World Academy of Science, Engineering and Technology
International Journal of Civil and Environmental Engineering
Vol:5, No:10, 2011

Development of Accident Predictive Model for Rural Roadway

Fajaruddin Mustakim, Motohiro Fujita

***Abstract*—**This paper present the study carried out of accident analysis, black spot study and to develop accident predictive models based on the data collected at rural roadway, Federal Route 50 (F050) Malaysia. The road accident trends and black spot ranking were established on the F050. The development of the accident prediction model will concentrate in Parit Raja area from KM 19 to KM 23. Multiple non-linear regression method was used to relate the discrete accident data with the road and traffic flow explanatory variable. The dependent variable was modeled as the number of crashes namely accident point weighting, however accident point weighting have rarely been account in the road accident prediction Models. The result show that, the existing number of major access points, without traffic light, rise in speed, increasing number of Annual Average Daily Traffic (AADT), growing number of motorcycle and motorcar and reducing the time gap are the potential contributors of increment accident rates on multiple rural roadway.

***Keywords*—**Accident Trends, Black Spot Study, Accident Prediction Model

I. Introduction

TRAFFIC accidents have been recognized as one of the major causes for human and economic losses both in developed and developing countries. In the year 2007, Malaysia has recorded 363,319 accidents, resulting in an average 18 deaths from road accident every single day. Furthermore, in year 2007 has caused 6,282 deaths, an estimated social cost of around 8 billion. Meanwhile in year 2006 there were 23.6 deaths in Malaysia for every 100,000 populations recorded according to (Royal Malaysia Police), which is among the highest figures in the world compare to Netherlands and Japan has 4.5 and 5.7 deaths per 100,000 people respectively. In Malaysia, motorcycles constitute nearly half than total registered vehicle in the country with 49% meanwhile cars stated second place consist 45%. Cars casualties contribute the highest rate with 67% followed by motorcycles recorded more than 16% of the total casualties in traffic crashes (Royal Malaysia Police, 2009)

The authorities, such as State Department of Transportation, may be interested in identifying black spot areas to propose safety treatment. Transportation engineer would be interested in identifying those factors (traffic flow, speed, road geometric and etc) that influence accident frequency to improve the roadway design and provide safer driving environment.

Fajaruddin Mustakim is with the Dept. of Scientific and Engineering Simulation, Nagoya Institute of Technology, Nagoya, Aichi Japan (phone: 080-4306-2277; fax: +81-52-735-5492; e-mail: fajardin@ uthm.edu.my).

Motohiro Fujita is Professor in Dept. of Scientific and Engineering Simulation, Nagoya Institute of Technology, Nagoya, Aichi Japan (e-mail: fujita.motohiro@nitech.ac.jp).

Previous studies have proved that improvement of highway design should effect significant reduction in the number of crashes.

The paper concerned with investigating major factors contributing to highway accident, this study concentrating the relationship between road condition, traffic flow, accident rates and their predicting, using multiple non-linear regression (MLR). The ability to predict accident rates is very important to transportation planner and engineers, because it can help in identifying hazardous location , sites which require treatment and as well as ranking the black spot locations.

II. Literature

This brief review of some of the existing literature purpose that a variety of traffic and design element such as AADT, cross section design, horizontal alignment, roadside feature, access control, pavement condition, speed limit, lane width and median width, affect accident rate. And most of the results have been based on Poisson or multiple linear and negative binomial (NB) regression model. [1], using data from Florida Department of transportation Roadway characteristic inventory (RCI) system, estimate (NB) regression for accident rates on various types of rural and urban highways with different traffic level. The result recommends that higher AADT levels and the presence of intersections are associated with higher crash frequency.

Numerous empirical relationship between vehicle accidents and these explanatory variables have been established in several previous studies [2] multiple linear regression models are used. In these models the independent variable (either number of accidents or accident rate) is a function of series of independent variables such as speed or traffic volume.

