IJBPAS, October, 2018, 7(10): 1852-1860

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

# SURVIVAL FITTING OF CORONARY HEART DISEASE USING A PIECEWISE EXPONENTIAL MODEL APPROACH

*MEL JEAN RICIO ESTOLANO AND JOHN CLOSTER FRONDA OLIVO

Department of Statistics, Central Luzon State University, Science City of Muñoz, Nueva Ecija

* Corresponding Author: E Mail: meljean_estolano@yahoo.com

Received 26th March 2018; Revised 11th April 2018; Accepted 16th May 2018; Available online 1st Oct. 2018

https://doi.org/10.31032/IJBPAS/2018/7.10.4540

## ABSTRACT

The paper focuses on the survival distributions that would best describe the length of survival of coronary heart disease (CHD) patients from the time of diagnosis to time of discharge. From among the exponential family of distributions considered in this study, namely, lognormal, exponential and Weibull distributions, results indicated that the survival function of CHD patients from the time of diagnosis to time of discharge was better estimated by Weibull distribution. Cox regression analysis revealed that three significant risk factors, namely, religion, patient's medical history of hypertension, and family history of hypertension. Likewise, the estimated median survival probability of CHD patients was 50%.

For further refinement or improvement of the survival function of CHD patients from the time of diagnosis to time of discharge, the piecewise exponential model (PEM) was applied. Results revealed the superiority of PEM in the sense that it improved the survival probability of patients to 73% on the average. Moreover, the use of PEM using a survival time of 3-day interval gave a much higher log-likelihood value.

**Keywords: Survival Function, Hazard Rate, Median Survival Time, Weibull Distribution, Piecewise Exponential Model Approach**

## INTRODUCTION

Coronary heart disease (CHD) also known as coronary artery disease is the most common type of heart disease. It happens when the arteries that supply blood to heart muscle hardens and narrows that can lead to chest pain or a heart attack (NIH, 2014).

According to USA's National Institutes of Health (NIH), CHD is the leading cause of death for males and females in the USA. 17.6 million Americans were thought to have had the condition in 2006. It caused the death of over 425,000 people in the USA in 2006.

According to the latest World Health Organization (WHO) data published in May 2014, the number one cause of death in the Philippines is coronary heart disease. Deaths in the Philippines reached 88,871 or 16.86% of the total deaths. The age-adjusted rate is 161.43 per 100,000 people. Currently, the Philippines is ranked #29 worldwide regarding deaths caused by coronary heart disease (World Life Expectancy, 2014).

Diseases of the heart had been consistently the number one cause of death for several years in the Philippines according to the Department of Health. The 2013 number of deaths from this condition is 5.4% higher than what was reported in 2012. 6 out of 10 males have a disease of the heart while the ratio for females is 4 out of 10, making men more prone to die due to this disease.

Medical and epidemiological studies are mostly conducted with interest in measuring the occurrence of an outcome event. Studies focused on measuring time to event or outcome; survival analysis is mainly used. Time to event could vary from time to fatal event such as death, or time to occurrence of a clinical endpoint such as disease, or attainment of a biochemical marker.

In the study of Han, et al. (2009), non-parametric models are easy to understand and apply because no assumptions on distribution are needed but generally, the uncertainty is high and less efficient than proper parametric models. Parametric models are parsimonious and make use of all the data. However, it is restricted by the distribution assumption, and more effort is needed to ensure parametric modeling. A middle ground to these models was studied piecewise parametric modeling. It gains most of the strength of both parametric and non-parametric models while minimizing the weaknesses.

Piecewise exponential distribution is a simple and flexible distribution for modeling time to event data (Kim and Proschan, 1991, Quio et al., 1999, Demarqui et al., 2008). According to Zhou (2000), piecewise exponential distribution is the most flexible among other distributions, since it may have many pieces and thus many parameters. It is also used to connect the parametric and nonparametric method: when the number of pieces/parameters grows to infinite, the parametric model becomes a non-parametric model. A recommendation from the study of Singh and Totawattage (2013) suggested the use of a piecewise constant exponential hazards model to allow more flexible modeling with weak parametric assumptions.

