Original Article

SURVIVAL ANALYSIS FOR THE RISK OF DEVELOPING HEART ATTACK

JACKIE D. URRUTIA[1], RAZZCELLE L. TAMPIS[1], JOSEPH MERCADO[2], AARON VITO M. BAYGAN[2]

Polytechnic University of the Philippines-Parañaque Campus Parañaque City, Philippines
Email: jackieurrutia20@gmail.com

ABSTRACT

Heart disease is the leading cause of death among Filipinos, accounting for 1 out of every 5 deaths in the past year. Each year, 170,000 Filipinos die from cardiovascular diseases, up from 85,000 more than 20 y ago. This paper aims to estimate the risk of developing Heart Attack using Survival Analysis. The data were obtained from the records of Quirino Memorial Medical Center with a total of 447 observations. The data includes the time t (patient's age), gender of the patient, covariates such as lifestyle (smoking and drinking alcohol); health conditions (Hypertension, Diabetes, High-density Lipoprotein (HDL) and Low-density Lipoprotein (LDL) level); and family history of Cardiovascular Disease, and the event of interest which is the occurrence of heart attack. The researchers applied two main statistical treatments in examining the data: (1) Cox Regression in formulating a model to estimate the risk of heart attack based on the given covariates; (2) Kaplan-Meier Estimates in calculating the probability of each patient to survive in accordance to their gender depending on the covariates the patients have. Results showed that females have more risks of developing heart attack than males for patients with hypertension, with diabetes, with a family history of CVD, and those who are smoking. However, for patients who are alcoholic, men are more prone to the risk of heart attack than women. The results were obtained with the help of SPSS.

Keywords: Cox Regression, Hazard Function, Heart Attack, Kaplan-Meier Estimates, SPSS, Survival Analysis, Survival Functions

INTRODUCTION

Heart disease is the leading cause of death among Filipinos, accounting for 1 out of every 5 deaths in the past year. The National Statistics Office (NSO) said that heart disease was cited as the cause of 100,908 deaths or 21% percent of the 480,820 deaths reported from January 2009 to March 2010 [1]. DOH said that 90% of Filipinos are exposed to one or more risk factors that trigger the progression of heart disease, and Myocardial Infarction, commonly known as Heart Attack, is the most common heart disease among the Filipinos [2].

Heart attack is the death of a segment of heart muscle caused by the loss of blood supply [3]. It occurs when one or more of your coronary arteries become blocked. This is caused by a coronary artery narrowed from the build-up of various substances, including cholesterol known as coronary artery disease [4]. A heart attack can also be caused by a severe spasm of a coronary artery. The spasm cuts off blood flow through the artery. Spasms can occur in coronary arteries that aren't affected by atherosclerosis. A spasm may be related to taking certain drugs, such as cocaine, emotional stress or pain, exposure to extreme cold and cigarette smoking [5].

Certain factors contribute to the unwanted build-up of fatty deposits that narrows arteries throughout the body [6]. There are modifiable risk factors (ones that can be changed) and non-modifiable risk factors (ones that can't be changed) that can contribute to having a heart attack. Modifiable risk factors include smoking, high total cholesterol, high blood pressure, diabetes, being physically inactive, being overweight or obese, depression, illegal use of drugs. Risk factors that can't be changed include increasing age, being male or female, and having a family history of heart disease [7].

Objective of the study

The main objective of this study is to formulate a mathematical model for the risk of developing a heart attack for patients with accordance to their age and gender, whether they are alcoholic or smoking, or if they have hypertension, diabetes, family history of Cardiovascular Disease (CVD), and based on the level of their high-density lipoprotein (HDL) and low-density lipoprotein (LDL). The researchers applied two main statistical treatments in examining the data which are Cox Regression Analysis and Kaplan-Meier Estimates. This paper will be helpful in aiding the health organizations in promoting awareness for the risk factors of a heart attack.
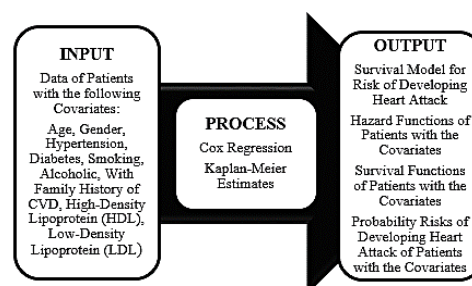

Fig. 1: Research paradigm

Research paradigm

The researchers input the data of the patients having the said covariates as shown in the fig. in conducting the survival analysis for the risk of developing a heart attack. They used two main processes which are Cox Regression and Kaplan-Meier Estimates. The expected outputs are the survival model for the risk of the occurrence of heart attack, the probability risks, the survival function and the hazard function of patients with the covariates.