A number of researchers have investigated this complex interaction in the past. One of the first such studies had analyzed accident and traffic flow on U.S Route 22 through the city of Newark, New Jersey [3]. Crash rates were plotted against hourly volume class, and the author found a distinct U-shape relationship, with more accidents observed at higher and lower traffic volumes. Roads with higher ADT and pedestrian traffic are associated with higher accident frequencies for all highway types [4]. In appropriate or excessive vehicle speed is a once of contributory cause of the traffic accident [5].

Although the connection between speeds and cars is complex, a considerable volume of research evidence shows that lower vehicle speeds result in fewer crashes. Recent work on speed and accidents has indicated that the relationship derived by Finch et al (1994) holds for the general case: i.e. every 1 km/h reduction in speed across the network leads to a 3% drop in accidents [6].

Numerous Studies have examined driver behavior at unsignalised intersections and their response to different gaps in the priority stream. Critical gap has been defined for instances as "the minimum time gap in the priority stream that a minor street driver is ready to accept for crossing or entering the major stream conflict zone [7]. Researchers as [8] used the parameter of gap acceptance in the accident model. The paper moves on to develop a number of statistical models by using the Multiple Linear Regression that can be used in the prediction of the expected number of accident on the Federal Route Malaysia. The statistical programs and software have been created to perform this task. This involved SPSS, XLStatistic and Excel, which will be used in this project for among others, obtaining and validating the coefficients and also the model.

III. Federal Route 50

Federal Route 50 is a four lane two ways undivided road that runs from Batu Pahat to Ayer Hitam shown in Figure 1 . The road has many access almost every kilometers, function in carrying approximately 69,510 veh/day with 4,197 veh/hr at the peak hour and the normal growth 7.7%. The road also having a high density of driveways and property access. Federal Route 50 from Batu Pahat to Ayer Hitam (KM1-KM40) experienced 7,078 road accidents between the years 2000 and 2007, killing 234 people and injuring 1,493 people. Meaning Malaysia has some of the most dangerous roads. The accident analysis process involves the identification of accident black spot locations, establishment of general patterns of accident, analysis of the factor involved, site studies and development of an accident prediction model using Multiple Linear Regression.

Fig. 1 Study Location Federal Route F050

A. Analysis of Accident Data

The data collected were analyzed to determine the nature of accident pattern of the studied area. The analysis provides more detail in order to rank the black spot sites such as ranking accident point weightage, number of accident at study location and accident black spot map.

B. Field Investigation

Field investigation involves site, route and area inspection. These include traffic counts, origin destination surveys, vehicle classification survey, spot speed studies and observation studies. Preceeding analysis work to identify possible causal factors of the accident as well as countermeasures option. The site route or area inspection includes both a drive-over and walk-over inspection. The drive-over allows correlating accident behavior and driver perception. Walk-over inspection is a detailed examination of the location and driver behavior.

IV. Road Accident Trends

Figure 2 shows the trend of accidents and casualties in Federal Route 50 KM 1- KM 40 from year 2000-2007. The figure represents the increasing number of accident from year 2000 to 2004, but the number of accident decreased in year 2005. This means the impact of upgrading the route from a two-lane road to a four-lane road increased the accident number especially during its construction stage from year 2002-2004. The numbers of accidents however have declined to 994 in year 2005 from 1115 in year 2004. Meanwhile in year 2005, fatal cases were at its highest at 40 compared to the previous years. A total of 7,078 accident cases occurred, of which 234 were fatal, 206 serious injuries, 1,287 slight injuries and 5,351 were damage only.

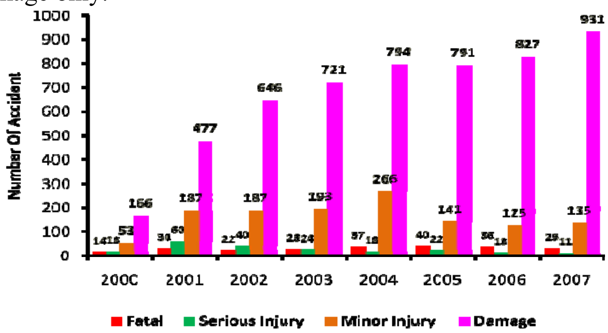

Fig. 2 Accident and casualty at F050

A. Accident by Hours of the Day

Figure 3, shows the worst accidents by hours of the day begin from the midday to midnight. The highest number of accidents happened from 16.01 to 18.00, recording 754 accidents. Second highest was at 18.01-20.00 involving 665 cases and the third highest at 18.01-20.00 with 636 accidents. The fatigues factor of the road users and high density of traffic plus too many conflict along the road probably contribute the accident to happen.

