ACCIDENT PREDICTION MODELS FOR UNSIGNALISED URBAN JUNCTIONS IN GHANA

Mohammed SALIFU

Principal Research Scientist, MSc., PhD, MIHT, MGhIE
Building and Road Research Institute (Council for Scientific and Industrial Research)
Kumasi, Ghana

(Received July 18, 2003)

The main objective of this study was to provide an improved method for safety appraisal in Ghana through the development and application of suitable accident prediction models for unsignalised urban junctions.

A case study was designed comprising 91 junctions selected from the two most cosmopolitan cities in Ghana. A wide range of traffic and road data together with the corresponding accident data for each junction for the three-year period 1996-1998 was utilized in the model development process. Potential explanatory variables, which were tested were largely identified from initial analysis of the accident characteristics and associated factors. Negative Binomial models of accident frequency were developed separately for T- and X-junctions.

The results showed that the best models based exclusively on traffic exposure functions (i.e. traffic flow) explained 50 per cent more of the systematic variation in accidents at T-junctions than at X-junctions. In the extended models that included road geometric and other traffic variables it emerged that the absence of street lighting and dedicated left-turning lanes and the average standard deviation of approach spot speeds of vehicles on the major road were all positively correlated with accident frequency at both T- and X-junctions. Significantly and contrary to expectation, T-junctions with YIELD control had a much lower accident potential than those with STOP control.

The accident prediction models developed have a potentially wide area of application and their systematic use is likely to improve considerably the quality and delivery of the engineering aspects of accident mitigation and prevention in Ghana.

Key Words: Ghana, Safety appraisal, Accident prediction models, Unsignalised urban junctions, Traffic control-type

1. INTRODUCTION

Road traffic accidents continue to be a major problem in Ghana, both from the public health and socio-economic perspectives. In the ten-year period 1991-2000, 85,867 traffic accidents were recorded and these resulted in 107,780 casualties of which 25,340 were fatalities[1]. And yet these figures may be much higher if it were possible to account for shortfalls in reporting. The overall annual cost of road traffic accidents to the national economy has recently been estimated as US$70 million[2]. Thus road accidents are as much a major threat to public health in Ghana as they are an enormous drain on the national economy. Improving safety on Ghana's roads is therefore a pressing national concern that has already found expression in the setting-up of the National Road Safety Commission (NRSC). The Commission has been tasked to initiate and oversee the implementation of a more proactive and structured programme of accident reduction. The initial efforts in this direction have culminated in the National Road Safety Strategy and Action Plan[3] both of which underscore the need for data-led interventions and innovative approaches to understanding the occurrence mechanisms and determining factors of road accidents and devising strategies to reduce the incidence of accidents on all manner of roads.

It is in furtherance of this strategic approach that this study was carried out with the express objective of developing accident prediction models that can be used in a proactive appraisal of accident potential and identification of accident-prone locations. In particular, it is also expected that the establishment of such quantified relationships between accidents on the one hand, and traffic flows and site characteristics on the other, would enable priorities for improvements to be more realistically assessed, thereby ensuring that more judicious use is made of the usually limited budgetary allocation to road safety activities. The focus of this study is unsignalised urban junctions, which currently account for more than 60 percent of all junction accidents in Ghana. In the long-term, however, it is envisaged that prediction models would be developed for all types of junction and link sections in rural as well as urban locations.

2. REVIEW OF PREVIOUS WORK

Although the single event of an accident is almost impossible to predict, due to its rare and random nature, researchers have found that aggregation of a large number of accidents over a sufficiently wide area and/or long period of time tends to exhibit a level of predictability that can be described by means of mathematical/statistical relationships[4]. Multivariate accident-prediction models represent a form of such relationships between accident frequency and a set of determining factors. These are empirically derived and vary in form, depending on the explanatory variables used.

2.1 Functional form of models

The relationship between accidents and traffic flow as a measure of exposure, in particular, has received considerable attention over the years. Tanner[5], for example, is credited with one of the earliest of such studies on intersections. He suggested that accident numbers were approximately proportional to the root of the product of the two-way major road traffic volume and turning flows from the minor road. Since then, numerous other forms of relationship, at times conflicting, have been proposed. For example, "the product of intersecting flow" model proposed by Hakkert and Mahalel[6] and the "product of flows, each raised to a power less than one" by Leong[7], and Hauer et al[8].

McGuigan[9] also investigated the "root product flow" after Tanner[5] and "throughput" or "sum of inflows" relationships and reported that preference for the former over the latter was not universally justified. The "sum of inflows" model form, however, has been associated with some logical inconsistencies[10]. These relate to the potential to predict more than zero accidents between conflicting streams of traffic even when one of the flows is zero and also the possibility of predicting equal numbers of accidents for a given value of total inflows, irrespective of the distribution of flows between the major and minor arms of a junction. In reality, accident frequency will depend on the relative balance of traffic flows between the major and minor approaches. Clearly, the variety of model forms mirrors the continuing confusion regarding the most appropriate form of exposure index and the differences may be rooted in the type, quality and manner of analysis of the data utilized. This means that conducting exploratory analysis of the specific data could provide useful clues as to the best functional form to adopt for new studies, a logical thing to do, considering that this particular work is a pioneering effort involving a dataset with potentially different attributes from those used in previous studies.

Model forms, which rely solely on traffic flows for predicting expected accidents are referred to as "coarse" models. Whilst such models have the advantage of being simple in form, they are useful only as a rough guide for identification of unusually hazardous locations, as well as for the prediction of the effect of traffic flow changes on accident occurrence. However, relationships of this nature are more likely to be associative rather than causal [11]. For the purposes of this study, comprehensive or causal accident prediction models are required, in order to quantify the effect of not only individual treatments but also the complete set of road characteristics, including traffic flows, site features and detailed geometry and traffic control variables. Therefore, flow-based models, although useful in their own way are developed and presented here as intermediate models towards comprehensive modelling.

