ACCIDENT PREDICTION MODELING OF FEDERAL ROUTE F050 MALAYSIA

Mustakim FAJARUDDIN** Motohiro FUJITA ***

Road accidents is one of the major contributors of human deaths in Malaysia. In the year 2007, 363,319 accidents were recorded, resulting in an average 18 deaths from road accidents every single day. Federal Route 50 from Batu Pahat to Ayer Hitam experienced 4,842 road accidents between the years 2000 and 2005, killing 244 people and injuring 1,644 people. The purpose of this study is to perform an accident analysis and to develop an accident prediction model for Federal Route 50 by using multiple linear regression analysis. The road accident trend and blackspot ranking were established at Federal Route (F050) Batu Pahat – Ayer Hitam. It revealed that the increment of accident rates can be explained by either the rises in speed, number of major access point and traffic volume.

1. INTRODUCTION

In Asia alone, 400,000 people are killed on the roads annually and more than four million injured. According to the World Health Organization, every year, nearly one million people are killed, three million are severely disabled for life and thirty million are injured in road accidents. In 1990, death by road accidents remained 9th in ranking and by the year 2020, it has been predicted that road accidents will be the third leading cause of death worldwide (World Disaster Report, 1998).

Malaysia has experienced a remarkable period of economic expansion and growth in population, industrialization and economic stability. Statistics released by the Transport Ministry showed that a total of 14,816,407 million vehicles were registered in the country until the end of 2005, which is nearly twice as those registered in 1996 (7,686,684 vehicle). Despite the marked increase in the number of vehicle over the last 10 years, there was a drop in fatality index with 8.2 deaths for every 10,000 vehicles in 1996, followed by 7.37 (1997), 6.28(1998), 5.83 (1999), 5.69 (2000), 5.17 (2001), 4.9 (2002), 4.9 (2003), 4.52 (2004) and 4.18 (2005). To be on par with developed countries, we need to reduce the average fatality index to 2 deaths for every 10,000 vehicles (Ministry of Transport Malaysia 2006).

The number of accidents on a given road section during a certain period of time is probabilistic in nature and is a non-negative integer. Despite the fact that accidents are random an unpredictable at micro level, statistical models can predict reliable estimates of expected accident by relating aggregates of accidents to various explanatory measures of flow, site characteristics, and road geometry at macro level. Numerous empirical relationship between vehicle accidents and these explanatory variables have been established in several previous studies (Miaou and Lum, 1993; Gwynn D.W ,1967; Maher and Summersgill, 1996 and Al-Masaeid, Hashem R, and Ghassan, 2004).

---

*Keywords: accident analysis, blackspot ranking, accident prediction model
** PhD, Dept. Scientific and Engineering Simulation Nagoya Institute of Technology
*** Member of JSCE, Prof. Dr .Eng. Graduate School of Engineering, Nagoya Institute of Technology
(Gokiso-cho, Showaku, Nagoya Aichi,Japan, E-mail: fujita.motohiro@nitech.ac.jp,FAX+81-52-735-5492)

A number of researchers have investigated this complex interaction in the past. One of the first of such studies had analyzed accidents and traffic flow on U.S Route 22 through the city of Newark, New Jersey (Gwynn, 1967). 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 average daily traffic (ADT) and pedestrian traffic are associated with higher accident frequencies for all highway types (Berhanu, 2004). In the report of the NHMRC Road Accident Research Unit, University of Adelaide, carried out a study on "Traveling Speed and The Risk of Crash Involvement"(CN Kloeden, AJ McLean, VM Moore and G Ponte, 1997) , the authors concluded that in a 60 kmph speed limit area, the risk of involvement in a casualty crash doubles with each 5 kmph increase in traveling speed above 60 kmph. This study will analyse the accident data and develop an accident predictive model, concentrating on Federal Route 50 in Parit Raja. The Multiple Linear Regression method was used to relate the discrete accident data with the road and traffic flow explanatory variables.

2. MATERIALS AND METHODS

The accident analysis process involves the identification of accident blackspots, establishment of general patterns of accidents, analysis of the factors involved, site studies, proposal of countermeasures and development of an accident prediction model using Multiple Linear Regression.