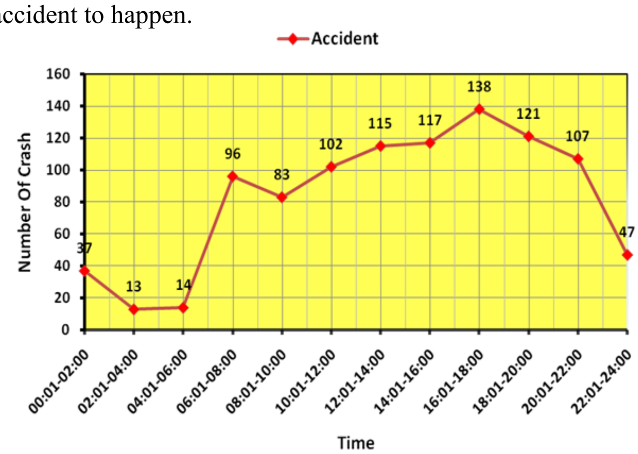

Fig. 3 Accidents by Hour of the Day at F050 (2003-2007)

B. Type of Vehicle involved Accident

Based on the Figure 4, it shows the type of vehicle involved in accidents on the federal route 50 in year 2007. The motorcars and motorcyclist formed a huge portion of the total number of accidents. Motorcars contributed the highest number of accidents with 65 % of the total vehicles involved in the accident, followed by motorcyclists 16% and third highest were van and small lorries with (6%) of vehicles involved.

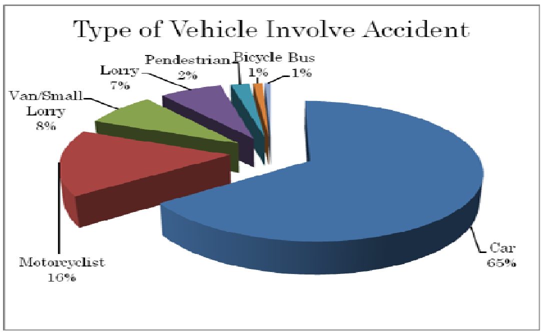

Fig. 4 Number of Vehicles involved by type of Accidents on F050

C. Accident Data By Type of Fault

According to Batu Pahat District Police Station, there are many factors contribute to the road accident along Federal Road F050. Figure 5 shows the type of faults data which involved number of accidents along F050 from year 2003 to 2007. The highest rank is driving to close with 2,759 accidents case and followed by accident at junction with 1,784 accidents case meanwhile speeding or out of control was in the third place with 982 numbers of accidents.

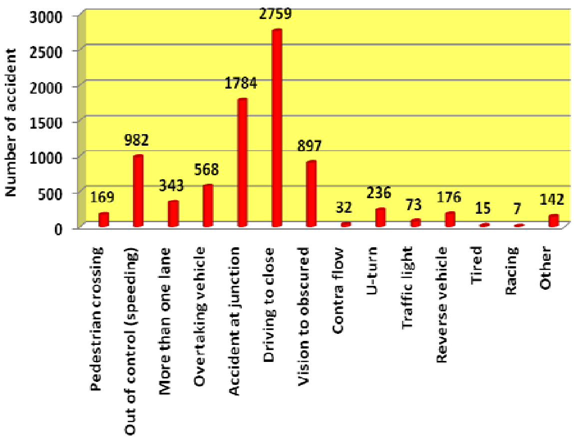

Fig. 5 Accident data by type of fault (2003-2007)

D. Black Spot Study

According to [10] some researchers rank locations by accident rate, same use accident frequency (accident per kilometer) and some use a combinationof the two. The Bureau of transport and Regional Economic of Australia (2001) location are in general classified as black spots after assessment of the level of risk and the likelihood of a crash occurring at each sites. Location that have an abnormally high number of crashes are describe as crash concentrated, high hazard , hazardous, hot spot or black spot sites.