Comprehensive prediction modelling of accidents at unsignalised urban junctions, in particular, remains largely unexplored and therefore, the results of the few reported studies[12,13] may be considered partly tentative and ought to form the basis for further independent study. For example, the approach of Summersgill *et al*[11] and many others in examining the effect of only speed limits, as opposed to actual speed, on accident occurrence could be improved because, in practice, speed limits are not necessarily indicative of actual levels of speed observed. In addition to addressing this particular issue, efforts were also made under the current study to include junctions with dual-carriageway arterial roads, few of which have been covered by the reported studies.

2.2 Statistical methods

The key tool in the model development process is multiple regression analysis, two types of which have been used in the literature surveyed; classical techniques and the generalised linear modelling approach. Classical least-squares (ordinary) regression techniques were used in developing the early accident predictive models[9]. However, recent research has shown that ordinary least-squares regression has some statistical properties that are undesirable for accident data analysis. These include the intrinsic assumption of homoscedascity (i.e. equal variance of the error terms for all values of the predictor variable) and the possibility of predicting accident frequency with negative values. In reality, accident counts are sporadic, discrete and non-negative and their occurrence pattern would be more akin to a Poisson process, like any count data.

Incidentally, an attribute of the Poisson distribution, namely that the mean of the predicted variable is equal to its variance, does not usually hold when a substantial proportion of a database comprises zero accident counts, as is often the case in accident prediction modelling. With over-dispersed data (i.e. when the mean is less than the variance), Miaou and Lum[14] observe that the Poisson model tends to produce inaccurate estimates. As a solution to this problem, the authors recommended the adoption of the Negative Binomial distribution, a more general probability distribution, which relaxes the constraints on the mean and variance. In other, more recent studies[10,12], the technique of "generalised linear models", using the software package GLIM[15] has facilitated the use of more generalised probability distributions like the Negative Binomial. The GLIM approach is preferable because, among other things, it allows the representation of accident counts as coming from the family of exponential distributions, from which one can be chosen to correspond to the data used and it yields maximum likelihood estimates of parameters, i.e. values of parameters that are most likely to have given rise to the accident data.

It is significant to note that most of the reported studies on accident prediction modeling have been carried out using data from industrialized countries, where vehicle-ownership levels are relatively high and road and traffic conditions vary significantly in many respects from that in a typical developing country. Therefore, it will be reasonable to anticipate that the significant explanatory variables and the size and direction of their influence are likely to be different in either case. This would underscore the need for the development of relevant "home grown" models for a country such as Ghana. It would appear that little or no work in this direction has been carried out in developing countries because road safety science in these countries remains largely rudimentary; little systematic attention is paid to road safety whilst accident databases are often not comprehensive and credible enough to meet the standard required for prediction modeling. These constraints are considerably diminished in the case of Ghana where sustained efforts over the last 15 years have led to the creation of a highly improved database and opened up possibilities for rigorous scientific safety analysis.

3. METHODOLOGY

3.1 Data collection

A judiciously selected sample of junctions, stratified mainly by traffic flow and junction features, was chosen to ensure that as wide a range of flows and junction features as possible would be captured. A purely random (and unstratified) sample of the same size, arguably, would not have guaranteed the inclusion of some key variables likely to have a significant impact on accidents. An initial list of 130 sites selected from desk studies were all visited in follow-up reconnaissance surveys during which some were discarded. Sites were dropped mostly because they were thought to have undergone some changes in features that could have affected their safety status during the study period 1996 to 1998 inclusive.

Other considerations were dictated by the need to have a critical mass of "typical" junctions for analysis. The final list of junctions numbered 91, comprising 57 T-junctions and 34 X-junctions. Three basic types of T-junction were captured; namely, two-way single-carriageway minor road without chanellisation / two-way single-carriageway major road (See Photo T-1), two-way single-carriageway minor road with channelisation / two-way single-carriageway major road (See Photo T-2) and

Photo T-1 2-way single carriageway major/minor

Photo T-2 2-way single carriageway major/minor, island on minor road

two-way single-carriageway minor road without channelisation / dual-carriageway major road (See Photo T-3). X-junctions on the other hand were of two basic types; two-way single-carriageway minor road without channelisation / two-way single-carriageway major road (See Photo X-1) and two-way single-carriageway minor road without channelisation / dual-carriageway major road (see Photo X-2). For each junction detailed information regarding accidents, traffic flow and geometric and traffic control features, among others was gathered, these are briefly described below.

3.1.1 Accident data

Accident data covering the period 1996-1998 inclusive for the selected junctions were retrieved from the national accident database at the Building and Road Research Institute. The database is painstakingly compiled from police files using a standard accident report form, which contains information on about 90 variables relating to the time, place, circumstances, the parties involved, etc. of the accident. Accident types covered include property-damage as well as person-involved collisions. Inevitably, the database is subject to some measure of under-reporting but since no extensive studies have been carried out to estimate the scale, it will be difficult to account for it in any systematic manner in the current study. Nonetheless, the data is quite comprehensive and operates on the Micro-computer Accident Analysis Package (MAAP5)[16] with immense possibilities for data manipulation and analysis. The very concise location coding system of the database using a combination of grid-referencing (X.Y. co-ordinates), a link-node system and strip maps makes it easy to accurately isolate and analyze data specific to any particular location on the road network. A total of 354 and 238 accidents respectively were recorded for all T- and X-junctions respectively. Thus the average number of accidents per junction was 6.2 and 7.0 in that order. Pedestrian accidents per junction averaged 40 per cent more at X-junctions than at T-junctions. Table 1 shows the overall frequency distribution of junctions by the number accidents recorded in the three-year period 1996-1998.

