Motorcycle crash prediction model for non-signalized intersections
S. Harnen, R.S. Radin Umar, S.V. Wong and W.I. Wan Hashim
1. INTRODUCTION

Crashes at intersections continue to pose a problem in many countries\(^1\). In the USA, there were 2,721,000 intersection related crashes in 2001, accounting for 43% of the total crashes in the country\(^2\). Of these, 8,490 were fatal while 970,000 crash victims sustained injuries. Meanwhile, in Japan, 3,813 fatal crashes occurred at intersections, constituting more than 45% of all fatal crashes in 2001\(^3\). Of these, almost 13% were fatal crashes involving motorcycles and mopeds. In the UK, traffic crashes at T, Y or staggered intersections and crossroads stood at 181,230, constituting about 43% of total crashes in 2001\(^4\). Of these, 8% involved motorcycle crashes. Given this scenario, many transportation and other agencies concerned with safety, are developing plans and programs to address such traffic safety problems\(^1,5,6\).

Previous studies on traffic crashes at intersections were largely concerned with all vehicle-related crashes\(^7\)–\(^10\). However, a breakdown of statistics from many studies revealed that motorcycle crashes at intersections are also significant\(^11\)–\(^15\). Furthermore, the probability of a fatality or injury to motorcyclists is much higher compared to passenger car occupants\(^12,13,16\). As such, a more extensive investigation on motorcycle crashes at intersections is strongly justified.

In Malaysia\(^17\), motorcycles constitute more than half of the total registered vehicles in the country. Motorcycle casualties constitute more than 60% of the total casualties in traffic crashes. Of these, 29% were fatal, 56% required hospitalization and 15% sustained slight injuries. In the period 1990-2000, almost 3,000 motorcyclists were killed annually in traffic crashes. Furthermore, motorcycle casualties are more than seven-fold compared to passenger car occupant casualties. Given the high number of motorcycle crashes, the Malaysian government provided for an exclusive lane for motorcyclists alongside trunk roads in the country. Since implementation of the exclusive motorcycle lane, a number of studies have been carried out to evaluate the impact of this facility on motorcycle crashes along the lane\(^12,18,19\). The reduction in motorcycle crashes was found to be highly significant (p<0.05) following the provision of this lane, with an average reduction of about 39%. However, not much work has been done on motorcycle crashes at in-
crashes, traffic flow, pedestrian flow, traffic speed, intersection geometry, number of intersecting legs and land use were assembled and used in this study. Motorcycle crash data over a 4-year period (1997-2000) were collected from the archives of the police crash recording forms, POL 27 (Pin 1/91). This form is designed for easy completion and fully compatible with the TRL's Microcomputer Accident Analysis Package, the MAAP.

Data were extracted from two complementary databases that are based on the POL 27 recording forms: (a) the Microcomputer Accident Analysis Package (MAAP-5) database for fatal and hospitalized crashes, and (b) the Computerized Accident Recording System (CARS-2000) database for slight injury crashes. The MAAP database is located at the Road Safety Research Center, Universiti Putra Malaysia, while the CARS-2000 database is located at the Royal Malaysian Police Headquarters.

Traffic flow data, as defined by Annual Average Daily Traffic (AADT), consists of total traffic entering the intersection through the major- and minor-road approaches. For analysis purposes, traffic flow recorded from each approach of the selected intersection was disaggregated by non-motorcycle and motorcycle flows. Records of hourly traffic flow were then converted to AADT by using hourly, daily and monthly factors. These factors were estimated using the 24-hour permanent traffic count station and traffic census data available at the Highway Planning Unit, Ministry of Works Malaysia. Established techniques commonly used to estimate the factors and the AADT were employed in this study. AADTs are expressed in terms of number of non-motorcycles per day (nmpd) and motorcycles per day (mpd). The term non-motorcycle refers to all types of motorized vehicles excluding the motorcycle.

