Accident Analysis Using Count Data for Unsignalized Intersections in Malaysia

Ashar Ahmed*a, Ahmad Farhan Mohd Sadullaha, Ahmad shukri Yahya

*aUniversity Sains Malaysia, Pulau Pinang 14300, Malaysia

Abstract

Accident data could be of various types. Count data being one of them. Effective accident analysis requires classification of count data with respect to geometric and control parameters. This paper focuses on the analysis of the effects of road width, land use, lane marking and traffic control on safety of unsignalized intersections. Intersection lying in non-urban areas with single line marking and no control on minor road were found to be the most vulnerable. The results are discussed and the recommendations are provided.

1. Introduction

Scientific study of accident occurrence and its causes has influenced the design of road infrastructure over the period of time, un-signalized intersections being one of them. As per this study they constitute a huge number among all the fixed control facilities provided on the Road Infrastructure of Malaysia. Hence they contribute equally in the total number of accidents that occur on Malaysian roads. This makes them an important area of
study. The type of data being analyzed dictates the model form used for the prediction of accidents and cause identification of their occurrence. Hence selection of appropriate data is vital in the modeling of specific parameters associated with accident occurrence. Accident data in Malaysia is recorded by Police on standard forms known as POL 27. It is then transferred to MIROS (Malaysian Institute of Road Safety Research) for digitization and development of accident database. The data recorded contains several attributes such as number of accidents, accident severity, vehicle type, road width, shoulder width, lane marking, control type, geometry, locale, weather and time of accident. This paper focuses on the analysis of count data of 448 unsignalized intersections in Malaysia. The geometric parameters that affect crash occurrence are examined and their results are discussed.

2. Literature Review

There are different ways to look into the subject of road safety and one of them is “if the user makes a mistake it does not mean he has to die for it”. Keeping in view the above statement as the primary goal of safety improvement of unsignalized intersections, many researchers had tried to make design improvements by modelling the geometric and traffic parameters that significantly affect the occurrence of crashes. Among them the prominent parameters are major and minor road volume, major and minor road approach speed, type of control such as Yield sign, Stop sign, Stop line or no control, gap/lag acceptance, traffic conflict, Post Encroachment Time (PET) and Time To Collision (TTC).

2.1. Volume as Measure of Safety

First, Reports on the accident analysis of cross roads and junctions are as old as 1950’s. The earliest accident analysis involved the measurement of indices such as number of accidents per left/right turning movement [01]. Apart from major and minor road volumes researchers started using STOP sign as a measure of intersection safety. As stop controlled intersections require the minor road drivers to come to a complete halt before entering an intersection, it was also argued that introduction of a stop sign also unnecessarily increases the number of accidents [02]. In a later study [03 and 04] it was concluded, using Generalized Linear Modelling (GLIM), that introduction of any kind of control measure to an uncontrolled intersection offers greater safety, especially at 4-leg intersections where STOP signs were introduced instead of no control. In late 1980s emphasis on other geometric features, along with major minor road volumes and STOP sign provisions, on accidents started taking place. These geometric features included clear sight distances from major and minor roads, grade, curve, type of median (raised, mountable, flush, none), raised pavement markers, rumble strips and separate turning lanes [05].

2.2. Speed as Measure of Safety

Use of gap acceptance as a parameter for modeling accidents had been argued in previous research [02], until it was used extensively by [06]. Unsignalized intersections are a facility which is usually provided on low volume roads. This provokes the drivers to increase their speeds. In a comprehensive study [07], involving the effect of time gap, speed and time to cross on the accident probability of minor stream vehicles a logit model was derived. The model was then extrapolated theoretically for conflict and accident probabilities respectively. It was found that the chances of a minor stream vehicle colliding with a major stream vehicle increases with the increase in major stream vehicles’ approach speed.

2.3. Alternative Measures

Since accidents occur rarely and those which occur are not 100% reported, therefore, an alternative was required to estimate the probability of accidents at intersections with little or no history of accident occurrence at all. Traffic conflicts were considered to be the solution to this problem; hence microsimulation technique was utilized to generate traffic conflicts at three legged and four legged unsignalized intersections in Italy using AIMSUN simulation software along with SSAM software [08]. Intersection related traffic parameters such as Post Encroachment Time (PET) and Time To Collision (TTC) were used to identify critical conflicts i.e. a collision is
very probable to occur if the values of TTC and PET lie within the range of 0 to 1.5 seconds and 0 to 5 seconds respectively. The number of accidents predicted by the conflict model was then compared with the conventional model which uses volumes of major and minor road as explanatory variables. Although major road time headway, which is a very important parameter, was not used in the modelling process but it was concluded that traffic conflicts can be successfully utilized as an alternative to actual crashes for estimating accidents per year at unsignalized intersections.