E. Accident Point Weightage

Total of accidents in every kilometer at federal road FT 050 can be rank by using accident point weightage formula. This method is based on the value that is been contribute by The Transport Research Laboratory (TRL) from Interim Guide on Identifying, Prioritizing and Treating Hazardous Locations on Roads in Malaysia which the points will be given based on type of the accident. Fatal accident (6 points), serious injury (3 points), slight injury (0.8 point), and damage only (0.2 point).The weightage formula:

$$APW = X_1\,(6.0) + X_2\,(3.0) + X_3\,(0.8) + X_4\,(0.2) \qquad (1)$$

Where,
X_1 = Number of fatal
X_2 = Number of serious injury
X_3 = Number of slight injury
X_4 = Number of damage only

This system can also be used as an alternative to rank black spot as shown in Table 1, using the accident data from the year 2004 to 2007. Section 5, Parit Haji Noor at Batu Pahat registered the highest with 129.8 weighting point, ranking it in first place based on the total number of accident. This was followed by Section 10 (Sharp factory, Batu Pahat), Section 2 (Masjid Baru Batu Pahat), Section 9 (Taman Ria, Batu Pahat), Section 20 (Pintas Puding, Parit Raja) and Section 24 (Taman Manis, Parit Raja).

TABLE I
RANKING ACCIDENT POINT WEIGHTING ALONG ROUTE F050 (2004-2007)

KM	Location	Type of Road Accident				Total	APW	Rank
		Fatal	Serious Injury	Slight Injury	Damage			
5	Parit Haji Noor	9	4	50	119	182	129.8	1
10	Sharp Factory	8	5	37	105	155	113.6	2
2	Masjid Baru BP	4	2	30	210	246	96	3
9	Tmn Ria Jaya	5	3	43	100	151	93.4	4
20	Pintas Puding	8	0	26	102	136	89.2	5
24	Taman Manis	6	3	28	85	122	84.4	6
8	Tmn Bahagia Baru	6	1	32	96	135	83.8	7
4	Maktab BP	2	2	26	172	202	73.2	8
6	Gillmill Industry	3	2	33	104	142	71.2	9
21	Taman Maju	2	3	23	142	170	67.8	10

F. Accident Maps

The accident blackspots map at F050 is shown in Figure 6, has been establish to represent spatial distribution of accident data. One method of identifying black spots location is the use of computerized Geographic Information system (GIS) software (ArcGIS 2001). As can be seen, the size of the circle is proportional to the accident point weightage. This feature allows a quick visual identification and ranking of the problematic location of area concerned.

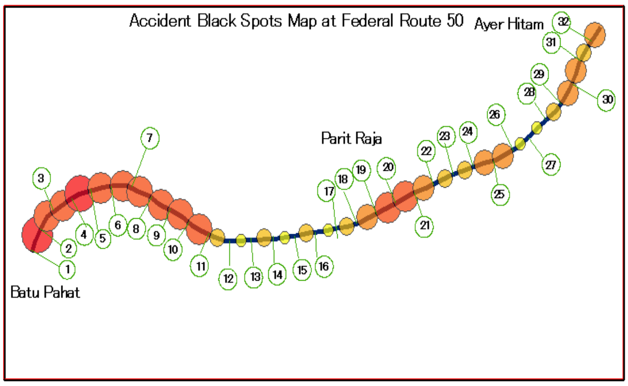

Fig. 6 Accident Black Spot Map on F050

G.Ranking of the Top Ten Accident Section

The simplest way of ranking sites, and the one currently recommended for use in Malaysia, is to list them in descending order of total accidents for each section in kilometer. Figure 7 shows the ten worst accident sections at F050. The analysis was based on data compiled over four year starting from 2004 and 2007. The highest accident frequency was KM 2 with 246 cases, followed by KM 4 with 202 cases, next KM 5 with 182 cases and the forth place KM21 with 170 cases while KM 20 and KM 24 recorded 136 and 122 cases respectively.

