IJABER, Vol. 14, No. 5, (2016): 3093-3112

SURVIVAL MODELLING AND ITS MARGINAL EFFECT TO PATIENTS HAVING CARDIOVASCULAR DISEASES

Hero Laurenciano Tolosa[1]

Abstract: *This study seeks to identify the risk factor that can predict the survival probability of patients having cardiovascular diseases. The study utilized data consisting of 300 records of patients who have been diagnosed with Angina Pectoris, Atherosclerosis and Myocardial Infarction for the year 2009. The researcher used multivariate analysis specifically logistic regression as a method to identify the risk factor that can predict the survival probability of the said patients. Subsequently performing logistic regression, the researcher found that the patients, who were diagnosed having angina pectoris, atherosclerosis, myocardial infarction, considered diabetic person, tobacco smoking, systolic, hematocrit & LDL cholesterol as predictors that have negative impact. On the other hand, the variable HDL cholesterol considered to have a positive impact in estimating the survival of the patients. The best model that was concluded by the researcher is model 4 which can present by*

$$probability\,(survive) = \frac{e^{(5.2848-0.0255*systolic-0.0311*hematocrit-0.5886*ldl+1.6469*hdl)}}{1+e^{(5.2848-0.0255*systolic-0.0311*hematocrit-0.5886*ldl+1.6469*hdl)}}.$$

Keywords: *Cardiovascular Diseases, Logistic Regression, Marginal Effect, LDL and HDL cholesterol, Angina Pectoris, Atherosclerosis and Myocardial Infarction*

1. INTRODUCTION

Heart is part of cardiovascular system that includes network of arteries, veins, and capillaries that transport blood throughout the body. It is also responsible for the flow of blood, nutrients, hormones, oxygen and other gases to and from cells (National Institutes of Health, 2015). Abnormality of circulatory system may cause the body aberrant conditions that is detrimental to disease fighting to maintain a stable internal environment such as proper temperature and pH known as homeostasis.

According to U.S National Library of Medicine (2015), the circulatory system or cardiovascular system is defined as simply a highway for blood and made up three (3) independent system that work together: the heart (cardiovascular), the lungs (pulmonary), and the arteries, veins, coronary and portal vessels (systemic). Based on the report of Arkansas Heart Hospital (2015), the average human capacity is about 2,000 gallons (7,572 liters) of blood travel daily through 60,000 miles (96,650 kilometers) of blood vessels. On the average adult, there were 5 to 6 quarts (4.7 to

[1] Polytechnic University of the Philippines, College of Science – Department of Mathematics and Statistics

5.6 liters) of blood which is made of red blood cells, white blood cells, plasma and platelets.

The heart, blood, and blood vessels make up the cardiovascular component of the circulatory system and includes pulmonary circulation which means a “loop” through the lungs where blood is oxygenated. Systemic circulation also incorporate in circulatory system which runs through the rest of the body to provide oxygenated blood (NLM, 2015). Arteriosclerosis is one of the most common disease of the circulatory system, in which the fatty deposits in the arteries causes the walls to stiffen and thicken the walls. Mayo Clinic (2015) reported that one the reasons of having this disease are the build-up of fat, cholesterol and other material in the artery walls. This restricts blood flow or, in severe cases, stop it all together which results in a heart attack or stroke.

According to Mitchell Weinberg of the North Shore-LIJ Health System (2015), stroke is another major condition of the circulatory system. It involves blockage of the blood vessels to the brain. Some of the risk factors of having stroke are smoking, diabetes and high cholesterol. Another is the occurrence of aortic aneurysm due to the damage, bulge or eventually tear of aorta, which can cause severe internal bleeding. This condition can be present at birth or a result of atherosclerosis, obesity, high blood pressure or a combination of these conditions.

Another circulatory disease is hypertension or commonly called high blood pressure that causes the heart to work harder or shortness of breath which leads to complications such as heart attack, stroke, or kidney failure. According to Radhakrishnan (2015) reported that peripheral arterial diseases (also known as PAD) typically involve areas of narrowing or blockage within an artery. Chronic venous insufficiency (also known as CVI) which involves areas reflux (or backward flow) within the superficial veins of the lower extremities is also a common cardiovascular disease. PAD is diagnosed through a non-invasive testing including ultrasound, CT scan, and/or MRI. Ultrasound is the least expensive of these methods but it gives the least amount of detail as CT and MRI show a much higher degree of anatomic detail when identifying areas of narrowing/blockage within an artery. CVI is diagnosed with ultrasound as the venous reflux can be measured accurately by ultrasound, which ultimately guides treatment.

According to World Health Organization (WHO) 2015, some leading causes of death are ischemic heart diseases and stroke with global mortality of 7.4 and 6.7 million respectively. Based on the study conducted by Department of Health in the Philippines in 2009, the disease of the heart and the vascular system have remained the top two causes of death in the county since 1993. According to DOH, it increased by 100% from 85,000 to 170,000 in the past 20 years. It was suggested that in order to diminish the chances of having heart disease and disorder’s risk factor level, proper diet, exercise and healthy lifestyle are the best regimen that are needed to be applied in our body.

Majority of the cardiovascular diseases may be explained by risk factors. These risk factors can be classified as uncontrollable, controllable and modifiable which are associated in survival of the individual. Some of the uncontrolled risk factors may be classified under demographic profile such as sex, age, and genetic makeup. For the controlled risk factors, it includes the medical treatment, medicine in-take while modifiable factors are lifestyle (tobacco use, alcoholic drinking, prohibited drugs, physical inactivity and obesity) and other biochemistry parameters (blood pressure, lipids and blood glucose), urinalysis (uric acid, blood urea nitrogen and creatinine), electrolytes (sodium and potassium) and rheological variables (hemoglobin, hematocrit, red blood cell and white blood cell count).

The objective of the researcher was to determine the predictors that can explain the probability of survival of patients with cardiovascular diseases and the effect of identified predictors using logistic regression analysis. Moreover, the researcher estimated the survival of the said patients who were diagnosed having angina pectoris, atherosclerosis, myocardial infarction, history of disease, history of surgery, person with diabetes, tobacco smoker, and prohibited drug taker using Kaplan-Meier method.

