Contributory Factors of Traffic Accidents in Lahore Using Generalized Linear Models (GLM)

Hajra Slam[1],Yasar Mahmood[1]*

[1,]Department of StatisticsGC University, Lahore (Pakistan)

* Corresponding Author: syed.yasar@gcu.edu.pk

Abstract

The geometric design of roads is the branch of highway engineering concerned with the positioning of the physical elements of the roadway according to standards and limitations with objectives to optimize efficiency and safety while minimizing cost and environmental damage. The present study aims to explore factors of geometric design and other factors which are root cause of accidents in Lahore. Data is carried out about demographic information, physical characteristics and geometric design of road over a period of 3 years. Poisson regression and negative binomial regression are used in analysis. The results show that most of accidents occur at office off timing and fatal due to reckless driving and over speeding. Mostly, cars and tralala hit the bikes and pedestrians. The poisson regression model gives good description of number of accidents which depend on various explanatory variables. Number of lanes, type of locations and roadway light are statistically significant factors. Narrow Shoulder width ($\boldsymbol{m}$), Median Width ($\boldsymbol{m}$) and Lane width ($\boldsymbol{m}$) increase accident occurrence. Three lanes and larger road structures increase accidents. Numbers of accident increase when roadway, type of locations, roadway light and traffic control signals decrease.

Keywords: Number of Accidents, Demographic Information, Physical Characteristics, Geometric Design of Road, Poisson Regression, Negative Binomial Regression

Introduction

A vehicle has a clash with another vehicle, animal, animal cart, pedestrian or fixed obstruction such as tree, road light and road signal is called traffic accident. Traffic accidents may result in main three categories i.e. non- fatality, fatality and property damage. Traffic accidents also called traffic crash, motor vehicle accident, car accident, automobile accident, road traffic crash, road traffic accident, car smash, or car destroy. There are many causes of traffic accidents e.g. slow driving, reckless driving, over speeding, driver sleepiness, personal disease, equipments failure, motor vehicle design, geometric design of road, road environment and driver experience and due to alcohol or drugs usage.

The geometric design of roads is most important cause of traffic accident. It is the branch of highway engineering. It concerned with the positioning of the physical elements of the road according to principles and limitations. Geometric design increases efficiency of roads and ensures safety from road accidents. It also results in reducing cost and environmental damage. Geometric design also has an effect on an access to job, schools, businesses and houses, contain a range of travel modes such as walking, bicycling, transport and vehicles, and decrease the fuel use, productions and environmental ruin.

There are three main parts of geometric design of roads: alignment, profile, and cross-section. They provide a three-dimensional outline for a roadway.

https://doi.org/10.24949/njes.v12i1.506

The sequence of horizontal tangents and curves is called alignment. The profile has a vertical feature of the road i.e. crest and sag curves, and the straight grade lines linking them. In cross section part of roadway, the location and number of motor vehicle and bicycle lanes and footways, along with their banking. Cross section as well consists on drainage features, pavement construction and many other objects outside the category of geometric design.

The purpose of this study is to evaluate the relationship between traffic accidents and geometric design of roads & other factors. This research would be beneficial for all higher authority (TEPA, NESPAK, CTP etc.) deals in planning & transport engineering and humans in overcoming the traffic accidents and for road safety. It would highlight contributory factors that cause of traffic accidents. It would help these organizations in well planning to cut down the number of accidents by looking for possible reasons of accidents.

Many researchers [1-7] have attempted some statistical approaches to relate accidents to geometric characteristics and traffic related explanatory variables. Multiple linear regression, poisson regression, negative binomial regression, multiple logistic regression and zero inflated poisson regression had been used in previous researches.

Jacobs [1] investigated relations among individual injury accident rates on rural infrastructure in Jamaica and Kenya aspects, for example road geometry and vehicle flow. Regression analysis has been used to obtain equations which could be useful for guesstimate modification into crash rates and next is development toward geometric plan for the road.

