An Overview of Crash Prediction Models and Their Suitability in Black Spot Identification

Mohit Nandal*[1], Navdeep Mor[2], and Dr. Hemant Sood[3]

[1]M.Tech Student, Civil Engg. Department, NITTTR, Chandigarh

[2] Research Scholar, Civil Engg. Department, NITTTR, Chandigarh

[3] Professor, Civil Engg. Department, NITTTR, Chandigarh

Abstract: Road accident is one of the major problems for most communities in developing countries, where 90% of the world road accident fatalities occur (WHO), which require serious attention in searching for preventive measures to minimize this problem. Next to human suffering, traffic accidents also result in huge social and economic costs. Road accidents can be the consequence of different factors like environmental, human, vehicle, and road factors. The most important factor is the experience and the alertness of the road user. The road characteristics and features of the immediate environment are a second important factor. Evaluation of road safety measures appears to be the weakest component of road safety management systems. To improve Road Infrastructure Safety Management, road authorities, road designers and road safety practitioners need prediction tools, commonly known as Accident Prediction Models (APMs), allowing them to analyse the potential safety issues, to identify safety improvements and to estimate the potential effect of these improvements in terms of crash reduction. This study aims to review of the development of crash prediction models, and their applications to analyse and identify black spots to improve road safety. Several modelling techniques have been reviewed in this study including, multiple linear regression, Poisson distribution, negative binomial, random effect technique, and multiple logistic regression models to identify their suitability to develop the crash prediction models. The models identified in this research are already being used but the modelling approaches can be further modified to include the latest technical application on roads, available post-crash management system or safety culture which are commonly related road safety outcomes.

IndexTerms **– Road Accident, APM, Black Spot mangement**

1. INTRODUCTION

Road accident is found to have a growing pattern with unpredictable level of seriousness mainly affected by road features, road users, characteristics and environmental factors. About 3.4 individuals killed in the roads consistently. A huge number of individuals are killed or injured each year. Kids, people on foot, cyclists and the elderly are the most victims of street accidents. According to World Road Statistics 2015 Report published by International Road Federation, Geneva, there is lower rate of deaths per 100,000 in the countries Australia, Canada, France, Germany, Japan, Republic of Korea, U.S.A, Poland, Portugal etc. except Russian Federation while paralleling it with India. This number of fatalities and injuries has a huge impact on the families affected, whose lives are often changed irrevocably due to these tragedies. Road crashes cannot be completely prevented, but through appropriate traffic engineering remedial actions and management approaches the crash rates can be reduced to certain acceptable societal limitations. Proper verification of the cause of crashes can help to suggest preventive and corrective measures in terms of traffic control and road design at potential crash locations, so for this reason, systematic studies of traffic crashes are required to be undertaken.