(1) Data Collection

Accident data were collected from Batu Pahat Traffic Police Station, Batu Pahat Public Work Department, Bukit Aman Royal Malaysia Police, Road Transport Department, Parit Raja Health Clinic Center, Universiti Tun Hussein Onn Malaysia (UTHM) and Road Safety Research Center UPM Serdang. The accident database was obtained from the POL27 and the crash records were collected from the year 2000 to 2005.

(2) Analysis of Accident Data

Accident data are required to determine the nature of the accident problem at the study area. The analysis of the accident needs to look for the accident pattern. Accident data analysis provides a more detailed ranking of the blacksport sites such as ranking accident point weightage at F0 50, ranking of the top ten accident sections and ranking of accident point weightage.

(3) 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. Preceding analysis work enable researchers to identify possible causal factors of the accident as well as countermeasure options.The site route or area inspection include both a drive-over and walk-over inspection. The drive-over allows to correlate accident behaviour and driver perception while the walk-over inspection is a more detailed examination of the location and driver behaviour.

(4) Accident Prediction Model

The model consists of several independent or explanatory variables, encompassing elements from road geometry to traffic condition. For this study, the variables which have considerable effect are 85th percentile speed, volume study and number of access points per kilometer. The data were collected in-field. The study section used for collecting data was about 5 kilometers 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 access points, the number of major access point per kilometer for every section is obtained. Traffic volume and spot speed were obtained over 2-hour time periods of field survey at each section, namely the morning (0800 - 1000 hrs), midday (1100 - 1300 hrs) and evening (1700 - 1900 hrs). Spot speed measurements were taken at every section using a speed radar equipment. The 85th percentile speeds were determined from spot speed measurement using SPSS program.

3. RESULTS AND DISCUSSIONS

Figure 1 shows the trend of accidents and casualties on Federal Route 50 (F050) KM 1-KM 38 from year 2000-2005. The figure shows an increase of number of accidents 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 number of accidents however declined to 905 in year 2005 from 1,084 in year 2004. Meanwhile in 2005, fatal cases were at its highest at 36 compared to the previous years. A total of 4,842 accident cases occurred, of which 152 were fatal, 182 serious injuries, 1010 slight injuries and 3,498 were damage only.

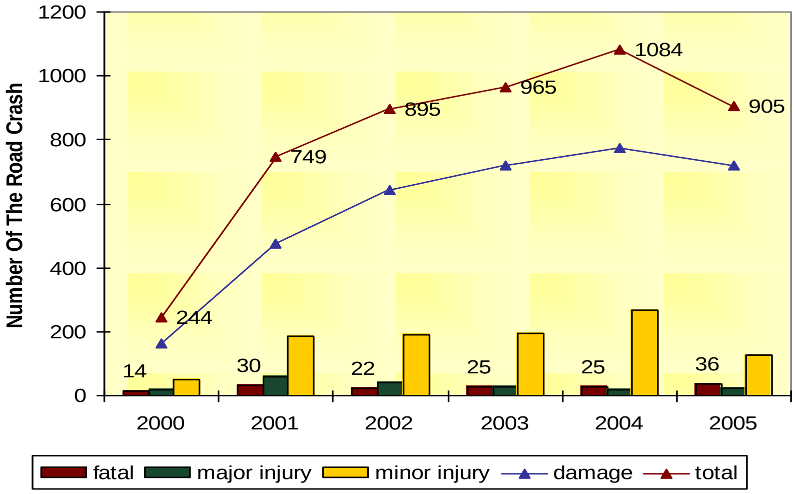

Figure 1: Accident and casualty along F050 (KM 1-38)

(1) Accident by Hours of the Day Year 2005

Figure 2 shows the worst accident by hours of the day begin from the midday to midnight. The highest number of accidents happened from 16.01 to 18.00, recording 138 accidents. Second highest was at 18.01-20.00 involving 121 cases and the third highest at 14.01-16.00 with 117 accidents. The fatigues factor of the road users and high density of traffic plus to many conflict along the road probably contribute the accident to happen.