3.1.2 Traffic flow data

Traffic flow data collected included vehicle counts classified by type of vehicle and turning movement, counts of pedestrians crossing all arms of the junctions and spot speeds of vehicles as they approached the junction area along the major arms. Vehicles were broadly classified into three categories, namely, cars, minibuses

Photo T-3 2-way single carriageway minor, dual-carriageway major

Photo X-1 2-way single carriageway major/minor

Photo X-2 2-way single carriageway minor, dual-carriageway major

Table 1 Frequency distribution of junctions by number of accidents recorded

All accidents recorded for the period 1996-1998 inclusive	T-junctions		X-junctions	
	Number of sites recording the given number of accidents	Proportion of all sites (%)	Number of sites recording the given number of accidents	Proportion of all sites (%)
0	10	17.6	5	14.7
1	1	1.7	3	8.8
2	2	3.5	0	0
3	5	8.8	3	8.8
4	4	7.0	4	11.8
5	12	21.0	5	14.7
6	3	5.3	0	0
7	5	8.8	4	11.8
8	5	8.8	1	3.0
9	2	3.5	1	3.0
10	1	1.7	0	0
11	1	1.7	1	3.0
12	0	0	2	5.9
13	1	1.7	0	0
14	0	0	1	3.0
15	1	1.7	3	8.8
16	1	1.7	0	0
18	1	1.7	0	0
29	0	0	1	3.0
32	2	3.5	0	0
TOTAL	57	100.0	34	100.0

("trotro") and Heavy Goods Vehicles and Buses. Traffic counts were carried out during the morning and evening peak periods from 0700 to 0900 hrs and from 1700 to 1900 hrs, respectively. The counts were subsequently converted into Average Annual Daily Traffic (AADT) using existing conversion factors. Pedestrian counts were also carried out concurrently as vehicle counts. No attempt was made to adjust pedestrian traffic for time trends due to the unavailability of the appropriate conversion factors.

Figure 1 shows the distribution of junctions by traffic flow (AADT) groups. Vehicle approach spot speeds were measured using a hand-held radar speed gun. Speeds were measured of vehicles selected at random until a total of 40 vehicles were covered for each arm. Significantly, a large proportion (56%) of speeds recorded at T-junctions were above the posted maximum limit of 50km/h. The corresponding figure for X-junctions was 40 percent.

3.1.3 Site and geometric data

Junction inventories were carried out to collect information relating to the site details. The information collected included junction layout, type of major and minor roads (i.e. whether single or dual-carriageway), numbers, type and widths of lanes, types of median or other island, if any, and dimensions. Other features were types of con-

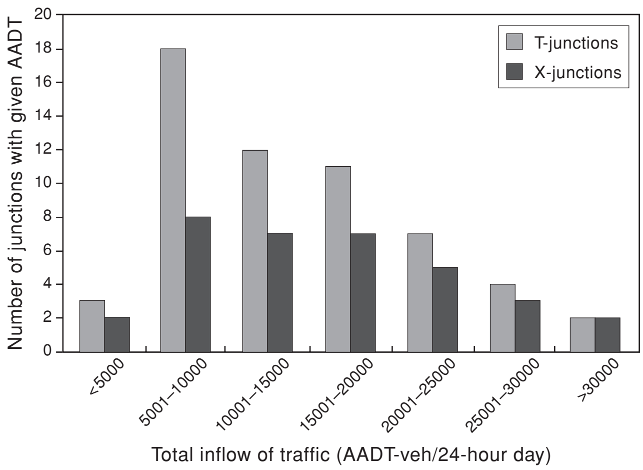

Fig.1 Distribution of junctions by total inflow of traffic (AADT)

trol, including road markings, street lighting and status of pedestrian crossing facilities. Due to the absence of as-built drawings for nearly all the sites, it was not possible to measure the radius of curvature of the entry kerb lines, which is considered important for junction safety. The width of the minor roads at the neck of the junctions was measured and used as a proxy for the latter. The site geometric and other traffic variables that were of some significance in the modeling process are shown in Table 2.

3.2 Model development

3.2.1 Regression analysis

The objective of modeling was to relate the average 3-year accident frequency at the junctions to the best set of explanatory variables. A multiple regression approach was therefore adopted within the framework of Generalised Linear Models (GLMs). The main advantage in doing this is that the theory of GLMs allows the variation in the dependent variable to be separated into the systematic and random parts[15]. As a consequence, it is possible to make structural and distributional assumptions, which describe these two types of variations respectively[12]. The structural assumption indicates that the expected value of the response variable can be related through a "link function" to a set of explanatory variables and their coefficients. On the other hand, random variation is described by a "random error term" associated with the model, which reflects the distributional properties of the response variable. The ordinary linear model tackles both the distributional and structural assumptions together and assumes the response variable to be Normally-distributed, quantitative and continuous and capable of taking any values. These run counter to the basic properties of accident counts, which are discrete, non-negative and generally governed by a non-stationary Poisson process[17]. Following the lead established from the review of previous work the general form of the models developed under this study was:

$$E(\mu_i) = kQ^{\alpha} \exp (\Sigma \beta_j X_{ij}) \quad \text{.. (1)}$$

where $E(\mu_i)$ is the expected number of accidents (in 3 years) at the i-th junction,

Q – a general traffic flow function,

k, β_j, and α – the model parameters or regression coefficients to be estimated (β_j represents the regression coefficient corresponding to the j-th explanatory

Table 2 Other traffic and road variables and factors for both T- and X-junctions

A. CATEGORICAL FACTORS				
Symbol	Description	Levels	Number of sites with given features: T-junctions	X-junction
ZEX	Zebra crossing	1 = present	17	12
		2 = absent	40	22
ILM	Island on minor road, entry/exit divided on either side	1 = present	13	0
		2 = absent	44	34
ITM	Triangular island on minor, two-way entry/exit on either side	1 = present	7	0
		2 = absent	50	34
STL	Street lighting	1 = present	27	15
		2 = absent	30	19
LFT	Left-turning storage lane on major	1 = present	8	9
		2 = absent	49	25
TCON	Traffic control on minor	1 = stop	27	29
		2 = yield	22	4
		3 = none	8	1
MED	Median on major road	1 = present	17	20
		2 = absent	40	14
LANE	Number of lanes on major in each direction	1 = one	28	14
		2 = two	29	20

B. NON-CATEGORICAL VARIABLES	
SSD	Average standard deviation of vehicles approach spot speeds (km/h)
JNEC	Average width of minor road at neck of junction (m)
MEDW	Average width of median on major arms (m)

variable other than traffic flow),

X_{ij} – the j-th explanatory variable other than traffic flow for the i-th junction (i = 1, n; n being the total number of junctions in the modeling database).