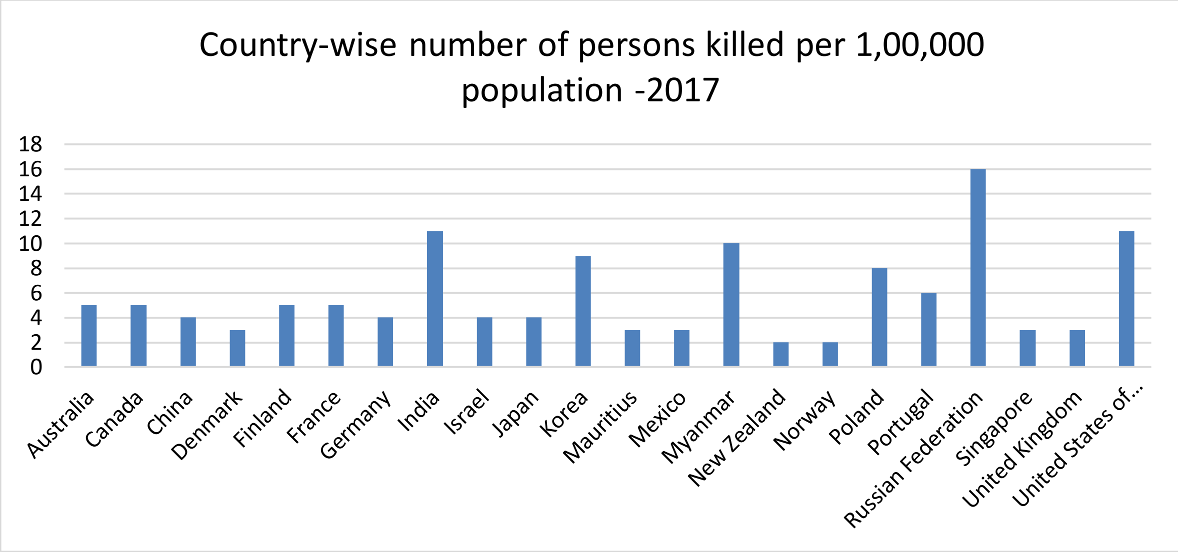


Figure 1 Country-wise number of persons killed per 1, 00,000 population

Traffic safety agencies in the past used various key measures such as the crash rate to traffic volume and/or the absolute crashes number at a location, to see whether that location had a traffic safety problem or not. Later on, the comparison with other locations of similar traffic and geometric conditions helped to improve the investigations. Thus, traditionally focuses have given on predicting the relation between the traffic crash frequency and other contributing variables which then propose strategies to alter the shape of traffic volume and road geometry. However, these methods tend to be subjective, short sighted, and an outdated view on road safety. The main reason for this could be the availability of data for developing the models. However the systematic integration of crash data with safety policy and focuses on the capability of collected information to meet the requirements to develop the remedial actions helped the development of new modelling approaches.

Road crashes occurs due to combination of several factors which includes the roadway, vehicle behaviour, roadway environment and road user's behaviour. Black Spot is a location on road network which is prone to road traffic accidents and a large number of crashes are taking place at that location. It may occur due to number of reasons like hidden junction on a fast road, improper locations of warning signs, sharp curve in a straight road etc. Black Spot can be identified using accidental data for that particular location. Black Spot Management (BSM) is a systematic approach for treatment, identification, evaluation and analysis of countermeasures used at hazardous locations on the road network. The Black Spot Identification involves statistical data analysis of accident record and mapping of accident locations. The BSM will decrease the risk of accidents at the treated locations.

2. DATA ANALYSIS

2.1 Black Spot Identification

Accident data collected from the police divisions are used for formulation of black spot identification. The following three indices (Accident Rate, Accident Frequency and Accident Severity) and corresponding critical values are used to identify the black spots. It contain three estimators which estimate the accident hot spots where the places highest number of accidents likely to be occurred. The used formulas are shown below.

i. **Accident Rate,**

$$R_j = \frac{A_j}{m_j}$$

Where,

Aj : Number of accidents on segment j during a certain time period

m_j : Number of vehicle kilometres in millions on segment j during the same time period

$$R_c = \lambda + K\alpha\sqrt{\frac{\lambda}{mj}} - \frac{0.5}{mj}$$

$$\lambda = \frac{\sum_{i=1}^{n} Ai}{\sum_{i=1}^{n} mi}$$

Where,

R_c : critical value for accident rate

λ : estimated average rate for segment belonging to same population

($0.5/m_j$): correction for continuity when approximating with normal distribution

K_a: constant that is chosen for significance test

Segment j is considered to be black spot from accident point of view, if $R_j > R_c$

ii. Accident Frequency

$$A_c = F_{ave} + K\alpha\sqrt{\frac{F_{ave}}{L_j}} - \frac{0.5}{L_j}$$

$$F_{ave} = \frac{\sum_{i=1}^{n} Ai}{\sum_{i=1}^{n} Li}$$

Where,

A_c : critical value for critical frequency

L_j : length of segment j

F_{ave} : average accident frequency for all road segments

Section j is considered to be a black spot from accident frequency point of view, if $A_j > A_c$

iii. Accident Severity

$$S_j = Fatalities * 9 + Injured\ persons * 3 + Damage\ vehicles * 1$$

$$S_c = \frac{\sum_{i=1}^{n} S_j}{n}$$

Where,

S_c : critical value for accident severity

Segment j is considered to be a black spot from accident severity point of view, if $S_j > S_c$

Selection criteria for a Black Spot is to satisfy all the three indices must satisfy the following conditions simultaneously:

Accident Rate : $R_j > R_c$

Accident Frequency: $A_j > A_c$

Accident Severity : $S_j > S_c$

In most studies regarding the identification of high-risk road locations two phases can be distinguished. A safety indicator is calculated in a first phase. In a second phase it can be investigated whether the value of this safety indicator significantly exceeds a threshold value. If this is the case, one may speak of a hazardous site. On the basis of these two phases, a great variety of possibilities to identify high-risk locations exists, ranging from relatively simple methods to more complex methodologies. Generally it can be stated that the required amount of data is likely to increase with the increasing complexity of the method.

