

# An accident prediction model for divided highways: a case study of Trabzon coastal divided highway

F. Ture Kibar, F. Celik & B. P. Aytac

*Department of Civil Engineering,  
Karadeniz Technical University, Turkey*

## Abstract

Traffic accidents in Turkey have been increasing every year. Although serious property damage and fatalities have occurred in accidents, Turkey hasn't sufficiently overcome this problem.

The main objective for this study is to investigate the factors which cause accidents and to create an accident prediction model which includes relationships between these factors. With this model, the expected number of accidents at divided highways can be predicted and suitable measures for providing road safety can be defined.

For this study, 5 years' (2002, 2003, 2004, 2006 and 2007) accident data of 113.5 km road sections of Trabzon coastal divided highway, traffic and highway characteristics of these sections were collected, then an accident prediction model was formed. The technique of generalized linear models (GLMs) was applied to the data. Because of over dispersion of Poisson regression model, a Negative Binomial regression model was found to be the most appropriate approach for analysis of this data. This model indicates that the vehicle kilometres of travel, the number of pedestrian crossing and average posted speed are significant variables on traffic accident occurrences.

*Keywords: traffic accidents, road safety, accident prediction model, Poisson regression, Negative Binomial regression.*

## 1 Introduction

Traffic accidents are a significant problem for society; because of them, almost 1.2 million people die and between 20 and 50 million people are injured



worldwide each year [1]. In addition, traffic accidents result in enormous economic costs; they create psychological problems for people who are injured in accidents or those who have lost their relatives. Turkey, like other countries, has suffered from traffic accidents. Each year, approximately 4500 people lost their lives and thousands more were injured by traffic accidents on Turkish roads. In Turkey, 90% freight and passenger transportation is made by highway transportation [2]. The main reason for the increasing number of traffic accidents is the concentration of freight and passenger transportations on highways in Turkey and also unsatisfactory condition of road infrastructures. Some preventive measures have been taken for decreasing accidents, but the annual number of traffic accidents has not yet significantly reduced.

This study focuses on traffic and highway characteristics which cause accidents. It is generally known that there is a complex relationship between traffic and highway characteristics and the occurrence of traffic accidents. A factor or combination of many factors can cause traffic accidents. To overcome this complexity, accident prediction models are preferred as suitable tools. Accident prediction models are used to predict the number of accidents on highways. With these models, it is aimed to establish a relationship between a dependent variable (accident frequency) and a number of independent variables (annual average of daily traffic, vehicle kilometres of travel, numbers of pedestrian crossings, lane width, average posted speed, etc.).

This paper begins with a review of previous research which includes hitherto developing accident prediction models. This is followed by general information about the methodology and data collection. Model development and estimation results are then presented and finally, conclusions are evaluated as a summary.

## 2 Review of previous research

Many studies have been done to establish the relationship between traffic accidents and traffic and highway characteristics. At first, Satterthwaite [3] found a relationship between traffic accidents and traffic volume. This fundamental relationship has been acknowledged the beginning of accident prediction models. After then, highway characteristics have been added to analysis as well as traffic volume.

The most common methodological approaches used in accident prediction studies are Multiple Linear regression, Poisson regression and Negative Binomial regression. Generally, Multiple Linear regression modelling with its assumption of normally distributed errors and homoscedasticity was used in early studies. In recent studies, Poisson regression and Negative Binomial regression modelling techniques have been used widely. Because, Jovanis and Chang [4] found that Multiple Linear regression model gave erroneous statistical results when applied to accident analysis. They discovered that as vehicle miles of travel increases, so does the variance of accident frequency. This result clearly violates the homoscedasticity assumption of linear regression. They offered that Poisson regression is superior alternative to conventional linear regression for establishing the relationship between accident frequency and explanatory

variables. In other work, Miaou *et al.* [5] used Poisson regression modelling for establishing the relationship between truck accidents and highway geometrics. Joshua and Garber [6], Ivan and O'Mara [7], Greibe [8], Al-Ghirbal and Al-Ghamdi [9], Zhang and Zhirui [10] also applied Poisson regression for prediction of accident frequencies. Poisson regression model has desirable statistical properties for describing vehicle accidents, but it has an important constraint. The mean and variance of the accident data are constrained to be equal. If the variance is greater than the mean, overdispersion occurs in data. Overdispersion does not affect the coefficient estimates but does cause their standard errors to be underestimated [5]. To overcome this problem, the Negative Binomial regression model has been employed to analyse traffic accidents [11–20].

In summary, previous research has shown that Multiple Linear regression is not a suitable method for applying the analysis of accident frequency. Poisson and Negative Binomial regression are suitable tools for establishing accident prediction models.