In accordance with the GLM framework Equation (1) is transformed into the prediction mode using a log-link function as follows:

$$\ln(E(\mu_i)) = \log(k) + \alpha \log Q + (\Sigma\beta_j X_{ij}) \qquad (2)$$

where, all parameters are as defined in Equation (1) above

By specifying the dependent variable, the model form, error distribution (in this case Poisson or Negative Binomial), the potential explanatory variables and the link function, the model is fitted, as the coefficients (model parameters) of the specified variables are estimated using the method of maximum likelihood.

3.2.2 Modelling procedure

To identify the best fitting models, different flow functions were initially tested, individually and in diverse combinations on the basis of a Poisson error structure. By this initial approach, it was possible to determine whether the fitted models were over-dispersed or not, following an assessment of the scaled deviance (SD) relative to the degrees of freedom (DF). Over-dispersion was considered indicated if the SD was at least 1.5 times the DF i.e. SD>>DF . As expected, in most cases, the models were over-dispersed relative to the Poisson error structure and so the next logical step was to specify a Negative Binomial error structure and refit. At this stage, the over-dispersion parameter (κ) was estimated automatically by maximum likelihood, using the GLIM macro NEGBIN. The model parameters were then assessed for their statistical significance and contribution to the reduction in deviance. Following a successful outcome of these assessments, parameters were accepted and the model's goodness-of-fit statistics calculated. This whole process resulted in the selection of the best 2 or 3 alternative flow-based models (i.e. models in which only the traffic flow function is the explanatory variable).

At the next level of modelling, the best flow-based models were each extended and tested, in turn, with the simultaneous addition of all other road and traffic variables. Starting with an initial value of the over-dispersion parameter equal to the one estimated during the first stage for the given flow-based model, the comprehensive model was fitted and the individual parameters assessed for their significance and contribution to the reduction in deviance. Insignificant parameters were excluded one by one, starting with the most insignificant and the remaining variables refitted and reassessed until only the significant variables were left in the model. Subsequently, the final value of the over-dispersion parameter (κ) was estimated iteratively. Starting with the residuals produced by the initial fit, a new value of κ was estimated and the model refitted and the process was repeated until satisfactory closure[8,10]. From this point, the model's goodness-of-fit statistics were calculated and the model was then added to the list of alternative models.

3.2.3 Model evaluation

Three types of objective assessments were always made as part of the process of selecting the most appropriate and best fitting models. These were tests of significance of individual parameters, contribution of the individual parameters to the reduction in deviance and the overall goodness-of-fit of the models. These assessments constituted the key basis for the acceptance or rejection of models. The specific objective criteria used are discussed below.

3.2.3.1 Assessment of individual model parameters

Individual model parameters were generally assessed at two levels. The first test was to ensure that the estimated parameter coefficients were statistically significant. Thus, the ratio of the estimated coefficient to its standard error was required to pass the t-test at the 5 per cent level of significance. The other aspect was to examine whether a parameter's contribution to the reduction in deviance was significant. In other words, this was to assess whether the addition of the said parameter to the model increased the explanatory power of the model significantly. According to Summersgill *et al*[11], the difference in scaled deviance between two nested models with degrees of freedom df_1 and df_2 will be distributed like χ^2 with (df_1 - df_2) degrees freedom and can be used to assess the significance of adding one or more terms to a model. This procedure was applied and, at the required level of significance (5 per cent), the drop in deviance following the addition of one parameter should have been at least 3.84 (χ^2 with 1.0d.f.).

3.2.3.2 Explanatory power of the models

To describe how well the developed models fitted the data overall, two global goodness-of-fit measures were used. These measures were part of an extensive list developed by Fridstrom *et al*[4] for generalised Poisson regression models, which give a measure of the percentage of systematic (explicable) variation in the response variable that is explained by the models. The measures, the log-likelihood ratio index (ρ^2) and the "Freeman-Tukey R^2" (based on the Freeman-Tukey transformation residuals), were applied in a similar way to the coefficient of determination (R^2), as used in ordinary least-squares regression.

(a) Log-likelihood Ratio Index (ρ^2)

This parameter is given by the expression:

$$\rho^2 = 1 - [LL(\beta) / LL(0)] \qquad (3)$$

where LL(β) is the log-likelihood value of the fitted model and LL(0) the corresponding value for the model in which only the constant term is used.

Both LL(β) and LL(0) are the result of the logarithmic transformation of the likelihood function of the Negative Binomial models, which is maximised to obtain the coefficient estimates for parameters in the models[15]. The value 2[LL(β) - LL(0)] is equivalent to the deviance value discussed in the previous section. By definition, therefore, ρ^2 represents the additional variation in accident frequency explained by the given model relative to the model with the constant term alone (the "null model").

(b) The "Freeman-Tukey R^2" (R^2_{FT})

Using the Freeman-Tukey variance stabilising transformation (f_i) and the mean of its normal distribution function (ϕ_i) for a Poisson variable y_i with mean λ_i, Fridstrom *et al*[7] provide the following expression in which the deviates ($e_i = f_i - \phi_i$) can be estimated from the corresponding residuals:

$$\hat{e} = \sqrt{y_i} + \sqrt{y_i + 1} - \sqrt{4\hat{y}_i + 1} \qquad (4)$$

where, y_i is the observed value of the dependent variable (in this case, the 3-year accident frequency) at the i-th junction; the corresponding predicted value being $\hat{y}$

Subsequently, the R^2_{FT} (Freeman-Tukey goodness-of-fit) measure is expressed as:

$$R^2_{FT} = \frac{\sum_i (f_i - \bar{f})^2 - \sum_i \hat{e}_i^2}{\sum_i (f_i - \bar{f})^2 - n} \qquad (5)$$

where, e_i is the deviate computed for the i-th junction and n - the total number of junctions. Other parameters are as described above.

Equation (5) is the result of dividing the ordinary R^2 goodness-of-fit measure for the transformed variables by the maximally obtainable fit in a perfect Poisson model. Thus, the ratio provides a measure of the proportion of the systematic variation in accident frequency that is explained by the fitted model. Although this is one of many well-established measures of the global goodness-of-fit of accident prediction models, it is useful to bear in mind that the derivation is founded on the key property of the Poisson distribution, which equates the variance to the mean. This means that the amount of expected random variation in the response variable is treated as though it was constant. This is important because the scope of random variation is variable and, according to Mountain *et al*[17], is larger when the expected accidents are smaller.

