a high injury risk in single vehicle accidents is predicted on a road with large traffic lanes. P_{d2} variation is also significantly explained by the number of SVA which occur on the roads with no paved hard shoulders [OR(UHS) = 31.80 %, Wald(1) = 3.961, p<0.05].

Other interesting finding of this paper regards the high explanatory power of the number of SVA in which are involved old vehicles whose mileage is between 100,000 km – 150,000 km. Each increase of the SVAs’ number which involve this type of vehicles leads to an increase by 153.20 % of the odds that this accident will result major injuries [OR(OD100-150) = 153.20 %, Wald(1) = 5.276, p<0.001]. The Romanian car park has a considerable average age (16.2 years) and most of these vehicles are passenger cars (78.80 %) having these mileages. As these kinds of old vehicles are not equipped with high performance active and passive safety systems, they are associated with a high injury risk of the passengers and drivers as it was stated in [12], [20-23]. The other finding that regards the car’s mileage variable is quite contrary to the previously mentioned papers. As we can see from table 5, each increase of a SVA in which a vehicle having the mileage > 150,000 km is involved, will lead to an increase by 38.90 % of the odds of major injuries [OR(OD>150) = 38.90 %, Wald(1) = 5.276, p<0.05]. Most of the vehicles having a mileage > 150,000 km have a considerable age, a small share in the national car park and are often used occasionally, during weekends or summer season.

A descriptive analysis conducted on P_{d2} frequencies generates the statistics presented in table 6.

Table 6. Statistical parameters of the Pd2 distribution curve.

Statistics

Pd2		
N	Valid	2190
	Missing	1
Mean		0.960441
Median		0.995920
Mode		1.0000
Std. Deviation		.0854170
Skewness		-3.688
Std. Error of Skewness		0.052
Kurtosis		16.782
Std. Error of Kurtosis		0.105

As the skewness value is a negative one (-3.688), the P_{d2} distribution curve is a strongly asymmetrical one (its value is far from zero) and there are many higher values in the detriment of the smaller ones, this showing higher values of the major injuries probabilities in the single vehicle accidents. The kurtosis coefficient has a positive value (16.782) and shows a leptokurtic tendency of P_{d2} frequencies distribution curve where homogenous values are found, most of them being grouped around the mean (figure 7).

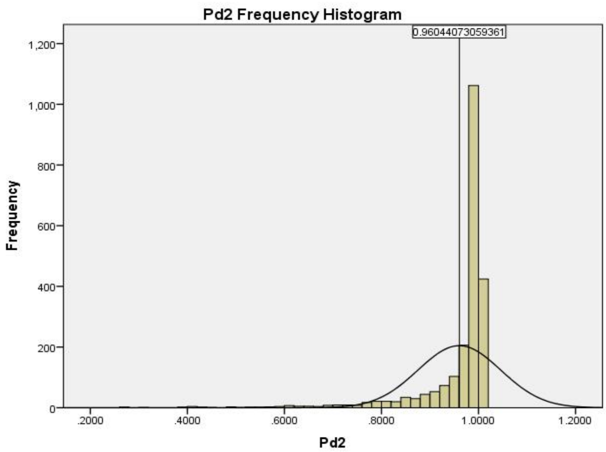

Figure 7. Frequency histogram of P_{d2} values and their grouping compared with the mean.

4. CONCLUSIONS

The probability of a SVA to be of fatality is increased by the number of SVA that occur outside or inside built-up areas, on those roads with paved hard shoulders, safety barriers and slides and no horizontal markings, and those vehicles with mileage between 25,000 km – 50,000 km that are involved in this collisions, as the logit coefficients of these predictors have positive values. Better roads maintenance that would consist in applying of more horizontal markings will decrease the probability of these accidents to be of fatality. As the odds of having a fatal SVA inside and outside the built-up areas are unusually higher, a better law enforcement and development of sustainable mobility plans will play a major role on decreasing the number of fatalities.

The car manufacturers, dealers and the insurance companies, as well, could contribute to the decrease of the number of fatalities as well, by development of educational programs, especially those focused on ADAS (Advanced Driving Assistance Systems) and of awareness campaigns for those who drive relatively new cars with mileages between 25,000 km – 50,000 km.

Other interesting finding of this paper regards the P_{d2} model, where all those 5 predictors included have positive values of the logit coefficients. This means that any decrease of these coefficients will lead to an important decrease of the victims' numbers that are seriously injured in these types of collisions. The simple "T" intersections are the most dangerous, as each SVA occurred in this intersections leads to an increase by 272.10 % of the odds of being seriously injured. The drivers of the those old vehicles whose mileage is between 100,000 km – 150,000 km have a greater risk of being seriously injured in the single vehicle accidents, as the odds values are very high.

Of the two P_{d1} and P_{d2} models, the second one is better significantly explained by the predictors in a proportion of 10.50 % - 37.20 % and has a prediction accuracy of 96.10 %. In addition, all the predictors of P_{d2} have positive values, a thing which gives a practical applicability to this model since any decrease of a predictor's values will lead to a decrease of the serious injuries' probability. Rehabilitation of the simple "T" intersections, reducing the speed around them and improving the lighting during nights, will decrease the number of accidents around these intersections and consequently will lead to a decrease of the numbers of the seriously injured victims.

We should remark that no other predictors related to weather or lighting conditions do significantly explain the probabilities of a SVA to be of fatality or to generate major injuries. These constraints are to be further researched since the daily level of data aggregation studied in this paper influences the "immediate effect" of random phenomena, as weather or lighting conditions.