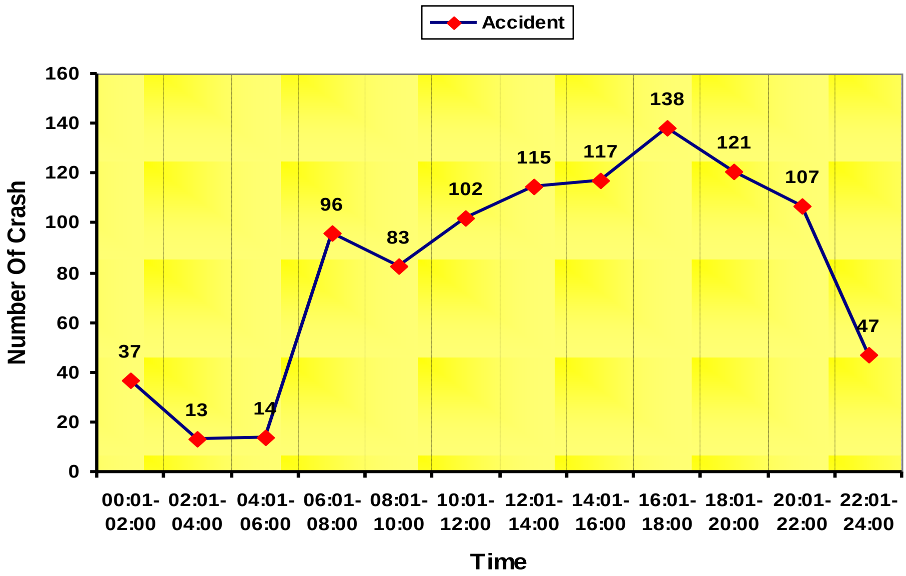

Figure 2: Accidents by Hour of the Day at F050 (2005)

(2) Ranking Accident Point Weightage

Taking this a step further, the severity of accidents is taken into account by weighting factors (which are normally related to the average accident cost of each severity level), and damage-only accident are also included (having a real cost), the results in ranking shown in Table 1.This system can also be used as an alternative to rank blackspots, using the accident data from the year 2002 to 2005. Section 5, Parit Haji Noor at Batu Pahat registered the highest with 146.8 weighting point, ranking it in first place based on the total number of accidents. This was followed by Section 2 (Mesjid Batu Pahat), Section 1 (Klinik Kesihatan Batu Pahat), Section 10 (SHARP Factory, Batu Pahat), Section 20 (Pintas Punding) and Section 6 (Gillmill Industry).

Table 1: Ranking Accident Point Weighting Along a Route F050 (KM 1-38) Over a 3 Years Period (2002-2005)

Section	Accident Severity					Weightage	
(KM)	Fatal	Serious	Slight	Damage	Total	Point	Rank
5	11	6	47	126	190	146.8	1
2	5	4	34	237	280	116.6	2
1	4	3	53	198	258	115	3
10	5	7	37	112	161	103	4
20	7	2	36	90	135	94.8	5
6	3	6	44	107	160	92.6	6
4	3	4	32	168	207	89.2	7
19	5	5	32	80	122	86.6	8
8	5	3	34	88	130	83.8	9
3	4	1	28	154	187	80.2	10

4. THE ACCIDENT PREDICTION MODEL

From the data shown in table 3,4 and 5 a regression analysis was performed using Microsoft Excel. Based on the results of the analysis, the accident prediction model for Federal Route 50 takes the following equation:

$$\ln(APW)=(1.1641\log AP+0.0017(HTV^{0.75}+V_{p85}^{1.25}))^2 \quad \ldots\ldots\ldots(1)$$

Where,

APW = Accident Point Weightage

AP = Number of Access Points per kilometer

HTV = Hourly Traffic Volume

V_{p85} = 85th percentile speed

The model has an R^2 of 0.995, which means that 99.5% of the variation in the number of accidents has been explained the regression line as given in the table 3. It proves that the regressed model does provide a good fit to the independent variable. 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 (Hair et al. 1995). The analyses items are further elaborated on as follows: The total sum of squares SS_{YY} can be written as:

$$SS_{YY} = SS_R + SS \quad \ldots\ldots\ldots (2)$$

Null hypothesis

$$H_0: \beta_1=\beta_2 = 0$$

Alternative hypothesis

$$H_1: \beta_i \neq 0$$

The null hypothesis is rejected if test statistic $F > F_{\alpha(k,n-p)}$. From the *F* distribution table with $F_{0.05}$, for k=2 (number of independent variables).

