Weibull Regression of Survival Rate of Patients with Coronary Artery Disease in Sarawak

Mohd Asrul Affendi Abdullah[1]*, Conny Lau Siew Jin [1], Rohayu Mohd Salleh[1], Siti Afiqah Muhamad Jamil[2], Shahiron Shahidan[3]

[1]Depertment of Mathematics and Statistics, Faculty of Applied Science and Technology, Univeristi Tun Hussein Onn Malaysia, 84600, Campus Pagoh, Muar, Johor, Malaysia.

[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[3]Faculty of Civil and Environmental Engineering, Universiti Tun Hussien Onn Malaysia, Parit Raja 86400, Batu Pahat Johor, Malaysia

*Corresponding author's email: afendi@uthm.edu.my

ABSTRACT: Survival analysis of Coronary Artery Disease (CAD) in Sarawak, Malaysia by parametric model, Weibull distribution. CAD is the principal cause of mortality in Malaysia while it is the second leading causes in Sarawak. According to Department of Statistic Malaysia in 2018, CAD is the leading cause of the death from the years of 2007 to 2017. The purpose of this research is to estimate the mortality rates due to CAD among Sarawakian population. With the consideration of Weibull distribution, it is known that the mortality rate of the patients with CAD is greater than 1 which indicates that the hazard ratio is increasing over time of study. Moreover, this study also been conducted to identify whether which age groups of patients have the highest death rate and to determine the affective factors on the disease and the rate of their effectiveness. Generally, the result shows that the patients who aged over 80 years old has the lowest survival proportion than other groups of patients. The patients of CAD with the presence of the high blood pressure level, high cholesterol level and high glucose level have the higher risk to death than the patients who absence with these risk factors..

Keywords: Coronary Artery Disease (CAD), Weibull distribution, Mortality rate

1. INTRODUCTION

Coronary Artery Disease (CAD) is a type of heart disease in this world which caused by a build-up of a waxy substance called plaque inside the coronary artery.. The risk factors of CAD can be classified as modifiable and non-modifiable [1][2]. Modifiable risk factors are the factors than can be controlled such as blood pressure and diabetes while unmodifiable factors are the factors that cannot take measure to change including age, gender and family history .There is about 10,000 people die with cardiovascular disease because of smoking behavior [3]. Depression is another reason that increases the mortality rate nowadays especially after CABG treatment [12]. [13] conclude that the obesity is the most influenced factor that increase the prevalence of CAD in Sarawak.

In Sarawak, CAD is the second leading cause of death in 2006 [13]. Among the region in Sarawak, the district of Kanowit has the highest death rate in 2017 with 9.1 per 1000 residents followed by Selangor and Negeri Sembilan [14].[15] states that the statistics on heart disease in Sarawak is worrying from years to years.

However, there is a lack of information and studies that related with this issue in Sarawak. Most of the researcher prefer to use the semi-parametric and nonparametric method in survival data analysis [4][5]. [6] indicate the survival t the survival times on regression analysis of lifetime data for leukemia patients to accelerate the failure time by using Weibull and log-logistic models.

2. METHODOLOGY

A. *Weibull distribution*

Weibull distribution is a type of parametric method that use to determine the failure time of the patients with CAD. It is the generalization function of exponential distribution. The probability density function of this distribution is

$$f(t;\theta) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^{\beta}} \qquad (1)$$

where θ is the vector of parameters of β and α, which the shape and the scale parameter respectively. The survival and hazard function of Weibull distribution are;

$$S(t;\theta) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}} \qquad (2)$$

and

$$h(t;\theta) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} \qquad (3)$$

respectively. Estimating μ by $\bar{x}$ and σ by s. By using the algebra and eliminating α, the following equations will be found.

$$2\ln\bar{x} - 2\ln\Gamma\left(1+\frac{1}{\beta}\right) = \ln(s^2+\bar{x}^2) - \ln\Gamma\left(1+\frac{2}{\beta}\right) \qquad (4)$$

$$\ln\Gamma\left(1+\frac{2}{\beta}\right) - 2\ln\Gamma\left(1+\frac{1}{\beta}\right) - \ln(s^2+\bar{x}^2) + 2\ln\bar{x} = 0 \qquad (5)$$

3. RESULTS AND DISCUSSION

Figure 1 is the graph of the overall mortality rate of CAD in Sarawak. Based on the graph, it is clearly to see that the death rate of the patients is increasing over time under the time of obeservation.
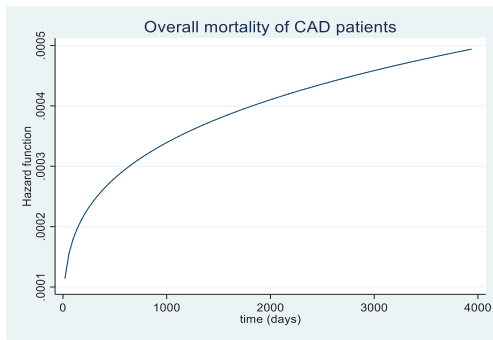
Fig 1: Overall Mortality rate of CAD patients

Based on the 2018 statistics on causes of death in Malaysia, it states that CAD with another name of IHD the main causes of death from the years of 2007 to 2017. The causes of death due to this disease has increased 54 percent as compared to 10 years ago. In the year of 2007, there are only 8,776 patients are died on this disease. This statistic has increased to 13,503 patient who die due to CAD in the year of 2017. In other words, there are around 37 deaths per day in 2017 as compared to 24 deaths in 2007.