2. STATEMENT OF THE PROBLEM

Using multivariate analysis specifically logistic regression, this study mainly focuses on creating a model that can identify the survival probability of patients with cardiovascular diseases given the risk factors. Specifically, this research seeks to answer the following questions:

1. What is the risk profile of patients having cardiovascular diseases based on the variables of angina, atherosclerosis, myocardial, history of disease, history of surgery, diabetes, smoking, alcohol drinking, and prohibited drug intake?
2. What is the behavior of the survival probability of the patients based on the variables of angina, atherosclerosis, myocardial, history of disease, history of surgery, diabetes, smoking, alcoholic drinker and taking prohibited drugs?
3. What is the predictors that can explain the survival probability of the patients having cardiovascular diseases using multivariate analysis?
4. Which model is the best in explaining the survival probability of the patients having cardiovascular diseases?

3. METHOD OF RESEARCH USED

The researcher utilized the quantitative method as a means of analysis. Quantitative method are the straightforward sequence for quantitative research flows: conceptualization, operationalization, application of the operational definition or measurement to collect the data (Neuman 2007). He also added that quantitative researchers established numerous ways thoroughly link abstract ideas to measurement procedures that will produce precise quantitative information about empirical reality. Additionally, quantitative data collection methods were used on the quantification of relationships between predictors and response variables. To prevent bias result and presentation, the researcher used quantitative approach. Descriptive and Inferential Statistics are both part of quantitative methods. Descriptive statistics is used to describe the basic features of the data in this study and provides summary about the sample; while in inferential statistics, the researcher tried to reach conclusions that extend beyond the immediate data alone (Trochim, 2006; Putri, E. I. Magdalena, R., Novamizanti, 2015).

3.1. The Data

The researcher employed the available data from the specific hospital in Metro Manila, Philippines. The said data consists of 300 cardiovascular diseases records of patients who have been diagnosed with Angina Pectoris, Atherosclerosis and Myocardial Infarction for the year 2009. The said records involved the patients' initial profile review and laboratory results in urinalysis, electrolyte and blood rheology test that consist of different factors affecting the cardiovascular diseases.

Table 1
Discrete and Continuous Variables
RISK FACTORS IN THE STUDY DATA SET (N=300)

Variable Name	*Description*	*Unit*
CATEGORICAL VARIABLES		
Angina	Angina Pectoris	1 = Prevalent 0= Free
Hypertension	Hypertension	1 = Prevalent 0= Free
Atherosclerosis	Atherosclerosis	1 = Prevalent 0= Free
Myocardial	Myocardial Infarction	1 = Prevalent 0= Free
Sex	Sex	1 = Male 0 = Female
Historyofdisease	History of CVD	1 = Yes 0 = No
Historyofsurgery	History of Surgery	1 = Yes 0 = No
Diabetes	Diabetes	1 = Diabetic 0 = Non-Diabetic
Smoking	Tobacco Smoking	1 = Smoker 0 = Non-Smoker
Alcholicdrinking	Alcohol Drinker	1 = Drinker 0 = Non-Drinker
Prohibiteddrugs	Prohibited drug intake	1 = Yes 0 = No

contd. table 1

Variable Name	*Description*	*Unit*
	CONTINUOUS VARIABLES	
Age	Initial age at diagnosis	
Bmi	Body Mass Index	kg/m²
Systolic	Blood Pressure Systolic	mmHg
Diastolic	Blood Pressure Diastolic	mmHg
Glucose	Blood Glucose	mmol/L
Cholesterol	Blood Cholesterol	mmol/L
Hdl	HDL Cholesterol	mmol/L
Ldl	LDL Cholesterol	mmol/L
triglycerides	Triglycerides	mmol/L
Uricacid	Uric Acid	umol/L
Vldl	VLDL	mmol/L
Bloodureanitrogen	Blood Urea Nitrogen	mmol/L
Creatinine	Creatinine	umol/L
Sodium	Blood Sodium	mEq/L
Potassium	Blood Potassium	mmol/L
Hemoglobin	Hemoglobin	g/dL
Hematocrit	Hematocrit	g/dL
plateletcount	Platelet Count	X10^9/L
Wbccount	While Blood Cell Count	X10^9/L
nooffollowups	No. of Check-ups	In days
Noofconfinementdays	No. of Confinement Days	In days

Statistical Treatment

The researcher used the descriptive statistics to describe the basic features of the data in a study. This includes the mode, and percentage. Moreover, the researcher also used inferential statistics to conclude from the sample data. Logistic regression was used to create, develop a model that could predict the survival probability of patients with cardiovascular diseases and diagnosed with Angina Pectoris, Atherosclerosis and Myocardial Infarction for the year 2009.

Based on the treatment mentioned above, the patient was identified alive if probability $(y=1) \geq 0.5$, otherwise patient was identified as dead. Another statistical treatment that may use for this study is cox proportional hazard function. This process need the complete data of patients involving the time or date of confinement which is did not all captured by the researcher.

3.2. Kaplan-Meier Method

Based from Hu 2013, this method is also called one-sample nonparametric method. Its survival function can be estimated with Kaplan-Meier plot, Life-table, or Cumulative hazard estimator. This method does not require any mathematical assumption for its hazard function or proportional hazard; it simply take into account with the empirical probability of surviving over certain time. But it assumes the value of survival function between successive distinct observations as constant. It is widely applied for studying survival function of population that is commonly used to compare the survival functions among groups. Its estimated survival function represents that of population, if the sample size is large enough.

It is mainly descriptive; this method does not take into account of covariates. The survival function usually shows as stepwise reduction plot. Therefore, it deals with categorical predictors or grouped continuous variables. It cannot take care of continuous variables directly, so it will not be able to accommodate time-dependent variables. It is still very popular even with this limitations because it treats censored data well, particularly right-censoring data.

