**10th National Convention on Statistics (NCS)**
EDSA Shangrila Hotel
October 1-2, 2007

# Parametric and Nonparametric Survival Analysis of Myocardial Infarction Patients Admitted in Dr. Paulino J. Garcia Memorial Research and Medical Center in Cabanatuan City

by

Maria Cristina V. Mingala and Marcial P. Estolano

For additional information, please contact:

| | | |
|---|---|---|
| Author's name | : | Maria Cristina V. Mingala |
| Designation | : | Part-time Lecturer |
| Affiliation | : | Central Luzon State University (CLSU) |
| Address | : | Department of Statistics |
| | | College of Arts and Sciences, CLSU |
| Tel. | : | (06344) 456-5192 |
| E-mail | : | kutenten@yahoo.con |

| | | |
|---|---|---|
| Co-author's name | : | Marcial P. Estolano |
| Designation | : | Associate Professor V |
| Affiliation | : | Central Luzon State University (CLSU) |
| Address | : | Department of Statistics |
| | | College of Arts and Sciences, CLSU |
| Tel. | : | (06344) 456-5192 |
| E-mail: | | mpestolano6@yahoo.com |

## Parametric and Nonparametric Survival Analysis of Myocardial Infarction Patients Admitted in Dr. Paulino J. Garcia Memorial Research and Medical Center in Cabanatuan City

by

Maria Cristina V. Mingala and Marcial P. Estolano

**ABSTRACT**

The paper focuses on parametric and nonparametric survival models that would describe the distribution of the length of survival myocardial infarction (MI) patients of Dr. Paulino J. Garcia Memorial Research and Medical Center from the date of diagnosis to the date of discharge. Patients' medical records from January 2004 to December 2005 which include the patients sex, age and civil status while the risk factors such as hyperlipidemia, diabetes, hypertension, smoking and family history of atheroscherotic disease of patient were used.

The average age of MI respondents was 58.62 $\pm$ 1.26 years old and the average survival time was 6.81 $\pm$ 0.31 days. Married males diagnosed with risk factors were more prone to MI disease.

Kaplan-Meier estimates of the median survival time of MI patients was 6 days while parametric Weibull distribution gives an estimated median survival time of 6.53 days.

Logrank and Wilcoxon tests indicated that survival times of married patients was significantly longer than patients who were single and widow. No significant differences exist on the survival times of MI patients classified according to age, sex and presence or absence of risk factors.

The Cox proportional hazard model indicated that patients with family history of heart disease is the only significant predictor risk variable that best describes the distribution of survival times of MI patients with an estimated model:

$\log_e(t) = 2.10185 - 0.22034(\text{family history}) + 0.43397\varepsilon_p$

where t is the survival time and $\varepsilon_p$ is the $p^{th}$ percentile of error distribution.

## I. Introduction

The myocardial infarction (MI), also known as heart attack, results when a blood clot completely obstructs the artery that supplies blood to the heart muscles resulting to necrosis of the cardiac muscle. It is caused by coronary atherosclerosis, a gradual process in which plaques of cholesterol are deposited in the walls of coronary artery. Cholesterol plaques cause hardening of arterial walls and narrowing of the inner channel of the artery. Arteries that are narrowed by atherosclerosis cannot deliver enough blood to maintain normal function of the parts of the body they supply. Eventually, when the inner channel is completely blocked, MI occurs.

In general, MI can occur at any age, but its incidence rises with age. The actual incidence is dependent upon the predisposing risk factors for atherosclerosis, such as hyperlipidemia (high blood cholesterol), diabetes, hypertension (high blood pressure), smoking, male gender, and family history of atherosclerotic arterial disease. The presence of any risk factor is associated with doubling the relative risk of developing atherosclerotic coronary artery disease (CAD)

In survival analysis, the behavior of generally terminal diseases as they affect the human body is determined by producing survival curves. The major variable of interest in this analysis is the number of days that the respective patients survive. Basically, the methods used/offered in survival analysis deal with the same questions as many of the other procedures, except that all methods in this type will handle censored data. In general, censored observations arise whenever the dependent variable of interest represents the time to a terminal event and the duration of the study is limited in time. That is, some of the respondents may have lost contact before the study ended, or the symptoms of the disease may have not reoccurred. In any way, the exact number of days of relapse is unknown, except that it must be greater than their last day of contact.

A common research question in medical research is to determine whether or not certain continuous (independent) variables are correlated with the survival times. Straightforward multiple regression techniques are not appropriate for survival analysis because survival times usually follow positively skewed distributions and some observations can be incomplete (censored). Accordingly, we use the Cox's Proportional Hazard Model.

**Statement of the Problem**

Generally, the study described the distribution of the length of survival from the date of diagnosis to the date of discharge of MI patients admitted at Dr. Paulino J. Garcia Memorial Research and Medical Center (PJGMRMC), Cabanatuan City.

Specifically, the study answered the following questions:

1. What are the social characteristics of the respondents in terms of sex, age and civil status?
2. What are the risk characteristics of the respondents in terms of hyperlipidemia, diabetes, hypertension, smoking and family history of atherosclerotic disease?

3. Are there significant differences on the occurrence of the disease between sexes, ages and civil status?
4. What is the distribution followed by the survival (in days) of MI patients?
5. Does the distribution of the length of survival the same within each social and risk characteristics?

**Scope and Limitation**

This study on the assessment of the length of survival for the day of diagnosis to time of discharge of Myocardial Infarction (MI) patients tackled mainly the concerns of MI disease to each respondent, regardless of other accompanied disease(s) he/she suffered.