Table 1 Comparison mortality rates between different age group of patients

Variables	Frequency	Coefficient	95% Confidence Interval	
Age II	68	0.1107	-1.3275	1.5489
Age III	66	0.0147	-1.4279	1.4573
Age IV	30	0.2574	-1.2035	1.7184

It's clearly known that age group II has the highest frequency to get the disease of CAD which has 40.24 percent (Frequency=68) patients diagnosed this disease while the age group I has the lowest chance to get the CAD disease where there is only 5 patients over 169 subjects have this disease during the time event. Age group III has the second highest frequency to diagnose with CAD disease with 39.05 percent (Frequency=66) and followed by age group IV which has 30 patients get the disease during the period of time study. The age group III has higher survival rate than age group II. However, the age group IV has the lowest survival rate (highest mortality rate) among the others group which this group of patients has the highest hazard ratio of 1.2936 at time=0 than others in the figure 2.
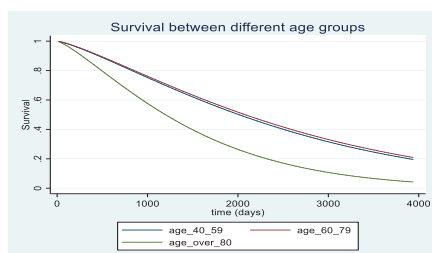

Fig 2: Survival rate between difference age groups

4. CONCLUSION

The mortality rate of the patients with CAD disease is monotonically increase under the time of observation. In this study, the mortality rate of the CAD patients is greater than 1 which indicates the hazard ratio is increasing over time of study. Furthermore, the patients with the disease which aged over 80 years old has the highest rate of death among the other patients in this study. There are three main risk factors that lead to the cardiovascular diseases in Sarawak, which are the presence of high blood pressure level, high cholesterol level and high glucose level. Larger sample size is needed when taking the survival analysis to get a more accurate result. The bigger the sample size, the smaller the sample error, the more precise the result.

REFERENCES

[1] American Heart Association Dallasm, Texas (2007). Heart Disease and Stroke Statistics.

[2] Heart and Stroke Foundation Canada (1997). Heart Disease and Stroke Statisitc

[3] Abdullah, W.M.S.W., Yusoff, Y.S., Nurlida, B. & Yusuf, M.M. (2017). Malaysia Rates Due to Coronary Heart Disease by Specific Sex and Age Groups among Malaysians. *World Congress on Engineering and Computer Science*, 6.

[4] Zaman, B.Q. & Pfeiffer K.P. Survival Analysis in Medical Research. Medical University Innsbruck.E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.

[5] Zhang, Z.H. (2017). Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translation Medicine*. 4(24). p. 484.

[6] Feigl, P. & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information, Biometrics 21. pp. 826–838.

[7] Saleem, M., Mahmud, Z. & Khan, H.K. (2012). Survival Analysis of CABG Patients by Parametric Estimations in Modifiable Risk Factors – Hypertension and Diabetes. *American Journal of Mathematics and Statistics*. 2(5). pp.120-128.

[8] Blumenthal, J.A., Lett H.S, Babyak, M.A, William, W., Smith, P.K., Mark D.B., Jones, R., Mathew, J.P. & Newman, M.F (2003). Depression as a risk factor for mortality after coronary artery bypass surgery. *The Lancet*. 362. pp. 604-609

[9] Chang, C.T, Lee. P.Y & Cheah, W.L (2011). The prevalence of Cardiovascular Risk Factors in the Young and Middle-Aged Rural Population in Sarawak, Malaysia. *Journal of Nutrition and Metabolism*.

[10] Bernama (2018). Ischemic heart disease remains No 1 killer in Malaysia. *Malaysia Kini.*

[11] Sman, C. (2016). Statistics on heart disease in Sarawak Worrying. *Borneo Post Online.*

[12] Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International journal of Ayurveda research*, *1*(4), pp. 274.

[13] Bland, J. M., & Altman, D. G. (2004). The logrank test. *Bmj*, *328*(7447), pp. 1073.

[14] Schober, P., & Vetter, T. R. (2018). Survival analysis and interpretation of time-to-event data: the tortoise and the hare. *Anesthesia and analgesia*, *127*(3), pp. 792.