Statement of the problem

The objective of this study is to undertake a survival analysis for the risk of developing heart attack for patients. Specifically, it attempted to find answers to the following questions:

What is the risk profile of patients that have experienced a heart attack and those who do not based on the covariates that lead to the occurrence of a heart attack?

a. Age

b. Gender

c. Hypertension

d. Diabetes

e. Smoking

f. Alcoholic

g. Family History of CVD

h. High-Density Lipoprotein (HDL)

i. Low-Density Lipoprotein (LDL)

What are the survival curves of patients using Kaplan-Meier Estimates according to their gender?

j. Hypertension Group

k. Diabetes Group

l. Smoking Group

m. Alcoholic Group

n. Family History of CVD Group

What is the hazard function and hazard curves of patients that can be computed using Cox Regression analysis and Kaplan-Meier Estimates?

What are the probability risks of developing heart attack of patients that has experienced a heart attack and those who do not in accordance to the covariates: age, gender, whether they have diabetes or hypertension, HDL and LDL levels, whether they are smoking, alcoholic or have a family history of CVD?

Scope and limitations

The study was limited by a small sample size as there were only 447 health data of patients available at Quirino Memorial Medical Center. The patients were admitted between January 2009 and June 2014. The researchers applied Cox Regression Analysis to identify the significant elements of a heart attack from the qualifying health survey data and medicinal clinic predictors of the patients. Measures did not assimilate interview.

MATERIALS AND METHODS

This section describes the data and the methods applied in this study to come up with the development of the survival model and the estimation of the probability of developing heart attack in a specified time for patients.

Data gathering procedure

The data used by the researchers consist of their personal profile: age, gender, either smoking or not, either an alcoholic or not; and medical records: diabetes, hypertension, family history of CVD, the level of LDL and HDL.

To compute the risk of the occurrence of heart attack, the status of each patient are needed to be observed. The status of the patients is 1 if the person failed, meaning the person has developed heart attack, and 0 if the person was censored or has not experienced a heart attack. Thereafter, the data must be organized to fit in statistical packages like Statistical Package for the Social Sciences (SPSS).

Kaplan-meier estimates

The Kaplan-Meier Estimates is a procedure of estimating time-to-event models in the presence of censored cases. It is a descriptive procedure for examining the distribution of time-to-event variables. The assumptions of Kaplan-Meier Method are the following [8]:

Probabilities for the event of interest should depend only on time after the initial event without covariates effects.

Cases that enter the study at different times (for example, patients who begin treatment at different times) should behave similarly.

Censored and uncensored cases behave the same. If for example, many of the censored cases are patients with more serious conditions, the results may be biased.

The survival probability at any particular time is calculated by the formula [9]:

$$S_t = \frac{\text{Number of subject living at the start} - \text{Number of subject died}}{\text{Number of subjects living at the start}}$$

For practical computational purposes, the same results can be obtained more efficiently by using the Kaplan-Meier product-limit estimator [10]:

$$S(t_i) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

Where $S(t_i)$ is the estimated survival probability for any particular one of the t time periods; n_i is the number of subjects at risk at the beginning of the time period t_i; and d_i is the number of subjects who die during the time period t_i.

Log-rank test

The Log-rank test is a non-parametric hypothesis test that is used to compare the survival distribution of samples. It is used when the survival curves do not cross each other (proportionality assumption), and to determine whether or not differences exist in the survival experiences of two groups [11]. The Log-rank test analyses the null hypothesis. The analysis is based on the times of events [12]. It also targets on the hazard function, not the survival functions. The Log-rank statistics is written as [13]:

$$\text{Logrank Statistic} = \frac{\sum_{j=1}^{K}(O_j - E_j)}{\sqrt{\sum_{j=1}^{k} V_j}}$$

Where O_j is the observed number of failures, E_j is the expected number of failures and, V_j is the variance.

H_0: There is no difference between the populations in the probability of an event at any time point.

H_a: There is a difference between the populations in the probability of an event at any time point.