1. **Safety indicator:** The starting point of black spot management is the calculation of a safety indicator that describes the level of safety of a road site. Usually this is done on the basis of the number of observed accidents during a specific period. Different indicators are often combined and used in accident analyses. Some of the most common safety indicators are accident rate, accident density, number of accidents.

2. **Black Spot Identification:** After the calculation of the safety indicator, one may proceed to the second phase of the process: the assessment of this indicator and thus the identification of the most dangerous road locations. For this purpose, there also exist various approaches. These can be subdivided into two categories, namely on the basis of a threshold value or on the basis of a value that is normal for similar sites.

2.1.2 Black Spot Management

Black Spot Management (BSM**)** is a systematic approach for treatment, identification, evaluation and analysis of countermeasures used at hazardous locations on the road network. The Black Spot Identification involves statistical data analysis of accident record and mapping of accident locations. The BSM will decrease the risk of accidents at the treated locations. Black spot management or black spot safety work can be described as "the task of improving road safety through alterations of the geometrical and environmental characteristics of the problematic sites or locations in the existing road network". In other words, the purpose of black spot management is to identify and remediate those intersections or road locations where the number of accidents is unusually high, or the so-called black spots. This section describes how these black spots are identified in the literature, and if there is any consensus about a preferred methodology for this identification. The methods used for black spot management are given below:

1. Traditional black spot management (BSM)

Black spot identification is based on a recorded accident frequency in a sections defined by sliding window of 250 m length, which meets a condition of at least three injury accidents in one year. Variables used are number of black spots and total accident frequency at these spots.

2. Empirical Bayes approach (EB)

The third approach, the empirical Bayes methodology, is based on an accident model and is currently being considered as the state of the art concerning the identification of dangerous links in the road network (Elvik, 2008a). When following the empirical Bayes method, the best prediction of the expected number of accidents can be obtained by combining the recorded number of accidents and the normal number of accidents, as determined by the accident prediction model. Accident prediction model of frequency of all accidents (both property-damage-only and injury) is obtained using explanatory variables AADT, length and CCR.

3. Preliminary road safety inspection (RSI)

Preliminary RSI is based on data collected by instrumented vehicle. During driving through the studied network, road safety auditor records the presence of risk factors, which may potentially increase accident occurrence and severity. For each risk factor, following features are evaluated i.e. risk type, forgiveness, and pavement quality.

2.1.3 Calculation of Safety Potential

The calculation of the safety potential is done using the following accident parameters: annual average accident cost, accident density, accident cost density, accident rate, accident cost rate, and basic accident cost rate. Then, the spots of the road network are ranked on the basis of the magnitude of the safety potential. Such ranking is of great use to further detailed studies so as to determine possible improvement measures. The higher the safety potential, the more societal benefits can be expected from improvements to the roads.

Table 1 Accident Parameters for determination of safety potential

Parameters	Formulas
AC_a	$AC_a = \frac{\sum_{i=1}^{4} nA(c_i) \times MCA(c_i)}{t}$ Where: AC_a : annual average accident cost [USD/year] $nA(c_i)$: number of accidents of specific accident category c_i, in $t \geq 3$ years $MCA(c_i)$: mean cost per accident of accident category c_i [USD/accident] t : period of time under review [years]
AD	$AD = nA/t$ Where: AD : accident density nA : number of accidents t : time period [years
ACD	$ACD = \frac{AC}{1000.t}$ Where: ACD : accident cost density [1000 USD/year] AC : accident cost t : time period [years]
AR	$AR = 10^6.nA/(365.AADT.t)$ Where: AR : accident rate nA : number of accidents AADT : average annual daily traffic t : time period [years]
ACR	$ACR = 1000.AC/(365.AADT.t)$ Where: ACR : accident cost rate [USD/1000.veh] AC : accident cost AADT : average annual daily traffic t : time period [years]
bACD	$bACD = \frac{bACR \times ADT \times 365}{10^6}$ Where: bACD : basic accident cost density [1000 USD/year] bACR : basic accident cost rate ADT : average daily traffic
SAPO	$SAPO = ACD - bACD$ Where: SAPO : safety potential [1000 USD/year] ACD : accident cost density bACD : basic accident cost density