### 3 Methodology

Since traffic accidents are sporadic, non-negative and discrete random events, it cannot be modelled with a normal distribution. In fact, the Poisson and Negative Binomial distributions are often more appropriate for discrete counts of events.

In this paper, the distributions of accident counts were assumed to follow the Poisson and Negative Binomial distributions. Poisson regression and Negative Binomial regression models which are developed the techniques of generalized linear model were considered to establish the relationship between traffic accidents and traffic and highway characteristics.

#### 3.1 Poisson regression model

In Poisson regression model which is assumed that the dependent variable  $y_i$ , the number of accidents in  $i^{\text{th}}$  highway segments during a period of time, follows the Poisson distribution with a parameter  $\mu_i$  which is the expected accident frequency (or number of accidents) for  $i^{\text{th}}$  highway section during a period of time.

$$P(y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \quad i = 0, 1, 2, \quad (1)$$

where,  $P(y_i)$  is the probability of  $y$  accidents occurring at  $i^{\text{th}}$  highway section during a period of time. In Poisson regression model, the expected accident frequency is assumed to be a function of explanatory variables such that

$$\mu_i = \exp(b_0 + b_1 * X_{i1} + b_2 * X_{i2} + \dots + b_q * X_{iq}) \quad (2)$$

where,  $X_{i1}, X_{i2}, \dots, X_{iq}$  are the explanatory variables which includes the traffic and highway characteristics of  $i^{\text{th}}$  highway section,  $b_0, b_1, b_2, \dots, b_q$  are the model

coefficients which are estimated by maximum likelihood methods [16, 21, 22]. The likelihood function is,

$$L(\mu_i) = \prod_{i=1}^n \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \quad (3)$$

A limitation of using Poisson regression model is that the variance of the data is constrained to be equal to the mean,  $E(y_i) = V(y_i) = \mu_i$ . In general, accident data is found to show overdispersion relative to Poisson model. To overcome overdispersion problem, accident data are assumed to be distributed by Negative Binomial distribution.

### 3.2 Negative Binomial regression model

The Negative Binomial regression model developed in this study has the following form [16],

$$P(y_i) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1) \Gamma(\frac{1}{\alpha})} \left( \frac{1}{1 + \alpha \mu_i} \right)^{1/\alpha} \left( \frac{\alpha \mu_i}{1 + \alpha \mu_i} \right)^{y_i} \quad (4)$$

$y_i = 1, 2, 3,$

where, the variance of  $y_i$  is [23],

$$\text{Var}(y_i) = E(y_i)[1 + \alpha E(y_i)] = \mu_i + \alpha \mu_i^2 \quad (5)$$

where  $\alpha$  is a measure of dispersion or overdispersion parameter and if  $\alpha$  is not significantly from zero, the Negative Binomial regression model is reduced to the Poisson regression model with  $\text{Var}(y_i) = E(y_i)$ .

As for Negative Binomial regression model, the coefficients,  $b_0, b_1, b_2, \dots, b_q$  and a measure of dispersion,  $\alpha$  are estimated by maximum likelihood methods [21, 22].

### 3.3 Goodness of fit test statistics

Goodness of fit test statistics are used to evaluate the accuracy of the selected model. To examine the goodness of fit, some statistical measures have been used. In this study, Pearson chi-square statistic (Pearson  $\chi^2$ ) is considered for goodness of fit test statistics.

Pearson chi-square statistic (Pearson  $\chi^2$ ) is calculated as [23]:

$$\text{Pearson } \chi^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\text{Var}(Y_i)} \quad (6)$$

## 4 Data

The data used to establish the accident prediction model consisted of two categories: Accident data and traffic and highway geometric data.

### 4.1 Accident data

Trabzon, which is a coastal city of Turkey, has a strategic importance by use of highway, airway and waterway transportation. But, in Trabzon, highway transportation is the most preferred transportation mode like in Turkey. Increased numbers of traffic accidents occur in each year (as can be seen in Table 1).

Table 1: Accident statistics of Trabzon per year.

| Years | Number of accidents | Number of fatalities | Number of injured persons |
|-------|---------------------|----------------------|---------------------------|
| 2002  | 2499                | 12                   | 1042                      |
| 2003  | 2247                | 21                   | 954                       |
| 2004  | 3019                | 20                   | 1052                      |
| 2005  | 3813                | 35                   | 1141                      |
| 2006  | 5007                | 31                   | 1427                      |
| 2007  | 5463                | 27                   | 1568                      |

Source: Trabzon Police Headquarters (2007) [24].

In this study, 5 years' (2002, 2003, 2004, 2006 and 2007) accident data of 113, 5 km road sections of Trabzon coastal divided highway were obtained from General Directorate of Turkish Highways (GDTH). Because of the accident statistics of year 2005 related to Trabzon coastal divided highway was not available, it was not added the analysis process. Each accident reports was analysed and all types of accidents (fatality, injured and property damage only accidents) were included in the accident prediction model. The total reported number of accidents that occurred on Trabzon coastal divided highway during 5 years was 3181 accidents.