Milton and Mannering [2] used negative binomial (NB) and poisson regression methods as a way to calculate accidents on the basis of traffic related and highway geometric factors. The prime causes are the land planning support database for geometric and traffic information. Accident occurrences are significantly interrelated with horizontal curvature, tangent length, speed, daily traffic and number of lanes between curves. The results showed that NB regression is most suitable way of analysis, when accident data are dispersed relative to the mean.

Iyinam et al. [3] illustrated road geometric design characteristics and elements which are taken into account and details are given on how and to which point they influence highway safety. This study describes relationship between some road geometric design elements and accident rates. The accident rates of these subsections are determined after the sections are divided into sub-sections. After this, geometric parameters about these sub-sections are made. Lastly, a regression analysis is conducted between the accident rates and geometric parameters and meaningful relationships are found among accident rates and some geometric design elements. It is observed that elements associated with horizontal geometry are more effectual on road safety than elements associated with road geometry in this segment. The impact of elements related with vertical geometry is probable to augment on roads with high traffic size.

Abdel- Aty and Radwan [4] used NB modeling technique for model the frequency of crashes rate and involvement. Three years period data has been taken, accounting for 1606 accidents to guesstimate the model on arterial in Central Florida. The model demonstrated the consequence of the degree of horizontal urban/rural curvature, lane, shoulder and, annual average daily traffic (AADT), section's length and the median widths on the rate of accident occurrence. The consequences exemplified that high speed, huge traffic volume, urban roadway sections, larger number of lanes; narrow lane width, reduced median width and narrow shoulder width raise the possibility for crashes.

Sharma et al. [5] examined the impact of traffic variables and road geometry on motorbike accidents using a statistical method called as zero inflated negative binomial regression. The independent variables preferred for this study includes heavy vehicle percentage (HVPER), speed variation from modal speed (VARMSP), access density (AD), shoulder width deficiency (SWDEF), standard deviation of speed (STDSP), and annual average daily traffic (AADT). Accident per year per km (ACCR) has been taken as dependent variable. Accident facts gathered from National Highway No.6 (NH-6) over a stretch of 100 km of road length between Lakhni (a township on NH-6) and Amravati (a city on NH-6) are used for modeling. It is noticed that percentage of heavy vehicles in traffic, speed variations and shoulder width deficiency have major effect on security of motorcyclist.

Chengye and Ranjitkar [6] examined accident prediction model of motorway safety that connect crash frequencies to their non-behavioral contributing features, including weather setting, traffic situation, geometric as well as operational characteristics of road. This study used a sample of crashes took place from 2004-2010 on a 74 km lengthy section of Auckland motorway. A number of crash prediction models are developed and evaluated for their extrapolative capability using NBR models. The outcomes revealed the security effects of different non-behavioral contributing factors, in which AADT per lane and the number of lanes segment length always have the deep impacts on crashes frequency.

Ratanavaraha and Suangka [7] utilized data of expressway networks in B.E.2550 (2007)-B.E.2553 (2010) according to this data legislatively directive as part, Authority of Expressway of Thailand, has implemented of 9 expressway paths as a duty, with a evidence of 2194 crashes. Main idea is to visualize the crash extremity through devising MLR to suppose the probability of serious crashes as well as injury crash in contrast with belongings harm only crash. It has been broadly judged statistical relationship with variables for example, period of time, weather situations, physical distinctiveness of crash locale, average speed on road segment, and average traffic V/D.

