ON THE ESTIMATION OF SURVIVAL TIME OF CARDIO-VASCULAR DISEASE PATIENTS WITH RANDOM NUMBER OF MYOCARDIAL INFARCTIONS USING PARAMETRIC AND SEMI-PARAMETRIC METHODS

Grover G.[(1)*], Gadpayle A.K.[(2)], Makhija N.[(1)]

[(1)]Department of Statistics, University of Delhi, India.
[(2)]Ram Manohar Lohia Hospital, Delhi, India.

Abstract: *Cardiovascular disease (CVD) is the leading cause of mortality in developing countries. CVD studies that involve the recording of two or more distinct and well-defined myocardial infarctions (MI) occurring over time in the same patient give rise to recurrent event data. In recent years a variety of fruitful statistical methods have been proposed for the analysis of recurrent events in medical areas. The present article is concerned with the estimation of the survival time of CVD patients, in the presence of recurrent myocardial infarctions followed by a terminal event death, under three different possibilities, i.e., the inter-event times between heart attacks follow gamma distribution, the number of heart attacks for an individual occur with time independent constant intensity λ , which is varying across individuals, and the hazard rates for the recurrent heart attacks vary from different attacks for the same individual. Cox's proportional hazard model has also been applied to study the effect of age at the time of first MI and the number of MI's experienced by the patient, on the survival time of CVD patients. Prior to that proportionality assumption has been tested. The methods are applied to a CVD patients data set obtained from Dr. Ram Manohar Lohia (R.M.L) Hospital, Delhi. The major advantage of developing models for estimating the survival time of CVD patients is that the treatment comparisons can be designed so that the expected survival time of new CVD patients can be predicted and improved after the first MI. Also it can be used as a baseline for further studies.*

Keywords: *Cardio-vascular disease, Myocardial Infarction, Recurrent Event, Survival Time.*

---

[*] Corresponding Author. E-mail: gurpritgrover@yahoo.com

1.   Introduction

Despite improved clinical care, heightened public awareness, and widespread use of health innovations CVD remains the leading cause of death in India and other developing countries [16,17]. Developing countries are undergoing a rapid epidemiological   transition – from infectious diseases such as diarrhea and pneumonia to chronic ones such as heart disease – that threatens to overwhelm their strapped health systems and cripple their fragile economies [Heidi Worley, January 2006]. Acute MI causes death or disability in many who are still in active years of life. Its personal and social costs are profound, both for patients and families involved and for the countries in which it is common [WHO Prevention of Coronary Heart Disease, TRS No. 678, 1982]. Cardiovascular diseases are manifested by the incidence of recurrent clinical events such as myocardial infarctions (MI). Often, the recurrence of these MI's, is interrupted by a terminating event, death during the course of the study.

A steady decline in the mortality rate from acute MI has been observed across several population groups since 1960. Although, its developing is still a   fatal   event   in   approximately one – third of patients [8]. A wide variety of approaches has been considered earlier for the analysis of recurrent event data including methods based on the inter-event times, marginal methods based on multivariate failure time data, general intensity based modeling, and methods based on event-counts [5]. In this paper, the research interest is focused on the generation of methods for estimating the survival time of CVD patients who had experienced at least one MI . This estimation can be used to predict the future lifetime of new CVD patients   and   thus   helps in developing treatment comparison designs [3].

The remainder of the paper is organized as follows. In section 2 developments of the models is discussed. It further consists of four parts. In the first part, we have estimated the survival time by taking the inter-event times between MI's to follow gamma distribution which has been verified to be a good fit to the data. In the second part we have estimated the survival time when N, the number of MI's , is a random variable and the intensity of MI's is constant for each interval but varying across individuals. In the third part we have estimated the survival time when MI's occur with varying intensities for the same individual. In the fourth part Cox's proportional hazard model [1,4,10] has been applied to study the effect of age at the time of first MI and the number of MI's experienced by the patient, on the survival time of CVD patients. Prior to that proportionality assumption has been tested. Section 3 applies the models to a data set of recurrent MI's in CVD patients obtained from R.M.L Hospital, Delhi, India. Although much work has been done for the analysis of recurrent events in the presence of a terminal event but to the best of our knowledge there is no study that has systematically estimated the survival time of CVD patients under the four possibilities mentioned above. Some concluding remarks are made in section 4.

2.   Development of the model

2.1   Part I: Estimation of survival time when the time between MI's is a random variable following gamma distribution.

Let T_{ij} (i=1,2,......,m; j=1,2,......,N) be the time of the jth MI for the ith patient measured from time zero. The zero point represents the time of birth of the individual. An individual is observed to