Cox regression

Cox Regression is a semi-parametric method for investigating the effect of several variables upon the time a specified event takes to happen [14]. The method represents the effects of explanatory variables as a multiplier of a common baseline hazard function, $h_0(t)$. The hazard function is the nonparametric part of the Cox proportional hazards regression function, whereas the impact of the predictor variables is a log-linear regression [15].

In assessing the validity of the assumptions of the Cox model, hazards must be proportional and that the effect of covariates is linear in the log hazard. One of the methods on verifying the proportional hazard assumption is the graphical examination of parallelism of plots. Another one is the comparison of stratified with unstratified models. One can also plot the differences in log against time [16].

The Cox Regression procedure is useful for modeling the time to a specified event, based upon the values of given covariates. One or more covariates are used to predict a status (event). The central statistical output is the hazard ratio. The data contain censored and uncensored cases. Cox regression assesses the relationship between survival time and covariates [8]. Mathematically, the Cox model is written as [17]:

$$H(t) = h_0(t) \times \exp\{b_1 x_1 + b_2 x_2 + \cdots + b_p x_p\}$$

Where the hazard function $H(t)$ is dependent on a set of p covariates $(x_1, x_2, ..., x_p)$, whose impact is measured by the size of the respective coefficients $(b_1, b_2, ..., b_p)$. The term h_0 is called the baseline hazard and is the value of the hazard if all the x_i is equal to zero (the quantity exp (0) equals 1). The 't' in $H(t)$ reminds us that the hazard may vary over time.

Probability risks

The risk of developing Heart Attack is given by:

$$R(X, \text{Event}) = 1 - e^{-H(t,X)}$$

Where $e^{-H(t,X)}$ is the survival function in terms of its cumulative hazard rate function $H(t, X)$. Since there were more covariates involved in the

study, the researchers used the Cox Regression to estimate the hazard rate of the event on patients. The formula is given by:

$$H(t, X) = h_0(t) \times \exp\{b_1x_1 + b_2x_2 + \cdots + b_px_p\}$$

Where $h_0(t)$ is the cumulative baseline hazard function. This term depends on time but not the covariates, X_p is possibly time–independent covariates, and β is the regression coefficients.

RESULTS AND DISCUSSION

Risk profile of quirino memorial medical center patients

The table above exhibits the comparison of percentages of the risk profile of Quirino Memorial Medical Center patients that has and has not experienced a heart attack. The researchers used a total of 327 patients who have experienced a heart attack and 120 patients who

do not, making a total of 447 observations. Males compose almost 60% of the population of patients with a heart attack while the females comprise the 60% of the patients without a heart attack. Also, with regards to age, the most frequency on the age of the patients is the range of 60 to 69 while the least is from the range of 90 to 99 with both having a single patient.

On alcohol consumption, almost 50% of the patients who have experienced a heart attack are alcoholic. On the contrary, approximately 50% of the patients who have not experienced a heart attack are non-alcoholic. For smoking, 55% of the patients that has experienced a heart attack are smoking while 70% of those who do not experience a heart attack are not smoking. Almost half of the patients with heart attack have a family history of CVD. On the other hand, 52% of the patients without heart attack reveal that they have no family history of CVD.

Table 1: Risk profile of patients

Factors	Classification	Percentage 447 patients	
		Has experienced heart attack	**Has not experienced heart attack**
	Total	**327**	**120**
Gender	Male	203 (62.08 %)	41 (34.17 %)
	Female	124 (37.92 %)	79 (65.83 %)
Age	20–29	8 (2.45 %)	5 (4.17 %)
	30–39	19 (5.81 %)	12 (10 %)
	40–49	41 (12.54 %)	16 (13.33 %)
	50–59	65 (19.88 %)	30 (25 %)
	60–69	98 (29.97 %)	35 (29.17 %)
	70–79	73 (22.32 %)	17 (14.17 %)
	80–89	22 (6.72 %)	4 (3.33 %)
	90–99	1 (0.31 %)	1 (0.83 %)
Alcoholic	Non–drinker	138 (42.20 %)	63 (52.5 %)
	Drinker	189 (57.80 %)	57 (47.5 %)
Smoking	Non–smoker	147 (44.95 %)	87 (72.5 %)
	Smoker	180 (55.05 %)	33 (27.5 %)
Family History of CVD	Without Family History	115 (35.16 %)	62 (51.67 %)
	With Family History	212 (64.83 %)	58 (48.33 %)
Hypertension	Normal	115 (52.29 %)	92 (76.67 %)
	Hypertensive	212 (64.83 %)	28 (23.33 %)
Diabetes	Non–Diabetic	171 (52.30 %)	103 (85.83 %)
	Diabetic	156 (47.71 %)	17 (14.17 %)
High–Density Lipoprotein (mmol/L)	Best(≥ 1.6)	33 (10.09 %)	6 (5 %)
	Better(≤ 1.5)	50 (15.29 %)	9 (7.5 %)
	Poor(≤ 1)	244 (74.62 %)	105 (87.5 %)
Low–Density Lipoprotein (mmol/L)	Best(≤ 1.8)	59 (18.04 %)	14 (11.67 %)
	Better(≤ 2.6)	70 (21.41 %)	31 (25.83%)
	Poor(> 2.6)	198 (60.55 %)	75 (62.50 %)

