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 © 2011 International Association of Traffic and Safety Sciences. Published by Elsevier Ltd. All rights reserved.
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.

The data collected for all sections for the study included: injury crash data (for a 3 year period, 2005–2007), traffic flow data, vehicle speed data, road characteristics and road geometry. The study was approved by the Institutional Review Board (IRB) of the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

3.2. Selection of road corridors

An extensive reconnaissance survey was carried out to identify suitable sites from which the sample of sites for the study was drawn. The road sections from which the samples were drawn were all two-lane single carriageway rural segments of road. Stratified random sampling technique was used in the road selection process in order to select segments from the population to include roadways from the three (3) different highway categories namely: national roads, N (which link regional capitals to the national capital); inter-regional roads, IR (which link regional capitals to one another); and regional roads, R (which link the district capitals to the regional capitals) so as to avoid bias.

Only road sections which had not undergone any modification during the period (2005–2007) for which crash data were studied were included in the sample. Since the objective of the study was to establish the effects of the different variables on crashes, selection of segments were done in order to achieve variety in traffic volumes, speeds, road characteristics and geometry.

3.3. Selection of road links

A total of fourteen (14) road corridors were selected for the project. These road corridors were subdivided into seventy-six (76) links and studied for the purpose of model development. Links were either taken as between two major junctions, between two towns, between a junction and a town or between a junction or town and any major landmark, for example, a river/bridge.

Thirty-one of the sites representing 40.8% were on National Roads, whilst 21(27.6%) were on Inter-Regional Roads and 24 (31.6%) were on Regional Roads. The total road length considered was 186.3 km long with National Roads, Inter-Regional Roads and Regional Roads constituting 41.3%, 31.4% and 27.3% respectively.

3.4. Road traffic crash and traffic flow data

In Ghana, information on road traffic crashes is available at the Building and Road Research Institute (BRRI). The database is compiled from police files using a standard crash report form. This form includes information about the crash location, the vehicle(s) and casualties involved. Most of the information has been coded and stored in computers at the BRRI using the Micro-computer Accident Analysis Package (MAAP) software. For the purpose of this study, the relevant information for each crash was the location and date. Crash data were retrieved and analysed with the help of the kilometre analysis facility available in the MAAP software. A total of 301 crashes were recorded over the period 2005–2007 on road links considered in the study of which 50.8% occurred on roads classified as National Highways, 30.2% occurred on the Inter-Regional Highways and 18.9% on the Regional Highways.

Traffic flow data including Average Daily Traffic (ADT) was obtained from the Traffic and Planning Unit of the Ghana Highway Authority (GHA) for the different road corridors for the study. Traffic volume for all the road categories considered ranged from 347 to 8948 vehicles per day.

3.5. Procedure for speed measurement

Speeds measurements were made with a radar gun at not less than three (3) locations within each segment in order to get a representative speed for the segments which oftentimes were un-homogenous. The data collection team was stationed in a parked vehicle and used the speed gun to measure spot speeds of vehicles. Effort was made to measure the speeds of all passing vehicles. In a situation where there was a platoon of vehicles, only the lead vehicle in the fleet was measured since vehicles following in a platoon do not travel at their desired speeds. Mean speeds of 83.5 km/h, 81.3 km/h and 69.2 km/h were recorded on National, Inter-Regional and Regional highways respectively.

3.6. Road characteristics and geometric data

Field surveys were carried out by a trained data collection team led by the Project Leader for all selected road segments for the purpose of establishing the road characteristics. Measurement of link length was done on site. Carriageway and shoulder widths were also measured on site and at more than two (2) locations within a link in order to get a representative figure for the entire segment. The variables presented in Table 3.1 were collected for the study.

Accesses were defined as any road connecting a private property to a public road. Junctions consisted of T-junctions, Y-junctions and X-junctions within segments.