Methodology

In this study, primary and secondary data have been used. Data is collected from different institutions of Lahore like, TEPA (Traffic Engineering and Planning Agency), NESPAK (National Engineering Services Pakistan), CTP (City Traffic Police) and Rescue 1122. The registered traffic accidents have been used from 01/01/2013 to 31/12/2015. Data set contains a sample of 356 registered traffic accidents. Two phase sampling technique has been used. On first phase, secondary data is carried out about demographic information and physical characteristics from CTP. Variables that are obtained i.e. place of incident, accident severity, time of accident, seasons, year, cause of accident and vehicle hit to, vehicle hit by for demographic information and physical characteristics of accidents. On second phase, primary and secondary data has been obtained about geometric design of roads from TEPA. Google earth and Auto Cad are also used for geometric design. All registered traffic accidents from Lahore has been taken as target population and all the registered traffic accidents on Ferozpur Road, Multan Road, Canal Bank Road and Grand Trunk Road as a sampled population. It holds traffic accidents of Ferozpur Road from Qurtaba Chowk to Purana Kahna, Multan Road from Chauburji to Manga Mandi, Canal Bank Road from Thokar Naiz Baig to Sky Land Water Park Bridge and Grand Trunk Road from Misri Shah to Wagha Cutt in Lahore. For model, total numbers of observations are 172. Here, GLM's for count is discussed: Poisson regression and negative binomial regression. Poisson regression analysis is applied only when equidispersion occurs. If over dispersion occurs, we apply both regressions and compare their results. Number of accidents used as dependent variable. Variables that are acquired i.e. lane width (m), median width (m), shoulder width (m), number of lanes, roadway, road structure, type of location and roadway / road / street lighting, traffic control signals for geometric design of roads as independent variables.

Generalized linear models (GLMs) is expanded form of ordinary regression models. In GLMs, response distributions are non-normal and have modeling functions of the mean. Generalized linear models [8] have three components. The random component has the dependent variable Y and has a specific probability distribution. The systematic component has all the independent variables for the model. The link function has a function of the mean of Y, which the GLM relates to the independent variables.

In GLMs, the random variable Y can take non-negative integer value like counts. The systematic component has a linear combination of all independent variables and link function of count Y is $g(\mu) = \log(\mu)$. The log link function has positive numbers, so the log link function is only used when μ has only non-negative numbers, alike count data. The GLM has the log link and known as log linear model. It is written as

$$\log(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k \quad (1)$$

There are many types of GLMs. If random component of GLMs have non- negative integer as count then Poisson regression is used. Poisson regression is form of GLMs. The Poisson log linear model has form for one independent variable x,

$$\log(\mu) = \alpha + \beta x \quad (2)$$

The mean μ shows the exponential relationship

$$\mu = \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x} \quad (3)$$

A one-unit increase in explanatory variable x has a multiplicative effect of e^{β} on μ. The mean of Y at $x + 1$ equals the mean of Y at x multiplied by e^{β}. If $\beta = 0$, then $e^{\beta} = e^{0} = 1$ and the multiplicative effect is 1. Then, the mean of Y does not change as x changes. If $\beta < 0$, then $e^{\beta} < 1$, and the mean decreases as x increases. If $\beta > 0$, then $e^{\beta} > 1$, and the mean of Y increases as x increases.

In Poisson regression [8], random component of GLMs Y has Poisson distribution. Its means that Y has same mean and variance. If variance of Y is larger than its mean, then data have greater variance than mean for a GLM and is called over-dispersion.

In ordinary regression Y is normally distributed so, there is not an issue of over-dispersion. The normal distribution has a separate parameter from the mean μ to tell location and variance σ^2 to explain variability. Over-dispersion is mostly occurring when Poisson regression of GLMs for counts is applied.

Negative binomial regression [8] is another form of GLMs. Negative binomial is used, when random component have non-negative integer as count and variance of random component is greater than its mean. It has an additional parameter denoted by D. The negative binomial distribution has

$$E(Y) = \mu, \qquad Var(Y) = \mu + D\mu^2 \tag{4}$$

The additional parameter D, which is non-negative, is called a dispersion parameter. If D has a larger value, it means that Poisson has a greater variability. As $D \to 0$, $Var(Y) \to \mu$ and the negative binomial distribution meets to the Poisson distribution. On the other hand, if D has a larger value then over-dispersion occurs.

The Wald method [9] introduced for significance testing and interval estimation for any GLM. It is used for a significance testing of a null hypothesis $H_0: \beta = \beta_0$ and individual testing of all the parameters. They all develop the large sample normality of ML estimators.