The data used were from the medical records of heart attack patients of Dr. Paulino J. Garcia Memorial Research and Medical Center (PJGMRMC) in Mabini Street, Cabanatuan City, Nueva Ecija hospitalized from January of 2004 up to December of 2005. This study focused only on the social characteristics namely age, sex and civil status of respondents; and risk characteristics such as hyperlipidemia, diabetes, hypertension, smoking and family history of atherosclerotic diseases.

## II. Research Methodology

### Description of Data

Patients' medical records from January 2004 to December 2005 were used in this study. It included the social characteristics of the patients namely sex, age and civil status. The risk characteristics, however, are hyperlipidemia, diabetes, hypertension, smoking and family history of atheroscherotic disease. Moreover, the dates of admission and discharges were also included.

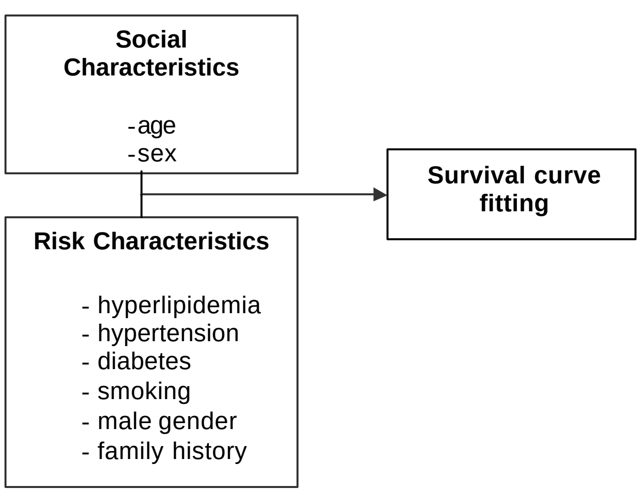
### Sampling Procedure

All MI patients were gathered from the master list of patients admitted for the years 2004 and 2005. All available data were obtained and subjected to data analysis.

### Research Analysis

Descriptive statistics and chi-square tests were used to describe the sample data and to determine if there exist a comparative difference between the survival times of the different social and risk characteristics. In estimating the survival curve of the data gathered, Life Table Method and Kaplan-Meier Product Limit Estimation were used. Parametric distributions such as lognormal, exponential and Weibull distributions were also utilized for the same objective. Likewise, Logrank and Wilcoxon tests were used in comparing the survival curves of the different classifications of the attributes within the social and risk characteristics. Cox's proportional hazards model was also used to determine the prognostic factors related to survival time.

### Conceptual Paradigm



## III. Results and Discussion

### Descriptive Analysis

From the total of 103 patient records, one was considered as censored observation. This particular patient was recorded as ‘transferred to other hospital’ and thus, his exact survival time was unknown.

To primarily determine if there existed a significant difference in the total number of MI cases from year 2004 to 2005, the chi-square goodness-of-fit test was used. Table 1 shows that there was an increase on the number of MI cases from the year 2004 to 2005. From 103 respondents, 66 or 64.1% were from year 2005 and 37 or 35.9% were from year 2004.

**Table 1.** Distribution of MI Patients between Years

| Year | Count | Percent |
|---|---|---|
| 2004 | 37 | 35.9 |
| 2005 | 66 | 64.1 |
| *Chi-square* | 8.165 | P=0.004 |

Figure 1 shows that there were relative increases on MI cases during January, April, June, August, October, November and December and decrease during July of 2005 as compared to year 2004.

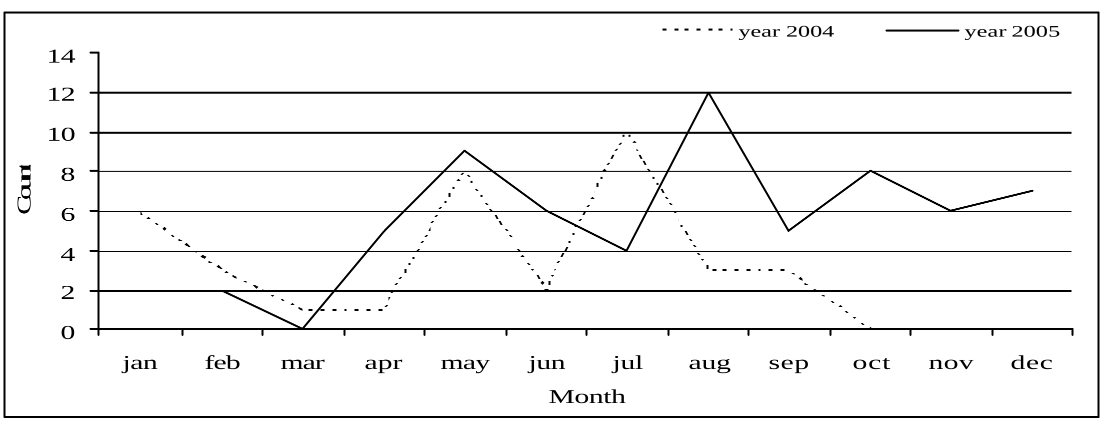



**Figure 1.** Distribution of MI Patients between Months

Descriptive statistics such as mean, standard deviation, variance, minimum and maximum values, and range were obtained. Table 2 shows that the survival time has a minimum of 1 day and has a maximum of 20 days, giving a range of 19 days with an average of 6.806 days and standard deviation of 3.190 days. The age of the respondents ranges from 35 to 87 years old with an average of 58.62 years and standard deviation of 12.78 years.