4. Model development

4.1. Multiple regression analysis

The main objective of modelling with several variables simultaneously was to permit greater insight into the relative effects of the different highway variables on crashes. Modelling was undertaken at two stages — the 'core' model and the 'full' model all within the framework of Generalised Linear Models (GLMs). The 'core' model included exposure variables only which in this case were the traffic flow and section length. Apart from the fact that 'core' models have the advantage of being simple in form, they are also useful as a rough guide for identification of locations with a high frequency of crashes (blackspots), as well as for the prediction of the effect of traffic flow changes on crash occurrence [9].

It is also important to know that there are a large number of variables apart from traffic flow and length that might contribute to crashes. These variables will not be in the list of factors included in the 'core' model. The 'full' model which was developed included a wider range of other important causal variables.

Following the knowledge established from the review of previous work the general form of the 'core' and 'full' models developed under this study were:

'Core Model' — exposure variables only.

$$E(Y) = a_o L^{a1} Q^{a2} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (4.1)$$

Table 3.1
Road characteristics and geometric data collected for the study.

Variable description	Range of values
Categorical data	
Terrain type	0 = Hilly/Rolling; 1 = Flat
Road marking	0 = Absent; 1 = Present
Shoulder	0 = Absent; 1 = Present
Should type	0 = Unpaved; 1 = Paved
Village settlement within road segment	0 = Absent; 1 = Present
Road surface type	0 = Surface dressing; 1 = Asphalt
Continuous data	
Length of section	0.8 km–6.7 km
Carriageway width	7.0 m–7.8 m
Shoulder width	1.0 m–2.5 m
Junction density within segment	0–1.7 per km
Access density within segment	0–17.5 per km
Horizontal curves density within segment	0–6.4 per km
Vertical curves density within segment	0–3.6 per km

'Full Model' — product of the powers of the exposure variables multiplied by an exponential function incorporating the remaining explanatory variables.

$$E(Y) = a_o L^{a1} Q^{a2} \exp\sum_j b_j x_j \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (4.2)$$

where:

E(Y)	predicted crash frequency,
L	section length (km),
Q	ADT (per day),
x_j	is any variable additional to *L* and *Q*, and
exp	exponential function, e = 2.7183
a_0, a_1, a_2, b_j	are the model parameters.

In accordance with the GLM framework, Eqs. (4.1) and (4.2) are transformed into the prediction mode using a log-link function as follows:

'Core Model'

$$\ln[E(Y)] = \ln(a_o) + a_1\ln(L) + a_2\ln(Q). \quad \ldots \quad \ldots \quad \ldots \quad (4.3)$$

'Full Model'

$$\ln[E(Y)] = \ln(a_o) + a_1\ln(L) + a_2\ln(Q) + \sum_j b_j x_j. \quad \ldots \quad \ldots \quad (4.4)$$

4.2. Modelling procedure

The mean and variance for the crash data were 3.96 and 8.28 respectively which indicated that the data set was over-dispersed since the variance is greater than the mean. Initial modelling using Poisson error structure also showed that the estimated dispersion parameter (Φ) defined as:

$$\Phi = \frac{Pearson\ \chi^2}{(N-p)} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (4.5)$$

where *N* is the total number of sections and *p* is the number of parameters in the model was greater than one (1) indicating that the data set was over-dispersed [27]. That means Poisson distribution is not capable of explaining the true distribution underlying the crash frequency.

Generalised Linear Model (GLM) was used to estimate the model coefficients using the STATA software package and assuming a Negative Binomial error distribution, all consistent with earlier research works in developing these models. By specifying the dependent variable, the explanatory variables, the error structure (in this case the Negative Binomial) and the link function (in this case log), the model is fitted. Model parameters (coefficients) were estimated using maximum likelihood approach. The procedure which was adopted in the model development was the forward procedure in which the variables were added to the model one by one.