The Wald test statistic have non-null standard error SE of $\hat{\beta}$ i.e.

$$z = \frac{\hat{\beta} - \beta_0}{S.E} \tag{5}$$

has standard normal distribution when $\beta = \beta_0$. Wald test can have two forms: z and other χ^2. z^2 has a chi-squared null distribution with 1 degree of freedom (df).

$$z^2 \sim \chi^2_{(1)} \tag{6}$$

Each method is constructing a confidence interval (C.I). Confidence intervals are more useful for parameters to test hypotheses about their values. For any test method, a confidence interval (C.I) results from capsizing the test. e.g., a 95% C.I for β is theset of β_0 for which the test of $H_0: \beta = \beta_0$ has a p-value greater than 0.05. The Wald confidence interval is the set of β_0 for which $|\hat{\beta} - \beta_0| \,/\, SE < Z_{\frac{\alpha}{2}}$. This provides the interval $\hat{\beta} \pm 1.96(SE)$. By statistical software, Wald confidence interval is usually calculated because ML estimates and standard error (S.E) are easily calculated.

The Akaike information criterion (AIC) tells about the quality of statistical models. Hence, AIC is used for only model selection. If we estimated more than one models at a time, estimated model that have smaller value of AIC will be best fitted model. The smaller value of AIC represents that this model is best.

$$AIC = -2\ln(L) + 2k \tag{7}$$

In above equation, the maximum value of the likelihood function for the model is denoted by L and k is number of estimated parameters in the model. Hence, AIC gives goodness of fit (as reviewed by the likelihood function).

Results and Discussion

Data set contains a sample of 356 registered traffic accidents. It holds traffic accidents of Ferozpur Road from Qurtaba Chowk to Purana Kahna, Multan Road from Chauburji to Manga Mandi, Canal Bank Road from Thokar Naiz Baig to Sky Land Water Park Bridge and Grand Trunk Road from Misri Shah to Wagha Cutt in Lahore. Variables that are obtained i.e. place of incident, accident severity, time of accident, seasons, year, cause of accident and vehicle hit to, vehicle hit by for demographic information and physical characteristics of accidents. For model, total numbers of observations are 172. Number of accidents used as dependent variable. Variables that are acquired i.e. lane width (m), median width (m), shoulder width (m), number of lanes, roadway, road structure, type of location and roadway / road / street lighting, traffic control signals for geometric design of roads as independent variables.

Most of accidents occur between 3 PM to 6 PM (office off time) i.e. 19.1% and 17.7% take place on Friday. Majority accidents happened in 2014 i.e. 44.4% and 41.6% occurred in summer (May to september1-15) season. A large number of accidents arise fatal which is 33.4% and due to reckless driving and over speeding which is 45.8%. 27.7 % of accidents occurred when vehicle hit by car and vehicle hit to bike which is 40.4%. From registered sample accidents, numbers of death are 197 and numbers of injuries are 311 from 2013 to 2015.

SPSS software has been used to perform analysis. For Poisson regression analysis, (i) Dependent variable consists of count data. (ii) Data have one or more independent variables. (iii) Data should have independence of observations. (iv) The distribution of counts (conditional on the model) pursues a Poisson distribution. (v) The mean and variance of the model are the same.

The dependent variable has mean 2.07 and the variance 3.1648 (1.779^2), which is a ratio of 3.1648 ÷ 2.07 =1.5289. A Poisson distribution supposes a ratio of 1 (i.e., the mean and variance are identical). There is a small amount of overdispersion before adding of any explanatory variable. When, all the independent variables have been added to the poisson regression. The value of "Pearson Chi-Square" is 154.558 and $d.f = 154$. For goodness of fit "$Value/df$", value of 1 indicates equidispersion, greater than 1 indicate overdispersion and below 1 indicate under dispersion. Here, $154.558/154 = 1.004$, indicates equidispersion.

Omnibus test is a likelihood ratio test of whether all the explanatory variables jointly improve the model over the intercept-only model (i.e., with no explanatory variables added). All explanatory variables have a *p*-value of 0.000, representing overall statistically significant model.

