Log-Rank Test on Road Traffic Accident Data in Batu Pahat, Johor, Malaysia

Jessintha A/P Lai* and Mohd Asrul Affendi Abdullah[2]

Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor Darul Takzim, Malaysia.

Abstract: Road traffic accident (RTA) is defined as an accident, which happens on the road between two or more objects, one of which must be any types of a moving vehicle. Nowadays, Malaysia went through rapid growth experienced in the population, economy, and motorization. Therefore, the accidents rates increased every year then become huge threat to public health and national authorities. Batu Pahat is one of the district in the south of Malaysia state, known as Johor that recorded the highest accident rate compared to other district". Thus, this study aims to determine if there are any significant difference on survival of RTA for different years in Batu Pahat district. Besides, this study identified the peak time of road network to determine associated measured could be adopt to reduce the likelihood and severity of accident. The log-rank test has been used to determine the Kaplan-meier curves of RTA for the different year. Data that been used in this study collected from Batu Pahat Police Station. Result shows that three peak times of RTA identified and the log-rank test concluded that there are no significant difference in survival of RTA for different years.

Keyword: Accidents; Survival; Peak time; Log-rank; Kaplan-Meier

1. Introduction

In Asia, 400,000 people are killed on the roads annually and more than four million people are injured [1]. Besides, developing countries such as European countries, United Kingdom, Japan and Malaysia account for 90% of global road traffic deaths, while accounting for only 20% of cars being driven worldwide. Road traffic injuries are the ninth leading cause of deaths globally, and are predicted to rank sixth by the year 2020 [2]. In 2007, Malaysia has recorded 363,319 accidents, which resulted in an average of 18 deaths from road accident every day [3]. Batu Pahat is one of the district in Johor that recorded the highest accident rates in Malaysia. Thus, while medical science has occupied the ravages of many diseases, accidents have become a new epidemic of public health importance that requires equal effort for control and prevention [4]. Accidents could occur at anytime and anywhere yet there must have some peak time that accident will occur at most. Basically, one not noticed of the peak times that accident occured.

2. Methodology

Study design

This study is a descriptive retrospective analysis of road accidents data in Batu Pahat. Phenomenon with measurable variables were explained and predicted by using quantitative method [5].

Data collation

The data of RTA were obtained from Traffic Police Station, Batu Pahat. The data obtained are according to the time of occurrence of the accident. The occurrences of road traffic accidents were recorded every two hours from morning to midnight. The accident data were collected from year 2009 to 2014. The data has been analysed by using Microsoft Excell and Statistical Package for the Social Science (SPSS) software. The graphical method was used to determine the peak time of the RTA.

Kaplan-Meier estimator

Kaplan-Meier (K-M) estimator is the standard nonparametric estimator of the survival function which is known as the product-limit estimator. This estimator comprise information from all the observations available, both censored and uncensored (event times), by consider survival to any point in times as series of step defined at the observed survival and censored times [6]. The survival function, $S(t)$ for the K-M estimator can be defined as:

$$S(t) = P(T > t) = \int_t^\infty f(x)dx. \qquad (1)$$

Based on Equation (1), T denotes the duration of road traffic accident, $f(x)$ is the probability density function of road traffic accident in duration T and $S(t)$ is the cumulative survival probability [7].

Log-rank test

Log-rank test is a large sample chi-square test which is use to evaluate whether or not the Kaplan-meier curves for two or more groups are statistically equivalent. The log-rank test calculates the chi-square for each event time for each group and sums the results. The summed results for each group are added to derive the ultimate chi-square to compare the full curves of each group. The null hypothesis being tested is that there is no overall difference between the two or more survival curves. The log-rank statistic is approximately chi-square with one degree of freedom. So, the p-value for the log-rank test will be obtained from tables of the chi-square distribution. The two or more survival group can be compared statistically by testing the null hypothesis. The hypothesis testing is:

H_0: There is no difference regarding survival among two interventions.
H_1: There is difference regarding survival among two interventions.

An approximation to the log-rank statistic is:

$$X^2 \approx \sum_i^{number\ of\ groups} \frac{(O_i - E_i)^2}{E_i}. \qquad (2)$$

From Equation (2), O_i denotes the observed value for each group and E_i denotes the expected value for each group. In the log-rank test for several groups, the test statistic will be more complicated mathematically. So, the log-rank test for several groups will be carried out by using computer program [8].

3. Results and discussions

Data sources

The secondary data was obtained from Batu Pahat Police Station, Johor. This study considered accident data from 2009 to 2014. The data obtained were according to the number of accidents occurred at specific time. The time period were recorded for every two hours start from the midnight which was twelve midnight. The number of road traffic accidents and the number of people killed in road traffic accidents were used as the response variable and the other variables such as the time the road traffic accident was occurred and the time people were killed by road accident as the explanatory variable.

Time Period in which Road Traffic Accident Occurred in Batu Pahat, Johor

The total number of road traffic accidents occurred in Batu Pahat at specific time period from year 2009 to 2014 was presented in Table 1.

Table 1 24-hours of RTA cases in Batu Pahat from 2009 to 2014.