There are more non-diabetic patients for both groups compare with the diabetic patients. The table also presents that majority of patients on both groups have poor HDL levels and LDL levels.

Survival curves of patients using kaplan-meier estimates according to their gender

Fig. 2 shows the survival curves of heart attack patients with hypertension, with diabetes, with a family history of CVD, who are smoking, and who are alcoholic. The survival curve is the statistical representation of the survival experience of the patients displaying the percentage of surviving against time.

From the five figures, patients with hypertension, with diabetes, with a family history of CVD, who are smoking, and those who are alcoholic reveals an inversely proportional relationship between the age and survival rates of the patients. It can be seen that as the age of the patient elevates, the probability of him surviving a heart attack decreases. The survival curves of females and males are very close to each other indicating no difference between the survival chances of the two. However, for long-term observation, the curve for the males is lower than females because of some factors, indicating that as men ages, they survive lesser chance than women, but both end up in developing a heart attack.

Among the five figures, patients who are alcoholic exhibit a different result. Males has a greater chance of surviving than

female which differs from the result of the other wherein women can survive more than men for patients with hypertension, with diabetes, with a family history of CVD and for those who are smoking.

Hazard function and hazard curves of patients using cox regression

The researchers aimed to formulate a model that can actually estimate the risk of a patient to experience a heart attack. Through Cox Regression Analysis, the researchers are able to come up with the hazard function written as:

$$H(t, x) = (0.00094944e^{0.0989t - 0.207x_1 + 0.301x_2 + 0.062x_3 + 0.337x_4 + 0.164x_5 - 0.037x_6 - 0.197x_7 + 0.006x_8})$$

Where t is the age of the patient, and x are the covariates: x_1 is the gender, x_2 is hypertension, x_3 is smoking, x_4 is alcoholic, x_5 is diabetes, x_6 is a family history of CVD, x_7 is HDL and x_8 is LDL. This means that the cumulative hazard a patient can have depends on their age, gender, and the covariates they have.

Like the survival curves of the patients, the hazard curves of females and males are very close to each other showing no difference among

the survival chances of the two. Though for a long period of observation, the curve for the females is lesser than males because of some factors, meaning, as women ages, they become more prone to the risk of heart attack than men.

Based on the results, males with hypertension, with diabetes, with a family history of CVD, and those who are smoking shows a higher risk than females. Apparently, among the five covariates, patients who are alcoholic reveal an opposite result from the others. Hazard curves presented that females have a greater risk of developing a heart attack. Additionally, males experienced heart attack at an earlier age than women do.

In an article entitled 'How does Heart Disease affect Women,' it is also stated that women are more likely than men to have Coronary Micro vascular Disease. Many researchers think that a drop in estrogen levels during menopause combined with other heart disease risk factors causes coronary MVD in females [18].
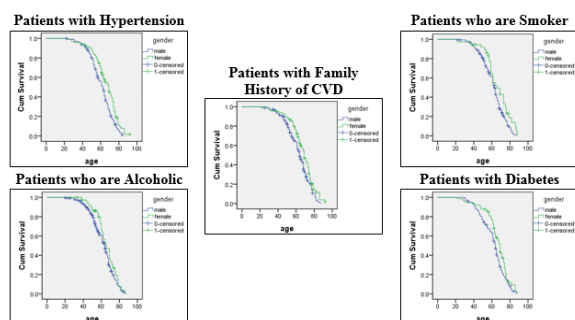