bACR	The basic accident cost rate (bACR) can be defined for many different types of roads and intersections which are derived from the detailed assessment of existing accident cost rates.

3. Accident Prediction Models (APMs)

Accident Prediction Models can be used for gaining more perception into the relationships between road and traffic conditions and number of accidents and estimating expected number of accidents on a particular road segment more reliably than by just counting the observed number of police reported accidents. Road crashes are caused by a combination of many factors, including the roadway, the roadway environment, vehicles and road users' behaviour. Crash Prediction Models (CPMs) have been employed as useful tools by road engineers and planners to identify the reasons hence to propose remedial actions to improve road safety. In this process, several modelling techniques have been used in crash prediction models including, multiple linear regression, Poisson distribution, negative binomial, random effect technique, and multiple logistic regression models. The models are discussed below:

3.1 Poisson Distribution Model

The vehicle arrivals can be modelled in two inter-related ways; namely modelling how many vehicle arrive in a given interval of time, or modelling what is the time interval between the successive arrivals of vehicles. Traditionally, Poisson distribution is used to model the random process, the number of vehicles arriving a given time period. Since crash occurrences are unavoidable, discrete and more likely random events, So many researchers argue that Poisson regression model appears to be more appropriate than multiple linear regression models. The study found that additional explanatory variables including speed limit, road environment, parking facilities, number of minor side roads and number of exits per km proved to be important and significant explanatory variables for estimating the number of crashes in road segments. Poisson models have some limitations and one of these limitations is that the mean must equal the variance of road crash number (dependent variable). In most crash data, the variance value of the road crash number exceeds the mean value and, in such case, the data would be over dispersed.

3.2 Multiple Linear Regression Model

The multiple linear regressions modeling technique has been used in several fields including engineering, health, agriculture and economics, to model the relationship between two or more explanatory variables and an out- come variable by fitting a linear equation to observed data. Detailed the creditability of the multiple linear regression models to describe relationships between continuous outcomes and explanatory variables. Although multiple linear regression models are used widely in road crash studies, they have limitations to describe adequately the random, non-negative, discrete, and typically sporadic events, which are all characteristics of road crashes.

3.3 Negative Binomial Model

Negative Binomial (NB) regression method is considered as an alternative which does not require the equal mean and variance assumption. The models suggested are also based on the generalized linear modelling (GLM) approach, assuming a negative binomial distribution error structure. In conclusion, three of the examined models were considered appropriate, based on practical considerations, statistical significance, and goodness of fit indicators. It is also apparent that for some special purposes the Poisson distribution is still a fairly good representation in those cases where the negative binomial is suspected or known to apply.

3.4 Random Effect Model

The random effect technique assumes that road crash data is hierarchical in nature. The hierarchy in road crash data is proposed as follows: the lowest level of the hierarchy represents the crashes themselves, while the type of location on road network at which the crash occurred represents the higher-level hierarchy. In this type of model, the main assumption is that an association may exist among crashes occurring at the same location, so these crashes may share unobserved or unrecorded characteristics related to the location. Over the last decade, this hierarchical modeling technique has been gaining an increasing amount of attention in accounting for the multilevel data structure in crash prediction. These unobserved characteristics might include low pavement friction, poor pavement condition, poor reflectivity of road signs, and other similar factors. The results from this technique may not be transferable to other data sets because the results are observation specific.

3.5 Multiple Logistic Modeling

The multiple logistic regression technique is used to analyze only crash binary outcomes, meaning the value of the dependent variable ranges between 0 and 1. For example, this technique can be used to build a model to provide a measure of the probability of injury or non-injury crash outcomes. Crashes occurring at the same location may share unobserved or unrecorded characteristics related to the location.