Meanwhile, data on approach speed and pedestrian flow were collected on site because of the non-availability of such data in the database. Data were obtained on site based on similar assumptions used in an earlier study. Approach speed is defined by the 85th percentile speed measured at a distance 50 m upstream from the corresponding stop line of the intersection. Such an approach speed was also considered in earlier studies on traffic crashes. The pedestrian flow at each intersection is defined as the total number of pedestrians crossing per hour, counted on major and minor roads.

Meanwhile, intersection geometry (lane width, number of lanes and shoulder width), number of intersecting legs and land use at each selected intersection were observed and recorded during site data collection. Of the 53 intersections, 36 were three-legged while 17

2. DATA ACQUISITION

In this study, a total of 53 intersections on urban roads in four districts of the state of Selangor, Malaysia were randomly selected. The selection of intersections was based on the criteria that the intersections had only marginal land use changes; had not undergone major modifications or upgrading; there was an equal number of lanes on the corresponding major- and minor-road approaches. Intersections located within commercial areas with access roads within 50-m distance from the intersection stop line were excluded from this study. Only those intersections with a history of Personal Injury Accident (PIA) statistics were included in this analysis. This is because of the reliability of data and higher weightage given to PIA cases compared to damage-only crashes. In the present study, intersection crashes involving motorcycles were defined as any motorcycle crash occurring within a 50-m distance from the corresponding stop line of the intersection.

For the selected intersections, data on motorcycle

intersections both along the exclusive and non-exclusive lanes. Hence, it is necessary to carry out an in-depth investigation into factors contributing to motorcycle crashes at intersections. Apart from addressing the exclusive lane criteria, a further outcome of such research would be the development of intersection treatment criteria dedicated to exclusive and non-exclusive motorcycle lane facilities.

Earlier models on traffic crashes used the classical linear regression approach that is based on normal error structure with a constant variance. Recently, there has been widespread acceptance of the use of Generalized Linear Models (GLMs) with a Poisson or Negative Binomial error structure considered as more appropriate. Earlier studies have reported the use of GLMs on the development of predictive models for traffic crashes using either the cross-sectional or the time series analysis.

This paper presents the development of a prediction model for motorcycle crashes at non-signalized intersections on urban roads in Malaysia using the Generalized Linear Modeling approach. Both the Poisson and negative binomial errors were incorporated to refine the model. The statistical software, namely GLIM, which was specifically designed for fitting Generalized Linear Models, was used to estimate the parameters and significance tests.

Meanwhile, intersection geometry (lane width, number of lanes and shoulder width), number of intersecting legs and land use at each selected intersection were observed and recorded during site data collection. Of the 53 intersections, 36 were three-legged while 17
were four-legged intersections. Land use was classified into two categories: (1) non-commercial areas, and (2) commercial areas. The non-commercial area was identified as being residential or unused. Commercial areas were identified as those areas with concentrations of offices, shops, railway stations and bus stations. Of the 53 intersections, 34 intersections were located in commercial areas and 19 in the non-commercial areas.

3. THE MODEL

The Organization for Economic Development and Cooperation (OECD) suggests that the choice of regressors should primarily be based on the theory used, the question to be answered, and on the professional knowledge rather than the multiple correlation and curve-fitting ambition. Taking the above suggestions into account, the variables of the model were defined. In this study, the response variable was the number of motorcycle crashes and the explanatory variables were traffic flow, pedestrian flow, traffic speed, lane width, number of lanes, number of intersecting legs, shoulder width and land use. The continuous variables were identified as traffic flow, pedestrian flow, traffic speed, lane width and number of lanes. The categorical variables were the number of intersecting legs (NL) with 2-factor levels, shoulder width (SHDW) with 3-factor levels, and land use category (LU) with 2-factor levels. It should be noted that the number of lanes for all approaches of the observed minor road was equal to one (LNn=1). The model therefore excluded this variable as one of the explanatory variables.

