An Overview of Challenges and Opportunities in Development and Use of Accident Prediction Models

Navdeep Mor[1*], Dr. Hemant Sood[2,] and Dr. Tripta Goyal[3]
[1]Research Scholar, Civil Engg. Department, NITTTR, Chandigarh
[2]Professor, Civil Engg. Department, NITTTR, Chandigarh
[3] Professor, CED, PEC (Deemed to be University), Chandigarh.

Abstract: Road crash prediction models are very useful tools in highway safety, given their potential for determining both the crash frequency occurrence and the degree severity of crashes. However, there is a gap between state-of-the-art and state-of-the-practice, with the practical application lagging behind scientific progress. This motivated a review of international experience with CPMs from perspectives of application by practitioners and development by researchers. The objective of the paper is to improve practitioner understanding of modelling road safety performance using CPMs for crash frequency estimation, leading to their greater uptake in improving road safety. In short, why and how should road safety practitioners consider CPMs?

From the review it is clear that developing CPMs is not a straightforward task: there are many available choices and decisions to be made during the process without definite guidance. This explains the diversity of approaches, techniques, and model types. This paper presents an overview of road crash prediction models used by transportation agencies and researchers to gain a better understanding of the techniques used in predicting road accidents and the risk factors that contribute to crash occurrence.

***Index Terms* – Accident Prediction Model, Road accidents, user safety.**

The developed world has become aware of the scale of the road safety problem and tried to remedy it. This has been successful, shown by the fall of death rate. The same doesn't apply to developing countries, where the number of road accidents has not stopped increasing. The forecasts are unfortunately pessimistic. The authorities of the majority of these countries do not provide sufficient measures to fight against this phenomenon.

According to the WHO, road traffic accidents are the number one cause of death through injury in the world, the 10th leading cause of death from any cause and the ninth largest contributor to the "burden of disease (BOD) 1 ". In 2001, the WHO estimated that there were more than one million road traffic accident fatalities, equivalent to approximately 3,000 people per day2. We can therefore effectively talk of carnage. In view of the predicted growth in population, urbanization and above all motorization, the projections for the future are grim: the WHO anticipates that, on a worldwide level, road traffic injuries will become the 3rd leading contributor to the BOD by 2020, behind only cardio-vascular accidents and cancer. The greater part of this growth will occur in the developing world (poor and transition countries) as the numbers of deaths in the majority of developed countries have been falling over the last few decades, thanks to the adoption of courageous policies. In developing countries, on the other hand, in spite of a much lower levels of motorization we note that crude death rates continue to rise and are thus already as high as in the rich countries.

Factors Affecting Road Traffic Accidents

A traffic accident may have many contributing factors, such as those related to driver behavior, road geometry, traffic volumes, vehicle, and environment. The influence of such variables on crash occurrence could significantly vary on a case-by-case basis, but in general, both behavioral factors related to the driver's errors, and non-behavioral factors

Related to road geometry, traffic flow conditions, vehicle, and environment are thought to significantly affect traffic crashes.

Problem of Perception of the Scale of the Problem

The traditional view of the accident as a "chance event" has led to the problem being neglected until very recently in the developed countries. During these last few decades, there has been a realization in these same countries, which has led to some extent to the reduction in the number of deaths recorded. Despite the improvements in the figures, we note that the problem is still perceived by some as being minor, while it is in fact a major public health challenge. This low awareness would appear to be due to various factors:

1. In people's minds, an accident is something that occurs randomly, that is a hypothetical danger.
2. Since the first death due to collision with a motor vehicle (1898), road traffic accidents have become routine. As noted by TJ Coats, the number of road traffic accident deaths in the United Kingdom, a country which is nevertheless considered to have a good road safety record, is equivalent to a Lockerbie plane crash once every four weeks.
3. The media bear some responsibility for the pervading ignorance. In spite of the alarming statistics, road traffic accidents rarely made the headlines of the newspapers.
4. Road safety is also in competition with much stronger economic interests that sometimes prevent courageous decisions from being taken: Drivers must arrive on time, alcohol must be drunk, sports cars must be sold, etc.
5. In our Judeo-Christian world, some people quickly tend to take shortcuts and only blame the "culprit" (he died because he was drunk), ignoring all the other factors that might have allowed the accident to be avoided, or limited its consequences.
6. Finally, data is sometimes lacking, which prevents an effective case from being mad.