4.3. Model evaluation

In consonance with earlier studies, the decision on which variables should be retained in the model was based on two criteria. The first criterion was whether the t-ratio of the estimated parameter was significant at the 95% confidence level (p-value less than 5%). The second criterion was whether the addition of the variable to the model cause a significant drop in the scaled deviance at the 95% confidence level.

Two statistical measures were used in assessing the validity of the model developed. These were the Pearson Chi-square statistic and Deviance statistic. The coefficient of determination (R^2) was also employed to determine the amount of variability in the response variable explained by the variation in the selected set of explanatory variables. The R-squared estimation was carried out by the method recommended by Miaou [24].

4.4. Variables included in the model

Although a large number of variables were collected and considered for inclusion in the 'full' model development, only variables with significant estimated parameter coefficients (p-values less than 5%) were maintained in the model. This method is similar to the one used by Vogt and Bared [28]. The variables shown in Table 4.1 were thus, used in the full model development.

5. Model results and interpretation

5.1. 'Core' model

Parameter estimations for the log-linear equation for the 'core' model using Negative Binomial error structure are as shown in Table 5.1.

The resulting 'core' model has been determined to be as follows:

$$E(Y) = 0.1 \times L^{0.3666} \times Q^{0.3711}$$

where:

E(Y)	expected crashes along the road segment for 3 years,
L	length (km) of road segment and
Q	Average Daily Traffic (ADT)

The goodness-of-fit statistics for the model shows that the model fits reasonably well with the data. The Pearson Chi-square and Deviance statistics divided by its degrees of freedom were estimated to be 1.09 and 1.06 respectively as shown in Table 5.1. The values are within the permissible range [25] of 0.8 and 1.2 indicating that the Negative Binomial distribution assumption is acceptable.

The Average Daily Traffic (ADT) and segment length were statistically significant ($p<0.05$) with positive estimated model parameters in the 'core' model. This indicates that the crash frequency increases with an increase in the traffic flow or segment length whilst the other variables are held constant. The exponent on segment length was 0.37. Since distance travelled is a measure of exposure, it follows that the more distance travelled, the more likely one is involved in a crash. The exponent on traffic flow (ADT) was also estimated to be 0.37. This figure seems to be small when compared with other studies elsewhere [25,28,29] and considering the importance of traffic flow as a major determinant of road traffic crashes. Qin et al. [29] when they analysed different crash types found the lowest exponent of 0.4 on Average Annual Daily Traffic (AADT) when they considered only single-vehicle crashes. In the crash statistics on rural highway segments sampled for the study, over two-thirds (68.1%)

Table 4.1
Description of variables and symbols used for crash prediction models.

No.	Variable name	Description	Variable type	Symbol
1	lnvolume	The logarithm of average daily traffic	Continuous	Q
2	lnlengthkm	The logarithm of link length (km)	Continuous	L
4	juncdensity	Junction density	Continuous	JDen
5	terraintype	Terrain type	Categorical (1 — flat, 0 — otherwise)	TerTyp
6	presenc~ment	Presence of village settlement	Categorical (1 — present, 0 — absent)	VSet

Table 5.1
Parameter estimation for 'core' model as estimated by STATA statistical software.

Variable	Coefficient	Standard error	Chi-square statistic, Z	Log likelihood = −163.89315		
				P> Z	95% conf. limits	
					Lower	Upper
lnvolume	0.37107	0.09064	4.09	0.000	0.1934	0.5487
lnlength	0.36661	0.14182	2.59	0.010	0.0887	0.6446
_cons	−1.92725	0.73210	−2.63	0.008	−3.3621	−0.4924

Goodness of fit measures: deviance (value/d.f.) = 77.7200 (1.0647); Pearson chi-square (value/d.f.) = 79.62985 (1.09082).

involved single-vehicles and this may partly explain the rather low exponent value for the ADT.

Using the R-squared value, the variables in the flow based model ('core' model), namely traffic flow and segment length could only explain 31.3% of the total variation in the crash data. According to Salifu [9], 'core' models may be regarded as relatively coarse and rough estimators of crash frequency.