The study generally aims to use piecewise exponential model (PEM) approach in fit the survival of patients with coronary heart disease and describe the distribution of the length of survival of the respondents admitted in Dr. Paulino J. Garcia Memorial Research and Medical Center (PJG-MRMC) compared with the parametric and semi-parametric methods.

**MATERIALS AND METHODS**

Secondary data were obtained from the medical records of patients with coronary heart disease admitted at Dr. Paulino J. Garcia Memorial Research and Medical Center (PJG-MRMC) in Cabanatuan City, Nueva Ecija, Philippines. The hospital, established in 1972, is the center of health researches in Nueva Ecija sponsored by the Department of Health.

Patients' records from January 2015 to December 2015 were used in the paper. The data were summarized in a monthly pattern to quickly determine the month the patient has been diagnosed and admitted to the hospital. The variables in the study were the demographic characteristics of the patients such as age, sex, religionand civil status. The date of admission to the hospital until the time of death was also included in the study. Risk factors such as past medical history and family history of diabetes and hypertension, alcohol intake, drug use, and smoking history were also considered in the study. The event was defined as the time of discharge or death. Survival time was the number of days from admission to the day discharged or died.

Descriptive statistics were used to describe the demographic and risk characteristics of the patients admitted with CHD.For the PEM approach, the first step was looking for the parametric approaches - exponential, Weibull, and lognormal - probability plotting and log-likelihood were used as criteria in determining which distribution fits the data best. Cox regression was also used as a criterion in determining the best-fitting distribution and the significant demographic and risk factors. The distribution with the largest log-likelihood was selected. Survival probabilities, hazard function, and survival curves for the significant variables were estimated.The time interval that would give the best fit to the data was also fitted to the data. Log-likelihood, deviance goodness-of-fit test and Pearson goodness-of-fit tests were obtained to determine the best fitting interval. Only the significant factors from the selected parametric approach were subjected to Poisson regression. These intervals and variables were subjected to Poisson regression for the computation of survival and hazard rates of each variable.

In selecting the best-fitting distribution for the data, the log-likelihood value and estimated median survival probabilities of the selected parametric distribution and piecewise exponential distribution were utilized to determine the goodness-of-fit of the distribution. The distribution with the highest log-likelihood value and estimated median survival probability was chosen as the best-fitting distribution for the survival of CHD patients.

**RESULTS AND DISCUSSION**

A total of 415 patients with CHD were considered in the study. Out of these patients, 2 (0.48%) respondents were considered censored or those who did not meet the event or the time of discharge. These patients were transferred to other hospital and information on their survival was unknown.

Figure 1 showed the monthly number of patients with CHD admitted at PJG-MRMC. Many cases had been recorded in January with 54 patients and April had the lowest number of cases with 21 patients admitted. The survival times of the patients were illustrated in Figure 2. It showed that most patients with CHD had a survival time of 3 days.

Descriptive statistics such as mean, standard deviation, minimum value and maximum value were obtained. Table 1 showed that the average age of patients admitted with CHD is 68.79 years with a standard deviation of 0.113 years. Minimum and maximum ages were 19 years and 100 years, respectively. The survival times of the patients ranged from 1 day and 27 days with an average of 5.5 days and a standard deviation of 4.08 days. The median and modal age were 69 years old.

Figure 3 summarizes the probability plot of the parametric distributions. Based on the graphs, the data most likely followed a lognormal or Weibull distribution since most of the points lied on the line. To support the subjective assessments on the graphical presentation, log-likelihood values were obtained. More substantial value of log-likelihood or value closer to zero implies that it gives a better fit to the data. Log-likelihood values for the parametric distributions and Cox regression were shown in Table 2. It showed that Lognormal and Weibull distributions had small variations regarding their log-likelihood values.

To further determine which would be the best-fitting distribution, the significance of the variables were considered. It can be considered, based on the assessments on the graph and log-likelihood, that Weibull distribution was the distribution that best fits the CHD data. The significant risk factors were religion group, patient's medical history of hypertension, and family history of hypertension.