A theoretical model containing all the terms used in this study was formulated as follows:

$$MCA = k QNMm^C QNMn^C QMm^C QMn^C QPED \exp\left[\beta_1\text{SPEED} + \beta_2\text{LNn} + \beta_3\text{LNm} + \beta_4\text{NL} + \beta_5\text{SHDW} + \beta_6\text{LU} + \epsilon \right]$$

(1)

where MCA is motorcycle crashes (response variable), while all the independent explanatory variables and their description is presented in Table 1. The $k$, $\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_4$, $\alpha_5$, $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$, $\beta_5$ and $\beta_6$ are the coefficients to be estimated and the $\epsilon$ term is the error representing the residual difference between the actual and predicted model. Note that model structure of Equation (1) consists of both additive and multiplicative forms of the model variables. This model structure has been used by authors in earlier studies on modelling of motorcycle crashes in Malaysia.

Using the natural logarithmic transformation, the log-linear version of the model is:

$$\ln(MCA) = \ln(k) + \alpha_1\ln(QNMm) + \alpha_2\ln(QNMn) + \alpha_3\ln(QMm) + \alpha_4\ln(QMn) + \alpha_5\ln(QPED) + \beta_1\text{SPEED} + \beta_2\text{LNm} + \beta_3\text{LNn} + \beta_4\text{NL} + \beta_5\text{SHDW} + \beta_6\text{LU} + \epsilon$$

(2)

It should be noted that total four-year crash frequen-

<table>
<thead>
<tr>
<th>Table 1 Descriptions and factor levels of the explanatory variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Explanatory Variables</strong></td>
</tr>
<tr>
<td>QNMm</td>
</tr>
<tr>
<td>QNMn</td>
</tr>
<tr>
<td>QMm</td>
</tr>
<tr>
<td>QMn</td>
</tr>
<tr>
<td>QPED</td>
</tr>
<tr>
<td>SPEED</td>
</tr>
<tr>
<td>LWm</td>
</tr>
<tr>
<td>LWn</td>
</tr>
<tr>
<td>LNm</td>
</tr>
<tr>
<td>LNn</td>
</tr>
<tr>
<td>NL</td>
</tr>
<tr>
<td>SHDW</td>
</tr>
<tr>
<td>LU</td>
</tr>
</tbody>
</table>
cies were used to fit the models. However, by introducing an offset variable in the fitting process, the final model would yield the number of crashes per year. This approach was also utilized in earlier studies on traffic crashes at intersections. To allow direct interpretation of the estimates of traffic flow terms produced by GLIM 4, the flow functions need to be transformed into logarithmic form i.e. using Ln(Q), while the non-flow functions need no transformation.

In the present study, the Poisson and negative binomial error distributions were incorporated to refine the model. The method suggested earlier was adopted to justify the error distribution used for the analysis. The study suggested that the choice of error distribution should be based on the goodness of fit test carried out on the observed crash frequencies under study. The deviance was used as a measure of the goodness of fit. The minimum deviance generated in the fitting process was considered, as the observed frequencies are the closest fit to the theoretical frequencies being analyzed. In addition, a hypothesis test at 95% confidence level was carried out on the selected error distribution.

The quasi-likelihood approach was used to overcome the dispersion problem, as this had been successfully used in earlier studies on motorcycle crashes. In this approach, the dispersion parameter was estimated from the mean deviance (scaled deviance over its degrees of freedom). This may lead to a model with scaled deviance equal to its degrees of freedom. Such an approach was extensively described in the GLIM System Release 4 Manual.