$$P = k + 1$$

$$F_{0.05,(2,27)} = 3.37. \quad \ldots\ldots\ldots(3)$$

and $F= 2969.475 > F_{0.05,(2,27)}$

Thus the *F* test for the overall fit of the regression model, reject the null hypothesis that state there is no relationship between the accident point weighatage (APW) and overall independent variable. The alternative hypothesis that state there is relationship between APW and overall independent variable is accepted. Therefore the model is significant and can be used for the prediction of the accidents. The coefficient table shows in the *t* test of each of the estimated coefficients. The critical value of *t* test (t_c) for significance level α = 0.05 and 30-1 = 29 degrees of freedom is 1.699. Thus three estimated coefficients namely number of access points, hourly traffic volume and 85th percentile speed succeed the *t* test; the null hypothesis states that there is no relationship between the accident weightage and those variables is rejected. Instead, the alternative hypothesis stating that there are relationships between accident weightage and those variables is accepted. Therefore all the explanatory variables have been found to be significant and hence, they can be used in the multiple linear regression equation. The variable access point has the strongest relationship with the accident weightage for having *t Stat*(5.6532) with low P- value of 0.000.

Table 3: Model Summary

Multiple R	R Square	Adjusted	Standard Error	Observations
0.9976	0.9953	0.9594	0.1411	30

Table 4: ANNOVA

	df	Sum of squares	Mean Square	*F*	Sig. *F*
Regression	2	118.0037	59.0018	2969.475	0.0000
Residual	28	0.5563	0.01987		
Total	30	118.56			

Table 5: Coefficient

	Coefficients	Standard Error	t Stat	P-value
Intercept	0	#N/A	#N/A	#N/A
log(AP)	1.1641	0.0102	5.6532	4.67E-06
$HTV^{0.75}+V_{P85}{}^{1.25}$	0.0017	0.0004	4.8669	3.99E-06

5. CONCLUSION

The objective of this research was to develop predictive model relating traffic accidents with the road environment and traffic flows. Multiple regression techniques were used to estimate the model parameters. The regression equation (1) can be used to predict accident rates as developed from this study. The result of the analyses provide sufficient evidence to support the hypothesis that the existence of a high density of junctions, an increase in traffic volume and vehicle speed in Federal Route 50 has contributed significantly to traffic accidents. Reduction of vehicle speed, access point and traffic volume are likely to have an influential effect on the road accidents. Based on this study, the percent accident reduction by changing the measures of each parameters are; one access point per kilometer reduction can reduce accidents by 28.05%, 5 kilometer per hour speed reduction can reduce accidents by 14.31%, and 100 vehicle per hour volume reduction can reduce accidents by 7.50%. More importantly, the significant accident predictive model developed in this study can be applied in road safety improvements and could serve as a basis for further research work of Federal Route in Malaysia.

References

1) World Disaster Report (1998); International Faderation of Red Cross and Red Crescent Societies Transport and Research Laboratory, UK.
2) Radin Umar Radin Sohadi and C.J Baguely, The Identification, Prioritising and Analysis Of Accident Blackspots In Malaysia, *REAAA Journal/January 1994.*
3) Abdul Kareem ;"Review Of Global Menace of Road Accidents with Special Reference to Malaysia"- A Social Perspective, Malaysia *Journal of Medical Science*.Vol10. No.2, (31-39) 2003.
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) Al-Masaeid, Hashem R, Suleiman, Ghassan (2004); Relationships between urban planning variables and traffic crashes in Damascus. Road & Transort Research, 2004.
6) Transportation Research Laboratory, TRL 421 Report ; The Effects of Drivers' Speed on the Frequency of Road Accidents,2000.
7) Radin Umar Radin Sohadi, Chai W.L., Hussain H, Law T.H (2001); "Modelling of Traffic Accidents Along Major Trunk Roads in Malysia", Journal IEM.vol. 62 No. 3 2001.
8) K.M.Lum, Hendry S.L. Fan, S.H.Lam and P.Olszewski (1998); "Speed-Flow Modelling of Arterial Roads in Singapore", Journal of Transportation Engineering./May/Jun 213-222,1998.