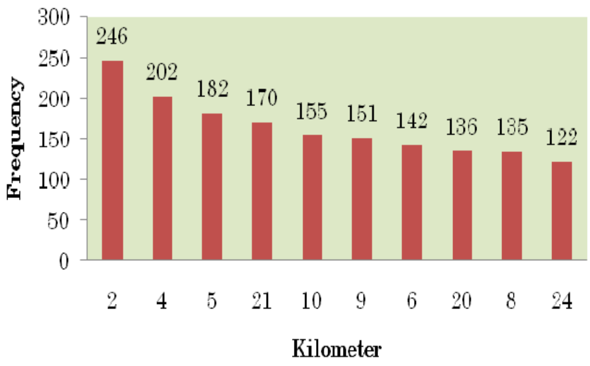

Fig. 7 Ranking of top ten section accident at F050 over4 year period (2004-2007)

H.Number of Accident at Study Location by Section (KM)

From Figure 8, shows that number of accident at 6 locations which being characterize by 4 number of accident which is fatal, serious injury, slight injury and damage only. The highest number of fatal accident is at KM 20 with 8 cases of deaths and the lowest number of fatal accident is at KM 22 and KM 23 with 1 case. Serious injury are occurred at certain study location which is KM 21, KM 22, and KM 24. For slight injury and damage only is happening at every study location.

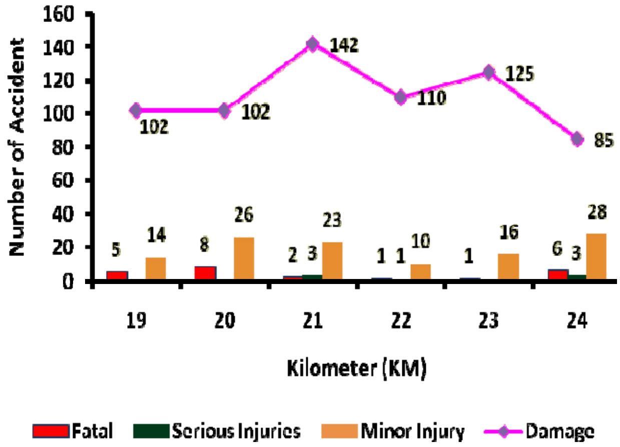

Fig. 8 Number of Accident according its type at FT 050 (2000-2007)

V. Model Development

The model consists of several independent or explanatory variables, encompassing elements from road geometry to traffic condition. Time series data of traffic and accident, over 4 years period (2004:2007) for the considering Federal Route, is utilized in the calibration of these predictive models. For this study, the variables which have considerable effects are approach speed, Annual Average Daily Traffic (AADT), number of access points per kilometer, traffic light and time gap. The data was collected on the field work. The study section used for collecting data was about 5 kilometer long, it Involves KM 19, KM 20, KM 21, KM 22 and KM 23 of Federal Route 50. By traversing the entire length of the road to observe the number of major access point and traffic light per kilometer for every related section kilometer is obtained. Traffic volume and approach speed were obtained over 2-hour time periods of field survey at each section, namely the morning (0800-1000 h), midday (1100-1300 h) and evening (1700-1900 h).

Traffic study data used in this paper were based on the hourly traffic volume. Motorcars and motorcycles data were collected from the hourly traffic volume (disaggregated by motorcars and motorcycles), counted on each kilometer section at the selected major access point. Motorcars(C per hour) were range between 767 to 1932, meanwhile motorcycles (MC per hour) range from 281-1192. Gap in this study defines as the time between the back of the successive motor vehicles and the in front of the following motor vehicles at the corresponding access point. The ranges for gap were 2.75 seconds to 3.93 seconds.

Approach speed is defines by the 85th percentile speed measured at a distance 50 meter upstream from the corresponding stop lines of the major access point. Approach speed measurement was taken at every section using speed radar equipment. At every section, more than 100 motor vehicles speed data was collected with ranges between 57.75 km/hr to 83.42 km/hr. The 85thpercentile speeds were determined from speed measurement by using SPSS program. Approach speed was also carried out in the previous studies on the traffic crash as one of the explanatory variable [11].