**Table 2.** Descriptive Statistics of Survival Time and Age

| Variable | Mean | Standard Deviation | Minimum | Maximum | Range |
|---|---|---|---|---|---|
| Survival Time | 6.806 | 3.190 | 1 | 20 | 19 |
| Age | 58.62 | 12.78 | 35 | 87 | 52 |

Table 3 gives the distribution of MI patients between the social characteristics. Males are known to be more prone to MI diseases as sustained by the sample, which has 69.90% male patients. This may be the effect of male attitude and lifestyle. They are more repressive on what they feel, which may contribute to the increase of their blood pressure. Most smokers are males and the physical capabilities of females like pregnancy and child birth need medical attention, thus their health are more monitored than the males'.

The age was categorized into two: the below and above the mean age, i.e., the 60 and below and the above 60. The first group is called the working group, wherein respondents belong in this group are economically productive and have such capacity to earn amounts for a living. On the other hand, the second group is called the retirement group where most of the patients are not economically productive. Chi-square statistic shows that there is no difference between the two age groups. This is explained by the fact that heart attacks can happen at any age.

Married individuals are more prone to MI disease than singles and widows, as shown by the 78.64% proportion by the sample and is supported by the chi-square statistic. This may be due to the fact that married people are more susceptible to stress than single or widow persons.

**Table 3.** Distribution of MI Patients Between Social Characteristics

| Variable | Category | Count | Percent | Chi-square |
|---|---|---|---|---|
| Sex | Male | 72 | 69.90 | 16.32** |
| | Female | 31 | 30.10 | |
| Age Bracket | 60 and below | 59 | 57.28 | 2.184 $^{ns}$ |
| | Above 60 | 44 | 42.72 | |
| Civil Status | Single | 4 | 3.88 | 98.00** |
| | Married | 81 | 78.64 | |
| | Widow | 18 | 17.48 | |

Table 4 shows the distribution of MI patients between the risk characteristics. Hypertension is not supported by chi-square statistics to contribute to the risk of having MI disease. This result may be a cause of a possible interaction of hypertension to other risk characteristics. On the other hand, other known risks such as hyperlipidemia, diabetes, smoking, and family history contribute to the risk of having MI disease as supported by the chi-square statistics.

**Table 4.** Distribution of MI Patients Between Risk Characteristics

| Variable | Category | Count | Percent | Chi-square |
|---|---|---|---|---|
| Hyperlipidemic | Yes | 66 | 64.08 | 8.165** |
| | No | 37 | 35.92 | |
| Diabetic | Yes | 77 | 74.76 | 25.252** |
| | No | 26 | 25.24 | |
| Hypertensive | Yes | 55 | 3.40 | .476 $^{ns}$ |
| | No | 48 | 6.60 | |
| Smoking | Yes | 41 | 39.81 | 4.282* |
| | No | 62 | 60.19 | |
| Has a Family History | Yes | 74 | 71.84 | 19.660** |
| | No | 29 | 28.16 | |

**Nonparametric Survival Distribution Analysis**

**Life Table Method**

The survival curve is used to find the 50th percentile and other percentiles of survival time. Steep survival curve represents low survival rate or short survival time, while a gradual or flat survival curve represents a high survival rate or longer survival. Given in Figure 2 is the survival plot of MI patients using life table with confidence intervals. The median survival time of MI patients is 7.0937 days. The third interval (6-8 days) has the largest survival time downfall followed by the second interval (4-6 days).

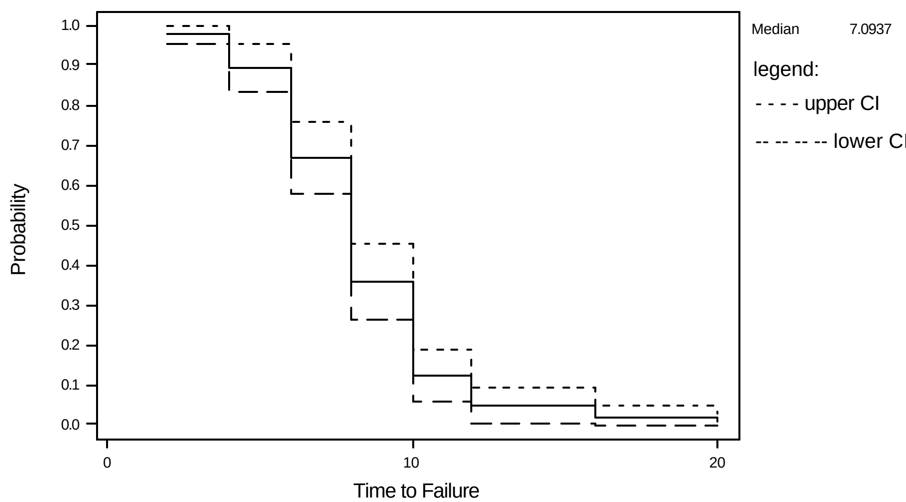



**Figure 2.** Survival Probability Plot of MI Patients Using Life Table Method

The hazard function gives the risk of failure per unit time (conditional failure rate) during the ageing process. It is estimated as the proportion of patients that are able to survive outside the hospital's premises in an interval per unit time, given that they have survived in the hospital to the beginning of the interval. Figure 3 gives the hazard plot of the survival time of MI patients. The hazard plot shows an increasing trend and reaches its peak during 9th to 11th days, then fluctuates.