undergo N illness events (MI's), N = 1,2,…..The inter-event times from T_{ij-1} until T_{ij} is denoted by $X_{ij}(i=1,2,........m; \ j=1,2,........N)$ [9, 11, 14] which are independently distributed random variables, each having gamma distribution with parameters, say γ_{ij} and $\lambda_{ij}(i=1,2,.......m; \ j=1,2,........N)$.
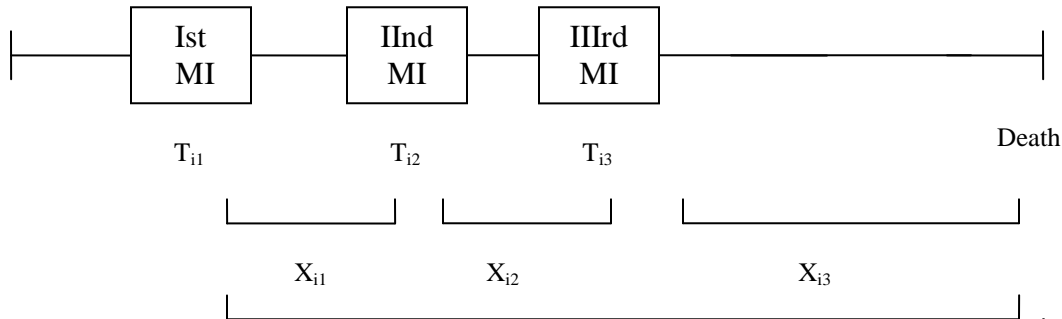
Figure 1. Representation of the History of a CVD patient who experienced at least one MI during the observed follow up.

The probability function of the time interval between MI's is given by:

$$f(x_{ij}) = \frac{\lambda_{ij}^{\gamma_{ij}}}{\sqrt{\gamma_{ij}}} e^{-\lambda_{ij}x_{ij}} x_{ij}^{\gamma_{ij}-1} ; \ \gamma_{ij}>1, \ \lambda_{ij}>0 \tag{1}$$

Here the shape parameter γ_{ij} describes that the hazard of the MI increases with time. Then $Y_i = X_{i1} + X_{i2} + X_{i3} +...........+ X_{iN}$ is the survival time of the ith patient with density function [15]:

$$g(y_i) = c\sum_{k=0}^{\infty} \delta_k \frac{y_i^{\rho+k-1}e^{-y_i/\beta_1}}{)\rho+k \, \beta_1^{\rho+k}} \tag{2}$$

where:

$$c = \prod_{j=1}^{N}\left(\frac{\beta_1}{1/\lambda_{ij}}\right)^{\gamma_{ij}}$$

$$\beta_1 = \min\left(1/\lambda_{ij}\right)$$

$$\delta_{k+1} = \frac{1}{k+1} \sum_{j=1}^{k+1} j\gamma_{ij} \delta_{k+1-j} ; k = 0,1,2,.......$$

$$\delta_0 = 1, \gamma_j = \sum_{j=1}^{N} \gamma_{ij} \frac{\left(1 - \frac{\beta_1}{\left(1/\lambda_{ij}\right)}\right)^k}{k} ; k = 1,2,3,.......$$

For the sake of simplicity we have done calculations when the numbers of MI's are 2 and 3 respectively. Parameters have been estimated by the method of maximum likelihood, by solving the likelihood function [7, 10].

$$L(\lambda_{ij}, \gamma_{ij}) = \prod_{k=1}^{m} \frac{\lambda_{ij}^{\gamma_{ij}}}{\overline{\gamma_{ij}}} t_k^{\gamma_{ij}-1} \exp(-\lambda_{ij} t_k) \tag{3}$$

where $t_k (k = 1,2,......,m)$ are the recorded survival times of m patients. It follows that:

$$\frac{\partial^2 \log_e L}{\partial \lambda_{ij}^2} = \frac{-n\gamma_{ij}}{\lambda_{ij}^2} \ , \ \frac{\partial^2 \log_e L}{\partial \lambda_{ij}\partial \gamma_{ij}} = \frac{n}{\lambda_{ij}} \ , \ \frac{\partial^2 \log_e L}{\partial \gamma_{ij}^2} = -n\Psi'(\gamma_{ij})$$

$$where \ \Psi(\gamma_{ij}) = \frac{d\log_e \overline{\gamma_{ij}}}{d\gamma_{ij}} \ and \ \Psi'(\gamma_{ij}) = \frac{d^2 \log_e \overline{\gamma_{ij}}}{d\gamma_{ij}^2}, \ \text{a trigamma function.} \tag{4}$$

The large sample variance-covariance matrix for $\hat{\lambda}_{ij}$ and $\hat{\gamma}_{ij}$ is given by:

$$\begin{pmatrix} \text{var}(\hat{\lambda}_{ij}) & \text{cov}(\hat{\lambda}_{ij}, \hat{\gamma}_{ij}) \\ \text{cov}(\hat{\lambda}_{ij}, \hat{\gamma}_{ij}) & \text{var}(\hat{\gamma}_{ij}) \end{pmatrix} = (n\Delta)^{-1} \begin{pmatrix} \Psi'(\gamma_{ij}) & 1/\lambda_{ij} \\ 1/\lambda_{ij} & \gamma_{ij}/\lambda_{ij}^2 \end{pmatrix} \tag{5}$$

where $\Delta = \left(\frac{\gamma_{ij}}{\lambda_{ij}^2} \Psi'(\gamma_{ij}) - \frac{1}{\lambda_{ij}^2} \right)$

Then the separate approximate 95% confidence interval for $\hat{\lambda}_{ij}$ and $\hat{\gamma}_{ij}$ are $\hat{\lambda}_{ij} \pm 1.96\sqrt{\text{var}(\hat{\lambda}_{ij})}$ and $\hat{\gamma}_{ij} \pm 1.96\sqrt{\text{var}(\hat{\gamma}_{ij})}$, respectively; where $\text{var}(\hat{\lambda}_{ij})$ and $\text{var}(\hat{\gamma}_{ij})$ are the diagonal elements of the variance-covariance matrix (5). Numerical values are obtained by replacing the parameter values with their maximum likelihood estimators [7].