The function called Product-Limit estimator that is proposed by Kaplan and Meier, which offers an efficient way of estimation its survival function for all the values of time (t) in the range with right censoring observations. It means before the time (t_1), there is no event occurred, so the survival function is 1. Once the event happens, its survival function will decrease. The time t1 is kind of checkpoint or cutting point for event occurring. Where, d_i is the number of event occurred in that time interval or at the point, and y_i is the number of subject at risk in that interval or at the time point. It includes all subjects who have died, dropped out, or not reached the event time yet at that time point or interval. They are applied for all the following equations.

$$\hat{S}(t) = \begin{cases} 1, & t < t_1 \\ \prod_{t_i \leq t} \left[1 - \frac{d_i}{y_i}\right], & t \geq t_1 \end{cases}$$

Its variance estimator follows the below formula,

$$\hat{V}\left(\hat{S}(t)\right) = \hat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{y_i(d_i - y_i)}$$

Since the survival function can be estimated, its cumulative hazard function can be calculated following formula: $\hat{H}(t) = -\ln(\hat{S}(t))$

3.3. Logistic Regression

Logistic regression is a specialized form of regression formulated to predict and explain a binary categorical variable or an ordinal variable rather than a metric dependent measure. It is also known as the logit analysis. The model yields regression-like coefficients that indicate the relative impact of each predictor variable. Logistic regression is useful for situations in which you want to be able to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous. Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model. Logistic regression is applicable to a broader range of research situations than discriminant analysis.

The logistic regression model

The relationship between the response variable and the predictors is given by the equation:

$$\text{Logit}(Y) = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{3+\cdots+} \beta_k X_k$$

where $Y = 1$ *for alive and* $Y = 0$ *for dead*

$\beta_o, \beta_1, \dots, \beta_k$: regression coefficient parameters
$X_1, X_2, \dots, X_k$: predictors

Then, the logistic regression model equation is transformed as:

$$P(Y = 1) = \frac{\exp[Logit(Y)]}{1 + exp[Logit(Y)]}$$

where can be used to predict the probability of being classified as alive.

Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable. In this way, logistic regression estimates the probability of a certain event to occur. Thus, it calculates changes in the log odd of the dependent, not changes in the dependent itself (Rencher, 1985; Rais, Suhadi, Maidin & Palutturi, 2015).

Johnson (2005) states that the difference between logistic regression and discriminant analysis is that logistic regression does not make any assumptions regarding the distribution of the independent variables. Therefore, it is preferred than discriminant analysis when the independent variables are a combination of categorical and continuous because in such cases the multivariate normality assumption is clearly violated. On the other hand, when the multivariate normality is not violated, then discriminant analysis is preferred because is computationally more efficient than logistic regression analysis.

3.4. Classification Functions

Classification of observations is done by first estimating the probabilities. These probabilities can be used to classify observations into two groups. Classification of observations in two groups will be based on a cut off value for predicted probabilities, which is usually assumed to be 0.5. All the observations whose predicted probabilities are greater than or equal to 0.5 will be classified as "alive," and those whose values are less than 0.5 will be classified as "dead".

3.5. Receiver operating characteristic (ROC) analysis

ROC analysis quantifies the accuracy of diagnostic tests or other evaluation modalities used to discriminate between two states or conditions, which are here referred to as normal and abnormal or control and case. The discriminatory accuracy of a diagnostic test is measured by its ability to correctly classify known normal and abnormal subjects.

The analysis uses the ROC curve, a graph of the sensitivity versus 1 - specificity of the diagnostic test. The sensitivity is the fraction of positive cases that are correctly classified by the diagnostic test, whereas the specificity is the fraction of negative cases that are correctly classified. Thus the sensitivity is the true-positive rate, and the specificity is the true-negative rate.

ROC analysis can be interpreted as a two-stage process. First, the control distribution of the classifier is estimated, assuming a normal model or using a distribution-free estimation technique. The classifier is standardized using the control distribution to 1 - percentile value, the false-positive rate. Second, the ROC curve is estimated as the case distribution of the standardized classifier values.

Covariates may affect both stages of ROC analysis. The first stage may be affected, yielding a covariate-adjusted ROC curve. The second stage may also be affected, producing multiple covariate specific ROC curves.

The global performance of a diagnostic test is commonly summarized by the area under the ROC curve (AUC). This area can be interpreted as the probability that the result of a diagnostic test of a randomly selected abnormal subject will be greater than the result of the same diagnostic test from a randomly selected normal subject. The greater the AUC, the better the global performance of the diagnostic test. Each of the ROC commands provides computation of the AUC.

Citing a lack of clinical relevance for the AUC, other ROC summary measures have been suggested. These include the partial area under the ROC curve for a given false-positive rate t [pAUC(t)]. This is the area under the ROC curve from the false-positive rate of 0 to t. The ROC value at a particular false-positive rate and the false-positive rate for a particular ROC value are also useful summary measures for the ROC curve (Pepe, 2003).

4. RESULTS AND DISCUSSION

This chapter presents the results of summary of data and statistical analysis. The discussion covers the following: 1) to determine the risk profile of patients having cardiovascular diseases and/or diagnosed with Angina Pectoris, Hypertension, Myocardial Infarction and Atherosclerosis. 2) to identify the behaviour of survival probability of the patients Angina Pectoris, Hypertension, Myocardial Infarction and Atherosclerosis using Kaplan-Meier method. 3) to determine the predictors that can explain the survival probability of the patients having cardiovascular diseases using multivariate analysis. 4) to identify which method is best in explaining the survival probability of the patients having cardiovascular diseases.

Risk Profile of Patients having Cardiovascular Diseases

The table 2, 3 and 4 show the frequency and percent of the risk profile from 300 patients with cardiovascular disease. With this, most of the patients diagnosed with myocardial with 119 or 39.67%. Most of this patients were male (158 or 52.67%), their age was 18-64 years old (268 or 89.33%), with history of disease (192 or 64.00%), no history of surgery (228 or 76.00%), with diabetes (208 or 69.33%), tobacco smoker (215 or 71.67%), alcoholic drinker (223 or 74.33%), and not taking prohibited drugs (254 or 84.67%).

Survival functions of patients with CVD using Kaplan-Meier Method

The figures below show the behavior of the survival probability of the patients with cardiovascular disease with respect to time of their confinement. Figure 1, 5, 6 and 7 show that the patients who were diagnosed to have angina pectoris, history of surgery, person with diabetes and tobacco smoker, respectively, got a lower survival probability as time increases. On the other hand, figure 2 shows some complement behaviors of the survival probability from the start. The behavior of the survival probability changed after 1500 days of confinement and it decreased

Table 2
Frequency and Percent Distribution by Angina Pectoris.