Haddon matrix

The Haddon matrix that presents the causal factors in their epidemiological and temporal dimensions, is commonly used to analyse causes and determine possible types of action.

Table 1: Haddon Matrix for user safety

Factors/ Phase	Human factors	Vehicle	Physical environment	Socio-cultural environment
Pre-event	Poverty, large gap between rich and poor, with the result that the poor have no other choice but to use hazardous public transport Carriers paid according to performance. They must transport as much as possible in as little time as possible Drivers poorly trained for driving on the road (driving licence fraudulently acquired, etc.) Low level of literacy Fatalism	Old, poorly maintained vehicles (brakes, lights, etc.) Large proportion of two wheeled vehicles compared to cars. Numerous pedestrians and public transport users Poorly regulated public transport (too many per vehicle, unscheduled stops)	Lack of road markings Poor lighting In some large cities: the road network is saturated (e.g., Lagos). Deficiency of alternative transport network (railway network)	Corruption, nepotism, difficulty enforcing the law (one of the traditional investments who are quickly successful is to buy busses). These self-made men are often above the law and their vehicles enjoy a certain impunity) No strong social pressure to stop drink driving and speeding
Event	Co-morbid conditions (AIDS, Tuberculosis, etc.)	Vehicles not in accordance with standards (no working seat belts, no air-bags, dangerous customisations) ; poor maintenance (brakes) Average number of passengers per vehicle	Condition of road surface	

		(There are always more than five deaths when a bus is involved in an accident)		
Post event	Impossibility for some to pay for high-quality care		Mediocrity or lack of emergency medical/evacuation service Evacuation by people with no first-aid training Poor health cover Obsolete, unsuitable equipment, shortage of medicine (blood), poor quality hospital infrastructure Little special training in medico surgical emergency treatment	Fatalism, esotericism

Crash Prediction Models

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. Over the last two decades, substantial research has been conducted on the development of CPMs for the estimation of the predicted crashes or crash rates on road network. Traffic accidents prediction models are very useful tools in highway safety, given their potential for determining both the frequency of accident occurrence and the contributing factors that could then be addressed by transportation policies. Vehicular crash data can be used to model both the frequency of crash occurrence and the degree of crash severity. Crash frequency refers to the prediction of the number of crashes that would occur on a specific road segment or intersection in a time period Crash severity methods generally explore the relationship between crash severity injury categories and contributing factors such as driver behavior, vehicle characteristics, roadway geometry, and road-environment conditions. Traffic accident related-fatalities and injuries can be prevented or at least minimized by a joint involvement from multiple sectors (i.e. transportation agencies, police, health departments, education institutions) that oversee road safety, vehicles, and the drivers themselves. Effective interventions include design of safer infrastructure and incorporation of road safety features into landuse and transport planning; improvement of vehicle safety features; improvement of post-crash care for victims of road crashes, and improvement of driver behavior, such as setting and enforcing laws relating to key risk factors, and raising public awareness.

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.

Crash Prediction Models (CPMs) and their uses

In Nutshell, CPMs may be used to accomplish various road safety management functions, such as:

1. Exploring and comparing combinations of individual risk factors that make some road locations unsafe.
2. Network safety screening, i.e. safety ranking road locations, or identification of hazardous locations.
3. Impact assessments, i.e. assessing safety of contemplated (re)constructions or safety treatments.
4. Economic analysis of project costs vs. safety benefits.

It is to be noted that point 1 is rather research-oriented; point 2, 3 and 4 represent typical practical tasks undertaken by many road agencies.

Regarding the selection of research for inclusion in the review, another distinction needs to be made. HSM introduces a set of CPMs (referred to as safety performance functions, SPFs) and crash modification factors (CMFs). Crash prediction in the HSM has two main steps:

(1) Prediction of a baseline crash rates using SPFs/ CPMs for nominal route and intersection conditions.