The expected survival time of the patients who received two MI's and the corresponding estimated variance and confidence interval are given by:

$$E(Y_2) = \frac{\gamma_{i2}}{\lambda_{i2}} \ , \ V(Y_2) = \frac{\gamma_{i2}}{\lambda_{i2}^2} \ , \ E(Y_2) \pm 1.96\sqrt{\text{var}(Y_2)\Big/m} \tag{6}$$

and the expected survival time of the patients who received three MI's and the corresponding estimated variance and confidence interval are given by:

$$E\ Y_3\ = \frac{\gamma_{i2}}{\lambda_{i2}} + \frac{\gamma_{i3}}{\lambda_{i3}} \ , \ V(Y_3) = \frac{\gamma_{i2}}{\lambda_{i2}^2} + \frac{\gamma_{i3}}{\lambda_{i3}^2} \ , \ E(Y_3) \pm 1.96\sqrt{\text{var}(Y_3)\Big/m} \tag{7}$$

2.2 Part II: Estimation of Survival Time when the number of MI's, and the hazard rate of MI's, are random variables.

As introduced in part I, T_{ij} (i=1,2,......,m; j=1,2,......,N) is the time of the jth MI and $X_{ij}(i=1,2,........m;$ j=1,2,........N) is the time interval between these MI's which are independent and identically distributed random variables, each having exponential distribution with parameter, say λ.

$$f(x_{ij}) = \lambda e^{-\lambda x_{ij}}; i = 1, 2, ..., m; j = 1, 2, ..., N; \lambda > 0 \tag{8}$$

Suppose that the number of MI's, N, is a random variable following Geometric distribution with parameter, say p. The probability function of N is given by

$$P(N) = q^{N-1}p; N = 1, 2, ...; 0 < p \leq 1; N \geq 1 \tag{9}$$

where p is the probability of experiencing a fatal MI. Let Y_i be the sum of time intervals between MI's.

Then $Y_i = X_{i1} + X_{i2} + ... + X_{iN} \sim Gamma\ N, \lambda$ \hfill (10)

In this part, we estimate the survival time of cardiac patients by assuming that the MI's occur with hazard rate λ which is constant for each individual but is varying from individual to individual according to the following probability law:

$$\phi(\lambda) = \lambda^{\alpha-1} e^{-\gamma\lambda} \frac{\gamma^{\alpha}}{\overline{|\alpha}}; 0 \le \lambda < \infty; \alpha, \gamma > 0$$

Then the conditional distribution of survival time Y_i given N, the number of MI's and λ, the intensity with which an MI is experienced, for the ith patient is given by:

$$g\left(y_i | N, \lambda\right) = y_i^{N-1} e^{-\lambda y_i} \frac{\lambda^N}{\overline{|N}}; \ N \ge 1, \ y_i \ge 0, \ \lambda > 0 \tag{11}$$

Integrating equation (11) with respect to λ over the range 0 to ∞, we get:

$$h(y_i | N) = \frac{y_i^{N-1}}{\overline{|N}} \cdot \frac{\gamma^{\alpha}}{\overline{|\alpha}} \cdot \frac{\overline{|\alpha+N}}{\left(y_i + \gamma\right)^{\alpha+N}} \tag{12}$$

which on summation over N (N = 1,2,……) gives:

$$f(y_i) = \frac{p\gamma^{\alpha}}{\overline{|\alpha}\left(y_i + \gamma\right)^{\alpha+1}} \sum_{N=1}^{\infty} \frac{y_i q}{\left(y_i + \gamma\right)^{N-1}} \cdot \frac{\left(\alpha+N-1\right)!}{\overline{|N}} \tag{13}$$

$$f(y_i) = \frac{\alpha p \gamma^{\alpha}}{\left(y_i p + \gamma\right)^{\alpha+1}}$$

$$E(Y_i) = \alpha p \gamma^{\alpha} \int_{0}^{\infty} \frac{y_i}{\left(y_i p + \gamma\right)^{(\alpha+1)}} dy_i \tag{14}$$

On integrating by parts we get:

$$E(Y_i) = \frac{\gamma}{(\alpha-1)p}$$

which is the general form of estimated survival time for the ith patient. But when α < 1 we get:

$$E(Y_i) = \frac{\gamma}{(1-\alpha)p} \tag{15}$$

Parameters α and γ have been estimated separately for each patient by the method of moments and the probability p of a fatal MI, has been estimated by the method of maximum likelihood, which consists in maximizing the following log likelihood equation:

$$\log L = \sum_{i=1}^{m} \log \alpha' + \log p + \alpha' \log \gamma' - (\alpha'+1) \log(y_i p + \gamma') \qquad (16)$$

where $\alpha' = \sum_{i=1}^{m} \alpha_i$ and $\gamma' = \sum_{i=1}^{m} \gamma_i$, using the software R.