This research applies the "Multiple Linear Regression" method in order to develop a model which relates accident point weighatge to the access point and traffic situation. The theoretical model of accident rate is represented by the following equation:

$$Y_i = b + b_1 X_{1i} + b_2 X_{2i} + ... + b_{ni} X_{ni} + \varepsilon_i \quad (2)$$

Where Y_i is accident point weighting (dependent variable) while all the independent variable are X_1, X_2 and X_{ni}. The b, b_1, b_2 and b_{ni} are the regression coefficients to be estimated and ε_i term is the error representing the residual difference between observed and predicted model value.

A. Significance of Regression Model

The coefficient of multiple of multiple determination (R^2) is often used as a convenient measure of the success of the regression equation in explaining the variation in the data. It is expressed as the percentage ratio of the SS caused by the regression to the total sum squares. Ideally, the value of R^2 could be 100%, however, practically it should not be no matter how good the model is.

The analyses of variance *ANNOVA* approach was used to test the statistical significance of the derived regression model. To test the amount of variation explained by the regression model is more than the variation explained by the average, the *F* ratio is used [10]. The analyses items are further elaborated on as follows: The total sum of squares SS_{YY} can be written as:

$$SS_{YY} = SS_R + SS_E$$

The null hypothesis is rejected if test statistic *F* -ratio > $F_{\alpha(p,\, n-p)}$. From the *F* distribution table with $F_{0.05}$, for p=4 (number of independent variables) and n = 30 (sample size)

$$F_{0.05,(4,26)} = 2.74$$

and,

$$F = 2646.552 > F_{0.05,(4,26)}$$

The (*Sig*), which indicates the meaningful level to obtained coefficient for the model parameters. Generally, variables with *Sig* value of 0.10, 0.05 and 0.01with confident interval (CI) of 0.90, 0.95 and 0.99 are statistically meaningful in the model.

TABLE II
DESCRIPTION OF THE STUDY VARIABLE

Abbr.	**Description**	**Coding**
APW	Accident point weighting	APW (per km)
AP	Access point	AP(per km)
AS	Approach Speed	AP (Km/hr)
AADT	Annual Average Daily Traffic	AADT (per hour)
GP	Vehicle gap	GP (second)
MC	Motorcyclist	MC (per hour)
C	Motorcar	C (per hour)
TL	Traffic light	TL (per km)

TABLE III
MODEL SUMMARY

Model	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.9973	0.9585	0.1104
2	0.9979	0.9576	0.0985

TABLE IV
MODEL ANNOVA

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	118.2695	4	29.560	2424.77	0.000
	Residual	0.317	26	0.0121		
	Total	118.56	30			
2	Regression	118.3175	5	23.664	2439.72	0.000
	Residual	0.2425	25	0.0097		
	Total		30			

TABLE V
COEFFICIENT

Model		Unstand. Coeff. Beta	Std. Error	T
1	Intercept	0	N/A	N/A
	AP	0.0350	0.0112	3.126***
	AS	0.0049	0.0009	5.477***
	AADT	0.00002	3.27E-06	2.070**
	GP	0.1898	0.0825	2.301**
2	Intercept	0	N/A	N/A
	AP	0.0402	0.00017	4.4219***
	AS	0.0150	1.35E-07	7.2489***
	(MC +C)	0.0001	0.0066	2.9608***
	GP	0.1092	0.0049	3.0864***
	TL	0.0830	0.0354	2.2244**

*,**,***=Significant at the 90%,95% and 99% level, respectively

B. The Accident Prediction Model

The SPSS and Excel Microsoft office was used in the regression analysis and the extract of the output is shown in Table II,III, IV and V whereas both model 3 and 4 were using natural log function. The final accident prediction model for Federal Route 50 takes the following equation:

$$\text{In}(APW)^{0.5} = 0.0350(AP) + 0.0049(AS)^{1.2} + 0.00002(AADT)^{0.7} + 0.1898(GP)^{0.7} \quad (3)$$

$$\text{In}(APW)^{0.5} = 0.0402(AP) + 0.0150(AS) + 0.0001(MC+C) + 0.1092(GP) + 0.0830(TL) \quad (4)$$

The model 3 and 4 has an R^2 of 0.9973 and 0.9979 respectively the result are given in table III respectively. These value means that 99.73% and 99.79% of the variation for model 3 and 4 in the number of accidents has been explained the regression line. It proves that the regressed model does provide a good fit to the independent variable. From the model development in this study, it is noted that the factors which contribute to accidents at four lane two way undivided rural roadway are, number of access points, vehicle speed, Annual Average Daily Traffic (AADT), motorcycle, motorcar and gap.