Both multivariate and univariate analyses were carried out in this study. The multivariate analysis was employed to assess which of the variable(s) had the most effect on the probability of motorcycle crashes. Meanwhile, the univariate analysis was carried out to obtain a complete picture of the effect of all explanatory variables on motorcycle crashes. Only those variables found significant at the 5% level in the univariate analysis were then included in the multivariate analysis. The development of the final model was based on the goodness of fit and the significance test carried out. They are the change in scaled deviance from adding or removing the terms, the ratio of scaled deviance to its degrees of freedom (mean deviance) and the 5% significant level of the t-statistic of the parameter estimates. The deviance is expressed in terms of a parameter D(y;μ), which is defined by:

\[ D(y;μ) = 2 \left( \frac{y}{μ} - 1 - \ln(μ) \right) \]  

The analysis carried out on two-error distributions revealed that the Poisson was slightly better in explaining the variation of crash occurrence than the negative binomial. The deviance (D) for the Poisson model was 4.46 with 11 degrees of freedom (df), while the deviance (D) for the negative binomial model was 4.52 with 10 degrees of freedom (df). The hypotheses tests also affirmed that the Poisson error was statistically significant (p<0.05) in representing the distribution of crashes under study.

The univariate analysis for the model showed that all terms, except the terms QPED and NL, were significant at the 5% level. As the terms QPED and NL were not significant, they were excluded from further analysis. Table 2 presents the multivariate analysis of the terms. It can be seen that all explanatory variables were significant at the 5% level. The scaled deviance was equal to its corresponding degrees of freedom, as the quasi-likelihood approach had been introduced in the fitting process. The scaled deviance changed from 12501.0 to 41.0 with a loss of 11 degrees of freedom. The mean deviance changed from 240.4 to 1.0.

Based on the multivariate analysis, the final model is as follows:

\[ MCA = 0.01315 \times QNMn_{0.1597} \times QNMn_{0.0975} \times QMn_{0.1071} \times QMn_{0.1356} \times EXP \left[ QPQPED - 0.0667 \times QPQPED - 0.0967 \times QPQPED - 0.0107 \times QPQPED - β_0SHDW + β_1LU \right] \]  

where MCA is motorcycle crashes per year, \( β_0 = 0.0, 0.01809 \) and 0.0502 for SHDW = 1, 2 and 3, respectively, \( β_1 = 0.0 \) and 0.01789 for LU = 1 and 2, respectively (Table 1). Figure 1 presents the actual motorcycle crashes compared with the ones modeled. It can be seen that the final model fits the data point satisfactorily.
Table 2 Multivariate analysis of the terms

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Estimates</th>
<th>Standard Error</th>
<th>Degrees of Freedom</th>
<th>Scaled Deviance</th>
<th>T-statistic</th>
<th>Sig.at 0.05</th>
<th>Mean Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-4.33100</td>
<td>0.40300</td>
<td>52</td>
<td>12501.0</td>
<td>-10.73</td>
<td>Yes</td>
<td>240.4</td>
</tr>
<tr>
<td>QNMm</td>
<td>0.15970</td>
<td>0.03400</td>
<td>51</td>
<td>773.4</td>
<td>4.70</td>
<td>Yes</td>
<td>15.2</td>
</tr>
<tr>
<td>QNMn</td>
<td>0.09730</td>
<td>0.01710</td>
<td>50</td>
<td>340.8</td>
<td>5.69</td>
<td>Yes</td>
<td>6.8</td>
</tr>
<tr>
<td>QMm</td>
<td>0.10710</td>
<td>0.01940</td>
<td>49</td>
<td>271.9</td>
<td>5.53</td>
<td>Yes</td>
<td>5.5</td>
</tr>
<tr>
<td>QMn</td>
<td>0.13360</td>
<td>0.03900</td>
<td>48</td>
<td>222.3</td>
<td>3.43</td>
<td>Yes</td>
<td>4.6</td>
</tr>
<tr>
<td>SPEED</td>
<td>0.02418</td>
<td>0.00286</td>
<td>47</td>
<td>92.6</td>
<td>8.44</td>
<td>Yes</td>
<td>2.0</td>
</tr>
<tr>
<td>LWm</td>
<td>-0.09670</td>
<td>0.03300</td>
<td>46</td>
<td>77.7</td>
<td>-2.93</td>
<td>Yes</td>
<td>1.7</td>
</tr>
<tr>
<td>LWn</td>
<td>-0.09070</td>
<td>0.03700</td>
<td>45</td>
<td>70.1</td>
<td>-2.45</td>
<td>Yes</td>
<td>1.6</td>
</tr>
<tr>
<td>LNm</td>
<td>-0.01079</td>
<td>0.00520</td>
<td>44</td>
<td>63.3</td>
<td>-2.07</td>
<td>Yes</td>
<td>1.4</td>
</tr>
<tr>
<td>SHDW (2)</td>
<td>-0.01809</td>
<td>0.00720</td>
<td>42</td>
<td>46.5</td>
<td>-2.51</td>
<td>Yes</td>
<td>1.1</td>
</tr>
<tr>
<td>SHDW (3)</td>
<td>-0.05020</td>
<td>0.01240</td>
<td>42</td>
<td>46.5</td>
<td>-4.06</td>
<td>Yes</td>
<td>1.1</td>
</tr>
<tr>
<td>LU (2)</td>
<td>0.01799</td>
<td>0.00765</td>
<td>41</td>
<td>41.0</td>
<td>2.34</td>
<td>Yes</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Note: Estimates for factors (2) and (3) are differences compared with the reference level (1)