2.3. Part III: Estimation of survival times when MI's occur with varying intensities for the same individual.

In this part, we estimate the survival time of cardiac patient by assuming that the jth MI occurs with intensity $\lambda_{ij} (i=1,2,........m;\ j=1,2,........N)$. Then the density function of Y_i given N, where Y_i is the sum of the inter MI times (since the first MI) is given by [6]:

$$f(y_i | N) = \frac{\sum_{j=1}^{N} e^{-\lambda_{ij} y_i} (-1)^{N-1} \prod_{j=1}^{N} \lambda_{ij}}{\prod_{\substack{k=1 \\ k \neq j}}^{N} (\lambda_{ij} - \lambda_{ik})} ;$$

which implies:

$$g(y_i) = \sum_{N=1}^{\infty} \frac{\sum_{j=1}^{N} e^{-\lambda_{ij} y_i} (-1)^{N-1} \prod_{j=1}^{N} \lambda_{ij}}{\prod_{\substack{k=1 \\ k \neq j}}^{N} (\lambda_{ij} - \lambda_{ik})} . pq^{N-1} \qquad (17)$$

Then using m.g.f property:

$$m.g.f(y | N) = \frac{\prod_{j=1}^{N} \lambda_{ij}}{\prod_{j=1}^{N} (\lambda_{ij} - t)}$$

On differentiating w.r.t t and substituting t=1 we get:

$$E(Y_i | N) = \sum_{j=1}^{N} \frac{1}{\lambda_{ij}} \qquad (18)$$

which implies:

$$E(Y_i) = \sum_{N=1}^{\infty} \sum_{j=1}^{N} \frac{1}{\lambda_{ij}} pq^{N-1}$$

On changing summation we get:

$$E(Y_i) = \sum_{j=1}^{\infty} \sum_{N=j}^{\infty} \frac{1}{\lambda_{ij}} pq^{N-1} = \sum_{j=1}^{\infty} \frac{1}{\lambda_{ij}} q^{N-1} \tag{19}$$

which is the estimated survival time for the ith patient. The parameters $\lambda_{i1}, \lambda_{i2}, \lambda_{i3}$ and p have been estimated as follows:

$$\lambda_{i1} = \frac{\text{Number of MI's}}{\text{Number of years survived}}$$

$$\lambda_{i2} = \frac{\text{Number of MI's after the first MI}}{\text{Number of years survived after first MI}}$$

$$\lambda_{i3} = \frac{\text{Number of MI's after the second MI}}{\text{Number of years survived after second MI}} \tag{20}$$

$$p = \frac{\text{Number of deaths due to CVD}}{\text{Total number of CVD patients}}$$

Student – t test has been used to ascertain the significance of differences between the observed and the estimated survival times obtained in part II and part III.

2.4 Part IV: **Cox's proportional hazard model.**
In this part Cox's proportional hazard model [1,4,10] has been applied on participants who have experienced MI's, to assess the influence of two covariates, viz., age at the time of first MI (years) and the number of MI's experienced by the patient, on the survival time of CVD patients [4]. Prior to that proportionality assumption has been tested graphically [2].
Here, we briefly describe the Cox's proportional hazard model in the notations used by Collet (2003). We consider a Cox's model of the form:

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi})h_0(t) = \exp(\beta' x_i)h_0(t) \tag{21}$$

where $h_i(t), i = 1, 2, \dots, n$ denotes the hazard function for the ith individual. The hazard of death at a particular time depends on the values $x_1, x_2, \dots, x_p$ of p explanatory variables, $X_1, X_2, \dots, X_P$, which are represented by the vector x, so that $x = (x_1, x_2, \dots, x_p)'$. $h_0(t)$ is the baseline hazard function for an individual for whom the values of all the explanatory variables

that make up the vector x are zero and β is the vector of coefficients of the explanatory variables $x_1, x_2, ..., x_p$ in the model.

Suppose that data are available for n individuals, among whom there are r distinct death times and n-r right-censored survival times. The r ordered death times are denoted by $t_{(1)} < t_{(2)} < .... < t_{(r)}$, so that $t_{(j)}$ is the jth ordered death time. The set of individuals who are at risk at time $t_{(j)}$ are denoted by the risk set $R(t_{(j)})$, so that $R(t_{(j)})$ is the group of individuals who are alive and uncensored at a time just prior to $t_{(j)}$. The relevant likelihood function for the proportional hazards model in equation (21) is given by:

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta' x_j)}{\sum_{l \varepsilon R(t_{(j)})} \exp(\beta' x_l)} \tag{22}$$

in which x_j is the vector of covariates for the individual who dies at the jth ordered death time, $t_{(j)}$. The maximum likelihood estimates of the β-parameters in the proportional hazards model can be found by maximizing the log-likelihood function using numerical methods. The calculations in this paper have been carried out using the package SPSS.

3.   Application

We now apply the methods developed in section 2 to CVD patients data obtained from Dr.R.M.L. Hospital, Delhi, India for the year 2005-2006. The records of patients admitted with diagnosis of CVD were reviewed. The CVD patients with co-morbidity like renal failure, pneumonia, obstructive respiratory disease, etc. were excluded from the study because their number was not sufficient to draw valid conclusion about these identities. Thus data records of 295 cases was collected. Out of these 295 cases there were only 35 cases (uncensored) who died after experiencing at least one MI. The demographic and risk variables recorded were : hospital discharged status (survived or died), sex, time of MI, diabetes (Y/N), hypertension (Y/N) and smoking (Y/N). In addition to these, two other covariates at baseline were measured: number of MI's experienced by the patient and age of the patient at the time of first MI (years).

The motivation of our model development was to estimate the survival time of CVD patients under various conditions. In the first part we have taken the inter-event times between MI's to follow gamma distribution. Under this condition the estimated survival time of the patients who received two and three MI's are given in Table 1 and Table 2 respectively along with the estimated parameters, their standard errors and confidence limits [7].