(2) Multiplying the 'baseline' models by crash modification factors (CMFs) to capture changes in geometric design and operational characteristics (deviations from nominal conditions). This approach has gained popularity, being incorporated into Interactive Highway Safety Design Model (IHSDM), and recently adopted in the European CPM.

Multiple Linear Regression Models

Multiple refers to many explanatory variables. Explanatory variables are characteristics whose effect on the outcome is being assessed. Multiple linear regression is a statistical methodology describing relationships between a continuous outcome and a set of explanatory variables (Kutner et al., 2005).

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 developed multiple linear regression models to investigate the effect of the roundabout geometry features on road crashes in urban and rural areas.

Poisson and Negative Binomial Models

Since accident occurrences are unavoidably discrete and more likely random events, the family of Poisson regression models appears to be more suitable than multiple linear regression models. The Poisson regression model is used when discrete response variables have counts as possible outcomes, for example, the number of accidents. However, Poisson models have potential problems; one constraint is that the mean must equal to the variance. If this assumption is not valid, that is, the accident data are significantly over dispersed (the variance is much greater than the mean), the standard errors usually estimated by the maximum likelihood method, will be biased and the test statistics derived from the model will be incorrect. This results in incorrect estimation of the likelihood of accident occurrence (Chin and Quddus, 2003).

To solve the problem of overdispersion, the negative binomial distribution has been employed instead of the Poisson. To establish the negative binomial regression model, an over dispersion parameter is introduced into the relationship of the mean and the variance. By relaxing the condition of mean equal to variance, the negative binomial regression models have more desirable properties than Poisson models to describe the relationship between accident occurrence and road characteristics (Chin and Quddus, 2003). Hence, in the successive sections Poisson and negative binomial regression models are presented as a more credible alternative to multiple linear regression analysis. The majority of the authors presented results from only the negative binomial models although the Poisson model was also fitted.

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. However, there are many studies in which crash outcomes are continuous (e.g., number of total crashes).

Random Effects Models

Models presented in the preceding sections (multiple logistic regression, multiple linear regression, negative binomial and Poisson regression) assume independent residuals across the number of accidents. These models are to some extent problematic to estimate when the data structure is characterized by correlated responses within clusters (intersections). The correlation within clusters violates the assumption of residual independence made by earlier statistical methods. Due to serial correlation in the accident data, non-hierarchical models seem to be inappropriate since accident data variables are likely to have location specific effects. Further, if significant correlation within clusters is not modelled, the consequence is attenuation of effects (parameter estimates tend toward zero), biased parameter estimates, under-estimated standard errors and incorrect statistical inferences. To overcome these problems, a more suitable alternative is random effects models which account for correlation within clusters by introducing random effects in the population based models (Kim et al., 2007). As a result, we describe random effects models in this section.

Conclusion

Greater uptake of state-of-the-art analytical techniques is necessary for continuing improvement in road safety. This study indicates that the traditional methods have now been replaced by more advanced modelling techniques to support the analysis for developing innovative counter measures to improve road safety. The system of data collection has also been flexed to meet the systematic integration of the data with the road safety strategies and policies. However, the future domain needs to break the barriers in providing additional information such as available advanced technology and communication, reliability of post-crash management system, and culture of road safety to the location as some of the key contributory factors for future studies.

This paper aimed to improve practitioner understanding of modelling road safety performance using CPMs, so that this useful analytical technique could become more accessible. The main consideration for the researches should be application of their models by intended practitioners. This applies equally in the context of basic research, such as seeking understanding of a new challenge, as in the context of applied research such as development of algorithms for inclusion in practitioner software. Either way the end users of CPMs are the practitioners, i.e. road agency engineers, policy makers, or data analysts. The review aimed to improve practitioner understanding of CPMs to bolster their use in improving road safety.