Fig. 1 Actual and modeled motorcycle crashes

5. DISCUSSION

A final model has been developed for motorcycle crashes at non-signalized intersections on urban roads in Malaysia. The model reveals that traffic flow entering the intersection, traffic speed, lane width, number of lanes, shoulder width and land use are significant in explaining motorcycle crashes at intersections. The results support an earlier study. The effect of examined variables on motorcycle crashes are briefly described in the following paragraphs.

An increase in non-motorcycle and motorcycle flows on major and minor roads is associated with an increase in motorcycle crashes (Figure 2). This is indicated by the estimates of QNMm, QNMn, QMm and QMn. For instance, doubling non-motorcycle flow on a major road (QNMm) is expected to cause a 12% increase in motorcycle crashes. If there was doubling of all vehicles entering the intersection, the model would predict a 36% increase in motorcycle crashes. It was also found that the non-motorcycle flow on a major road (QNMm) had the
highest effect on the probability of motorcycle crashes. The effect of traffic flow on motorcycle crashes is in agreement with the finding reported in earlier studies on traffic crashes at intersections.\textsuperscript{7,9,23-36}

Meanwhile, the estimate of SPEED shows that an increase in approach speed is associated with an increase in motorcycle crashes. For instance, a 10 km/h increase in approach speed is expected to cause a 27% increase in motorcycle crashes. The result supports the findings of earlier studies on the relationships of traffic speed to traffic crashes.\textsuperscript{9,22,24,25,37}

A wider lane is associated with a reduction in motorcycle crashes. This is indicated by the estimates of LWm and LWn. For instance, widening the lane on major and minor roads by 0.50 m is expected to reduce motorcycle crashes by about 5% and about 4% respectively. This result is in line with the findings reported in an earlier study on traffic crashes at intersections.\textsuperscript{9}

The estimate of LNm implies that an increase in the number of lanes on a major road is associated with a reduction in motorcycle crashes. However, the effect of number of lanes on motorcycle crashes is marginal, only about 1%. The result appears to be in line with the finding of an earlier study on traffic crashes at intersections.\textsuperscript{9} This reduction is probably a result of the presence of an exclusive turning lane on the intersection, as out of the 53 intersections, 29 intersections were furnished with an exclusive turning lane on a major road. The presence of such lanes may reduce the rear-end crashes involving motorcycles. However, a more accurate explanation would be elicited by developing a separate model to explain the effect of exclusive turning lanes on any type of motorcycle crashes at intersections. The benefit of such lanes towards the reduction of crashes has been confirmed by earlier studies on traffic crashes at intersections.\textsuperscript{9,28,29} and traffic crashes at road segments.\textsuperscript{40}