5.2. 'Full' model

Parameter estimations for the log-linear equation for the 'full' model using negative binomial error structure are as shown in Table 5.2.

The resulting 'full' model has been determined to be as follows:

$$E(Y) = 0.1847 \times L^{0.5060} \times Q^{0.2479} \times EXP^{(0.2779\ JDen + 0.4551\ TerTyp + 0.4717\ VSet)}$$

where:

E(Y) expected crashes along the road segment for 3 years,
L length (km) of road segment and
Q Average Daily Traffic (ADT)
JDen Junction Density
TerTyp Terrain Type (1 — flat, 0 — otherwise)
Vset Presence of Village Settlement (1 — present, 0 — absent)
EXP Exponential function, e = 2.718282.

The goodness-of-fit statistics for the full model shows that, the model fits reasonably well with the data. The Pearson Chi-square and Deviance statistics divided by its degrees of freedom were estimated to be 0.81 as shown in Table 5.2. In other words, the estimated values of the Pearson Chi-square and Deviance divided by the degrees of freedom were within the permissible range (i.e. between 0.8 and 1.2) indicating that the Negative Binomial distribution assumption is acceptable.

The R-squared value for the full model seems to be reasonable as it could explain 61.2% of the variation in the crash data. Dissanayake and Ratnayake [25] have said the coefficient of determination is significant if found to be greater than 0.45. Approximately 39% of the systematic variations in the data set remain unexplained by the model developed. The 'percentage unexplained' may be attributed to human behaviour which has increasingly been acknowledged as one of the predominant factors in road traffic crashes and factors relating to the vehicle.

Table 5.2
Parameter estimation for 'full' model as estimated by STATA statistical software.

Variable	Coefficient	Standard error	Chi-square statistic, Z	log likelihood = −143.43250		
				P> Z	95% conf. limits	
					Lower	Upper
lnvolume	0.24793	0.81300	3.05	0.002	0.0886	0.4073
lnlengthkm	0.50599	0.13304	3.80	0.000	0.2452	0.7667
juncdensity	0.27786	0.14124	1.97	0.049	0.0010	0.5547
terraintype	0.45512	0.12866	3.54	0.000	0.2030	0.7073
presenc~ment	0.47174	0.13591	3.47	0.001	0.2054	0.7381
_cons	−1.68896	0.64160	−2.63	0.008	−2.9465	−0.4315

Goodness of fit measures: deviance (value/d.f.) = 56.8476 (0.8121); Pearson chi-square (value/d.f.) = 56.67047 (0.80958).

5.2.1. Significant explanatory variables

The following sections consider the significant variables identified through the 'full' model results.

5.2.1.1. Traffic flow and safety. The Average Daily Traffic (ADT) in the 'full' model was also significant with positive estimated model parameter. From the model, an increase in the traffic flow by 50% is expected to cause an increase in road traffic crashes by 11% whilst doubling the number of vehicles on the road is expected to cause an increment of 19%. Meanwhile, the proportions of heavy goods vehicles were found not to have a significant effect on crashes in this study.

5.2.1.2. Road segment and safety. The length of road segment, the other important exposure variable aside traffic volume was the most significant predictor in the 'full' model with a positive estimated exponent of 0.51. This was as expected as distance travelled is a measure of exposure therefore it follows that the longer the distance travelled, the more likely one is involved in a crash.

5.2.1.3. Junction density and safety. The number of junctions per unit length within a road segment has a considerable influence on the crash risk as indicated by the estimated model parameter. In general, the more the number of junctions per unit length of segment, the higher the crash risk. For example, for all other variables held constant, adding one junction to 1 km section of road segment increases crashes by 32.0%. The number of junctions on a highway should therefore be limited to carefully designed ones to improve road traffic safety in Ghana.