Since it was concluded that Weibull distribution as the best-fitting distribution, survival and hazard functions were estimated using the functions below

$$\boldsymbol{S}\ (\boldsymbol{t}) = \mathbf{exp}\ \{-(\lambda t)^{p}\} \text{ and } \boldsymbol{h}\ (\boldsymbol{t}) = p\lambda t^{p-1}$$

where $t$ is the survival time, scale parameter $\lambda = 0.6614$ , and shape parameter $p = 1.5120$.

Figure 4 illustrated the behavior of the hazard plot and survival times of patients discharged with CHD. Hazard plot showed that there was a constant hazard rate. This implied that the hazard experienced by the patients was decreasing over time since the slope of the hazard function was decreasing over time. Looking at the survival curve, the estimated median survival time was 4.80873 days and a probability of approximately 0.5. These results implied that using a Weibull distribution, the probability of being discharged within 5 days or less was 0.5.

Selection of intervals was made for the fitting of PEM approach on the data. Each time interval had insignificant deviance goodness-of-fit test and Pearson goodness-of-fit test implying that data fitted. Table 3 showed that the three days interval had the highest log-likelihood and lowest deviance and Pearson goodness-of-fit test value, thus, the survival time in this study was split into the 3-day interval. Poisson regression was used to estimate the hazard functions that were used to compute for the hazard rate of the respondents.

Figure 6 presented the survival curve of the significant risk factors using piecewise exponential distribution. Results showed that the estimated median survival time of patients' medical history of hypertension was approximately 5 days with a probability of 0.74. Looking at Table 4, the survival curve of the patient's medical history of hypertension using piecewise exponential distribution had an estimated median survival time of approximately 5 days with a probability of 0.73. The survival curve of family history of hypertension of the respondents using piecewise exponential distribution showed that the estimated median survival time was approximately 5 days with a probability of 0.78.

Median survival probabilities of the significant risk factors using Weibull and piecewise exponential distribution were compared to determine the best-fitting data. Higher probabilities would indicate a better fit of the distribution to the data. It was found out that the estimated median survival probabilities of the risk factors were higher when using piecewise exponential distribution.

To further support the goodness-of-fit of the PEM approach, Table 5 showed that the log-likelihood of the PEM was larger than the Weibull distribution. It implied that the former was more flexible and gave better fit to the data than that of the latter.

