IATSS Research 35 (2011) 34–40

Contents lists available at ScienceDirect

IATSS Research

Crash prediction model for two-lane rural highways in the Ashanti region of Ghana

Williams Ackaah [a,*], Mohammed Salifu [b]

[a] CSIR – Building and Road Research Institute, Kumasi, Ghana
[b] Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

ARTICLE INFO

Article history:
Received 28 September 2010
Received in revised form 3 January 2011
Accepted 19 January 2011

Keywords:
Ashanti region of Ghana
Crash prediction model
Injury crashes
Rural highways

ABSTRACT

Crash Prediction Models (CPMs) have been used elsewhere as a useful tool by road Engineers and Planners. There is however no study on the prediction of road traffic crashes on rural highways in Ghana. The main objective of the study was to develop a prediction model for road traffic crashes occurring on the rural sections of the highways in the Ashanti Region of Ghana. The model was developed for all injury crashes occurring on selected rural highways in the Region over the three (3) year period 2005–2007. Data was collected from 76 rural highway sections and each section varied between 0.8 km and 6.7 km. Data collected for each section comprised injury crash data, traffic flow and speed data, and roadway characteristics and road geometry data. The Generalised Linear Model (GLM) with Negative Binomial (NB) error structure was used to estimate the model parameters. Two types of models, the 'core' model which included key exposure variables only and the 'full' model which included a wider range of variables were developed. The results show that traffic flow, highway segment length, junction density, terrain type and presence of a village settlement within road segments were found to be statistically significant explanatory variables ($p<0.05$) for crash involvement. Adding one junction to a 1 km section of road segment was found to increase injury crashes by 32.0% and sections which had a village settlement within them were found to increase injury crashes by 60.3% compared with segments with no settlements. The model explained 61.2% of the systematic variation in the data. Road and Traffic Engineers and Planners can apply the crash prediction model as a tool in safety improvement works and in the design of safer roads. It is recommended that to improve safety, highways should be designed to by-pass village settlements and that the number of junctions on a highway should be limited to carefully designed ones.

1. Introduction

Road Traffic Crashes (RTCs) continue to be a major socio-economic problem for most developing countries [1,2]. In Ghana, it has been estimated that, road traffic crashes cost the nation 1.6% of its Gross Domestic Product [3]. Research on road traffic crashes have shown that crash severity tends to be mostly higher on rural highways than on urban roads [4–6]. During a three year period 2005–2007, a total of 21,709 injury road traffic crashes occurred in Ghana resulting in 48,605 casualties and yet these figures may have been higher if under-recording and non-reporting which have been established to be considerable [7] were accounted for in the data. The statistics further indicated that crashes in the non-urban environment (mostly on rural highways) accounted for about 70% of all road traffic fatalities. It is also evident from the data that whilst fatalities on roads in urban settlement areas follow a gradual upward trend, that for the non-urban areas follow a steep trend [4].

Crash Prediction Models (CPMs) have been used elsewhere as a useful tool by road Engineers and Planners. Fletcher et al. [8] found that due to wide differences in traffic mix, road quality, design and road user behaviour, it would be neither valid nor useful to apply simple multiplicative factors or even devise more complex conversion formulae for models developed elsewhere for another country. In Ghana, Salifu [9] developed a crash prediction models for unsignalised urban junctions. Afukaar and Debrah [10] also developed models for predicting crashes for signalised urban junctions. However, there is no study on predicting crashes on rural highways in Ghana.

The study was carried out in the Ashanti Region, the most populous and third largest in terms of land area (24,390 km [2]) of the ten administrative Regions of Ghana. The population of the Region is projected to be about 5,000,000 inhabitants [11] and lies approximately at the centre of the country. The Ashanti Region is one of the Regions with above average Public Health Risk (PHR) of road traffic crashes with a fatality index of 8.8 fatalities/100,000 population per year as against a national average of 8.5 fatalities/100,000 population per year [12].

The objective of the research work was to develop a crash prediction model to identify contributory factors that are likely to affect injury crashes on the rural highways in the Ashanti Region of

* Corresponding author.
E-mail address: ackaahwillie@yahoo.com (W. Ackaah).

0386-1112/$ – see front matter
doi:10.1016/j.iatssr.2011.02.001

Ghana. The model was developed for all crashes resulting in death or injury over a three (3) year period 2005–2007 on selected rural highways in the Ashanti Region of Ghana. Damage only crashes were not included in the study because of the high level of under-reporting associated with this type of crashes in Ghana [7]. In this report, a rural highway is defined as a highway passing through the non-urban environment i.e. passing through the village (population of not more than 5000 inhabitants) and rural non-settlement areas.

2. Literature review

Substantial research has been conducted over the years on the development of Crash Prediction Models (CPMs) for highway facilities. From the extensive review of literature, the elements considered in this chapter have been identified to be necessary for any model development.