Table 2. Advantages and Disadvantages of each model

Model type	Advantages	Disadvantages
Poisson	Most basic model; simple to estimate	negatively influenced by the low sample-mean and small sample size bias; Cannot handle over- and under-dispersion
Negative Binomial/ Poisson-gamma	simple to estimate and can account for over dispersion	can be adversely influenced by the low sample-mean and small sample size bias; Cannot handle under-dispersion
Poisson-lognormal	More flexible than the Poisson-gamma to handle over-dispersion	can be adversely influenced by the low sample-mean and small sample size
Zero-inflated Poisson and Zero-inflated negative binomial	handle datasets that have a large number of zero-crash notes	bias cannot estimate a varying dispersion parameter; Cannot handle under dispersion
Artificial Neural Networks and Fuzzy Logic models	First check really ANN is required to model the given problem. The model need not assume any model structure before starting the ANN model. It can be used for non-linear problems. ϖ It is a non-parametric method, thus	It cannot extrapolate the results

	eliminates the error in parameter estimation.	
Conway–Maxwell–Poisson	be able to handle under- and over dispersion or combination of both using a variable dispersion parameter	No multivariate extensions available to date; Could be negatively influenced by the low sample-mean and small sample size bias
Multiple Linear Regression	Ability to determine the relative influence of one or more predictor variables to the criterion value. Ability to identify outliers, or anomalies.	Outputs of regression can lie outside of the range [0, 1]. It has limitations in the shapes that linear models can assume over long ranges. The extrapolation properties will be possibly poor. It is very sensitive to outliers It often gives optimal estimates of the unknown parameters
Negative multinomial	Can account for over-dispersion and serial correlation; panel count data	Cannot handle under-dispersion; can be adversely influenced by the low sample-mean and small sample size bias
Zero-inflated Poisson and Zero-inflated negative binomial	handle datasets that have a large number of zero-crash notes	zero-inflated negative binomial can be adversely influenced by the low sample-mean and small sample size bias; Can create theoretical inconsistencies
Poisson – Weibull	It account for over-dispersion	Cannot handle under-dispersion; can be adversely influenced by the low sample-mean and small sample size bias
Gamma	be able to handle under-dispersed statistics	Dual-state model with one state having a long-term mean equal to zero
Generalized estimating equation	be able to handle under-dispersed statistics	Dual-state model with one state having a long-term mean equal to zero
Random-effects	be able to handle temporal relationship	Determine or evaluate the type of temporal correlation a priori; results sensitive to missing values

References

[1] NHTSA—National Center for Statistics and Analysis (NCSA) (2016) NHTSA. https://www.nhtsa.gov/press-releases/us-dot-announces-steep-increase-roadway-deaths-based-2015-early-estimates

[2] Lord, D., and Mannering, F. (2010) The Statistical Analysis of Crash Frequency Data: A Review and Assessment of Methodological Alternatives. Accident Analysis and Prevention, 44, 291-305.

[3] Mohan, D. (2002) Road Safety in Less-Motorized Environments: Future Concerns. International Journal of Epidemiology, 31, 527-532.

[4] Kiran, Bangaram Naga, Nekkanti Kumaraswamy, and Chundupalli Sashidhar. "A review of road crash prediction models for developed countries." American journal of traffic and transportation engineering 2, no. 2 (2017): e10-e10.

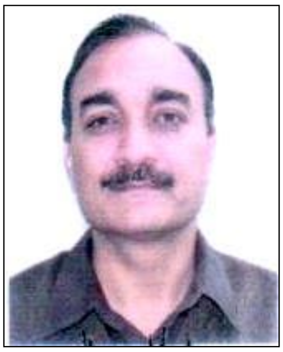
[5] Chang, L.Y., and Wang, H.W., (2006). Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis & Prevention. 38 (5), pp. 1019-1027.

[6] Chin H.C., and Quddus, M.A., (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis & Prevention. 35, pp. 253-259.

[7] NAMBUUSI, Betty, Tom BRIJS, and Elke HERMANS. "A review of accident prediction models for road intersections." (2008). [Available at: https://pdfs.semanticscholar.org/65b2/1e7f07fdb634e091d8692dc394716f64b443.pdf].

[8] Ambros, Jiří, Chris Jurewicz, Shane Turner, and Mariusz Kieć. "An international review of challenges and opportunities in development and use of crash prediction models." European Transport Research Review 10, no. 2 (2018): 35.

Author's Profile



1* Navdeep Mor
Research Scholar, NITTTR

2* Dr. Hemant Sood
Prof., CED, NITTTR

3* Dr. Tripta Goyal
Prof., CED, PEC, Chd.