Table 1. Parameter estimates and mean survival time (years) of patients who experienced 2 myocardial infarctions (MI's), with corresponding standard errors and 95% confidence intervals

Parameter	Number of MI's experienced - 2		
	Estimate	**Standard Error**	**95% Confidence Limits**
Shape λ_{i2}	5.2585337	2.281000394	0.78777292; 9.72929446
Scale γ_{i2}	2.6292669	1.196747279	0.28364218; 4.97489151
Mean survival time(years)	2.0000015	0.275802161	1.45942926; 2.54057374

Table 2. Parameter estimates and mean survival time (years) of patients who experienced 3 Myocardial infarctions (MI's), with corresponding standard errors and 95% confidence intervals.

Parameter	Number of MI's experienced -3		
	Estimate	**Standard Error**	**95% Confidence Limits**
Shape γ_{i2}	2.9108545	0.872903736	1.199963137; 4.621745783
γ_{i3}	4.8532587	1.484991627	1.942675114; 7.763842294
Scale λ_{i2}	0.8820771	0.288675716	0.316272705; 0.499083112
λ_{i3}	3.5950065	1.158936893	1.323490137; 5.866522757
Mean survival time(years)	4.65	0.453689894	3.760767806; 5.539232194

The estimated survival time of the patients who received 2 MI's came out to be 2.0000015 years since birth and the observed survival time was 2 years. The estimated survival time of the patients who received 3 MI's came out to be 4.65 years since birth and the observed survival time was also 4.65 years. Thus the estimated and the observed survival times in both the cases were found to be remarkably similar.

Next, in part II, we have taken the number of MI's N to be a random variable following geometric distribution with parameter p and also the MI's occur with constant intensity λ which is varying across individuals. The parameter p as estimated by the method of maximum likelihood using the software R came out to be 0.3339609. The 95% confidence interval for the estimated mean survival time came out to be (3.154386, 4.4528543). The paired Student's t statistic gave an observed value of 1.827 (p>0.05) on 25 degrees of freedom. Thus we conclude that there is not enough evidence to reject the hypothesis that there is no significant difference between the estimated and the observed survival times.

Figure 2 shows the estimated and observed survival times (years) of patients (as obtained in part II) who received two and three MI's.

In part III, we have estimated the survival time when MI's occur with varying intensities for the same individual. The parameter p is estimated to be 0.118644067. The 95% confidence interval [7] for the estimated mean survival time came out to be (2.610309184, 4.822947958). The paired Student's t statistic gave an observed value of 1.835 (p>0.05) on 33 degrees of freedom. Thus we

conclude that there is not enough evidence to reject the hypothesis that there is no significant difference between the estimated and the observed survival times. Figure 3 shows the estimated and observed survival times (years) of CVD patients who received two and three MI's (as obtained in part III).
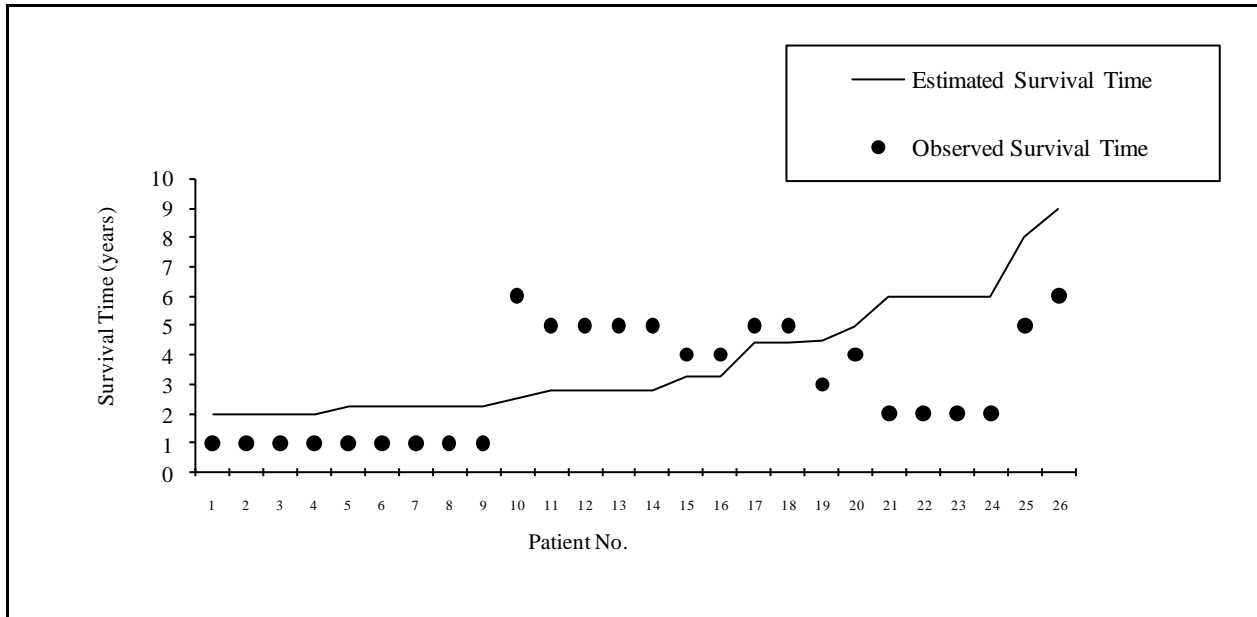