Factors	*Classifications*	*Agina Pectoris*			
		Dead		*Alive*	
		F	*P*	*F*	*P*
Sex	Female	12	4.0	18	6.0
	Male	21	7.0	29	9.7
Age Group	below 18 years old	0	0.0	0	0.0
	18-64 year old	32	10.7	44	14.7
	65 years old and above	1	0.3	3	1.0
History of Disease	No	16	5.3	24	8.0
	Yes	17	5.7	23	7.7
History of Surgery	No	22	7.3	36	12.0
	Yes	11	3.7	11	3.7
Diabetic Person	No	3	1.0	15	5.0
	Yes	30	10.0	32	10.7
TobaccoSmoker	No	2	0.7	5	1.7
	Yes	31	10.3	42	14.0
Prohibited Drugs Takers	No	26	8.7	37	12.3
	Yes	7	2.3	10	3.3

Table 3
Frequency and Percent Distribution by Atherosclerosis

Factors	*Classifications*	*Atherosclerosis*			
		Dead		*Alive*	
		F	*P*	*F*	*P*
Sex	Female	11	3.7	18	6.0
	Male	3	1.0	19	6.3
Age Group	Below 18 years old	0	0.0	0	0.0
	18-64 years old	13	4.3	37	12.3
	65 years old and above	1	0.3	0	0.0
History of Disease	No	4	1.3	11	3.7
	Yes	10	3.3	26	8.7
History of Surgery	No	9	3.0	26	8.7
	Yes	5	1.7	11	3.7
Diabetic Person	No	5	1.7	11	3.7
	Yes	9	3.0	26	8.7
TobaccoSmoker	No	7	2.3	15	5.0
	Yes	7	2.3	22	7.3
Prohibited Drugs Takers	No	14	4.7	32	10.7
	Yes	0	0.0	5	1.7

Table 4
Frequency and Percent Distribution by Myocardial Infarction

Factors	*Classifications*	*Myocardial Infarction*			
		Dead		*Alive*	
		F	*P*	*F*	*P*
Sex	Female	13	4.3	39	13.0
	Male	13	4.3	54	18.0
Age Group	below 18 years old	0	0.0	1	0.3
	18-64 year old	16	5.3	77	25.7
	65 years old and above	10	3.3	15	5.0
History of Disease	No	1	0.3	2	0.7
	Yes	25	8.3	91	30.3
History of Surgery	No	17	5.7	68	22.7
	Yes	9	3.0	25	8.3
Diabetic Person	No	6	2.0	23	7.7
	Yes	20	6.7	70	23.3
TobaccoSmoker	No	5	1.7	23	7.7
	Yes	21	7.0	70	23.3
Prohibited Drugs Takers	No	22	7.3	73	24.3
	Yes	4	1.3	20	6.7

continuously. In Figure 3 and 8 show different survival probability compare to others. Both of this have a common pattern which overlap after 1500 days of confinements. In addition, figure 8 also shows almost the same survival probability between those patients who took and did not take prohibited drugs.

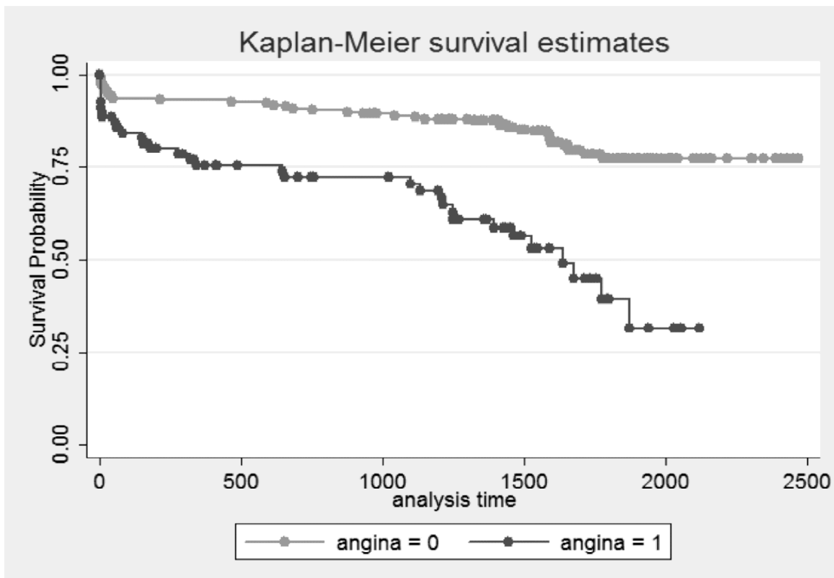

Figure 1: Survival Probabilities of Observed Patients with Cardiovascular Disease who were diagnosed having Angina Pectoris

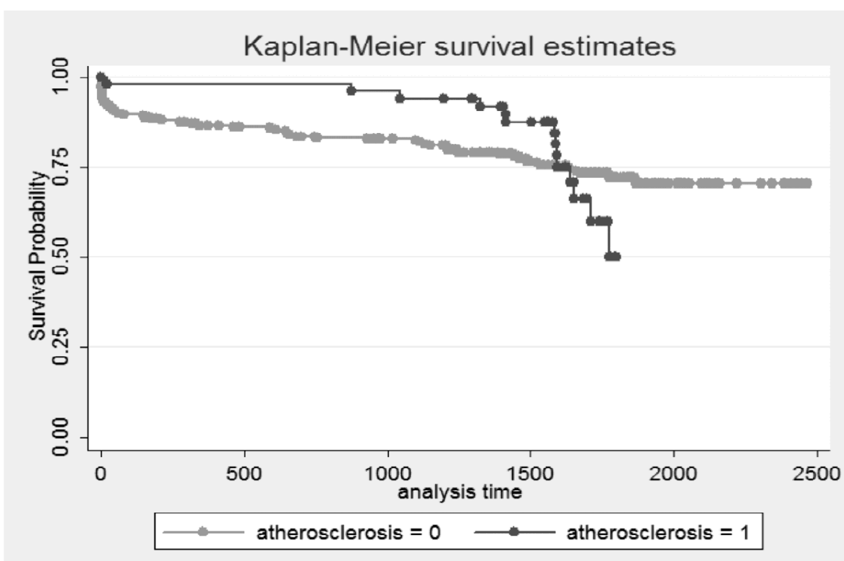

Figure 2: Survival Probabilities of Observed Patients with Cardiovascular Disease who were diagnosed having Atherosclerosis