2.4. Current State of Practice

Incorrect specification of statistical models, defining the relationship between accidents and their causes, is the core issue, which leads to erroneous crash frequency predictions and incorrect inferences relating to the factors that determine the frequency of accidents [09]. Therefore, it is important to select models relevant to the type of data being analyzed. The current state of practice in the field of crash data modelling is defined by prominent accident studies summarised in table 1 below.

Table 1. Modelling Technique used in different studies for accident analysis.

<table>
<thead>
<tr>
<th>Study</th>
<th>Model Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>A micro-simulation approach for predicting crashes at unsignalized intersections using traffic conflicts [08]</td>
<td>Microsimulation using AIMSUN (Version 6.1.2)</td>
</tr>
<tr>
<td>Analysis and methods of improvement of safety at high-speed rural intersections [10]</td>
<td>Multivariate Ordered Probit Model</td>
</tr>
<tr>
<td>Using a reliability process to reduce uncertainty in predicting crashes at unsignalized intersections [11]</td>
<td>Negative Binomial and Bayesian Updating Model</td>
</tr>
<tr>
<td>Examining traffic crash injury severity at unsignalized intersections [12]</td>
<td>Ordered and Binary Probit Model, Nested Logit Model</td>
</tr>
<tr>
<td>Predictive model for motorcycle accidents at three-legged priority junctions [14 and 15]</td>
<td>Poison and Negative Binomial</td>
</tr>
</tbody>
</table>

3. Study Data

Accident data could be of various types such as binary, binomial, categorical, count and real valued. Example of binary data could be the occurrence or non-occurrence of accidents. Binomial data represents series of outcomes such as crash occurrence on a site in each time slice over a given period. Categorical data could be accident severity such as fatal, severe injury, slight injury, property damage only or road geometry such as 3-leg/ 4-leg intersections or 2 lane, 3 lanes, 4 lane facilities. Count data is that type of data in which the observations can take only the non-negative integer values (0, 1, 2, 3 ...)[16]. Examples of count data in road safety are number of accidents per year or number of vehicles per hour. Real valued data comprises of real numbers (1.68, 2.95, 45.1………..) such as road width and speed. With relevance to the context of this paper the count data is used as response variable.

3.1 Data Collection

In Malaysia the Royal Malaysian Police records the accident information on standard forms known as POL 27 which are then transferred to MIROS (Malaysian Institute of Road Safety Research) to update the accident database. The information recorded consists of numerous types of data, such as crash information which includes number of accidents, accident severity and vehicle type, field information such as major/minor road width, lane marking, control type, geometry and location, and miscellaneous information such as weather, time of accident,
and Average Annual Daily Traffic (AADT). Since most of the above information is recorded by Police and later digitized in raw form, it is required to reduce the above information customizing it for specific analysis.

### 3.2 Data Reduction

The raw data collected from MIROS contains several attributes such as driver characteristics, vehicle characteristics, intersection characteristics, traffic control, landuse and environmental characteristics. It was reduced to four specific attributes relevant to the context of intersection geometry for analysis. The attributes selected are road width, locale, lane marking and control. The landuse comprised of four subtypes namely city, town, small town and rural. The lane marking comprised of five subtypes namely single, double, one-way, divider and no marking. The traffic control comprised of three subtypes namely stop sign/line, yellow box and no control. Table 2 gives a brief description of the attributes and their subtypes.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Subtype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Width</td>
<td>--------</td>
<td>Major road width measured in meters</td>
</tr>
<tr>
<td></td>
<td>City</td>
<td>Urban area comprising of large business centres and shopping plazas</td>
</tr>
<tr>
<td></td>
<td>Town</td>
<td>Urban area comprising of smaller business centres and housing plazas</td>
</tr>
<tr>
<td>Locale</td>
<td>Small Town</td>
<td>Semi-urban area chiefly comprising of housing societies</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>Rural area with vegetation being the dominant landuse</td>
</tr>
<tr>
<td></td>
<td>Single</td>
<td>Single line lane marking at the center of the major road for separation of traffic moving in opposite direction</td>
</tr>
<tr>
<td></td>
<td>Double</td>
<td>Double line lane marking at the center of the major road for separation of traffic moving in opposite direction</td>
</tr>
<tr>
<td>Lane Marking</td>
<td>One-Way</td>
<td>Traffic moving in only one direction on the major road</td>
</tr>
<tr>
<td></td>
<td>Divider</td>
<td>Two way divided major road separated by concrete or grass median</td>
</tr>
<tr>
<td></td>
<td>No Marking</td>
<td>No marking on major road</td>
</tr>
<tr>
<td></td>
<td>Stop Sign/Line</td>
<td>Traffic controlled by Stop sign or Stop line on the minor road</td>
</tr>
<tr>
<td>Traffic Control</td>
<td>Yellow Box</td>
<td>Traffic controlled by Yellow Box on the major road</td>
</tr>
<tr>
<td></td>
<td>No Control</td>
<td>No control of traffic on minor road</td>
</tr>
</tbody>
</table>