4. MODEL RESULTS AND INTERPRETATION

The models were developed separately for X- and T-junctions. Such grouping of the junctions was to ensure that the models would capture more accurately the apparent differences in accident patterns and risks associated with different layout designs. Also, because different exposure functions and variables were usually involved, it was not always possible to identify one single "best fit" model. Therefore, as much as possible, a number of alternative "best models" were selected and are presented to enable comparison.

4.1 X-junction models

A total of 238 accidents were recorded for all the 34 X-junction sites included in the database for the study period 1996 to 1998 inclusive. The average number of accidents per junction was therefore 7.0. The best fitting models identified for X-junctions are presented in their linear form in Table 3.

For the coarse (flow-based) models it was observed that most traffic flow functions tested produced reasonably good statistical fit to the data. However, as evident from Table 3, flow-functions involving interacting traffic streams like the sum of crossing flow products (CFPD) and encounter flow products (ENCP) appeared to fit the data a good deal better than the rela-

Table 3 Accident prediction models for X-junctions (Total number of accidents=238; number of sites=34)

Model Description	Model Terms*	Estimated Coefficient	Standard Error of Estimate	*t*-statistic	Freeman-Tukey R^2	Log-Likelihood ratio (ρ^2)
1. Null model	Lk	1.946	0.179	10.872		LL(0) =
Dispersion parameter	κ	*1.054*	*0.310*	*3.400*		-204.9**
2. Flow-based models						
(a)	Lk	-6.444	2.834	-2.274	0.27	0.050
	LTINF	0.965	0.308	3.133		
	LMRSH	0.669	0.262	2.553		
Dispersion parameter	κ	*1.595*	*0.536*	*2.976*		
(b)	Lk	-6.758	2.695	-2.508	0.24	0.043
	LCFPD	0.496	0.155	3.200		
Dispersion parameter	κ	*1.518*	*0.503*	*3.018*		
(c)	Lk	-6.257	2.660	-2.352	0.21	0.040
	LENCP	0.465	0.152	3.059		
Dispersion parameter	κ	*1.472*	*0.483*	*3.048*		
3. Full Models						
(a)	Lk	-5.988	2.529	-2.368	0.89	0.198
	LTINF	0.453	0.293	1.546		
	LMRSH	0.949	0.319	2.975		
	LFT(2)	1.319	0.428	3.082		
	MEDW	0.335	0.174	1.925		
	HGV	0.185	0.076	2.434		
	JNEC	0.134	0.039	3.436		
Dispersion parameter	κ	*3.595*	—			
(b)	Lk	-9.419	2.230	-4.224	0.91	0.223
	LCFPD	0.370	0.132	2.803		
	STL(2)	0.580	0.239	2.427		
	LFT(2)	0.661	0.286	2.311		
	HGV	0.190	0.071	2.676		
	JNEC	0.134	0.036	3.722		
	SSD	0.100	0.042	2.381		
Dispersion parameter	κ	*4.650*	—			
(c)	Lk	-9.111	2.277	-4.001	0.89	0.215
	LENCP	0.349	0.134	2.604		
	STL(2)	0.640	0.246	2.602		
	HGV	0.183	0.073	2.507		
	JNEC	0.135	0.038	3.553		
	SSD	0.100	0.043	2.326		
	LFT(2)	0.666	0.295	2.258		
Dispersion parameter	κ	*4.250*	—			

* The prefix "L" indicates that the parameters are still in their logarithmic forms, e.g. LTINF = Log (TINF).
** LL(0) - Log-likelihood value for null model

tively simple ones like total inflow (TINF) or major road flow (MAJF) and minor road flow (MINF). Crossing flow products (CFPD), is obtained by summing up all products (multiplication) of each pair of traffic flows (expressed as AADT), whose normal paths through the junction cross each other. Encounter flows on the other hand, include crossing as well as diverging and merging flows. Total inflow produced a considerably better fit only when specified alongside a flow-function reflecting the proportion of minor road traffic inflow. Thus the three best fitting flow-based models (see section 2 of Table 3) in their exponential form were:

$$A = 1.59x10^{-3}\ TINF^{\ 0.965}\ MRSH^{\ 0.669} \quad (6)$$

$$A = 1.16x10^{-3}\ CFPD^{\ 0.496} \quad (7)$$

$$A = 1.92x10^{-3}\ ENCP^{\ 0.465} \quad (8)$$

where A is the 3-year accident frequency,

TINF – total 24-hour traffic inflow to junction,

CFPD – crossing flow products (i.e. sum of the products of all crossing flows),

ENCP – encounter flow products (sum of the products of all encounter flows), and

MRSH – minor road's share of total junction traffic (i.e. MINF/TINF)

Although these are "all accidents" models (i.e. for estimating total junction accidents), it was interesting to observe that the pedestrian flow function (PEDF), a key exposure variable for pedestrian accidents, was not found to be significant. This probably had to do, in part, with the rather rough estimates of pedestrian flows used (only peak hourly counts of pedestrians) as well as the fact that pedestrian accidents comprised only 1 in 5 of all accidents at X-junctions. Nonetheless, it was assumed that pedestrian accidents within the model estimate would be accounted for by the relevant or associated vehicular traffic as there cannot be a pedestrian accident unless it was a collision with a vehicle.

Since most variables with potentially significant impact on accidents are not included in the flow-based models, such models may only be regarded as relatively coarse and rough estimators of accident frequency. That the selected models explained between 20-30 per cent of the systematic variation in accident frequency underscores the importance of traffic flow as a major determinant of accidents. It is evident from all the models that accident frequency generally increases at a decreasing rate with traffic flow. In the model represented by Equation 6 accident frequency was almost proportional to total junction vehicle inflow (the exponent for this parameter was close to 1.0) at the same time as it followed the general trend with respect to the minor road's share of traffic. In order to determine causal models, as we set out to do under this study, an extensive list of other traffic variables and factors describing the junction environment and geometry had to be tested simultaneously with the best flow-based models.