VI. Discussion and Conclusion

The writing of this paper primarily motivated by the need to identify, Malaysian rural multilane roadway, the safety effect on the expected number of crashes of following parameters: traffic flow, approach speed, gap, access point and traffic light with a propose of countermeasure.

The findings from this research were established the accident point weighting as the ranking tool of the black spots section by kilometer along the 4-lane undivided rural roadway. It will be benefit for the authority as a tool to prioritizing the hazardous location with appropriate treatment.

Base on the 4 years observation period from 2004 to 2007 carried out on undivided dual carriageway by using Model 4, the percent accident reduction through changing the measures of each parameter are, one access point per kilometer reduction can reduce accident by 16%, 5 kilometer per hour reduction in approach speed equated to a 27 % drop in crash, per long the 0.1 second in gap will cut down number of accident by 1%, whilst 100 vehicles of motorcycle and motorcar per hour can reduce accidents also by 5%. Furthermore installing of 1 signalized opening at the access point would reduce number of crash by 44%.

In the light of the above outcome, the model developed for Malaysian rural roadway in this paper appear to be useful for many application such as the existing number of access point, increasing number of motorcycle and motorcar, rise in speed and shorted in gap are among the potential contribute of increment accident rate. This paper revealed that by installing the traffic light at the hazardous access point would drastically reduce the number of accident. On the top of that segregation between the motorcycle and other road user. Thus, these studies are realistically convinced that this latter may contribute a point of reference for the engineer in improving or designing multilane rural roads. More importantly, the significant accident predictive model developed in this study is applicable in road safety improvement and could serve as a basis for further research work in Malaysia and other developing countries.

References

[1] Hadi, M.A., Aruldhas, J., Chow, L.F., Wattleworth, J.A., Estimating Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression. Transportation Research Record, 1500, TRB, National Research Council,1993.

[2] Miaou, S.-P. and Lum, H.. "Modeling Vehicle Accidents and Highway Geometric Design Relationships." Accident Analysis and Prevention 25(6): 689-709(1993).

[3] Gwynn, D.W. ; Relationship between road accident and hourly volumes. Traffic Quartely, pp.407-418, (1967).

[4] Berhanu, G.; Model relating traffic safety with road environment and traffic flow on arterial roads in Addis Ababa. Adis Ababa University, pp 697-704(2004).

[5] Quimby, A., Maycock, G., Palmer, C., & Grayson, G.B. (1999 b). *Drivers speed choice: an indepth study*. Transport Research Laboratory TRL, Report 326, Crowthorne.

[6] Taylor, M.C., Lynam, D.A., and Baruya, A. (2000) *The effects of drivers' speed on the frequency of road accidents*, Transport Research Laboratory Report 421, Crowthorne, Bucks: Transport Research Laboratory

[7] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[8] Brilon, W., Koenig, R. and Troutbeck, R.J. :Useful estimation procedures for critical gaps. Transportation Res.- *A* 33 pp 161-186(1999).

[9] Miller A.J., "Nine estimators of gap acceptance parameters". *Proceedings of the 5th International Symposium on the Theory of Traffic Flow, pp. 215-235* (1972).

[10] Hauer, E.: Identification of sites with promise, Transportation Research Record, 30, 54-60 (1996).

[11] S.Harnen, R.S.Radin Umar, S.V.Wong, W.I.Wan Hashim;Motorcycle Crash Prediction Model For Non-Signalized Intersections.IATSS Research Vol.27 No.2,(2003).pp.508-65.