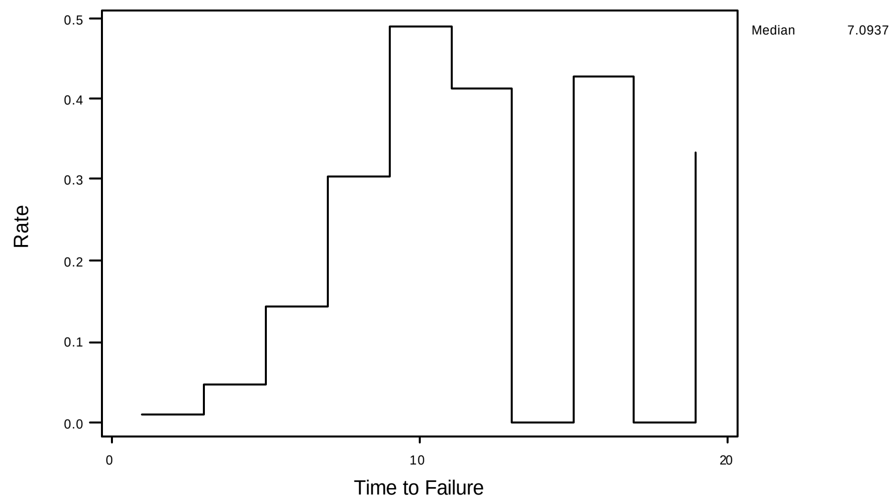



**Figure 3.** Hazard Plot for Survival Time using Life Table

**Kaplan-Meier Product Limit Estimation**

Kaplan-Meier (KM) product limit estimation can be considered as a special case of life table estimation where each interval contains only one observation. The only difference is that KM estimate is based on individual survival times while in the life table method survival times are grouped into intervals. The estimated median survival time from Figure 4 is 6 days. This is smaller compared to the median time of life table since the latter is based in ranges or intervals while the former is based in points. Figure 5 shows that there is a constant increasing trend in the hazard plot as time increases.

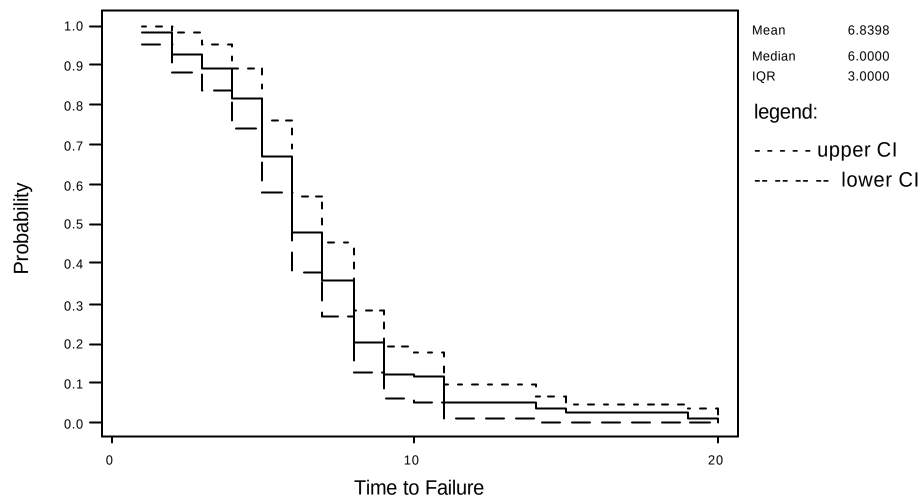



**Figure 4.** Survival Probability Plot of MI Patients Using Kaplan-Meier Estimation

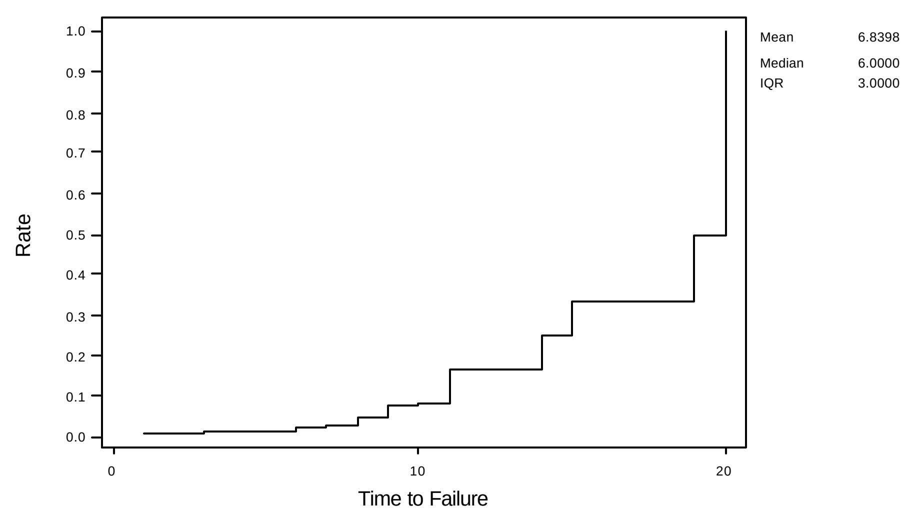



**Figure 5.** Hazard Plot for Survival Time Using Kaplan-Meier Estimation

## Parametric Survival Distribution Analysis

### Survival Distribution Fitting

Three types of parametric distributions were used: Lognormal, Weibull and Exponential Distributions for two reasons. First, they are the frequently used distributions in survival distribution fitting; and second, they are right-skewed distributions, the usual shape of survival functions.

Graphical methods such as probability plotting have long been used for display and interpretation of data for they are simple and effective. They are frequently used in place or in combination with numerical analysis; a plot of data simultaneously serves a number of purposes, which no numerical method can. Thus, the author chose this method for it provides estimates of the parameters of the distribution by simple graphical means. Moreover, it allows one to asses whether given data follow a particular theoretical distribution.