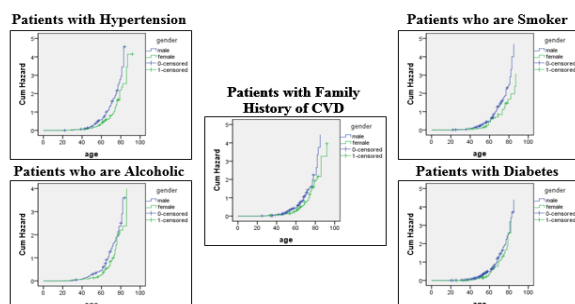**Fig. 2: Survival curves of patients according to gender**

Fig. 3: Hazard curves of patients according to gender

Probability risk of patients to develop heart attack

The probability risk of developing heart attack is identified by the model $R(x, \text{Heart Attack}) = 1 - e^{-H(t,X)}$ where $e^{-H(t,X)}$ is the survival probability in terms of a cumulative hazard rate function $H(t, X)$ which is derived from Cox Proportional Hazard Model and can be confirmed in the appendices.

Substituting the Cox proportional hazard rate function to the model, the probability risk function is written as:

$$R(x, \text{Heart Attack})$$
$$= 1$$
$$- e^{-\left(0.00094944\, e^{0.0989t - 0.207x_1 + 0.301x_2 + 0.062x_3 + 0.337x_4 + 0.164x_5 - 0.037x_6 - 0.197x_7 + 0.006x_8}\right)}$$

Where t = age of the individual

$$x_1 = \begin{cases} 1, \text{if female} \\ 0, \text{if male} \end{cases}$$

$$x_2 = \begin{cases} 1, \text{if he has hypertension} \\ 0, \text{if he has no hypertension} \end{cases}$$

$$x_3 = \begin{cases} 1, \text{if smoker} \\ 0, \text{if non-smoker} \end{cases}$$

$$x_4 = \begin{cases} 1, \text{if he is alcoholic} \\ 0, \text{if he is not alcoholic} \end{cases}$$

$$x_5 = \begin{cases} 1, \text{if he is diabetic} \\ 0, \text{if he is not diabetic} \end{cases}$$

$$x_6 = \begin{cases} 1, \text{if he has family history of CVD} \\ 0, \text{if he has no family history of CVD} \end{cases}$$

Table 2: Probability risk

Age	Probability risks percentage
61	54.216
62	51.033
63	64.538
64	51.499
65	69.105
66	73.695
67	75.712
68	80.172
69	83.804
70	82.574
71	82.875
72	91.352
73	89.143
74	87.775
75	88.443
76	92.243
77	93.809
78	95.877
79	97.309
80	96.906
81	99.716
82	99.625
83	99.821
84	99.925
85	99.140
86	99.840
87	99.936
88	99.639
92	99.982

Using the formulated model, the probability risks of Heart Attack patients were computed. Some of the patients with tremendous results are shown below together with their age. The table indicates that elderly are really prone to the development of Heart Attack.

CONCLUSION

Heart attack kills many lives without any notice. It is important to monitor this disease that exterminates people silently. Founded by the results, the researchers concluded that there is a 99 % possibility of developing a heart attack when a patient is positive of the given covariates: (hypertension, diabetes, family history of CVD, smoking and alcoholic) and when the patients are within the range of 60 to 90 y old.

Furthermore, in comparing the probability risks of patients according to their gender, men showed a higher record than women meaning males has more chances of experiencing a heart attack than females. In addition, Kaplan-Meier Estimates showed that females could survive more, than males.

Among the five covariates, patients who are alcoholic reveal an opposite result from the others. Survival curves of alcoholic patient's display that males have more possibility of surviving, and their hazard curves presented that females have a greater risk of developing a heart attack. For patients with the other covariates, they showed the opposite result both in their survival and hazard rates. Also, males experienced heart attack at an earlier age than women do.

Additionally, patients must have a high level of HDL and low level of LDL to avoid having a heart attack. An HDL greater than 1.6 mmol/l is considered as the best level of HDL while an HDL of less than one is considered as poor. For LDL level, it is good to have a total of less

than 1.8 mmol/l. If the patient exceeds to 2.6 mmol/l, it is measured as dangerous for the heart.