### 4.2 Traffic and highway geometric data

The traffic and highway geometric data were also obtained from General Directorate of Turkish Highways (GDTH). Annual average daily traffic was considered as traffic volume and vehicle kilometres of travel, the number of pedestrian crossing, average posted speed, the number of lanes and the number of intersections were considered as highway geometric data for each highway section.



## 5 Model development and results

A total number of 3181 accident records related to Trabzon coastal divided highway were used to develop the accident prediction model. At first, the highway, whose length is 113.5 km, was divided into homogeneous highway sections. There is a total of 70 homogeneous road sections during the five years period. The section lengths vary from 0.5 to 37.5 km.

The mathematical model form that was used as follows:

$$E(A) = \exp(b_0 + b_1X_1 + b_2X_2 + \dots + b_qX_q) \quad (7)$$

where,

$E(A)$  = The number of predicted annual accidents

$X_i$  = Explanatory variables

$b_i$  = Model parameters

In modelling process, the number of annual accidents on each highway segments was considered as the dependent variable and annual average daily traffic, vehicle kilometres of travel, the number of pedestrian crossing, average posted speed, the number of lanes and the number of intersections as the explanatory variables (or the independent variables). Poisson regression modelling was used at first step of the modelling process. Overdispersion parameter,  $\alpha$  is found 1,217. Owing to overdispersion of the model, it was formed by Negative Binomial regression. All available modelling variables were tested to find the significant variables at 95% confidence level. As a result, Negative Binomial regression model was found the best model.

Table 2: Negative Binomial estimation results for annual accident frequency.

| n = 70    df=n-p= 66    Pearson $\chi^2 = 68.921$ $\chi^2_{0.05,66} = 95.631$ |             |                |         |         |
|---|-------------|----------------|---------|---------|
| Variable  | Coefficient | Standard error | z ratio | P >   z |
| cons  | 5.2019      | 0.8959         | 5.81    | 0.000   |
| lnL   | 0.5750      | 0.1630         | 3.53    | 0.000   |
| aps   | -0.0536     | 0.0159         | -3.37   | 0.001   |
| pc  | 0.0427      | 0.0200         | 2.13    | 0.033   |

From Table 2, the model was formed as:

$$E(A) = \exp(5.2019 + 0.5750 \cdot \ln L - 0.0536 \cdot \text{aps} + 0.0427 \cdot \text{pc}) \quad (8)$$

where,

|      |   |
|------|---|
| n    | = the number of observation                   |
| p    | = the number of parameter                     |
| df   | = degrees of freedom                          |
| cons | = constant                                    |
| lnL  | = natural log of vehicle kilometres of travel |
| aps  | = average posted speed                        |
| pc   | = the number of pedestrian crossing           |

Table 2 indicates that, constant, natural log of vehicle kilometres of travel, average posted speed and the number of pedestrian crossing are the significant variables at the 95% confidence level since p value is much smaller than 0.05. When the Pearson  $\chi^2$  is considered for goodness of fit, it seems that the model is suitable for the Negative Binomial regression model. Because,  $\chi^2_{0.05,66}$  (the value of critic  $\chi^2$ ) is greater than the Pearson  $\chi^2$ .

The positive coefficient of the vehicle kilometres of travel (lnL) indicates that as the vehicle kilometres of travel increases, so does the number of accidents. Similarly, the increase of the number of pedestrian crossing on highways causes the increase of traffic accidents. Because, the conflict points between vehicles and pedestrians occur at these places.

The other variable of the model is average posted speed (aps). In the model aps' coefficient was found negative. This indicates that, the numbers of traffic accidents that occur on rural highways are less than urban highways. The reason for the negative coefficient of aps is that rural highways have fewer conflict points and also high aps than urban highways.

## 6 Conclusion

The objective of this study was to develop the statistical model of the annual accident frequency for divided highways. The Poisson regression model and Negative Binomial regression model was proposed for establishing the relationships between traffic accidents and traffic and highway characteristics. The technique of generalized linear models (GLMs) was applied to the data. Because of overdispersion of Poisson regression model, Negative Binomial regression model was found the most appropriate approach for the analysis of these data.

In developed model, it was found that the highway segments with higher vehicle kilometres of travel and higher the number of pedestrian crossings would have more traffic accidents than fewer ones. The highway segments with high speed limits would have fewer accidents than the highway segments with low speed limits. For road safety, the number of pedestrian crossings must be reduced on highways and they can be transformed to footbridges or underpasses. Besides, in urban highways which have more accidents than rural highways, the number of accidents can be reduced by strict traffic controls.