The three alternative full (comprehensive) models obtained out of this process are also presented in their linear form in Table 3. There is, apparently, not much to choose between these three models. All of them consistently produced very good *t*-statistics for individual parameters, at the same time as explaining about 90 per cent of the systematic variation in accident frequency. On account of the explanatory power and fewer degrees of freedom utilized, model 3b (see Table 3) was the most preferred. In the exponential form, this model is:

$$A = 8.12x10^{-5}\ CFPD^{\ 0.370}\ e^{(0.580STL(2)+0.661LFT(2)+0.190HGV+0.134JNEC+0.100SSD)} \quad \text{.......... (9)}$$

where A is the 3-year accident frequency at X-junctions and the parameters as defined in Table 2.

Apart from the traffic flow function, all the other variables, which appeared in this model, were consistently significant in most of the models explored for estimating total accidents at X-junctions. These variables were, left turn lane on the major road (LFT), proportion of heavy goods vehicles and buses as a percentage of total traffic inflow (HGV) and the standard deviation of average spot speeds on the major approaches (SSD). The others were streetlights (STL) and the average width of the minor road at the neck of the junction (JNEC). Given their stability and consistency, these variables could be considered as representing causal rather than associative effects.

The preferred model (Equation 9) showed that, when the impact of the other variables was considered, the absence of dedicated left-turn lanes on the major road (LFT(2)) increased accident frequency by a factor of 1.94, whilst the absence of street lights (STL(2)) resulted in an increase of 1.79 times. Not surprisingly, the full models had a much better fit to the data than the flow-based models. By fitting the extra variables, the explanatory power of the models increased from between 20 and 30 per cent to about 90 percent.

Although the log-likelihood ratio values for the models appeared low, they nonetheless compare rather favorably with what has been widely reported in the literature. It is also useful to remember that the log-likelihood ratio statistic measures the extra amount of variation in accident frequency explained by the given model, relative to the model with only the constant term. The deviance measure, proportion of systematic variation explained and the log-likelihood ratio statistics, as used above, are important tools that helped to identify generally good quality models that represented the key features of the overall data using as few parameters as possible. Important as they were, these indicators reflected only the global (overall) fit of the models and might not necessarily have reflected a good local fit to all individual data points as well. It was necessary therefore, to test how well the model fitted the individual data points. This could have been done by plotting a graph of predicted values against observed and these values should generally be similar and follow the line of equality. However, such straightforward comparison can be misleading since the values of the raw residuals (differences between predicted and observed values) could merely be a reflection of the size of the original observation, thus large residuals could emerge solely because the original observed values were large.

A more reliable approach involves standardization of the residuals under a common scale and plotting them against their Normal ordered statistics with a mean of 0 and standard deviation of 1. In this case residuals (standardized) lying outside plus or minus 2 (95 percent confidence interval) could be considered poor fits and coming from a potential outlier[18]. Such a plot is provided by the "Normal Q-Q" plot of the GLIM Macro Library and is presented in Figure 2 for the chosen X-junction full model. The plot amply demonstrates that the selected model equally well fitted all individual data points as all of them lie within the 95 percent confidence interval of the Normal order statistics. The other important piece of information from the plot is that the generalized Poisson assumption used in the modeling process was very appropriate. This is supported by the Filliben correlation coefficient of 0.95, evidence of a straight-line relationship between the Normal ordered statistics and the transformed residuals.

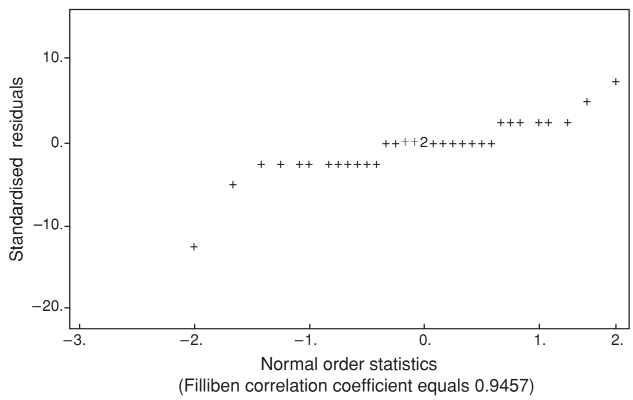

Fig.2 Normal Q-Q plot for X-junction full model to demonstrate the fit to individual data points

4.2 T-junctions models

354 accidents were recorded at all the 57 T-junction sites in the database for the period 1996-1998 inclusive. This translated into an average 3-year accident frequency per junction of 6.21. Similar procedures for model development as used for X-junctions were applied for selecting the best accident models for T-junctions and as was the case with the "all accident" models for X-junctions, most of the large variety of vehicle flow functions tested for T-junctions yielded statistically significant fits to the data. The pedestrian flow function (PEDF) was again not significant when combined in the appropriate form with the vehicle flow functions. The selected models are presented in their linear form in Table 4 and discussed below.

The flow-based models were:

$$A = 5.09x10^{-4}\, XPDF^{\,0.552} \qquad (10)$$

$$A = 6.37x10^{-4}\, MAJF^{\,0.501}\, MINF^{\,0.583} \qquad (11)$$

$$A = 7.99x10^{-4}\, TINF^{\,1.032}\, MRSH^{\,0.505} \qquad (12)$$

where A is the expected 3-year accident frequency at T-junctions,

XPDF – cross product of flows (i.e. product of the major (MAJF) and minor (MINF) road daily in flows),

TINF – the total 24-hour traffic inflow to the junction), and

MRSH – the minor road's share of total junction traffic

These models mean that the expected total accident frequency at T-junctions increased approximately as a function of the square root of the vehicle exposure functions XPDF, MAJF, MINF and MRSH. The exception was total junction traffic inflow (TINF), to which the expected accident frequency was almost directly proportional. The exponent value for the latter variable, as in Equation (model) 12, does not differ substantially from 1.0.