When an appropriate distribution is chosen, the probability and hazard plots result to a straight line fit to the data. Parameters of the chosen distribution can be estimated from the plot without tedious numerical calculations. Such estimates are adequate and useful for most purposes.

To support the subjective assessments of the above method, the chi-square goodness of fit test was used. A calculated chi-square value can be very small if the fit is good, or very large if the fit is bad.

Probability plotting and chi-square test were used to choose and compare the three theoretical distributions. Figure 6 to Figure 8 show the different probability plots of the survival time using the three right-skewed distributions to the survival data. It is clearly shown in Figure 6 that the survival function does not follow exponential distribution for there are many points that are too far from the straight line. Looking at a glance on the plots of Lognormal and Weibull distributions, it appears that both are appropriate for the data. But looking at the Chi-square goodness –of –fit test results in Table 5, we can see that Weibull distribution best fits the data.

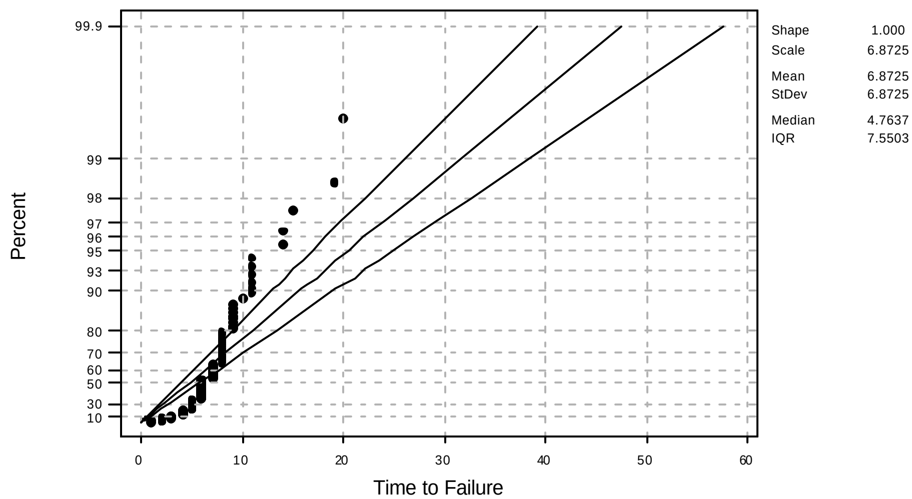



**Figure 6** . Exponential Probability Plot for MI Patients with Confidence Intervals

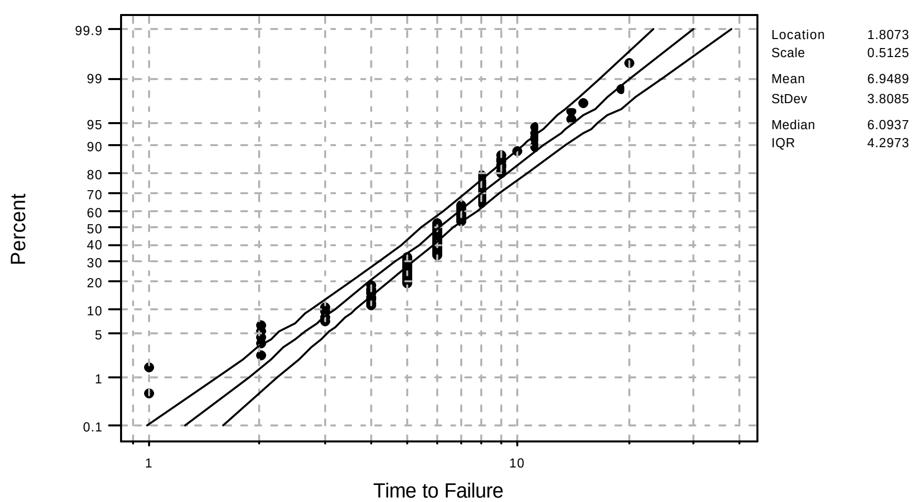



**Figure 7.** Lognormal Probability Plot with Confidence Intervals

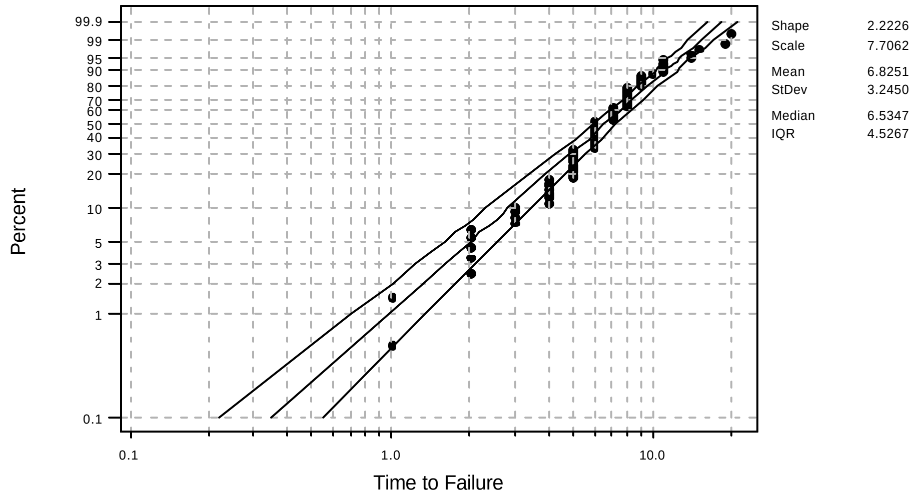



**Figure 8.** Weibull Probability Plot with Confidence Interval

**Table 5. Chi-square Goodness-of-Fit Test**

| **Distribution** | **Chi-Square Statistic** | **Critical Value** | **DF** |
|---|---|---|---|
| Exponential | 112.00 | 34.8053 | 18 |
| Weibull | 29.008 | 33.4087 | 17 |
| Lognormal | 489.146 | 33.4087 | 17 |