5.2.1.4. Terrain type and safety. The analysis suggests that, flat terrains record more injury crashes when compared with rolling or hilly terrains. An increase of 57.6% in injury crashes occurred at links with flat terrain compared with links with rolling or hilly terrain. This probably had to do in part, with the fact that drivers tend to travel at higher speeds on straight and flat road sections than on rolling or hilly terrains.

5.2.1.5. Presence of village settlement and safety. The explanatory variable for presence of village settlement proved to be very important in the model. An increase of 60.3% in injury crashes were found for sections which had a village settlement within them compared with segments with no settlements.

From the crash data used in the model development, 90 (29.9%) were 'hit pedestrian' collisions of which 81.1% occurred in the settlement areas. It was presumed that the variable 'presence of village settlement' represents the level of pedestrian interference with vehicles on the highways. The interaction between fast moving vehicles and pedestrians is undesirable and contribute to the high number of injury crashes in the settlement areas. More fatal pedestrian collisions have been reported in the village settlements along the major highways in Ghana [30].

5.2.2. Non-significant explanatory variables

The following sections also look at the variables which were considered but were not significant in the 'full' model results.

5.2.2.1. Horizontal and vertical curves. Horizontal curvature has been found to influence road traffic crashes in other studies [31,32]. Sharper and longer curves results in more crashes because of the additional effort required by the driver to control the stability of the vehicle. Again, studies have shown that reducing gradients reduces the number of crashes [31,32]. The absolute number of curves per segment length used in this study did not capture the effect of the individual degrees of curvatures of the curves. Construction plans (horizontal and vertical profiles) for most of the roads selected for the study were not readily available for the curvatures to be extracted from them. Attempt to use the tract function of a Global Positioning System (GPS) device to measure tracks along the road segments from which the vertical and horizontal curvatures could have been calculated proved futile. It was found that the data were not of sufficient completeness or accuracy to be included in the analysis as in particular, the vertical curvature of relatively flat sections was often lost. Fletcher et al. [8] also experienced a similar difficulty. This may explain why both vertical and horizontal curvatures did not show to be statistically significant in the study.

5.2.2.2. Speed. Recent studies have concluded that crash involvement increases with speed on any particular road or for any particular driver [18,33,34]. Mean speed was not a useful explanatory variable in this study as it correlated highly with the variables presence of village settlement and terrain type. Standard deviation speed and 85th percentile speed were also tested in the model but were not statistically significant.

5.2.2.3. Presence of shoulder, shoulder type, shoulder width and road width. The effect for the variable presence of shoulder on crash rates could not be assessed as there was no variability in the data collected as the entire road links sampled had shoulders. Shoulder type has also been found in other studies to be important in terms of stability of vehicles when a vehicle goes off the road. However, the variable shoulder type did not prove to be a useful explanatory variable as it correlated with traffic flow. Adequate shoulder width has also been established to be necessary for safe recovery when vehicles veer off the road [32]. The variable shoulder width was not statistically significant when data from all segments (which included paved and unpaved shoulders) was used in the model. This may be explained by the fact that paved and unpaved shoulders have different effect on crashes. Two-thirds ($n=51$) of the segments had paved shoulders while one-third ($n=25$) had unpaved shoulders. Efforts to develop models for the sub-groups (paved and unpaved) also did not yield any significant results as the range of paved shoulder width encountered was limited (2.0 m, 2.2 m or 2.5 m) and the sample size for unpaved shoulders was small.

Generally it has been found that crash rates decrease when lane width increases [28,32]. However, road width was not found to be a useful explanatory variable in this study as it also correlated highly with traffic flow.

5.2.2.4. Road marking. Marked roads have been found to be safer than unmarked roads [8]. Most of the road segments sampled were delineated ($n=62$, 82%) while only a few ($n=14$, 18%) had no marking. The comparative small sample size may have resulted in this variable not being a significant explanatory variable in this study.

6. Conclusions and recommendations

6.1. Conclusions

Statistical models have been developed to predict crashes on rural highways in the Ashanti Region of Ghana. 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 traffic exposure variables only and the 'full' model which included a wider range of variables were developed.