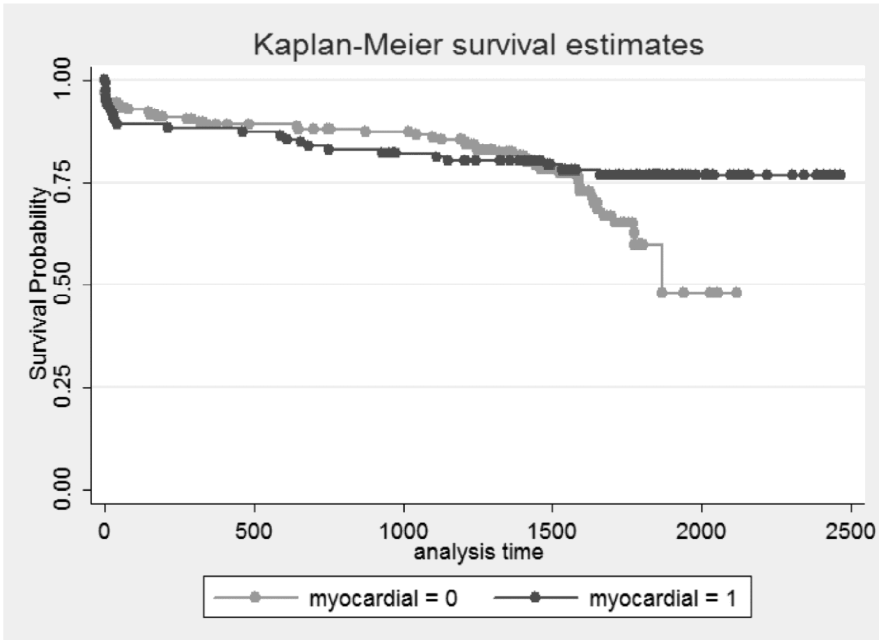

Figure 3: Survival Probabilities of Observed Patients with Cardiovascular Disease who were diagnosed having Myocardial Infarction

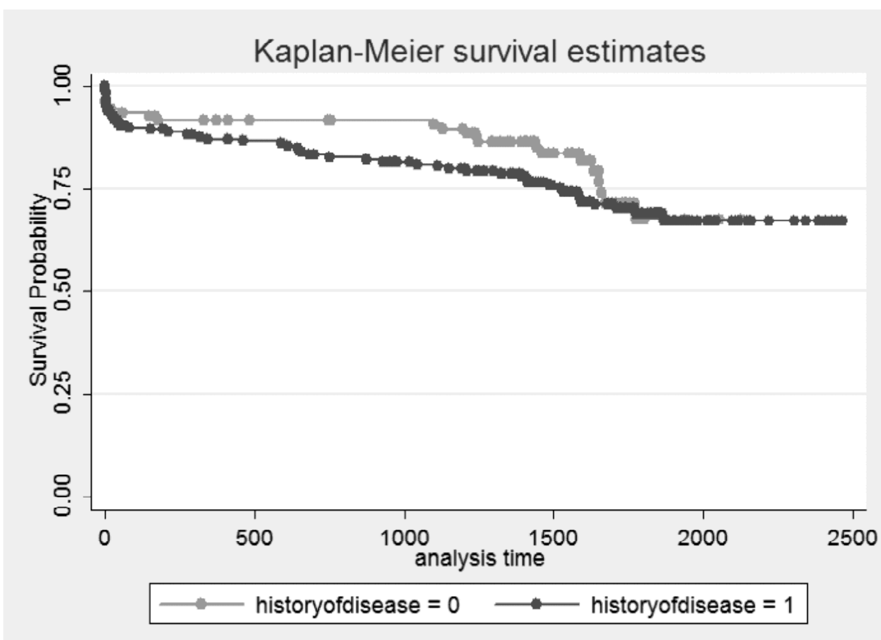

Figure 4: Survival Probabilities of Observed Patients with Cardiovascular Disease who have History of Disease

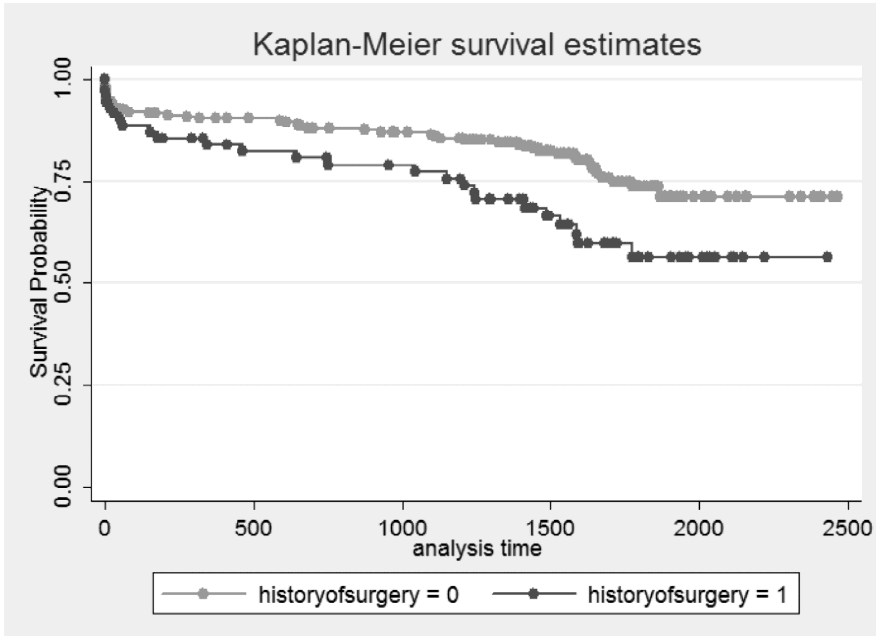

Figure 5: Survival Probabilities of Observed Patients with Cardiovascular Disease who have History of Surgery