Figure 2. Estimated and observed survival time of CVD patients when the number of MI's, N, is a random variable and also the MI's occur with constant intensity which is varying across individuals.
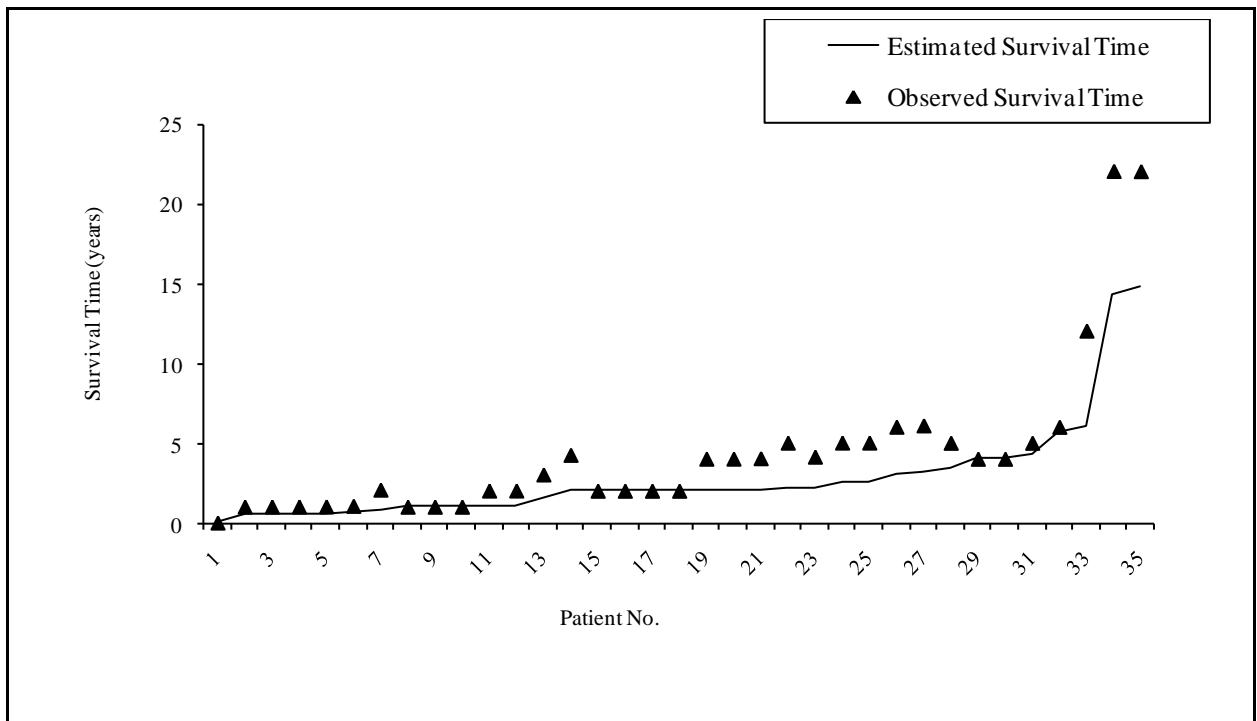

Figure 3. Estimated and observed survival time of CVD patients when MI's occur with varying intensities for the same individual.

The fourth part in this paper is devoted to investigate the effect of the age at the time of first MI and number of MI's experienced on the survival time of CVD patients. In this study, the primary response variable is the time for first MI. The data, obtained from Dr. R.M.L. hospital, Delhi, India, relate to 105 patients, who were aged between 30 and 90 years. Some of these patients had not died by the time the study was terminated, and so these individuals contribute right-censored survival times. At the time of first MI, two covariates at baseline were measured for each patient: number of MI's experienced by the patient and age of the patient at the time of first MI (years). The age of a patient at the time of first MI has been classified according to whether the patient is ≤ 60 years or > 60 years. Also, the number of MI has been classified according to whether the patient experiences 1, 2 or 3 MI's. Here we are considering three models for these data depending on whether hazard function is related to neither, one or both of these factors [4]. Checking of the model's assumption of proportional hazards is done graphically using log minus log hazard plot as shown in figure 4 and figure 5. The graphs revealed that the hazards for patients with ≤ 60 years of age and with > 60 years of age are proportional. Also, the hazards for patients who experienced 1, 2 and 3 MI's are also found to be proportional [2, 4].

First we consider a model (Model I) where only the number of MI's are assumed to influence the survival time of the patient and we define the time-independent covariates as:

- $Z_1 = 1$ if the individual experienced one MI and zero otherwise.
- $Z_2 = 1$ if the individual experienced two MI's and zero otherwise.

The proportional hazard model for the ith individual is then:

$$h_i(t) = \exp(\beta_1 Z_1 + \beta_2 Z_2) h_0(t) \tag{23}$$

where $h_0(t)$ is the baseline hazard function that corresponds to the hazard of death at time t for an individual who experienced three MI's.

The value of $-2 \log \hat{L}$ for the null model is 481.868. The reduction in $-2 \log \hat{L}$ on adding the effect due to number of MI's is 10.101 on 2 degrees of freedom which is significant at 5% level of significance (p = 0.006). Thus we conclude that the hazard function does depend on the number of MI's experienced by the patient. The estimates for Model I are given in table 3. The hazard ratio given in table 3 suggests that the hazard of death at any given time is greatest for patients who experienced only one MI, but there is little difference in the hazard functions for patients who experienced two and three MI's respectively.