From the model developed, traffic flow, segment length, junction density, terrain type and presence of village settlement within road link are significant explanatory variables that influence the prediction of injury crashes on the rural highways in the Ashanti Region of Ghana. The study proved that, passing a highway through settlement areas, traffic volume, provision of long straight and flat sections and increasing number of junctions per unit road length tend to worsen road traffic safety.

Road and Traffic Engineers and Planners can apply the crash prediction model as a tool in safety improvement works and in the design of new safer roads.

6.2. Recommendation

From the model developed and subsequent discussions above, the following recommendations are drawn:

- Safety situations on the highways should be improved through the provision of by-passes around the village settlements.
- The number of junctions on a highway should be limited to carefully designed ones to improve road traffic safety in Ghana.
- Road Engineers should as much as possible avoid provision of long straight and flat sections on highways to improve traffic safety. Intermittent graceful curves could be introduced to break the monotony.

Acknowledgements

The authors would like to acknowledge the Road Traffic Injury Research Network (RTIRN) for partly sponsoring this research work through their scholarship programme — "Grants for Junior Researchers" with funding support from the World Bank Global Road Safety Facility. The opinions, findings and conclusions expressed here are those of the authors.

References

[1] G.D. Jacobs, A. Aeron-Thomas, A. Astrop, Estimating global road fatalities, TRL Report 445, Transport Research Laboratory, Crowthorne, Berkshire, 2000.
[2] C. Ghee, D. Silcock, A. Astrop, G. Jacobs, Socio-economic aspects of road accidents in developing countries, TRL Report 247, Transport Research Laboratory, Crowthorne, Berkshire, 1997.
[3] F.K. Afukaar, W. Agyemang, E.K. Debrah, W. Ackaah, Estimation of the cost of road traffic accidents in Ghana, Ministry of Transportation, National Accident Management Project, Final Report, 2006.
[4] F.K. Afukaar, W. Agyemang, W. Ackaah, I. Mosi, Accident Statistics 2007. Building and Road Research Institute, Council for Scientific and Industrial Research, Kumasi, 2008.
[5] K. Hamilton, J. Kennedy, Rural road safety: a literature review, Scottish Executive Social Research (2005).
[6] G.D. Jacobs, A study of accident rates on rural roads in developing countries, TRRL Report, Transport and Road Research Laboratory, Crowthorne, Berkshire, 1976, p. 732.
[7] F.K. Afukaar, W. Ackaah, W. Agyemang, Estimation of levels of under-reporting of road traffic accidents in Ghana, Ministry of Transportation, National Accident Management Project, Final Report, 2006.
[8] J.P. Fletcher, C.J. Baguley, B. Sexton, S. Done, Road accident modelling for highways development and management in developing countries, Main Report: Trials in India and Tanzania, Project Report No: PPR095, , 2006.
[9] M. Salifu, Accident prediction models for unsignalised urban junctions in Ghana, International Association of Traffic and Safety Science (IATSS) Research 28 (1) (2004) 2004.
[10] F.K. Afukaar, E.K. Debrah, Accident prediction models for signalized intersections in Ghana, Ministry of Transportation, National Accident Management Project, Final Report, 2007.
[11] Ghana Statistical Service, 2000 population and housing census, Special report on 20 largest localities, 2002.
[12] M. Salifu, I. Mosi, K. Addae-Bofah, K. Sarpong, Accident Statistics 2005. Building and Road Research Institute, Council for Scientific and Industrial Research, Kumasi, 2006.
[13] S.-P. Miaou, H. Lum, Modelling vehicle accidents and highway geometric design relationships, Accident Analysis and Prevention. 25 (6) (1993) 689–703.
[14] S.-P. Miaou, The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, Accident Analysis and Prevention. 26 (4) (1994) 471–482.