A wider shoulder is associated with a reduction in motorcycle crashes. This is indicated by the estimates of SHDW. For instance, motorcycle crashes at intersections without a shoulder is about 57% higher than those at intersections with a shoulder width wider than 1.00 m. While motorcycle crashes at intersections without a shoulder is about 1% higher than those at intersections with a shoulder width of between 0.00 m and 1.00 m. This finding seems plausible since motorcyclists utilize available shoulders as a traveled path when approaching an intersection. This condition is common in countries with a high population of motorcycles such as Malaysia. In this situation, the rate of rear-end and sideswipe crash types between motorcycles on the shoulder and other vehicles on the traveled path is probably reduced. However, a better explanation can now be made since a separate model was developed to explain the effect of shoulders on any type of motorcycle crash at intersections. The finding, however, may encourage traffic engineers to enhance the benefits of paved shoulders in reducing motorcycle crashes at intersections. The investigation should focus on segregating motorcycles from other larger vehicles at in-
tersections by means of the provision of non-exclusive motorcycle lane facilities. Further analysis is also suggested to find out the magnitude of traffic flow at intersections where most motorcyclists preferred to use shoulders for riding rather than sharing the traveled way with other larger vehicles. This could reflect the demand for non-exclusive motorcycle lanes at intersections by motorcyclists. The result seems to be in line with the finding reported in an earlier study on traffic crashes at intersections. However, a better justification can be made since a separate model is developed to explain the effect of a shoulder on any type of motorcycle crashes at an intersection.

The estimate of LU indicates that the non-signalized intersections located within commercial areas are associated with an increase in the number of motorcycle crashes. This finding is in line with the results of an earlier study on traffic crashes at intersections. However, the effect of land use categories on motorcycle crashes showed only a marginal difference (1.8%). This is because there was no access road to the adjacent land within the distance of 50 m from the intersection stop line for the selected intersections located within commercial areas. As such, the number of conflicts between vehicles entering or leaving the intersection and vehicles turning in or turning out onto the adjacent land use may be reduced, hence reducing crashes. The effect of access control or the number of accesses on traffic crashes at intersections has also been reported in earlier studies.

6. CONCLUDING REMARKS

Based on the findings of this study, the following concluding remarks may be drawn:

(a) Traffic flow is significant in explaining motorcycle crashes at non-signalized intersections. An increase in non-motorcycle and motorcycle flows on major and minor road is associated with an increase in motorcycle crashes. Non-motorcycle flow on major road has the highest effect on the probability of motorcycle crashes.

(b) Approach speed is significant in explaining motorcycle crashes at non-signalized intersections. An increase in approach speed is associated with an increase in motorcycle crashes.

(c) Lane width, number of lanes and shoulder width are also significant in explaining motorcycle crashes at non-signalized intersections. Wider lanes, a greater number of lanes, and wider shoulders are associated with a reduction in motorcycle crashes.

(d) Land use at intersections is significant in explaining motorcycle crashes at non-signalized intersections. Motorcycle crashes at non-signalized intersections located within commercial areas is higher than those located within non-commercial areas.

The model developed in this study can be used to determine the appropriate intervention level for intersection treatment with respect to motorcycle crashes. Using the model developed in this study, appropriate design parameters of non-signalized intersections could be specified. The treatment could be the provision of non-exclusive motorcycle lanes facilities at intersections. However, this model might only be valid for a typical mixed traffic environment in developing countries like Malaysia, where the proportion of motorcycles using non-signalized intersections constitutes 20% to 40% of all vehicles.

REFERENCES


MOTORCYCLE CRASH PREDICTION MODEL FOR NON-SIGNALIZED INTERSECTIONS

S. HARNEN, R.S. RADIN UMAR, S.V. WONG, W.I. WAN HASHIM


ACKNOWLEDGEMENTS

This paper reports part of a study conducted under an IRPA Project entitled Development of Design Criteria and Standard for Malaysian Motorcycle Lanes. The financial support from the Ministry of Science, Technology and Environment Malaysia is gratefully acknowledged. The authors would like to thank the Royal Malaysian Police and the Highway Planning Unit, Ministry of Works, Malaysia for providing the data.