Time	Accident cases
00:01-02:00	1925
02:01-04:00	1218
04:01-06:00	871
06:01-08:00	3381
08:01-10:00	3805
10:01-12:00	4128
12:01-14:00	5532
14:01-16:00	5476
16:01-18:00	5793
18:01-20:00	4748
20.01-22:00	4031
22:01-24:00	2595

This figure comprised the total number of accidents occurred at specific time in Batu Pahat from year 2009 to 2014.
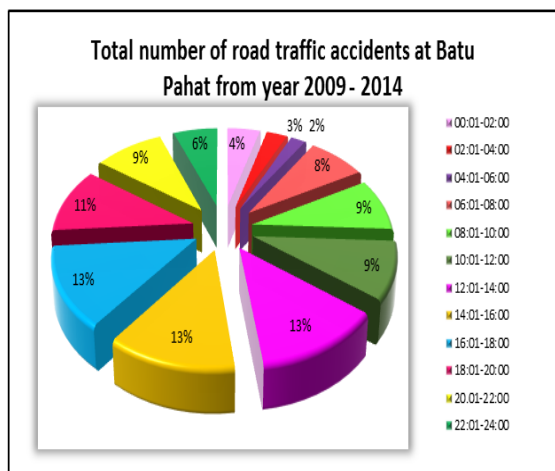

Fig. 1 Total number of RTA occurred at specific time in Batu Pahat.

It is clearly showed that the peak time which road traffic accidents occurred at most were at duration 12.01pm to 14.00pm, 14.01pm to 16.00pm and 16.01pm to 18.00pm with the percentages of approximately 13% for these three time periods. However, from Table 1, the time that road traffic accident occurred at most was at 16.01pm to 18.00pm with 5793 total number of accidents occurred in this time period followed by 5532 number of accidents occurred within 12.01pm to 14.00pm and 5476 cases within 14.01pm to 16.00pm.

The time that RTA occurred at most were the time where people on their way home from their work. People typically having a bad traffic jam at this time which will influence the road traffic accident to occur. Besides that, at 12.01pm to 14.00pm was the lunch time for the public so people were busy going out.

There was least number of road traffic accidents occurred at 04.01am to 06.00am with the total of 871 (2%) followed by 1218 (3%) road traffic accidents occurred at 02.01am to 04.00am. People were rarely traveled at this time. The peak time obtained from the analysis was supported by publication of Goswami and Sonowal [9] which they stated that the present study recorded more than 60% of the accidents during day time (6am to 6pm) and the peak time was between 12pm to 6pm. These times coincide with the hours when people are more active and mobile.

Log-rank test

The two survival groups can be compared statistically by testing the null hypothesis. The purpose of this test is to determine whether there any difference regarding survival between years 2009 to 2014.

H_0: There is no difference of survival RTA for different years.
H_1: There is some difference of survival RTA for different years.

Table 2 Overall comparisons for the different years

	Chi-square	df	Significant
Log-Rank (Mantel-Cox)	4.180	5	0.524

Based on the results above, there was no difference regarding survival in road traffic accidents from year 2009 to 2014. The survival rate of road traffic accidents from year 2009 to 2014 were almost the same.

4. Conclusion

Data analyse shows that RTA have three peak hours of RTA which are from 16:01pm to 18:00pm due to the highest number of road traffic accident occurred from year 2009 to 2014, secondly from 12:01pm to 14:00pm and lastly from14:01pm to 16:00pm. This is three main time period which road traffic accidents occur at most, so this is the most considerable time to take an appropriate action to reduce the road traffic accidents. Road traffic accidents were rarely occurring at the middle of the night till dawn. Based on the results of log-rank test, there are no significantly difference of survival RTA from 2009 to 2014. The survival curves of road traffic accident were almost similar for every year.

References

[1] Kareem, A. (2003). "Review of Global Menace of Road Accidents with Special Reference to Malaysia a Social Perspective" in *Malaysian Journal of Medical Sciences*, Vol. 10. No. 2. pp. 31-39.

[2] Sharma, G., Upadhyay, M. and Ramaboot, S. (2004). "The Challenge of Road Traffic Injury in South-East Asia: Moving Beyond Rhetoric" in *Regional Health Forum WHO South-East Asia Region*, Vol. 8. No.1.

[3] Razzak, J.A. and Luby, S.P. (1998). "Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan, through the capture-recapture method" in *International Journal of Epidemiology*, pp. 866-870.

[4] Che Zakaria, N.A. (2010). "Road Safety Level: A Case Study of Major Road Linked Terengganu and Kuantan Town". Unpublished Bachelor Degree dissertation, Faculty of Civil Engineering & Earth Resources, Universiti Malaysia Pahang.

[5] Lu, P.H. (2011). "Customer Satisfaction towards Retailer". Unpublished Master dissertation, Gotland University.

[6] Hosmer, D.W., Lemeshow, S. and May, S. (2008). Applied Survival Analysis: Regression Modelling of Time-to-Event Data, Willey Series in Probability and Statistics, Second Edition.

[7] Shi, Y., Zhang, L. and Liu, P. (2014). "Survival Analysis of Urban Traffic Incident Duration: a Case Study at Shanghai Expressways" in *Journal of Computer*s, Vol. 26. No. 1.

[8] Kleinbaum, D.G. and Klein, M. (2005). Survival analysis. A self-learning text. Second Edition.

[9] Goswami, A. and Sonowal, R. (2011). "A Statistical Analysis of Road Traffic Accidents in Dibrugarh city, Assam, India" in Division of Epidemiology and Nutrition, Regional Medical Research Centre.