4. CONCLUSION

This study comprehensively reviewed the simplified statistical models used for predicting road crashes and hence their application to identify the black spot locations to improve the road safety in urban and rural roads. Several studies show that the relationship between road crash frequency and explanatory variables are having a very good relationship. This finding has further led most road safety researchers to use statistical models in which the dependent variable is the crash frequency. The statistical methods such as Poisson and Negative binomial regression have traditionally been used as suitable techniques used for developing road crash models. This is due tothe ability of these techniques to analyses data while preventing the possibility of having a negative integer crash value over some time period.

Applying a suitable crash prediction model is very important to engineers and transportation planner, because it can help in identifying the black spot locations that require treatment and as well as ranking the hazardous locations by calculating Potential for Safety Improvement (PSI). In general, the CPMs and observed crashes do not account for Regression to mean effect (RTM) associated with crash data. RTM is the tendency of crash data to regress back to the mean. For instance, a site may have high crashes at a given period and low crashes the next period without any road safety implementations. In addition, a high-risk site may have a certain period of randomly low frequency crashes and therefore be over- looked during road safety evaluation. The Empirical Bayes (EB) approach has been introduced by researchers as a means of solving the RTM problem. This approach identifies high crash locations (black spots) based on their Potential for Safety Improvement (PSI), calculated as the difference between predicted and expected crashes at the location.

Safety potential is the difference between the actual accident cost and the expected accident cost conforming to the best-practice design standard. The expected accident cost depends on the basic accident cost rate. In ideal circumstances this expected accident cost contains no influence of the infrastructure on the accidents any more but represents the accident cost caused only by the other two components of the transport system –vehicle and road users. The best way to estimate the target values would be to calculate the accident cost rate for a sample of spots with best practice design. Another possibility would be to use a specific percentile of the overall distribution of the accident cost rates. Also statistical tests must be done to make sure that the random variation is not the decisive factor in the process of identifying the high accident frequency locations. Random variation in accident count at the identified locations can be estimated by Poisson probability distribution.

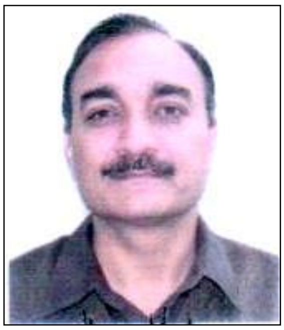
REFERENCES

1. Al-Marafi, Mohammad, and Kathirgamalingam Somasundaraswaran. "Review of crash prediction models and their applicability in black spot identification to improve road safety." Indian Journal of Science and Technology 11, no. 5 (2018): 1-7.
2. Greibe P. Accident prediction models for urban roads. Accident Analysis & Prevention. 2003 Mar; 35(2):273-85. Crossref.
3. Hauer E, Ng JC, Lovell J. Estimation of safety at signalized intersections. Transportation Research Record Journal. 1988; 1185:48-61.
4. Mountain L, Fawaz B, Jarrett D. Accident prediction mod- els for roads with minor junctions. Accident Analysis & Prevention. 1996 Apr; 28(6):695-707. Crossref.
5. Kutner MH, Nachtsheim CJ, Neter J, Li W. New York, McGraw-Hill Irwin: Applied linear statistical models, 4th edn. 2005; p. 1-1415.
6. Chin HC, Quddus MA. Applying the random effect nega- tive binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis & Prevention. 2003 Mar; 35(2):253-9.
7. Nguyen, Huy Huu, Pichai Taneerananon, and Paramet Luathep. "Approach to identifying black spots based on potential saving in accident costs." *Engineering Journal* 20, no. 2 (2016): 109-122.
8. Ambros, Jiří, Pavel Havránek, Veronika Valentová, Zuzana Křivánková, and Radim Striegler. "Identification of hazardous locations in regional road network–comparison of reactive and proactive approaches." *Transportation research procedia* 14 (2016): 4209-4217.
9. Sivakumar, Thillaiampalam, and Dinusha Amarathung. "Development of Traffic Accident Prediction Models Using Traffic and Road Characteristics: A Case Study from Sri Lanka." In *EASTS 2015 Conference.*

Author's Profile

1* Er. Mohit Nandal

2* Er. Navdeep Mor



3* Dr. Hemant Sood