The researchers recommend that governments and heart associations must introduce new plans and policies in order to tackle about the awareness of heart attack and reduce the frequency of patients having a heart attack. This requires the understanding of the conventional risk factors and also the less known and new risk factors and ways which they may be prevented.

CONFLICT OF INTERESTS

Declared none

REFERENCES

1. ABS-CBN News. Heart disease is PH's top killer: NSO. Available from http://www.abs-cbnnews.com/lifestyle/10/20/11/heart-disease-phs-top-killer-nsowww.abscbnnews.com/lifestyle/10/20/11/heart-disease-phs-top-killer-nso. [Last accessed on 10 Oct 2011].
2. GMA News Online. Philippine Heart Association releases updated guidelines for Coronary Artery Disease. Available from: http://www.gmanetwork.com/news/story/390361/lifestyle/healthandwellnes/philippine-heart-association-releases-updated-guidelines-for-coronary-artery-disease. [Last accessed on 29 Nov 2014].
3. Living Heart Foundation. Heart Attack. Available from: http://www.livingheartfoundation.org/heartattack.pdf. [Last accessed on 29 Nov 2014].
4. Mayo Clinic. Causes of Heart Attack. Available from: http://www.mayoclinic.org/diseases-conditions/heart-attack/basics/causes/con-20019520. [Last accessed on 29 Nov 2014].
5. National Heart, Lung and Blood Institute. What Causes a Heart Attack? Available from: http://www.nhlbi.nih.gov/health/health-topics/topics/heartattack/causes. [Last accessed on 28 Nov 2014].
6. Mayo Clinic. Risk Factors of Heart Attack. Available from: http://www.mayoclinic.org/diseases-conditions/heart-attack/basics/risk-factors/con-20019520. [Last accessed on 25 Nov 2014].
7. Better Health Channel. Heart Attack–Risk Factors. Available from:http://www.betterhealth.vic.gov.au/bhcv2/bhcarticles.nsf/pages/Heart_disease_your_risk-factors_explained?open. [Last accessed on 29 Nov 2014].
8. Hui Bian. (nd). Survival Analysis using SPSS. Available from: http://core.ecu.edu/ofe/StatisticsResearch/Survival%20Analysis%20Using%20SPSS.pdf. [Last accessed on 29 Nov 2014].
9. Manish Kumar Goel, Pardeep Khanna, and Jugal Kishore. Understanding Survival Analysis: Kaplan-Meier Estimate. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059453/. [Last accessed on 15 Jan 2011].
10. Vassar Stats. Kaplan-Meier Survival Probability Estimates. Available from: http://vassarstats.net/survival.html. [Last accessed on 15 Jan 2011].
11. MBA Skool. Log-Rank Test. Available from: http://www.mbaskool.com/business-concepts/statistics/8764-log-rank-test-.html. [Last accessed on 15 Jan 2014].
12. Martin J Bland, Douglas G Altman. The log-rank test. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/ PMC403858/. [Last accessed on 01 May 2004].
13. Lu Tian, Richard Olshen. (nd). Survival Analysis: Log rank test. Available from: http://web.stanford.edu/~lutian/coursepdf/survweek3.pdf. [Last accessed on 01 May 2004].
14. Stats Direct. Cox (Proportional Hazards) regression. Available from: http://www.statsdirect.com/help/default.htm# survival _analysis/cox_regression.htm. [Last accessed on 09 May 2014].
15. Math Works. Cox proportional hazards regression. Available from:http://www.mathworks.com/help/stats/cox-proportional-hazard-regression.html. [Last accessed on 08 May 2014].
16. Harrell Jr, Frank E. Regression modeling strategies. Available from: www.springer.com/gb/book/9781441929181. [Last accessed on 01 May 2015].
17. MJ Bradburn, TG Clark, SB Love, DG Altman. Survival analysis part II: multivariate data analysis–an introduction to concepts and methods. Available from: http://www.ncbi. nlm.nih. gov/pmc/articles/PMC2394368/. [Last accessed on 29 Jul 2003].
18. Gibbons, Gary H. How does heart attack affect women? Available from: http://www.nhlbi.nih.gov/health/health-topics/topics/hdw. [Last accessed on 11 May 2015].

How to cite this article

- Jackie D Urrutia, Razzcelle L Tampis, Joseph Mercado, Aaron Vito M Baygan. Survival analysis for the risk of developing heart attack. Int J Pharm Pharm Sci 2016;8 Suppl 2:6-10.