The Weibull distribution is characterized by the shape parameter ? and the scale parameter ?. Thus, the survivorship, hazard and probability functions of MI patients are given by $S(t) = \exp\left[-(7.7062t)^{2.2226}\right]$, $h(t) = (7.7062)(2.2226)(7.7062t)^{1.2226}$ and $f(t) = (7.7062)(2.2226)(7.7062t)^{1.2226} \exp\left[-(7.7062t)^{2.2226}\right]$, respectively.

The survival function of MI patients using Weibull distribution gives the median survival time which is 6.5347 days.

There exists a relationship between the parameters ? and ?. When ?=1, the hazard remains constant as time increases, this is the exponential case. When ?>1, the hazard rate increases but it decreases when ?<1 as time increases. Therefore, this distribution may be used to model the survival distribution of a population with increasing, decreasing, or constant risk. Figure 10 gives the hazard plot of the survival time of MI patients. The hazard plot shows a constant increasing trend throughout the distribution.

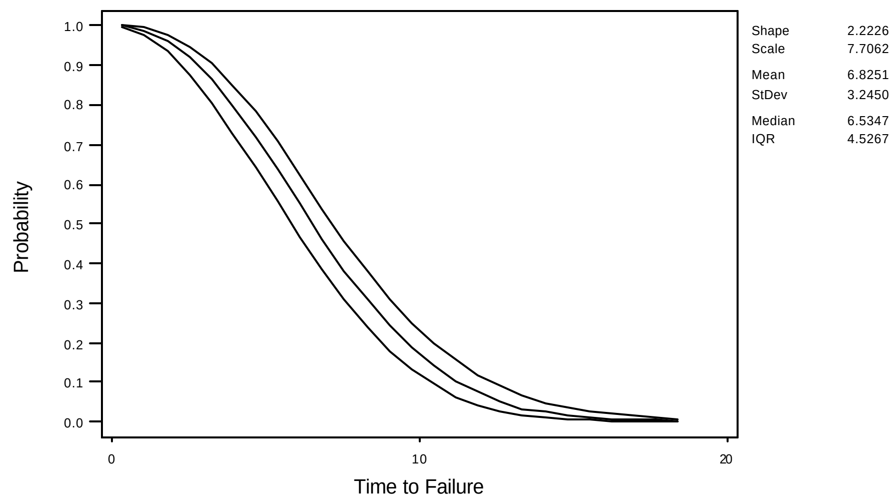



**Figure 9** . Survival Plot of MI Patients using Weibull Distribution

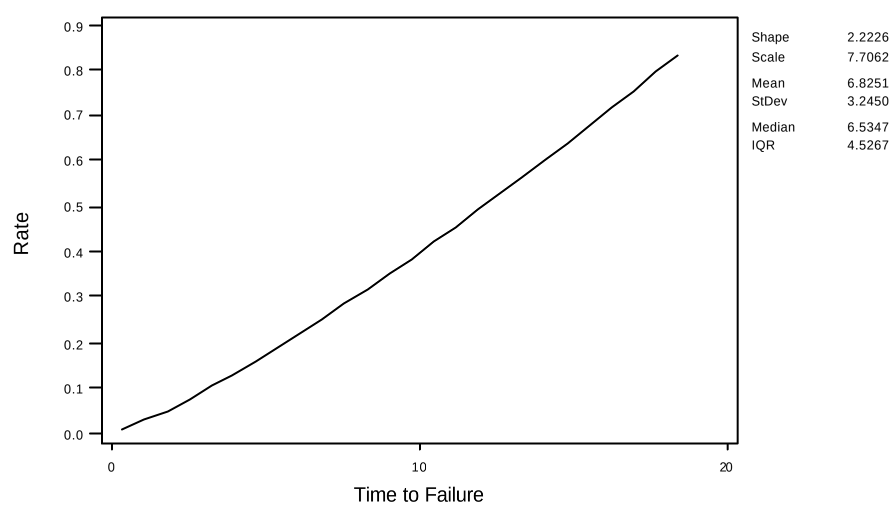



**Figure 10**. Hazard Plot of MI Patients using Weibull Distribution

**Comparison of Survival Distributions of the Different Social and Risk Characteristics**

It is necessary to use rank order-based nonparametric tests because survival times follow non-normal distributions. Logrank and Wilcoxon tests were used in this study to compare the survival curves of the different classifications of the attributes within the social and risk characteristics.