Second we consider a model (Model II) where only the age of the patient at the time of first MI is assumed to influence the survival time of the patient and we define the time-independent covariate as:

$Z_3 = 1$ if the patient is ≤ 60 years and zero otherwise.

The proportional hazard model for the ith individual is then:

$$h_i(t) = \exp(\beta_3 Z_3) h_0(t) \tag{24}$$

where $h_0(t)$ is the baseline hazard function that corresponds to the hazard of death at time t for an individual aged > 60 years.
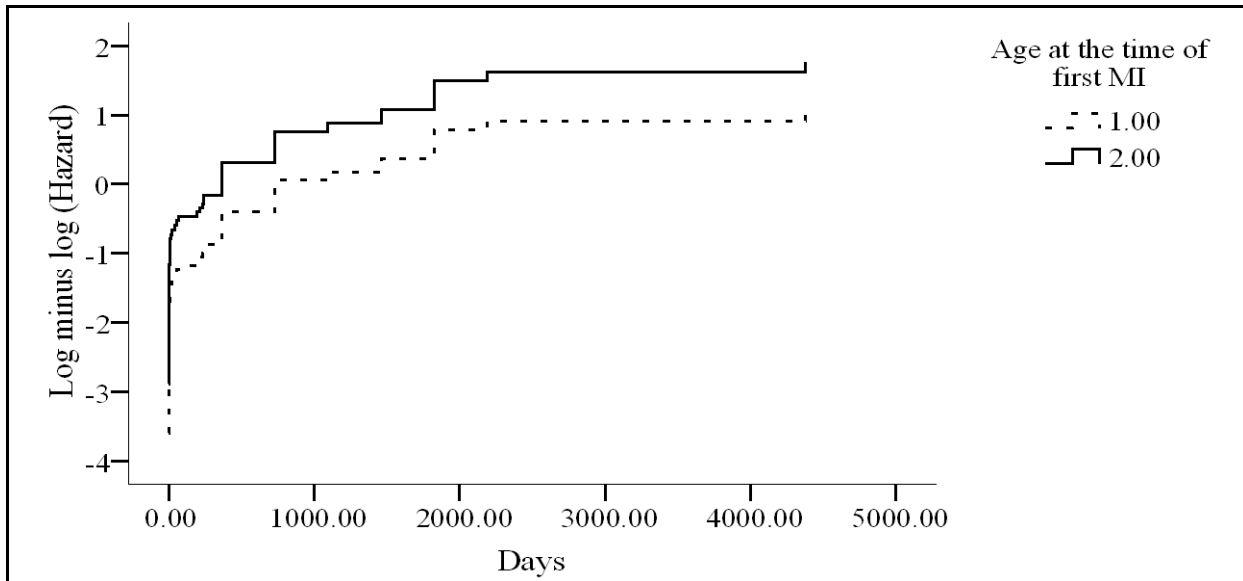

Figure 4. Goodness-of-fit plot for age when scored 1 for age less than or equal to 60 years and 2 for age greater than 60 years.
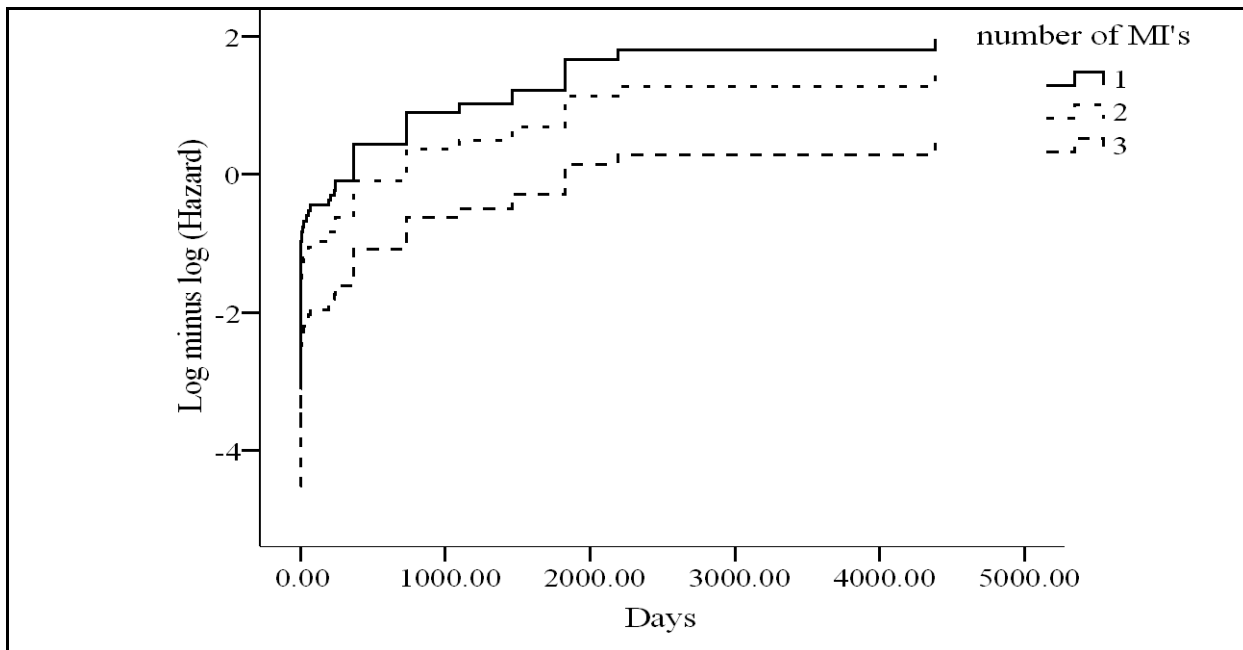

Figure 5. Goodness-of-fit plot for number of MI's when scored 1 for number of MI =1, 2 for number of MI's = 2 and 3 for number of MI's = 3.