[15] S.H. Kim, S.B. Chung, K.H. Song, K.S. Chong, Developing of an accident prediction model using GLIM (generalized log-linear model) and EB method: a case of Seoul, Journal of the Eastern Asia Society for Transportation Studies 6 (2005) 3669–3682.

[16] N.J. Garber, L. Wu, Stochastic models relating crash probabilities with geometric and corresponding traffic characteristics data, Research Report No. UVACTS-5-15-74, 2001.

[17] Z. Sawalha, T. Sayed, Traffic accident modeling: some statistical issues, Canadian Journal of Civil Engineering 33 (2006) 1115–1124.

[18] M.C. Taylor, A. Baruya, J.V. Kennedy, The relationship between speed and accidents on rural single-carriageway roads, TRL Report TRL511, Transport Research Laboratory, Crowthorne, Berkshire, 2002.

[19] S. Harnen, R.S. Radin Umar, S.V. Wong, W.I. Wan Hashim, Development of prediction models for motorcycle crashes at signalized intersections on urban roads in Malaysia, Journal of Transportation and Statistics V7 (2004) N2/3.

[20] E.T. Donnell, J.M. Mason, Predicting the frequency of median barrier crashes on Pennsylvania interstate highways, Accident Analysis and Prevention, 38, 2006, pp. 590–599.

[21] D. Lord, S.P. Washington, J.N. Ivan, Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory, Paper Accident Analysis and Prevention, 2004, p. 03225.

[22] J.T. McClave, T. Sincich, Statistics, Tenth Edition. Pearson Education, Inc0-13-200302-3, 2003.

[23] L. FridstrØm, J. Ifver, S. Ingebrigsten, R. Kulmala, L.K. Thomsen, Measuring the contribution of randomness, exposure, weather, and daylight to the variation in the road accident counts, Accident Analysis and Prevention 27 (1) (1995) 1–20.

[24] S.-P. Miaou, Measuring the goodness-of-fit of accident prediction models, FHWA-RD-96-040, Federal Highway Administration, Washington, D.C, 1996.

[25] S. Dissanayeke, I. Ratnayake, Statistical modelling of crash frequency on rural freeways and two-lane highways using negative binomial distribution, Advance in Transportation Studies an International Journal (2006), Section B 9.

[26] P. Resende, R. Benekohal, Effects of roadway section length on accident model, Traffic Congestion and Traffic Safety in the 21st Century Conference, American Society of Civil Engineers (ASCE), Chicago, IL, 1997.

[27] P. McCullar, J.A. Nelder, Generalized Linear Models, Chapman and Hall, London, 1989.

[28] A. Vogt, J.G. Bared, Accident models for two-lane rural roads: segments and intersections, Publication No. FHWA-RD-98-133, 1998.

[29] X. Qin, J.N. Ivan, N. Ravishanker, Selecting exposure measures in crash rate prediction for two-lane highway segments, Accident Analysis and Prevention 36 (2004) 183–191.

[30] W. Ackaah, D.O. Adonteng, Analysis of fatal road traffic crashes in Ghana, First published on 20 may 2010, International Journal of Injury Control and Safety Promotion (2010) 1–7, iFirst.

[31] R. Elvik, T. Vaa, The Handbook of Road Traffic Safety Measures, Elsevier Ltd978-0-08-044091-0, 2004.

[32] S.-P. Miaou, P.S. Hu, T. Wright, S.C. Davis, A.K. Rathi, Development of relationship between truck accidents and geometric design: phase I, Publication No. FHWA-RD-91-124, 1993.

[33] C.N. Kloeden, G. Ponte, A.J. McLean, Travelling speed and the risk of crash involvement on rural roads, Road Accident Research Unit, Adelaide University, 2001.

[34] A. Quimby, G. Maycock, C. Palmer, S. Buttress, The factor that influence a driver's choice of speed — a questionnaire study, TRL Report 325, Transport Research Laboratory, Crowthorne, Berkshire, 1999.