The tests of model effects display the statistical significance of each of the independent variables. However, roadway (i.e., $p = .670$), road structure (i.e., $p = .368$), traffic control signal (i.e., $p = .330$), SW (Shoulder width (m)) (i.e., $p = .901$) and MW (Median width (m)) (i.e., $p = .339$), LW (Lane width (m)) (i.e., $p = .449$) are not statistically significant, but number of lane (i.e., $p = .0.034$), type of location (i.e., $p = .000$) and roadway light (i.e., $p = .042$) are statistically significant. This test is frequently useful for categorical explanatory variables because it is the only test that generally considers the result of a categorical variable, unlike the parameter stimates (table No. 1) as given:

Table No. 1: Parameter Estimates

Parameter Estimates										
			95% Wald Confidence Interval		Hypothesis Test				95% Wald Confidence Interval for $Exp(B)$	
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.	$Exp(B)$	Lower	Upper
(Intercept)	2.188	.8788	.466	3.910	6.200	1	.013	8.918	1.593	49.922
[Numberoflane=2]	.276	.1704	-.058	.610	2.623	1	.105	1.318	.944	1.840
[Numberoflane=3]	-.150	.1400	-.424	.125	1.142	1	.285	.861	.655	1.133
[Numberoflane=4]	0^a	.	.	.	.	.	.	1	.	.
[Roadway=1]	-.164	.3856	-.920	.592	.181	1	.670	.849	.399	1.807
[Roadway=2]	0^a	.	.	.	.	.	.	1	.	.
[RoadStructure=1]	.328	.7443	-1.131	1.786	.194	1	.660	1.388	.323	5.968
[RoadStructure=2]	.480	.7681	-1.025	1.986	.391	1	.532	1.616	.359	7.284
[RoadStructure=3]	0^a	.	.	.	.	.	.	1	.	.
[Typeoflocation=1]	-1.333	.7486	-2.800	.134	3.170	1	.075	.264	.061	1.144
[Typeoflocation=2]	-1.802	.8035	-3.377	-.227	5.027	1	.025	.165	.034	.797
[Typeoflocation=3]	-1.200	.7559	-2.681	.282	2.520	1	.112	.301	.068	1.325
[Typeoflocation=4]	-.665	.7681	-2.171	.840	.750	1	.386	.514	.114	2.316
[Typeoflocation=5]	-.902	.7577	-2.387	.583	1.416	1	.234	.406	.092	1.792
[Typeoflocation=6]	-.808	.6221	-2.027	.412	1.686	1	.194	.446	.132	1.509
[Typeoflocation=7]	0^a	.	.	.	.	.	.	1	.	.
[Typeoflocation=8]	-.233	.6650	-1.536	1.071	.123	1	.726	.792	.215	2.917
[Typeoflocation=9]	0^a	.	.	.	.	.	.	1	.	.
[Roadwaylight=0]	-1.105	.5443	-2.172	-.038	4.122	1	.042	.331	.114	.962
[Roadwaylight=1]	0^a	.	.	.	.	.	.	1	.	.
[Trafficcontrolsignal=0]	-.173	.1771	-.520	.175	.949	1	.330	.841	.595	1.191
[Trafficcontrolsignal=1]	0^a	.	.	.	.	.	.	1	.	.
SW	-.009	.0747	-.156	.137	.015	1	.901	.991	.856	1.147
MW	-.010	.0100	-.029	.010	.914	1	.339	.990	.971	1.010
LW	-.163	.2147	-.583	.258	.573	1	.449	.850	.558	1.295
(Scale)	1^b									

Dependent Variable: Numberofaccidents
Model: (Intercept), Numberoflane, Roadway, RoadStructure, Typeoflocation, Roadwaylight,Trafficcontrolsignal, SW, MW, LW

a. Set to zero because this parameter is redundant.
b. Fixed at the displayed value.

The parameter estimates (table no. 1) give both the coefficient estimates (the β column) of the Poisson regression and the exponential values of the coefficients (the exp (β) column). It is generally the latter that are more useful. These exponential values can be used for interpretation.