2.1. Model form

Miauo and Lum [13] investigated the statistical properties of two conventional linear regression models and identified potential limitations of these models in developing vehicle crashes and highway geometric design relationships. It was demonstrated that the conventional linear regression models lack the distributional property to describe adequately random, discrete, non-negative, and typically sporadic, vehicle crash events on the road. Several other literatures have supported the unsatisfactory property of ordinary linear regression models in developing vehicle crashes, traffic flow and highway geometric design relationships [14–16].

Currently, Generalised Linear Regression Model (GLM) is used almost exclusively for the development of Crash Prediction Models [17–20]. Sawalha and Sayed [17] said the mathematical form used for any CPM in the framework of GLM should satisfy the following conditions: yield logical results and there must exist a known link function that can linearise the model for the purpose of coefficient estimation.

2.2. Error structure

The occurrences of vehicle crashes are discrete random events. That is, the number of vehicles involved in crashes on a given road section during a period of time is probabilistic in nature. Researchers in this field say the GLM assumes an error structure (probabilistic structure) that is used in the data generation process (crashes). This distributional assumption is also used to obtain tests and confidence statements about the estimated regression coefficients.

According to Lord et al. [21] a crash is in theory, the result of a Bernoulli trial. From literature, the appropriate probability model that accounts for a series of Bernoulli trials is known as the binomial distribution [21,22]. For typical motor vehicle crashes which are rare and therefore the event has a very low probability of occurrence and a large number of trials exists, the binomial distribution is approximated by a Poison distribution. However, one limitation of the Poisson distribution is that the mean is assumed to be equal to its variance. Many previous studies have found that crash data tend to be over-dispersed in many situations with the variance being significantly higher than the mean. In such cases, any inferences made based on Poisson model estimation may lead to wrong conclusions. As a result of this, many researchers recommend using alternative methods such as Negative Binomial distribution which does not require the equal mean and variance assumption [13,14,20,21].

2.3. Model goodness of fit assessment

Researchers in this area have identified two major statistical measures used in assessing the goodness of fit in both Poisson and Negative Binomial approaches. These are Pearson Chi-square statistic and Deviance.

To measure the overall goodness-of-fit in linear regression models, the coefficient of determination, R-squared is often used. The ordinary R-squared cannot be used for Poisson and Negative Binomial models because R-squared is maximised by ordinary least squares whiles in the Poisson and Negative Binomial models, the Maximum Likelihood estimation method is usually used. FridstrØm et al. [23] developed several alternative goodness-of-fit methodologies for generalised Poisson regression models. Miaou [24] also investigated different approaches to calculate R-squared values for different regression techniques using different distribution assumptions including Poisson and Negative Binomial.

2.4. Outlier analysis

Data collected may contain odd or extreme observations called outliers either because errors took place during data collection and recording or they are really different from the rest of the data [17]. These extreme observations may have effect on the model equation if not managed well. Outlier analysis could be carried out to identify these influential observations. However, Sawalha and Sayed [17] emphasised that excluding influential outliers from the development of a crash model is not synonymous with neglecting these outliers or removing them from the database. Rather, they should be investigated to determine what makes them different and whether any information can be extracted from them.

2.5. Selecting the model explanatory variables

According to Sawalha and Sayed [17] model generality requires that a model be developed in accordance with the principle of parsimony, which calls for explaining as much of the variability of the data using the least number of explanatory variables. Deciding on which variables to include in a model is a multi-step process. First, variables believed to relate to the phenomenon in question have to be determined. This is usually based on previous studies, experience and engineering judgement based on available information [17,25]. The manner in which these variables will be used in the model must also be determined. Last is the statistical determination as to which variables should remain in the model. Many studies have suggested that the t-ratio of its estimated parameter should be significant at the 95% confidence level, i.e. p-values <0.05[17,19,20,25].

2.6. Delineation of road sections

Miaou and Lum [13] considered two (2) methods, i.e. fixed-length sections, or homogeneous sections. They found out that, short road sections had undesirable impacts on the estimation of their linear regression models. Longer sections were, therefore, suggested to be preferred. However, given that most of the curved and graded sections were relatively short, the analysts oftentimes were unable to find a sufficient number of long and homogeneous road sections, which exhibited wide enough variations in geometric design variables to study the relationships. To overcome this problem, many analysts according to Miaou and Lum [13] have chosen to keep long road sections and not to insist on having homogenous road sections.

Resende and Benekohal [26] reached the conclusion that, to get a reliable crash prediction models crash rates should be computed from 0.8 km or longer sections.

3. Methods for data collection

3.1. Range of data collected

Variables believed to relate to crashes on rural highways were identified. This was primarily based on literature and engineering judgement based on available information and exploratory analysis.