P_{d2} model has a good applicability in the identifying, prioritizing and treating of the black spots. Knowing the predicted probabilities on the road segments where a higher accident concentration is found would help the road administrators to implement suitable countermeasures in order to mitigate the major injury risk at lower costs. The road features with a higher major injury probability can be improved in terms of safety by implementing different countermeasures or by a better maintenance.

The Romanian road authority could also use it in driver education and injury risk identification in order to mitigate the severity of the accidents, including in the process of road law improvement.

REFERENCES

[1] Adminaite D and Jost G and Stipdonk H and Ward H 2017 Reducing *Deaths in Single Vehicle Collisions*. PIN Flash Report 32. European Transport Safety Council https://etsc.eu/wp-content/uploads/PIN_FLASH32-FINAL.pdf (accessed 03.07.2020)

[2] Ministry of Internal Affairs, General Inspectorate of the Romanian Police 2018 *Road Safety Bulletin: 2018 Annual Report* https://www.politiaromana.ro/files/pages_files/RAPORT_2020_COMPLET_LOW.pdf (accessed 03.07.2020)

[3] Rifaat S M and Chin H C 2005 *Accident severity analysis using ordered probit models* J. Adv. Transp 41 (1), pp 91-114

[4] Essam Dabbour and Murtaza Haider and Eman Diaa 2019 *Using random-parameter and fixed-parameter ordered models to explore temporal stability in factors affecting drivers' injury severity in single-vehicle collisions* Journal of Traffic and Transportation Engineering, Volume 6, Issue 2, pp 132-146, ISSN 2095-7564

[5] Dong Bowen and Ma Xiaoxiang and Chen Feng and Chen Suren 2018 *Investigating the Differences of Single-Vehicle and Multivehicle Accident Probability Using Mixed Logit Model* Journal of Advanced Transportation

[6] Lord D and Manar A and Vizioli A 2005 *Modeling crash-flow-density and crash-flow-v/c ratio relationships for rural and urban freeway segments* Accid. Anal. Prev., 37 (1) pp 185-199

[7] Nawaf Alnawmasi, and Fred Mannering 2019 *A statistical assessment of temporal instability in the factors determining motorcyclist injury severities* Analytic Methods in Accident Research Volume 22

[8] Grigorios Fountas and Achille Fonzone and Niaz Gharavi and Tom Rye 2020 *The joint effect of weather and lighting conditions on injury severities of single-vehicle accidents* Analytic Methods in Accident Research Volume 27

[9] Peng Y and Jiang Y and Lu J and Zou Y 2018 *Examining the effect of adverse weather on road transportation using weather and traffic sensors* PLoS ONE 13(10)

[10] Pokorny Petr and Jensen Jan and Gross Frank and Pitera Kelly 2020 *Safety effects of traffic lane and shoulder widths on two-lane undivided rural roads: A matched case-control study from Norway* Accident Analysis and Prevention 144 105614

[11] Chang H L and Yeh T 2006 *Risk factors to driver fatalities in single-vehicle crashes: Comparisons between non-motorcycle drivers and motorcyclists* Journal of Transportation Engineering 132 (3) pp. 227-236

[12] Yau K K W 2004 *Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong* Accident Analysis and Prevention 36 (3) pp. 333-340

[13] Chen C and Zhang G and Liu X 2016 *Driver injury severity outcome analysis in rural interstate highway crashes: a two-level Bayesian logistic regression interpretation* Accident Analysis and Prevention,97 pp. 69-78

[14] Lee C and Li X 2014 *Analysis of injury severity of drivers involved in single- and two-vehicle crashes on highways in Ontario* Accident Analysis and Prevention 71 pp 286-295

[15] Qiong Wu and Guohui Zhang and Xiaoyu Zhu and Xiaoyue Cathy Liu and Rafiqul Tarefder 2016 *Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways* Accident Analysis and Prevention, Volume 94 pp 35-45

[16] Park E S and Carlson P J and Porter R J and Andersen C K 2012 *Safety effects of wider edge lines on rural, two-lane highways*. Accident Analysis and Prevention 48, pp 317–335

[17] Sava F 2011 *Analiza datelor în cercetarea psihologică* [*Analysis of data in psychological research*] Cluj Napoca ASCR Publishing House

[18] Gitelman V and Doveh E and Carmel R and Hakkert S 2019 *The influence of shoulder characteristics on the safety level of two-lane roads: a case-study* Accident Analysis and Prevention 122 pp 108–118

[19] Behnood A and Mannering F 2015 *The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: some empirical evidence* Analytic Methods in Accident Research 8 pp 7-32

[20] Behnood A and Roshandeh A and Mannering F 2014 *Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities*. Analytic Methods in Accident Research 3–4 pp 56-91

[21] Joon Ki Kim and Gudmundur F Ulfarsson and Sungyop Kim and Venkataraman N Shankar 2013 *Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender* Accident Analysis and Prevention Volume 50 pp 1073-1081

[22] Dabbour E 2017 *Investigating temporal trends in the explanatory variables related to the severity of drivers' injuries in single-vehicle collisions*. J. Traffic Transp. Eng 4 (1) pp 71-79

[23] Wen H Y and Xue G 2019 *Exploring the relationships between single-vehicle traffic accident and driver's route familiarity on the mountainous highways*. Cogn. Technol. Work 21 pp 1-10

Intentionally blank