**Table 6**. Logrank and Wilcoxon Tests Results for Social Characteristics of MI Patients

| Variable | Method | test statistic | df | p-value |
|---|---|---|---|---|
| Age Group | Logrank | 0.02334 | 1 | 0.8786 |
| | Wilcoxon | 0.05077 | 1 | 0.8217 |
| Sex | Logrank | 2.4390 | 1 | 0.1184 |
| | Wilcoxon | 2.8403 | 1 | 0.0919 |
| Civil Status | Logrank | 24.6179 | 2 | 0.000 |
| | Wilcoxon | 23.7788 | 2 | 0.000 |

The distribution of survival times between age groups were the same, as explained by the p-value greater than 0.05. Thus, there is no enough evidence to state that the survival curve of MI patients ageing 60 and below differs from that of the patients above 60. As seen in Figure 11, the survival plot of 60 years old and below and the above 60 years of age seems to overlap, which means that there is homogeneity of their distributions. Maybe this is due to similar body conditions of the two groups having this ailment. Also, this is explained by the fact that heart attack occurs at any age and is not affected whether the person is neither still working nor already retired. Figure 12 shows that there is some what heterogeneity of the survival distributions of sexes but do not differ significantly as the Logrank and Wilcoxon tests show. Thus, though males are more prone to MI disease than females the survival of the two sex classifications do not differ significantly. The survival curves of single, married and widow MI patients differ significantly. This result may be due to the dominancy of the observations of married MI patients in the sample data. The lack of the observations of single MI patients is reflected in Figure 13.

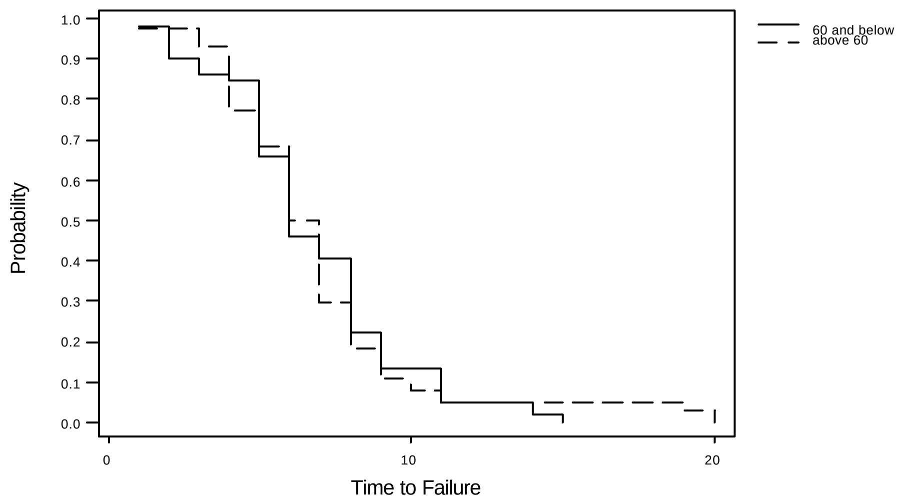



**Figure 11**. Survival Curves for Age Groups of MI Patients

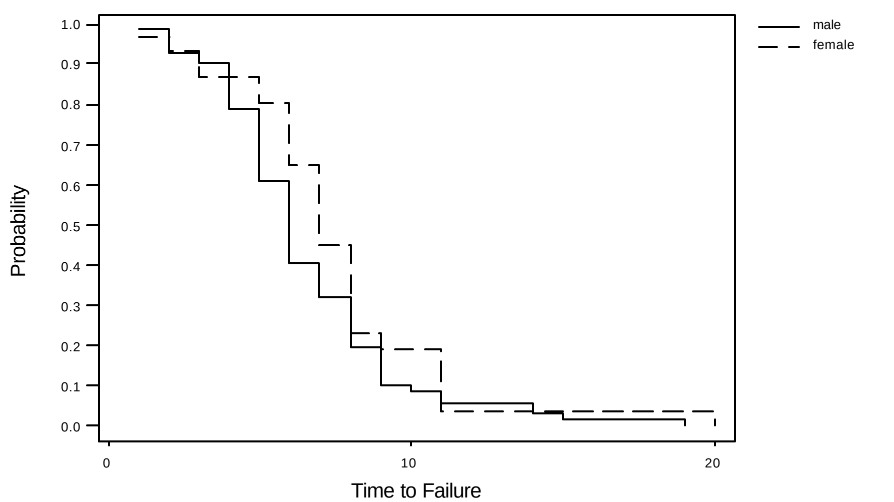



**Figure 12.** Survival Curves for Sexes of MI Patients

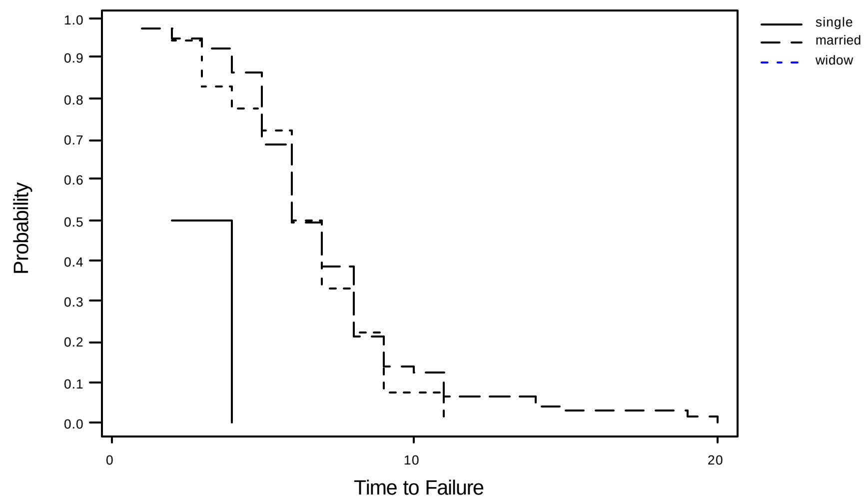



**Figure 13.** Survival Curves for Civil Statuses of MI Patients

It is clearly shown in Table 7 that there are no significant results of Logrank and Wilcoxon tests for homogeneity of survival curves between the MI patients with risks from the patients without risks. Thus, there is no enough evidence to state that there existed significant differences between the distributions of survival times of patients with risks and patients without risks.