Of the three alternative models, the one based on the cross product of flows function (Equation 10) was the most preferred, because it used one less degree of freedom than the others and still managed to produce one of the highest proportion of systematic variation explained (i.e. 37 per cent). The model's log-likelihood ratio statistic was also relatively high. The alternative full models involving extensions of the flow-based models shared similar characteristics as the core flow-based models. Thus, based on similar considerations as before, the full model built on the cross product of flows exposure function (XPDF), emerged as the preferred one. This model was:

$$A = 1.01x10^{-3}\, XPDF^{\,0.514}\, e^{(0.0694SSD-0.465TCON(2)-0.952TCON(3)-0.151MEDW)} \qquad (13)$$

where A is the expected 3-year frequency of accidents at T-junctions,

Table 4 Accident prediction models for T-junctions (Total number of accidents = 354; number of sites = 57)

Model Description	Model Terms*	Estimated Coefficient	Standard Error of Estimate	*t*-statistic	Freeman-Tukey R^2	Log-Likelihood ratio (ρ^2)
1. Null model	Lk	1.826	0.131	13.918	—	LL(0) =
Dispersion parameter	κ	*1.218*	*0.298*	*4.089*		-330.3**
2. Flow-based models						
(a)	Lk	-7.583	1.837	-4.128	0.37	0.063
	LXPDF	0.552	0.108	5.102		
Dispersion parameter	κ	*2.154*	*0.631*	*3.414*		
(b)	Lk	-7.358	1.957	-3.760	0.37	0.064
	LMAJF	0.501	0.184	2.720		
	LMINF	0.583	0.141	4.146		
Dispersion parameter	κ	*2.159*	*0.633*	*3.412*		
(c)	Lk	-7.132	2.068	-3.449	0.33	0.059
	LTINF	1.032	0.226	4.564		
	LMRSH	0.505	0.149	3.385		
Dispersion parameter	κ	*2.065*	*0.596*	*3.465*		
3. Flow- geometry-factors						
(a)	Lk	-7.158	1.881	-3.805	0.50	0.096
	LMAJF	0.573	0.198	2.894		
	LMINF	0.480	0.132	3.636		
	SSD	0.0691	0.0360	1.919		
	TCON(2)	-0.480	0.224	-2.143		
	TCON(3)	-0.953	0.338	-2.820		
	MEDW	-0.166	0.082	-2.024		
Dispersion parameter	κ	*3.008*	*0.972*	*3.095*		
(b)	Lk	-6.897	1.695	-4.069	0.49	0.096
	LXPDF	0.514	0.098	5.245		
	SSD	0.0694	0.0358	1.939		
	TCON(2)	-0.465	0.223	-2.085		
	TCON(3)	-0.952	0.338	-2.817		
	MEDW	-0.151	0.0703	-2.148		
Dispersion parameter	κ	*2.996*	*0.967*	*3.098*		
(c)	Lk	-6.962	2.001	-3.479	0.46	0.091
	LTINF	1.001	0.216	4.634		
	LMRSH	0.390	0.148	2.635		
	SSD	0.0677	0.0367	1.845		
	TCON(2)	-0.493	0.230	-2.143		
	TCON(3)	-0.971	0.342	-2.839		
	MEDW	-0.160	0.084	-1.905		
Dispersion parameter	κ	*2.829*	*0.895*	*3.161*		

* The prefix "L" indicates the natural logarithmic form of the variable, i.e LXPDF = Log (XPDF).
** LL(0) is the log-likelihood value of the null model

TCON(2) – parameter representing traffic control level 2 (i.e.YIELD) on the minor road,
TCON(3) – level 3 of traffic control on the minor road (i.e. no control),
MEDW – average width of the median on the major road and
SSD – average standard deviation of vehicle spot speeds on the major approaches.

The Normal Q-Q plot for this model as shown in Figure 3 also demonstrates that apart from meeting the global goodness of fit measures, the model also fitted the individual data points very well. It is significant to observe that two of the three additional variables included in the model, i.e. traffic control on the minor road and the average width of the median on the major, had negative signs to their parameter estimates. This means that the said parameters were negatively correlated to the expected accident frequency at T-junctions, in which case the presence of the stated traffic control type and increas-

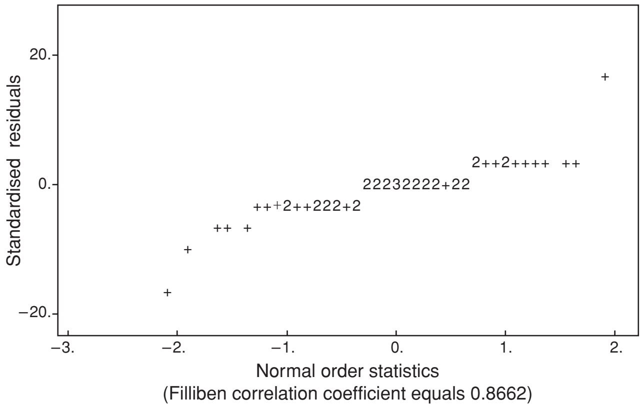

Fig. 3 Normal Q-Q plot for T-junction full model to demonstrate the fit to individual data points

ing values of MEDW would lead to less accident frequency. The specific effects of traffic control as captured in the model was that, when the type of control at the minor road was TCON(2), which represented the YIELD sign, the accident frequency reduced by a factor of 0.63. On the other hand, TCON(3) (i.e. no control) on the minor road was associated with a reduction in accident frequency by a factor of 0.39. These effects were relative to level 1 of traffic control (i.e. TCON(1)), which represented STOP control on the minor approach road.

This finding, controversial as it might appear, confirmed observations made about the relative safety records of the three types of unsignalised junction control following initial analysis of the accident characteristics and associated factors[27]. Thus, the confounding question about the safety record of STOP control remains unanswered. The model confirmed that, at least as far as the modelling database was concerned, STOP control was associated with the worst impact on accident potential at T-junctions. It may as well be that the level of control at the particular junctions might have been stepped up to STOP control in response to a bad accident situation in the first place. But since, as is apparent, the intervention appears not to have improved the situation, it is entirely appropriate to question the effectiveness of the STOP control as an accident remedial measure. This is an important concern as it touches at the heart of long-established codes of practice, as set out in safety/accident warrants, which until now, have taken for granted the relative safety benefits of increasing the level of control at unsignalised junctions from no control at all, through Yield to Stop.