The estimated model is:

$$log(\hat{\mu}) = 2.188 - 0.009(SW) - 0.010(MW) - 0.163(LW) + 0.276(NL2) - 0.150(NL3) - 0.164\,(RW) + 0.328(RS1) + 0.480(RS2) - 1.333(TL1) - 1.802(TL2) - 1.200(TL3) - 0.665(TL4) - 0.902(TL5) - 0.808(TL6) - 0.233(TL8) - 1.105(RL) - 0.173(TS) \quad (8)$$

Notations

SW: Shoulder Width, MW: Median Width, LW: Lane Width, NL2: Two Lane, NL3: Three Lane, RW: Roadway, RS1: Straight Ahead, RS2: Horizontal Curve, TL1: Straight, TL2: Curve, TL3: Intersection, TL4: Junction, TL5: Bus stop, TL6: Upgrade, TL8: Bridge, RL: Roadway Light, TS: Traffic Control Signal

In parameter estimates (table no. 1), shoulder width (m), median width (m) and lane width (m) used as continuous variables. Shoulder Width (m) (i.e., SW row), the exponential value of β is 0.991. This means that number of accidents (response variable) will decrease by 0.991 times for shoulder width increases by 1m. Median Width (m) (i.e., MW row), the exponential value of β is 0.990. This means that number of accidents will decrease by 0.990 times for median width increases by 1m. Lane Width (m) (i.e., LW row), the exponential value of β is 0.850. This means that number of accidents will decrease by 0.850 times for lane width increases by 1m. A negative sign of all these widths show that the accidents will be reduced by increasing width. These findings confirm previous results [1, 3-6].

Number of lanes used as categorical variable. Number of lanes (i.e., number of lanes = 2 row in table no. 1), the exponential value of β is 1.318. This means that average number of accidents will increase by 1.318 times for three lanes road. It can also be interpreted in the way that there is 31.8% increase in the number of accidents for three lanes road. These findings confirm previous results [2,4,6,7]. But number of lanes (i.e., number of lanes = 3 row in table no.1), the exponential value of β is 0.861. This means that average number of accidents will decrease by 0.861 times for four lanes and there is almost 14% decrease in accidents.

Roadway used as categorical variable. Roadway (i.e., roadway=1 row in table no.1), the exponential value of β is 0.849. This means that average number of accidents will decrease by 0.849 times at one way road as compare to two way road.

Road structures used as categorical variable. Road structure (i.e., road structure =1 row in table no.1), the exponential value of β is 1.388. This means that average number of accidents will increase by 1.388 times for horizontal curve road structure. Another way of saying this is that there is 38.8% increase in the number of accidents for horizontal curve road structure. Road structure (i.e., road structure =2 row in table no.1), the exponential value of β is 1.616. This means that average number of accidents will increase by 1.616 times for elongated horizontal curve. A positive sign of road structures exhibit that accidents increase due to straight road and elongated horizontal curve but more increased rate attached with horizontal curve. These findings confirm previous results [1,3,4].

Types of location used as categorical variable. Type of location (i.e., type of location=1 row in table no.1), the exponential value of β is 0.264. This means that average number of accidents will decrease by 0.264 times for long straight area. Type of location (i.e., type of location=2 row in table no.1), the exponential value of β is 0.165. This means that average number of accidents will decrease by 0.165 times for more curve area and also statistically significant. Type of location (i.e., type of location=3 row in table no.1), the exponential value of β is 0.301. This means that average number of accidents will decrease by 0.301 times for more intersections. Type of location (i.e., type of location=4 row in table no.1), the exponential value of β is 0.514. This means that average number of accidents will decrease by 0.514 times for more junctions. Type of location (i.e., type of location=5 row in table no.1), the exponential value of β is 0.406. This means that average number of accidents will decrease by 0.406 times for more bus stops. Type of location (i.e., type of location=6 row in table no.1), the exponential value of β is 0.446. This means that average number of accidents will decrease by 0.446 times for more upgraded area. Type of location (i.e., type of location=8 row in table no.1), the exponential value of β is 0.792. This means that average number of accidents will decrease by 0.792 times for more bridges. A negative sign of locations show that more the conveniences less will be the accidents. These findings confirm results [2,3,7].

Roadway light used as categorical variable. Roadway light (i.e., roadway light=0 row in table no.1), the exponential value of β is 0.331. This means that average number of accidents will decrease by 0.331 times for more roadway light. A negative sign of roadway light tell that more is the roadway light less will be the accidents and it is logical acceptable. It is also statistically significant.