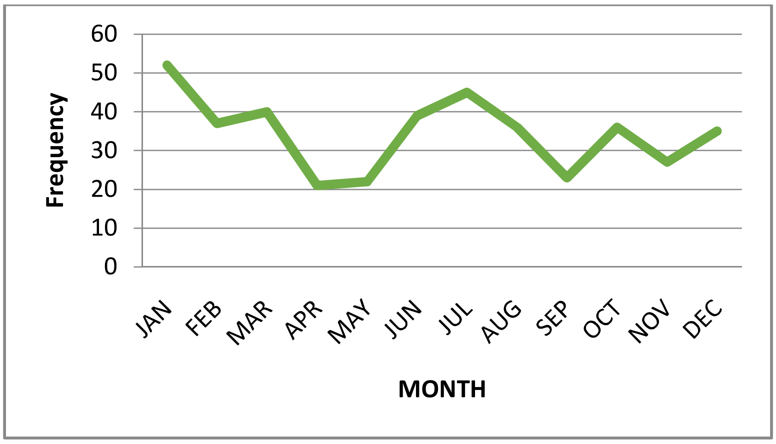


**Figure 1: Number of patients with CHD admitted every month in 2015**

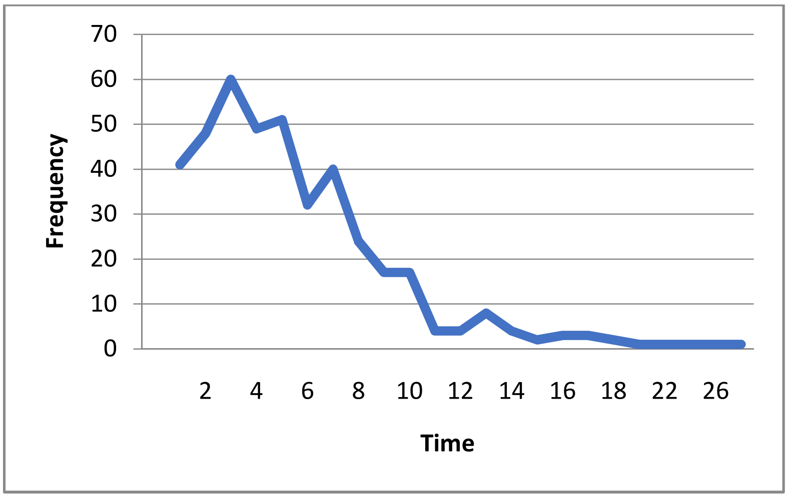


**Figure 2: Number of patients with CHD in different survival times**

**Table 1: Descriptive Statistics of Age and Survival Time**

| Variable | Mean | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|
| Age | 68.7880 | 11.3185 | 19 | 100 |
| Survival Time | 5.5277 | 4.041 | 1 | 27 |

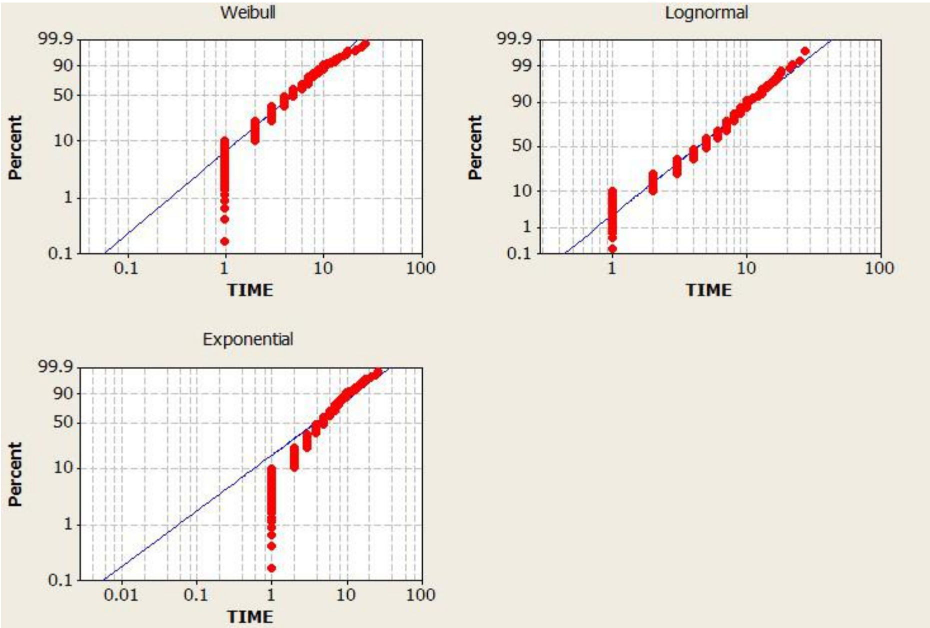


**Figure 3: Probability plot of Weibull, Lognormal, and Exponential distribution**

**Table 2: Log-likelihood Values of Parametric Distributions in Distribution Fitting**

| Distribution | Log-likelihood | |
|---|---|---|
| | Distribution Fitting | Cox Regression |
| Exponential | -1121.132 | -1119.031 |
| Weibull | -1075.009 | -1068.722 |
| Lognormal | -1065.127 | -1062.048 |