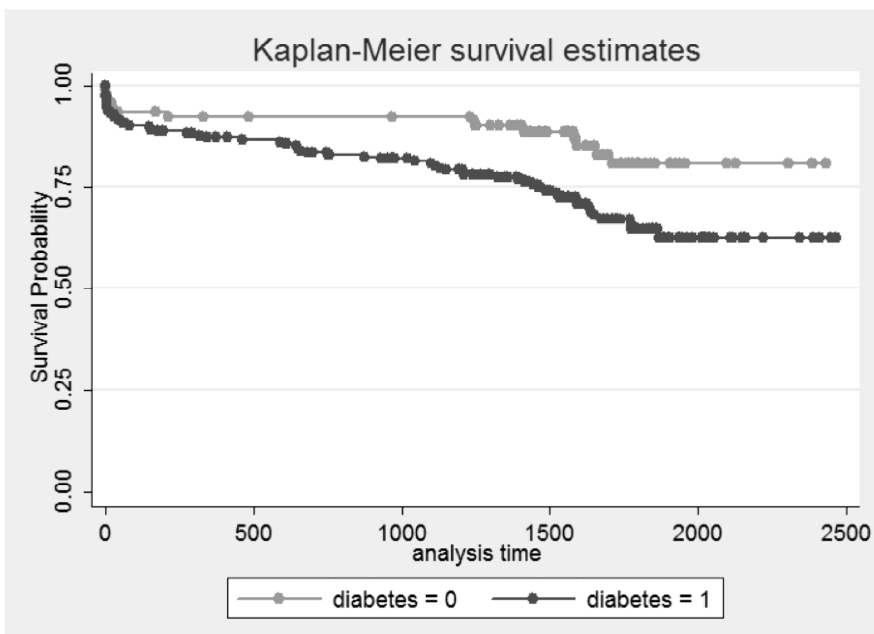

Figure 6: Survival Probabilities of Observed Patients with Cardiovascular Disease who were diagnosed having Diabetes

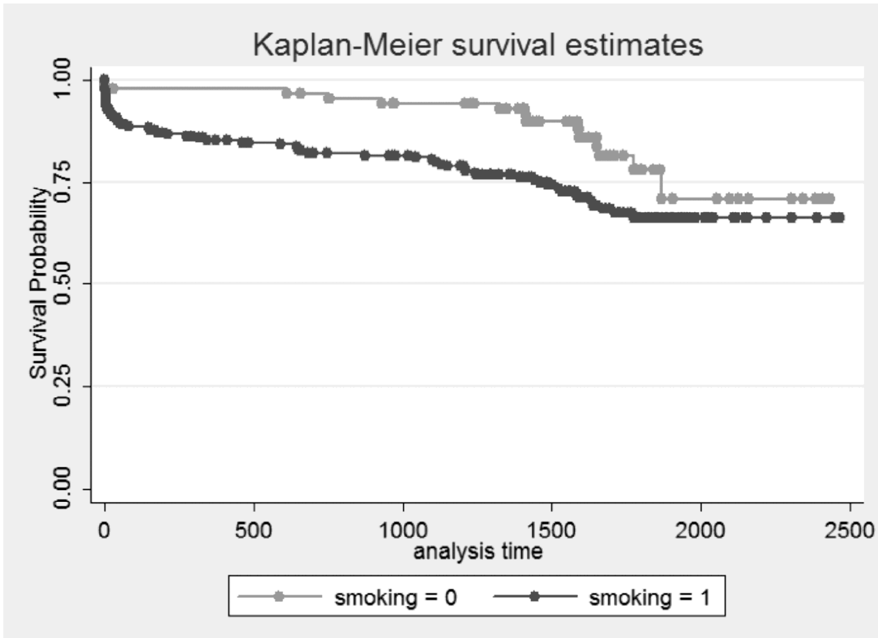

Figure 7: Survival Probabilities of Observed Patients with Cardiovascular Disease who Smoke Tobacco

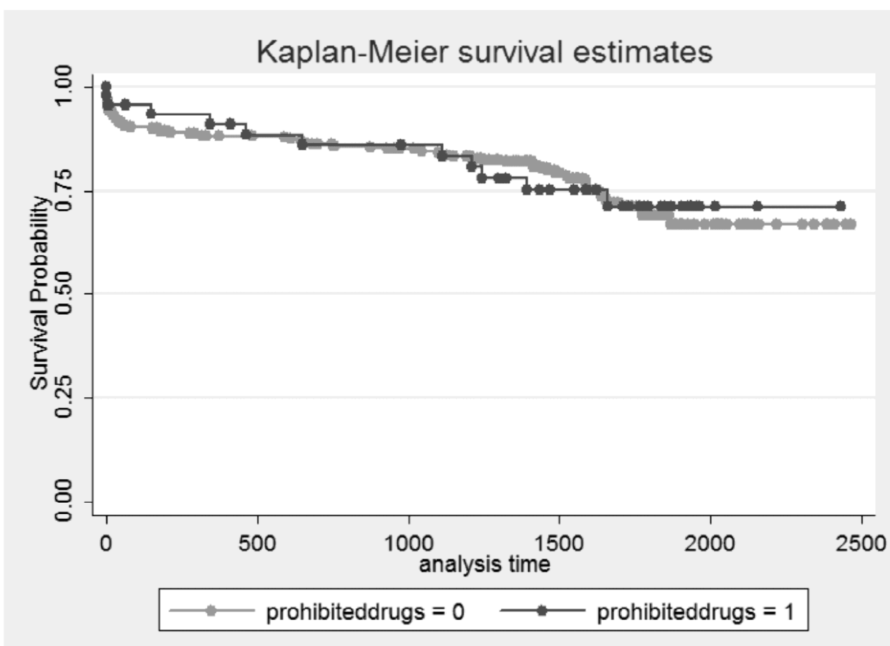

Figure 8: Survival Probabilities of Observed Patients with Cardiovascular Disease who took Prohibited Drugs

Predictors that can explain the survival probability of the patients having cardiovascular diseases using Logistic Regression

The following tables show the predictors that have significant relationship to explain the survival probability of the patient with cardiovascular disease. The Model 1 shows that the patient with CVD and diagnosed of having diabetes and identified as tobacco smokers can affect the probability of the patients survival. Additionally, patients have diabetes were 14.46% less likely to survive (in comparison with those that are not diabetic patients). Tobacco smokers were 11.91% less likely to survive than non-tobacco smokers. According to Davidson, et. al. (2012), diabetes is a metabolic disorder, which means a problem with the process by which food is digested and used as energy by the body. It is a chronic (long-term) condition characterized by high levels of glucose in the blood (hyperglycemia). If not treated, it can cause long-term complications such as coronary heart disease, kidney damage, stroke, circulatory problems and damaged vision. It can occurs as the result of medications, certain medical conditions, trauma or surgery, and some genetic disorders. Likewise, smoking is an unhealthy behavior that become an addiction. Smoking includes all forms of smoking, such as cigar smoking, cigarette smoking, pipe smoking, and exposure to second hand smoke. All forms of smoking are harmful and there is no form of safe or safer smoking. Since smoking constrict blood vessels and negatively affects circulation, smoking cessation is very beneficial for people who have other serious disease that affect blood vessels and circulation, such as diabetes, congestive heart failure, and other cardiovascular disease (McDonough, 2011). These definition support the result of the created model which is statistically significant and have negative impact to estimate the survival probability of the patients.