Traffic control signal used as categorical variable. Traffic control signal (i.e., traffic control signal=0 row in table no.1), the exponential value of β is 0.841. This means that average number of accidents will decrease by 0.841 times for more traffic control signal. A negative sign of traffic control signal show that the more the traffic control signal less will be the accidents and it is logical acceptable.

Data satisfies assumption no. 5 because equidispersion occurs. So, Poisson regression is applied on this data. If data do not satisfy this assumption, it is either under- or overdispersed. With overdispersion, negative binomial regression should also be applied according to the objective and there should be comparison of both regressions results but equidispersion occurs so, Poisson regression can only be applied. If negative binomial regression is applied on this data, it will give insignificant results.

Conclusions and Recommendations

The results of this study show that most of accidents occur at office off timing. Number of accidents increased in 2014 as compare to 2013 but in 2015 frequency of accidents decreases. A large number of accidents arise fatal due to reckless driving and over speeding. In mostly accidents, cars and tralala hit the bikes and pedestrians. Most of accidents occur on Friday and Monday in summer season. From registered sample of accidents, number of deaths are 197 and numbers of injuries are 311 from 2013 to 2015.

Generalized linear models for count data: Poisson regression and Negative binomial regression are discussed in this research. Poisson regression has been applied only because of equidispersion occurs. If negative binomial regression is applied on this data, it will give insignificant results. The Poisson regression model gives good description of number of accidents depending on various explanatory variables. Number of lanes, type of locations and roadway light are statistically significant. Narrow Shoulder width (m), Median Width (m) and Lane width (m) cause in increasing accident occurrences. Accidents increase when there are three lanes and larger road structures. Numbers of accident increase when Roadway, type of locations, roadway light and traffic control signals decrease.

This research can be beneficial for higher authority (TEPA, NESPAK, CTP etc.) deal in planning & engineering and humans in overcoming the traffic accidents for road safety. The accident prediction model can be used to assess the safety performance and identify hazardous locations needing treatment for both existing and proposed roads. Hence, this research can be helpful for these organizations in well planning to cut down the number of accidents by looking for possible reasons of accidents.

It is recommended that the accident prediction model may develop by using statistical techniques that have been explaining successfully traffic accident factors of next year’s 2016 to onward in the future research. The safety impact of different factors has always been a focus of interest for traffic engineers. Future studies should be conducted in order to identify risk factors for traffic accidents to improve traffic safety.

References

1. G.D. Jacobs, "A study of accident rates on rural roads in developing countries. Transport and Road Research Laboratory Crowthorne, Berkshire (TRRL laboratory report 732)", Available from: "http://docshare01.docshare.tips/files/7803/78031750.pdf", 1976.
2. J.C. Milton & F.L. Mannering, "The relationship between highway geometric, traffic related elements and motor vehicle accidents", *Transport Research Board*, 74, 1996.
3. A.F. Iyinam, S. Iyinam, & M. Ergun, "Analysis of relationship between highway safety and road geometric design elements: turkish case", Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.513.7741&rep=rep1&type=pdf, 1997.
4. M. Abdel-Aty & E. Radwan, "Modeling traffic accident occurrence and involvement", *Accident Analysis and Prevention*, 32(5), 2000, pp. 633–642.
5. A.K. Sharma & V.S. Landge, "Zero inflated negative binomial for modeling heavy vehicle crash rate on Indian rural highway", *International Journal of Advances in Engineering & Technology*, 5(2), 2013, pp. 292-301.
6. P. Chengye & P. Ranjitkar, "Modelling motorway accidents using negative binomial regression", *Proceedings of the Eastern Asia Society for Transportation Studies*, 9, 2013.
7. V. Ratanavaraha & S. Suangka, "Impacts of accident severity factors and loss values of crashes on expressways in Thailand", *IATSS Research,* 37(2), 2014, pp. 130–136.
8. A. Agresti, "A*n introduction to categorical data analysis*", 2nd ed., New Jersey: A John Wiley & Sons, Inc., Publication: Wiley Interscience, 2007.