**Table 7.** Logrank and Wilcoxon Tests Results for Risk Characteristics of MI Patients

| Variable | Method | test statistic | df | p-value |
|---|---|---|---|---|
| Hyperlipidemia | Logrank | 0.03399 | 1 | 0.8537 |
| | Wilcoxon | 0.7760 | 1 | 0.3784 |
| Hypertension | Logrank | 0.7079 | 1 | 0.4001 |
| | Wilcoxon | 2.8327 | 1 | 0.0924 |
| Diabetes | Logrank | 0.2654 | 1 | 0.6064 |
| | Wilcoxon | 0.4264 | 1 | 0.5137 |
| Smoking | Logrank | 0.3257 | 1 | 0.5682 |
| | Wilcoxon | 0.004616 | 1 | 0.9458 |
| Family History | Logrank | 2.6501 | 1 | 0.1035 |
| | Wilcoxon | 1.4297 | 1 | 0.2318 |

## Identification of Prognostic Factors Related to Survival Time

The method used in regression model building was backward elimination wherein a regression on all 8 variables was done first and then variables are deleted from the regression equation one by one until the regression is satisfactory. The elimination of variables was based on their ranking, i.e., the variable with the lowest significance level or with highest p-value using a specialized Z statistic was eliminated first from the model, the next lowest was the second eliminated, and so on.

Among the models presented, only the model with a predictor family history had a significant coefficient and log-likelihood. Thus, the model with family history as a predictor best describes the distribution of survival times of MI patients. Hence, the model that describes the relationship between family history and survival time of MI patients is given by:

$$\log_e(t) = 2.10185 - 0.22034(FamilyHistory) + 0.43397\boldsymbol{e}_p$$

where t is the survival time and $e_p$ is the $p^{th}$ percentile of error distribution.

## IV. Summary and Conclusion

### Summary

Summary statistics such as mean, standard deviation, minimum and maximum values, and range of the variables were taken. The average survival time is 6.806 days with a standard deviation of 3.190 days and the average age is 58.62 years old with a standard deviation of 12.78 years. The social and risk characteristics were put into chi-square tests to determine if the proportion of the number of MI patients was the same in the different independent variables. Under social characteristics, sex and civil status were found to be significantly unequal in proportion, while the age was found to be insignificant. Thus, the proportion of MI patients under the working group (60 years old and below) and the age of the MI patients under the retirement group (above 60 years old) is equal. Similarly, hyperlipidemia, diabetes and family history contribute to the risk of having MI disease as supported by the chi-square statistics. Remarkably, hypertension was not found by chi-square to contribute to the risk of having MI disease. This result may be explained by possible interaction of hypertension to the other risk characteristics like hyperlipidemia. In other words, hypertension may contribute to the risk of having MI disease when combined to the other risk factors.

Nonparametric survival distribution fitting was then performed. Life Table method and Kaplan-Meier Product Limit Estimation were utilized to compute and plot the survival and hazard probabilities for each interval. As seen in the life table results, the median survival time of MI patients is 7.0937 days. The third interval (6-8 days) has the largest survival time downfall followed by the second interval (4-6 days). The hazard plot shows an increasing trend and reaches its peak during $9^{th}$ to $11^{th}$ days, then fluctuates. The estimated median survival time from Kaplan-Meier estimation results is 6 days. This is smaller as compared to the median time of life table since the latter is based in ranges or intervals while the former is based in points. Its hazard plot shows that there is a constant increasing trend in the hazard plot as time increases.

Probability plotting and chi-square goodness-of-fit test were used to compare the three theoretical distributions. Among those three, the Weibull distribution is the most appropriate. The survival function of MI patients using Weibull distribution gave the median survival time of 6.5347 days.

Logrank and Wilcoxon tests were used in this study to compare the survival curves of the different classifications of the attributes within the social and risk characteristics. The tests revealed that there were no differences between the survival curves of MI patients in terms of their ages, civil statuses, and risk factors such as hyperlipidemia, hypertension, diabetes, smoking and family history.

Cox's proportional hazards model was used to determine the prognostic factors related to survival time. Backward elimination method was done. The model with a predictor family history had a significant coefficient and log-likelihood. Thus, the model with family history as a predictor best describes the distribution of survival times of MI patients.

## Conclusion

Chi-square test showed that sex and civil status were significant factors that affected the ability of a person to catch the disease. Males were more likely to catch the disease than females. Married people were more prone than singles or widows. Moreover, hyperlipidemia, diabetes and family history were found to contribute to the risk of having MI diseases. Hyperlipidemic people were more likely to catch the disease. Diabetic people were more prone than non-diabetic people. People with family history of atherosclerotic diseases were also found to catch the disease than people without this history.

Life table hazard plot showed that 48.98% of the patients survived during the $9^{th}$ day. Moreover, Kaplan-Meier showed that there is a constant increasing trend in the hazard plot as time increases.

Probability plotting showed that the distribution of survival times of MI patients follow the Weibull distribution.

Logrank and Wilcoxon tests showed that there were no significant differences between the survival curves of MI patients in terms of their social characteristics such as age, civil status and sex; and risk characteristics such as hyperlipidemia, hypertension, diabetes, smoking and family history.

The model that describes the relationship between family history and survival time of MI patients is given by:

$$\log_e(t) = 2.10185 - 0.22034(FamilyHistory) + 0.43397e_p$$

where t is the survival time and $e_p$ is the $p^{th}$ percentile of error distribution.