For future studies, it is recommended that the model can be developed by more explanatory variables such as other highway geometric features (curves,

slope and ext.), environmental features, driver characteristics, socio-economic and demographic data. In modelling process, different methods such as neural networks can be used to see if the prediction performance could be improved.

## References

- [1] WHO, Prevention of traffic accidents on highways: World report 1 Summary, World Health Organization, 2004.
- [2] GDTH, <http://www.kgm.gov.tr>
- [3] Satterwaite, S., P., *A Survey of Research into the Relationships Between Traffic Accidents and Traffic Volumes*, Transport and Road Research Laboratory, TRRL, Supplementary Report 692, United Kingdom, 1981.
- [4] Jovanis, P.P. & Chang, H.L., Modeling the relationship of accidents to miles travelled, *Transportation Research Record 1068*, Transportation Research Board, National Research Council, pp. 42-51, 1986.
- [5] Miaou, S-P., Hu, P.S., Wright, T., Rathi, A.K. & Davis, S.C., Relationship between truck accidents and geometric design: A Poisson regression approach, *Transportation Research Record 1376*, Transportation Research Board, National Research Council, pp. 10-18, 1992.
- [6] Joshua, S. & Garber, N., Estimating truck accident rate and involvement using linear and Poisson regression models, *Transportation Planning and Technology*, 15(1), pp. 41-58, 1990.
- [7] Ivan, J. & O'Mara, P., Prediction of traffic accident rates using Poisson regression. *Presented at the 76<sup>th</sup> annual meeting of the transportation research board*, Washington D.C., paper no. 970861, 1997.
- [8] Greibe, P., Accident prediction models for urban roads, *Accident Analysis and Prevention*, 35(2), pp. 273-285, 2003.
- [9] Al-Ghirbal, A. S. ve Al-Ghamdi, A. S., Predicting severe accidents rates at roundabouts using Poisson distribution, *TRB Annual Meeting*, TRB paper 06-1684, 2006.
- [10] Zhang, Y. & Zhirui, Y., Investigating goodness-of-fit test statistics for generalized linear crash models with low mean values, *Presented at the 86th Annual Meeting of the Transportation Research Board*, TRB, Washington D. C., paper no. 07-2640, 2007.
- [11] Miaou, S.P., The relationship between truck accidents and geometric design of road sections: Poisson versus Negative Binomial regressions, *Accident Analysis and Prevention*, 26(4), pp. 471-482, 1994.
- [12] Maher, M. J. & Summersgill, I., A comprehensive methodology for the fitting of predictive accident models, *Accident Analysis and Prevention*, 28 (3), pp. 281-296, 1996.
- [13] Poch, M. & Mannering, F., Negative Binomial analysis of intersection-accident frequencies, *Journal of Transportation Engineering*, 122(2), pp. 105-113, 1996.
- [14] Milton, J. C. & Mannering, F. L., The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies, *Transportation*, 25, pp. 395-413, 1998.





- [15] Abdel-Aty, M.A. & Radwan, A. E., Modelling traffic accident occurrence and involvement, *Accident Analysis and Prevention*, 32(5), pp. 633–642, 2000.
- [16] Hadayeghi, A., *Accident prediction models for safety evaluation of urban safety transportation*, the thesis of Master of Science, Civil engineering department of Toronto University, Canada, 2002.
- [17] Chang, L.Y., Analysis of freeway accident frequencies: Negative Binomial regression versus artificial neural network. *Safety Science*, 43(8), pp. 541–557, 2005.
- [18] Hadayeghi, A., Shalaby, A. S., Persaud, B. N. & Cheung, C., Temporal transferability and updating of zonal level accident prediction models, *Accident Analysis and Prevention*, 38(3), pp. 579-589, 2006.
- [19] Caliendo, C., Guida, M. & Parisi, A., A crash prediction model for multilane roads, *Accident Analysis and Prevention*, 39(4), pp. 657-670, 2007.
- [20] Ramírez, B. A., Izquierdo, F. A., Fernández, C. G. & Méndez, A. G., The influence of heavy goods vehicle traffic on accidents on different types of Spanish interurban roads, *Accident Analysis and Prevention*, 41(1), pp.15-24, 2009.
- [21] Cramer, J., S. *Economic applications of maximum likelihood methods*, Cambridge University press: New York, 1986.
- [22] Lawless, J. F., Negative Binomial and mixed Poisson regression, *The Canadian Journal of Statistics*, 15(3), pp. 209-225, 1987.
- [23] McCullagh, P. & Nelder, J.A., *Generalized Linear Models*, 2<sup>nd</sup> edition, Chapman and Hall, London, 1989.
- [24] Trabzon Police Headquarters, Trabzon, Turkey, <http://www.trabzon.pol.tr/>