The Model 2 shows the result of logistic regression which consider the patients with CVD and have an angina pectoris and diabetes. This indicates that the patients having angina are 18.41% less likely to survive than patients do not have angina pectoris. Moreover, patients having diabetes are 12.53% less likely to survive than patients do not have a diabetes.

The Model 3 shows the logistic regression involving the patients having angina pectoris, myocardial infarction and atherosclerosis. Through this, it shows that the patients with CVD having angina pectoris are 59.96% are less likely to survive than patients who does not have. Angina pectoris is the medical term for chest pain or discomfort due to coronary heart disease. It is typically describe as squeezing, pressure, heaviness, tightness or pain in your chest. It occurs when the heart muscle does not get as much blood as it needs (Mayo Clinic 2016). Similarly, patients with CVD of having myocardial infarction are 44.36% to survive than patients who does not have a myocardial infarction. Acute myocardial infraction is the medical name for a heart attack. A heart attack is a life-threatening condition that occurs when blood flow to the heart is abruptly cut off, causing tissue damage.

This is usually the result of blockage in one or more of the coronary arteries. A blockage can develop due to build-up of plaque, a substance mostly made of fat, cholesterol, and cellular waste products (Macon, *et al.* 2015). Also, patients with CVD having atherosclerosis are 49.49% less likely to survive than patients who does not have atherosclerosis. Atherosclerosis is the medical name for hardening and narrowing of the arteries. This progressive silently and slowly blocks arteries, putting blood flow at risk. It is the usual cause of heart strokes, and peripheral vascular disease or known as cardiovascular disease (Heart Disease Health Center 2015).

The Model 4 shows the logistic regression involving the systolic, LDL cholesterol and HDL cholesterol from lipid and biochemistry profile, hematocrit from rheological profile. It indicates that for every unit change to the measurement systolic there are 0.39% less likely to survive assume all other things held constant. The desired systolic pressure is from 90 to 119 mmHg while diastolic pressure is from 60-79 mmHg. There is isolated systolic high blood pressure (isolated systolic hypertension, or ISH) is elevated (above 140 mm Hg), but diastolic blood pressure stays below 90 mm Hg. A systolic blood pressure over 140 mm Hg is an important risk factor for stroke and heart disease. ISH can cause damage to organs such as the kidneys, brain, heart, or eyes. And it should be treated (WebMD Medical Reference from Healthwise 2015).

It also shows that, for every unit change to the measurement of hematocrit there are 0.47% less likely to survive assume all other things held constant. Hematocrit (HCT) is the percentage of your blood that is made of red blood cells~ in other words, HCT reveals the volume of red blood cells in the blood. A complete blood count includes hemoglobin concentration, platelet count, and white blood cell count. The HCT percentage is about 45% to 50% in men and between 36% and 45% in women. Blood is mostly made up of plasma, which is made up of water, proteins, and blood glucose, among others, followed by red blood cells. Red blood cells are carriers of oxygen, and hematocrit levels represent the capacity of oxygen the blood can carry. If HCT levels are above the normal values, it means that the oxygen carrying capacity of the blood has increased which is good, but if it increases more than the needed amount, it can cause the blood to thicken~ this makes it difficult for the heart to pump blood, eventually leading to a heart attack (Home Remedies for You, 2010).

Furthermore, for every unit change to the measurement of LDL Cholesterol there are 8.99% less likely to survive assume all other things held constant and for every unit change to the measurement of HDL cholesterol there are 25.16% more likely to survive assume all other things held constant. According to the American Academy of Family Physicians Cholesterol (2015) is a waxy material that is produced naturally by the liver. It protects the nerves, produces hormones and makes cell tissues. However, too much cholesterol can be a bad thing which is why it's important to manage it and keep it at reasonable levels.

People can check their cholesterol levels by getting a simple blood test. The test measures total cholesterol, HDL (high density lipoprotein), LDL (low density lipoprotein) and triglycerides, another type of fat in the blood stream.

According to the American Heart Association (2015) HDL is the "good" cholesterol that keeps LDL, the "bad" cholesterol, down. Too much LDL cholesterol can cause deposits to build up in the blood vessels, known as plaque, which decreases the amount of blood and oxygen going to the heart. This in turn can lead to heart disease and heart attack (Geggel, 2015).

Table 4
Logistic Regression (Model 1)

	Coefficient	*Average Marginal Effect (dy/dx)*	*Computed p-value*
Diabetes	-0.8066	-0.1446	0.0150
Smoking	-0.6643	-0.1191	0.0460
Constant	2.2111		0.0000

Correctly Classified = 75.33 Area Under ROC Curve = 0.6153;
AIC = 330.0806; BIC=341.1919

Table 5
Logistic Regression (Model 2)

	Coefficient	*Average Marginal Effect (dy/dx)*	*Computed p-value*
Angina	-1.0650	-0.1841	0.0000
Diabetes	-0.7247	-0.1253	0.0300
Constant	1.9909		0.0000

Correctly Classified = 75.33 Area Under ROC Curve = 0.6479;
AIC = 320.9789; BIC=332.0903

Table 6
Logistic Regression (Model 3)

	Coefficient	*Average Marginal Effect (dy/dx)*	*Computed p-value*
Angina	-3.5382	-0.5996	0.0010
Myocardial	-2.6173	-0.4436	0.0110
Atherosclerosis	-2.9200	-0.4949	0.0060
Constant	3.8918		0.0000

Correctly Classified = 75.33 Area Under ROC Curve = 0.6839;
AIC = 311.1363; BIC= 325.9515

Table 7
Logistic Regression (Model 4)