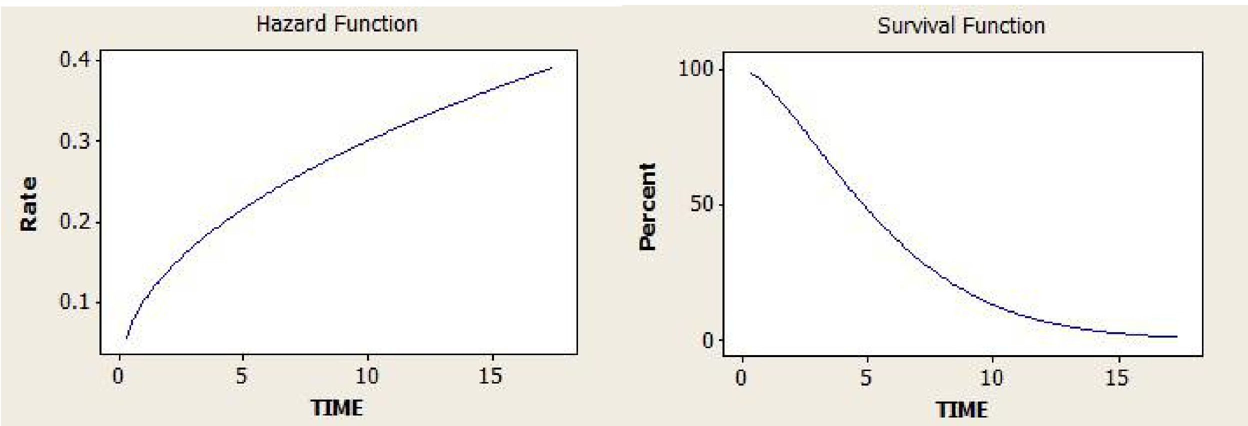


**Figure 4: Hazard plot and survival curve using Weibull distribution**

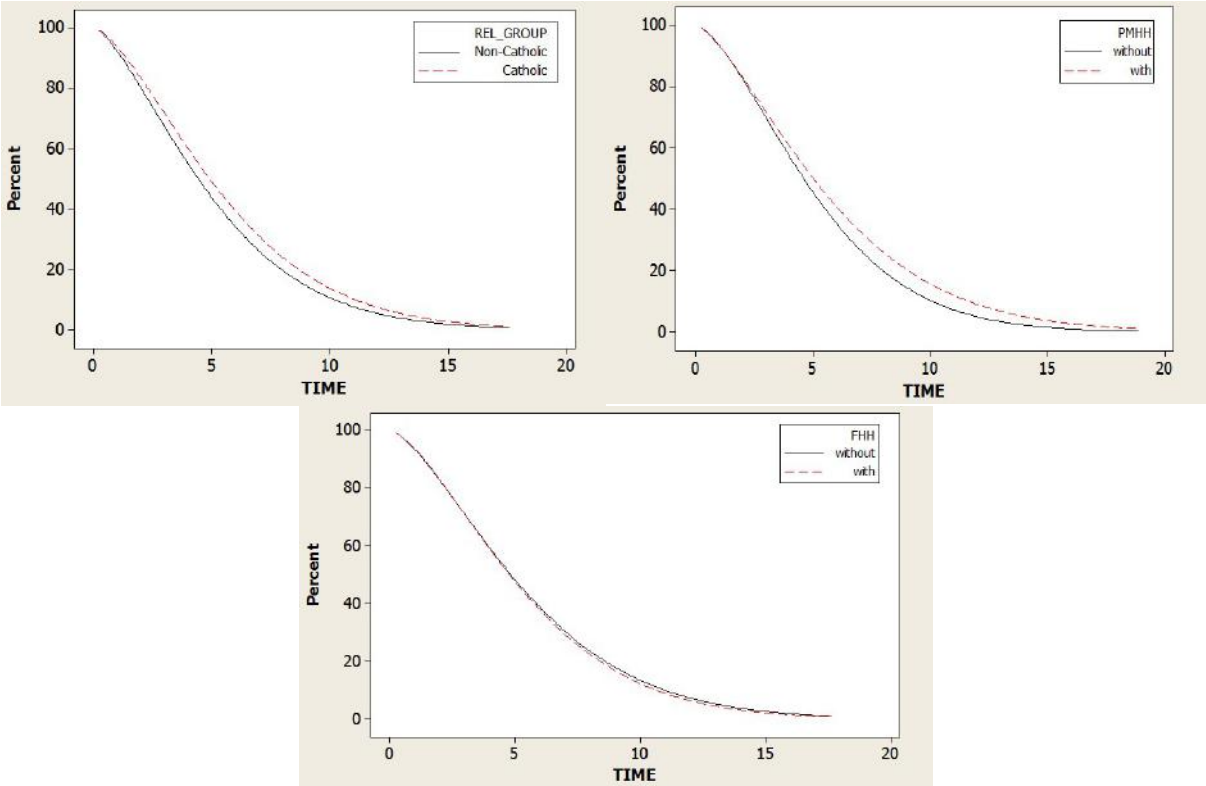


**Figure 5: Survival curve of significant factors using Weibull distribution**

**Table 3: Log-likelihood Values of Selected Time Intervals**

| Interval | Log-likelihood | Pearson Goodness-of-Fit | Deviance Goodness-of-Fit |
|---|---|---|---|
| 3 days | -11.0843 | 204.2319 | 18.1686 |
| 5 days | -11.9235 | 287.4766 | 19.8470 |
| 7 days | -11.6501 | 252.3886 | 19.3002 |

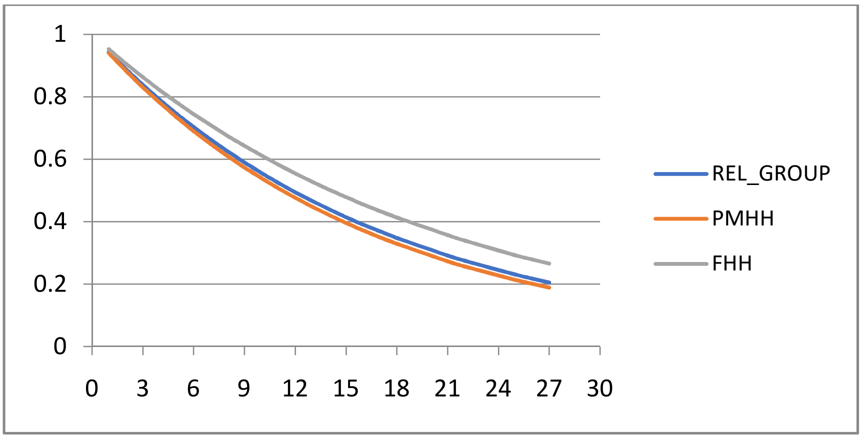


**Figure 6: Survival curve of significant risk factors using piecewise exponential distribution**

**Table 4: Median Survival Probabilities of Risk Factors using Weibull and Piecewise Exponential Distribution**

| Variable | Median Survival Probability | |
|---|---|---|
| | Weibull | Piecewise Exponential |
| REL_GROUP | 0.50 | 0.75 |
| PMHH | 0.53 | 0.73 |
| FHH | 0.57 | 0.78 |

**Table 5: Log-likelihood Values of Piecewise Exponential and Weibull Distribution**

| Distribution | Log-likelihood |
|---|---|
| Piecewise Exponential | -11.084 |
| Weibull | -1075.009 |