Since the data obtained from MIROS contained several readings in which some or all of the attributes required for analysis were missing. Therefore, all data points with missing values were omitted reducing them to only those which were complete in every respect. Thus, a total of 448 data points were selected out of 180,000 readings. Five year accident data, from 2006 to 2011, was acquired for this study.

### 4. Analysis of the Data

The data was analyzed separately for each attribute. The effect of each subtype on the overall behaviour of accident occurrence was studied. Since, road width is one of the governing factors in the geometric design of unsignalized intersections; therefore, all other attributes were analyzed further with respect to this attribute to elaborate its effect on crash frequency.

#### 4.1 Effect of locale

Malaysia is a country that has a vast road infrastructure spread over a diversified locale. Although most of the population is saturated into cities and towns but small residential towns and rural areas are also inhabited by considerable number of residents. Since the rural road infrastructure is equally developed in comparison to urban,
it was expected that share of accidents will be identical. But the analysis revealed that the number of accidents occurring in cities is far more less then towns, small towns and rural areas. Detailed investigation was performed by dividing the accident data with respect to major road width as shown in fig. 1. Spearman’s Rank Correlation value reveals that locale is very weakly correlated (-0.146, p-value 0.123) with number of accidents and weakly correlated (-0.333, p-value 0.000) with major road width, however, road width has significant impact on the accidents occurring at different locales. In cities unsignalized intersections are mostly provided in areas with lesser traffic volume and smaller major road width while in other landuse they are provided on every location where the volume was less irrespective of the width of the major or minor road. This proved to be the reason for the under representation of cities in the overall number of accidents occurring in different landuse. In towns most of the accidents occurred on intersections having major road width between ten to fourteen meters. The range was broader in small towns with accidents occurring on roads having width between six to fourteen meters. In rural areas the range was skewed towards the intersections with lesser major road widths ranging between six to twelve meters.

![Fig.1. Number of accidents per locale versus road width](image)

4.2 Effect of lane marking

Lane marking is an important parameter of traffic control. It reflects the situation of traffic prevalent on the major road. Roads with smaller width and lesser volume usually have no marking while single line and double line lane marking is provided where the volume is higher. Roads with substantial volume are either provided with concrete or grass median to separate the traffic moving in opposite direction or converted to one way facilities. Analysis of the data showed that roads with single line lane marking are overrepresented in the total number of accidents that occurred on various lane markings. In order to investigate further the data was sliced into different road widths as shown in fig. 2. Spearman’s Rank Correlation value reveals that lane marking is very weakly correlated (-0.086, p-value 0.364) with number of accidents and weakly correlated (-0.211, p-value 0.025) with major road width, however, road width has significant impact on the accidents occurring at intersections with various types of lane markings. Intersections with divider have slightly higher number of accidents as compared to no marking ones because they were located in high volume areas with fast moving traffic. The effect of road width was not significant in the overall accident occurrence in the context of major road lane markings because there was no variation in term of number of accidents with respect to road width. It was found that single line lane marking is futile in terms traffic control. Because the number of accidents that occurred on them were very high as compared to other types of lane markings provided. This study also proves that lane marking can be a key parameter in accident investigation and black spot analysis. The examination of the data also elaborates the effectiveness of one way traffic.
4.3 Effect of traffic control