	Coefficient	*Average Marginal Effect (dy/dx)*	*Computed p-value*
Systolic	-0.0255	-0.0039	0.0010
Hematocrit	-0.0311	-0.0047	0.0080
LDL Cholesterol	-0.5886	-0.0899	0.0190
HDL Cholesterol	1.6469	0.2516	0.0400
Constant	5.2848		0.0010

Correctly Classified = 78.33 Area Under ROC Curve = 0.7552;
AIC = 288.3942; BIC= 306.9131

CONCLUSIONS

The following conclusions are derived from the findings presented:

1. Most of the patients are diagnosed with myocardial infarction and are identified as male with age under 18-64 years old category. Most of the patients have: a history of disease, no history of surgery, with diabetes, smoker, alcoholic drinker, and does not taking prohibited drugs.
2. Using Kaplan-Meier Method it shows that patient diagnosed with angina has a better discrimination of the survival probability among the other factor. This means that if that patient diagnosed with angina, there is a low probability of survival than the patients who do not have it.
3. After performing logistic regression, the following variables are significant determinants of the survival probability of the patients with cardiovascular diseases. The following variables such as angina pectoris, myocardial infarction, atherosclerosis, diabetes, tobacco smoking, systolic, hematocrit and LDL cholesterol considered as predictors that has negative impact while the variable HDL cholesterol considered has a positive impact in determining the survival of the said patients.
4. Given the four models created, model three (3) and four (4) have a greater impact in explaining the survival probability of the patients with CVD because of lower RMSE, MAE and MAPE.

RECOMMENDATIONS

Based on the foregoing findings and derived conclusions of the study, the following recommendations are formulated:

1. The Department of Health and/or Word Health Organization may use these results as basis to improve and develop treatment to cure the following factors that may cause heart attack.

2. Results may be used for an advocacy program to improve health through healthy food choices, physical activity, managing stress/stress reduction, and stopping of smoking.
3. The researcher also suggests that the government should improve hospital facilities to be able to prevent and treat the chronic complications of diabetes. Also, to modify nutrient intake and lifestyle appropriate to the prevention and treatment of obesity, dyslipidaemia, cardiovascular disease, hypertension, and nephropathy.
4. Address individual nutritional needs taking into consideration the personal and cultural preferences and lifestyle while respecting the individual's wishes and willingness to change.
5. The researcher suggests to the future researchers to include the treatment conducted to the patients with cardiovascular diseases. This process will help to determine which treatment can help to increase the survival probability of patients.
6. Further research on the same topic in other locales should be conducted in order to verify, amplify, or negate the findings of the study.

References

Books

Cox, DR (1958). "The regression analysis of binary sequences (with discussion)". J Roy Stat Soc B 20: 215–242.

Draper, N.R.; Smith, H. (1998) Applied Regression Analysis, Wiley. ISBN 0-47117082-8 (Chapter14) "Interpreting the Coefficients on Dummy Variables".

Journals

American Heart Association (2015). Good vs Bad Cholesterol. Retrieved from *http://www.heart.org/HEARTORG/Conditions/Choles erol/AboutCholesterol/Good-vs-Bad Cholesterol_UCM_305561_Article.jsp*

Brindles Lee Macon, Winnie Yu, and Lauren Reed-Guy (2015). Acute Myocardial Infarction. Copyright © 2005 - 2016 Healthline Networks, Inc.

Davidson, A.M., Cumming, A.D., Swainson, C.P. & Turner, N. (1999) Diabetes Mellitus. In C.Haslett, E.R. Chilvers, J.A.A.

Geggel, Laura (2015). Cholesterol Levels: High, Low, Good & Bad. Retrieved from *http://livescience.com/*

Home Remedies for You (2010). Low Hematocrit Levels: Symptoms, Treatments & Diagnosis. Copyright © 2016 Home-Remedies-for-You.com. Retrieved from *http://www.home-remedies-for you.com/blog/category/common-ailments*

Mayo Clinic (2016). Angina Pectoris Definition. © 1998-2016 Mayo Foundation for Medical Education and Research.

McDonough, Brian MD (2011). Medical Reviewer, © 2015 Healthgrades Operating Company, Inc.

Neuman, W. Lawrence (2007). Basics of Social Science Research: Qualitative and Quantitative Approaches (2nd Edition). Pearson Education, Inc. U.S.A.

WebMD Medical Reference from Healthwise (2015). Hypertension/ High Blood Pressure Health Center. © 1995-2015 Healthwise, Incorporated. Retrieved from *http://www.webmd.com*

World Health Organization. (2011). Global Atlas on cardiovascular disease prevention and control, France.

World Health Organization. (2002). Reducing Risks, Promoting Healthy Life.

Pepe, M. S. 2003. The Statistical Evaluation of Medical Tests for Classification and Prediction. New York: Oxford University Press.

Internet Sources

Cardiovascular Diseases. Retrieved from *http://www.webmd.com/heart-disease/guide/diseases-cardiovascular*

Logistic Regression. Retrieve from *http://en.wikipedia.org/wiki/Logistic_regression*

Mayo Clinic, "Risk Factors" *http://www.mayoclinic.org/diseases-conditions/heart-disease/basics/risk-factors/con-20034056*

National Institutes of Health (2015). Retrieved from *http://www.nih.gov/*

Philippine Daily Inquirer, Retrieve from *http://www.philstar.com/breaking-news/2012/07/09/826043/5-out-10-filipinos-die-heart-disease-nso*

US-National Library of Medicine (2015). Retrieved from *www.nlm.nih.gov/medlineplus/ency/anatomyvideos/000023.htm*

World Health Organization, (2015). Retrieved from *http://www.who.int/mediacentre/factsheets/fs317/en/*

World Heart Federation (2011). "World Heart Federation: Cardiovascular disease risk factors" retrieved from *http://www.world-heart-federation.org/cardiovascular-health/cardiovascular-disease-risk-factors/*

Putri, E. I. Magdalena, R., Novamizanti, L. (2015). The detection of cervical cancer disease using an adaptive thresholding method through digital image processing. *Journal of Advances in Health and Medical Sciences, 1*(1), 30-36.

Rais, M. K., Suhadi, Z., Maidin, A., & Palutturi, S. (2015). Transformation study of implementation of national health insurance (NHI) at Mamuju Regency, Indonesia. *International Journal of Health and Medical Sciences,* vol. 1, no. 2, pp. 27-31, 2015.