## CONCLUSION

In the comparison of the distributions, the piecewise exponential model (PEM) approach had higher log-likelihood value than parametricapproaches. Regarding the estimated median survival probability, PEM yielded higher probabilities of survival than parametric approaches. Therefore, it can be concluded that the PEM approach gave the best fit for the survival of CHD patients as well as to time-censored data.

## REFERENCES

**[1]** National Institute of Health (2014). What is coronary artery disease? https://www.nlm.nih.gov/medlineplus/coronaryarterydisease.html

**[2]** World Life Expectancy (2007). Philippines: coronary heart disease. http://www.worldlifeexpectancy.com/philippines-coronary-heart-disease

**[3]** Han G, Schell MJ, Kim, J. Improved survival modeling using a piecewise exponential approach. National Center for Biotechnology Information, Stat Med. 2009; 33(1):59-73.

**[4]** Kim and Proschan (1991), Quio et al., (1999), Demarqui et al., (2008) found at the article "Improved survival modeling using a piecewise exponential approach" by Han, G. et al. (2009).

**[5]** Zhou M. Exponential and piecewise exponential distribution. 2000. Date Accessed:May.2016.http://www.ms.uky.edu/~mai/sta635/expo2.pdf

**[6]** Singh RS, Totawattage DP (2013). The statistical analysis of interval-censored failure time data with applications. Open Journal of Statistics. 2013; 3:155-166.