There are two types of traffic control mostly provided on unsignalized intersections in Malaysia. Stop signs or Stop lines being the most common while Yellow box is also provided where the major road traffic volume is higher. Usually no control on minor road is provided where the volume is less. This proved to be a disaster in terms of road safety. Because almost all the accidents that occurred on unsignalized intersections were on locations where there was no control of traffic on minor road. With the purpose of investigating further the data was alienated into different road widths as shown in fig.3. Spearman’s Rank Correlation reveals that traffic control is very weakly correlated (-0.058, p-value 0.538) with number of accidents and weakly correlated (-0.194, p-value 0.040) with major road width, however, road width has significant impact on the accidents occurring at intersections with various types of traffic control. No control on minor road gave the drivers an open invitation to take risk and drive carelessly. Therefore, lack of control on minor road was found to be the sole cause of accidents. Intersections with very weak control, such as stop line only, performed very well in terms of road safety as compared to intersections with no control at all.

4.4 Effect of road width

The width of major road plays an important role in the overall safety of unsignalized intersections along with speed and sight distance. The greater will be the road width the more will be the time required by the minor road vehicles to complete their manoeuvre. Thus, the risk of accident associated with unsignalized intersections increases with the increase in major road width. The S-shape curve of cumulative percent number of accidents
versus road width proves the width dependency of accident occurrence. Analysis of the data shows that almost 40% of the accidents that occurred on unsignalized intersections were on major road widths between nine to twelve meters as shown in fig.4. Since, the numbers of intersections that fall into the range of the width class mentioned above are more, they contribute equally in terms of number of accidents. The graph also indicates the target range of intersections for improvement in terms of road safety. In general the numbers of accidents were found to be normally distributed over the width of the major road.

Fig.4. Cumulative % number of accidents versus road width class

5. Results and Discussion

The analysis of the data has shown the effect of road width, locale, lane marking and traffic control on the accident occurrence of unsignalized intersections in Malaysia. Only 8% of the total accidents that occurred happened in the city while the rest were in town, small town and rural areas. Majority of the accidents occurred in small town where the traffic volume is less. This gives rise to over speeding by major road drivers. Coupled with no control on minor road, makes them the most vulnerable site for accident occurrence. 70% accidents happened at intersections with single lane marking and only 2% and 4% occurred on intersections with one way traffic and divider respectively. This shows that only painted markings are ineffective in terms of traffic control for unsignalized intersections. While, intersections that contain one way traffic or divider are the safest. 93% of the accidents that occurred were on intersections with no control on minor road. Thus provision of any kind of control can greatly reduce the accident risk associated with unsignalized intersections.

Table 3 shows the behaviour of accident occurrence and the share of each subtype in the total number of accidents for each attribute. The overall results provide a very good insight into the factors that contribute towards the accident vulnerability of unsignalized intersections. Most accidents happened on intersections having width b/w 11.0 to 12.0 metres lying in town/small town/rural areas with single lane marking and no control on minor road. Therefore, the results not only highlight the target intersections requiring immediate improvement but also indicate their cause of accident occurrence.

Table 3. Number of accidents and their percentages per attribute’s subtype.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Locale</th>
<th>Lane Marking</th>
<th>Traffic Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtype</td>
<td>City</td>
<td>Town</td>
<td>Single</td>
</tr>
<tr>
<td>Number of Accidents</td>
<td>34</td>
<td>140</td>
<td>164</td>
</tr>
<tr>
<td>Percentage</td>
<td>7.59</td>
<td>31.25</td>
<td>36.61</td>
</tr>
</tbody>
</table>
6. Conclusion and Recommendation

It is concluded that count data should be classified in terms of geometric and control parameters for effective utilization in accident analysis. Attributes such as road width, locale, lane marking, and traffic control provide the key factors responsible for accident occurrence. A more detailed investigation can be performed by dividing the data with respect to road width for each attribute. It is recommended that localized treatment can be provided to intersections with single lane marking by constructing concrete or grass median in the effective area of the intersection. Another option is to convert the major road into one way facility if the volume is high. Since, provision of Stop sign and/or Stop line greatly reduces the number of accidents occurring at unsignalized intersections. Therefore, all intersections with no control should be furnished with stop signs and stop lines. This will contribute heavily towards their risk reduction.

This study has shown that accident occurrence at unsignalized intersections can be effectively analyzed if the right type of data is selected and correctly reduced for examination. The results thus obtained will be highly relevant and will highlight the core cause of accident occurrence that makes an intersection vulnerable. Hence, the necessary action taken in response will be effective and will greatly reduce the accident risk. Thus, improves the overall safety of the road infrastructure.

References