The impact of the other parameters on accident frequency in the model appeared fairly straightforward and logical. It is not incomprehensible, for example, that, the width of the median on the major road would be related to fewer accidents, considering junctions of equivalent traffic with and without the median. On the other hand, large values of SSD would suggest wide variability and extremes in approach speeds of vehicles, leading to less predictability and poor mutual anticipation between drivers. Such an atmosphere would breed more conflicts and potentially lead to more accidents.

Interestingly, the proportion of systematic variation in accident frequency explained by the full models was only about 10 per cent more than their corresponding flow-based models and generally only about half the percentages achieved for the models for X-junctions. This was most probably due to the fewer additional parameters accepted on account of their significance into the full models for T-junctions. Also, the contribution of the individual parameters to the reduction in model deviance, although statistically significant, was generally less than the levels attained for X-junctions. By implication, therefore, accidents at T-junctions are much less dependent on road geometric and other traffic variables outside the traffic exposure function. Nonetheless, the 50 per cent proportion of systematic variation explained by the full models was still good by most standards reported in the literature.

5. CONCLUSION

On the whole, the results of modeling showed that traffic exposure functions such as the cross product of flows (XPDF), sum of crossing flow products (CFPD) and the sum of encounter flow products (ENCP) produced much better fit to the accident data than simpler flow functions like the total junction traffic inflow (TINF). The most influential traffic exposure function for X-junction accidents was the sum of the crossing flow products (CFPD), whilst the cross product of minor and major road traffic inflows (XPDF) influenced accidents at T-junctions most. The best flow-based models for T-junctions had about one-and-a-half times more “proportion explained” than those obtained for X-junctions. The three most consistent additional variables that featured in the extended accident models for X-junctions were street lighting and dedicated left–turning lanes, as well as the average standard deviation of approach spot speeds of vehicles on the major road. Those for T-junctions were level of traffic control, average width of the median on the major road and the average standard deviation of vehicle approach spot speeds on the major road. The absence of street lighting and dedicated left-turning lanes and the average standard deviation of vehicle approach spot speeds were all positively correlated with accident frequency.

Interestingly, the accident potential of T-junctions that had YIELD or no control was adjudged to be much lower than that of similar sites with STOP control. This particular result has cast doubt on the prudence in taking for granted the safety benefits of increasing junction control from no control at all through Yield to Stop control as recommended by the relevant accident/safety warrants. Given the large variety of variables tested in the model estimation process, the quality of the models obtained and the consistency of the additional variables, it can be concluded that the full models developed represented causal rather than associative relationships. The models can be used subsequently, therefore, for the reliable prediction of accident frequency associated with the junction types and features described. Their use in this manner will facilitate a more proactive and cost-effective management of traffic safety and accident blackspots in the urban environment.

REFERENCES

1. Road traffic accidents in Ghana – Statistics 2000, Building and Road Research Institute. (2002).
2. Ghee, C., Silcock, D., Astrop, A. and Jacobs, G., Socio-economic aspects of road accidents in developing countries, Transport Research Laboratory Report 247, pp29. (1997).
3. Ministry of Roads and Transport, National Road Safety Strategy 2001-2005, National Road Safety Commission, pp41. (2000).
4. Fridstrom, L., Ifver, J., Ingebrigsten, S., Kulmala, R. and Thomsen, L.K., Measuring the contribution of randomness, exposure, weather and daylight to the variation in road accident counts, Accident Analysis and Prevention, Vol.27, No.1, pp1-20. (1995).
5. Tanner, J.C., Accidents at rural 3-way junctions, Journal of Institution of Highway Engineers, Vol.2, No.11, pp56-67. (1953).
6. Hakkert, A.S. and Mahalel, D., Estimating the number of accidents at intersections from the knowledge of traffic flows from the approaches, Accident Analysis and Prevention, Vol.10, No.1, pp69-79. (1978).
7. Leong, H.J.W., Relationship between accidents and traffic volumes at urban intersections, Australian Road Research, Vol.5, No.3, pp72-90. (1973).
8. Hauer, E., Ng, J.C.N. and Lovell, J., Estimation of safety at signalized intersections, Transportation Research Record 1185, Transportation Research Board, Washington D.C., pp48-61. (1989).
9. McGuigan, D.R.D., The use of relationships between road accidents and traffic flow in “blackspot” identification, Traffic Engineering and Control, Vol.22, Nos.8/9. pp448-453.
10. Mountain, L. and Fawaz, B., Estimating accidents at junctions using routinely available input data, Traffic Engineering and Control, Vol.37, No.11, pp624-628. (1996).
11. Summersgill, I., Kennedy, J.V. and Baynes, D., Accidents at three-arm priority junctions on urban single-carriageway roads, TRL Report 184, Transport Research Laboratory, Crowthorne, pp132. (1996).
12. Kulmala, R., Safety at rural three- and four-arm junctions, development and application of accident prediction models, Technical Research Centre of Finland, VTT 233, 104p. (1995).
13. Poch, M. and Mannering, F., Negative binomial analysis of intersection accident frequencies, Journal of Transportation Engineering, Vol.122, No.2, pp105-113. (1996).
14. Miaou, S. and Lum, H., Modelling vehicle accident and highway geometric design relationships, Accident Analysis and Prevention, Vol.25, No.6, pp689-709. (1993).
15. McCullagh, P. and Nelder, J.A., Generalised Linear Models: Monographs on Statistics and Applied Probability, Vol.37, Chapman and Hall. (1989).
16. Hills, B., Elliot, G. and Clarke, D., Microcomputer accident analysis package v5.0 (MAAP5) user guide. Transport Research Laboratory, Crowthorne. (1994).
17. Jovanis, P.P. and Chang, H-L, Modelling the relationship of accidents to miles traveled, Transportation Research Record 1068, Transportation Research Board, Washington D.C., pp42-51. (1986).
18. O’Brian, L., Introduction to quantitative geography: Measurements, methods and generalized linear models, Chapman & Hall, New York, pp202-266. (1992).