Table 3. Cox's proportional hazard model when only the number of myocardial infarctions (MI's) are assumed to influence the survival time of the CVD patients.

Number of MI's	B	Standard Error	Wald	d.f.	Sig.	Exp(B)
1	1.515	0.516	8.619	1	0.003	4.548
2	0.984	0.451	4.758	1	0.029	2.676

The reduction in $-2\log \hat{L}$ on adding the effect due to the age of the patient at the time of first MI is 7.323 on 1 degree of freedom which is significant at 5% level of significance (p = 0.007). Thus we conclude that the hazard function does depend on which age group the patient is in. The estimates for Model II are given in table 4. The negative sign of the regression coefficient indicates that there is a decreased relative risk for the patients ≤ 60 years of age. The hazard of dying for patients ≤ 60 years of age is about 49% less than that of the hazard of dying for patients > 60 years of age when no other covariates are considered.

Table 4. Cox's proportional hazard model when only the age of the patient at the time of first myocardial infarction (MI) is assumed to influence the survival time of the CVD patients.

	B	Standard Error	Wald	d.f.	Sig.	Exp(B)
Age	-0.708	0.259	7.466	1	0.006	0.493

Third we consider a model (Model III) where both the age of the patient at the time of first MI and the numbers of MI's experienced are assumed to influence the survival time of the patient. The impact of number of MI's on the hazard is independent of the age group of the patient. The proportional hazard model for the ith individual is then:

$$h_i(t)= \exp(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3) h_0(t) \tag{25}$$

The change in the value of -2log $\hat{L}$ when the effect due to the age of the patient at the time of first MI is included in the model that contains the effect due to the number of MI's is 8.286 on 1 degree of freedom. This is significant at 5 % level of significance and so there is some evidence that effect due to the age of the patient at the time of first MI is needed in a model that contains the effect due to the number of MI's. The estimates for model III are given in table 5.

The change in the value of -2log $\hat{L}$ when the effect due to the number of MI's is included in the model that contains the effect due to the age at the time of first MI is 5.286 on 2 degrees of freedom. Thus the test statistic is significant and so there is some evidence that effect due to the number of MI's is needed in a model that contains the effect due to the age of the patient at the time of first MI. Thus putting these two results together we conclude that the hazard function depends on both the patient's age group at the time of first MI and the number of MI's experienced by the patient. The hazard ratio given in table 7 suggests that the hazard of death at any given time is greatest for patients who have experienced one MI. Also the number of MI's (2) is found to be no longer significantly related to the survival time of the CVD patients. Here again the negative sign of the regression coefficient indicates that there is a decreased relative risk for the patients ≤ 60 years of age. The hazard of dying for patients ≤ 60 years of age is about

53% less than that of the hazard of dying for patients > 60 years of age, when the effect due to the number of shocks experienced by the patient is also considered.

Table 5. Cox's proportional hazard model when both the age of the patient at the time of first myocardial infarction and the number of myocardial infarctions (MI's) experienced are assumed to influence the survival time of the patients.

	B	Standard Error	Wald	d.f.	Sig.	Exp(B)
Age	-0.628	0.267	5.532	1	0.019	0.534
No. of M I's			7.809	2	0.02	
No. of M I's (1)	1.418	0.528	7.205	1	0.007	4.13
No. of M I's (2)	0.771	0.468	2.717	1	0.099	2.162

4.   Discussion

The aim of this study is to estimate the survival time of CVD patients under three different conditions. The major use of estimating the survival time of CVD patients is that the current as well as the future life time of new CVD patients can be predicted. Also it can be used as a baseline for further studies. It may also allow one to gain a deeper insight into the various differences that may exist between the treatments given to CVD patients.

In the first part we have assumed that the times between recurrent events for subject i follow gamma distribution. In the second part the hazard rate λ of the recurrent events is assumed to be constant for each individual but is varying from individual to individual. Also the recurrent events for subject i follow geometric distribution with parameter p, which is the probability of getting a fatal MI. For the sake of simplicity we have considered the probability of developing an MI to be a constant but in many applications this probability is impacted by covariates. This concept is suggested for further studies. In the third part we have assumed that the recurrent events occur with intensities λ_{ij} (i = 1,…., m ; j = 1, 2,…,N) for the ith patient. In all the above three models we have considered only uncensored cases. Clearly, considering only uncensored cases will increase the mean of the survival time. Also there are many potential applications in which this assumption might be questionable [13]. For example, if the occurrence of an event causes some physiological damage to an individual and this effect is cumulative, the occurrence of several events might in fact be prognostic for death (which will be considered as a censoring because we focus on the recurrent events process). This feature will induce dependence between censoring and the event process [13]. In the analysis of recurrent event process the investigators are often interested in estimating the frequency of recurrences over time as well as assessing the effects of covariates on the recurrence times [12].To deal with such complications it is essential to adopt more elaborate methods. Thus further modeling should be developed accordingly.

In the fourth part of the paper we have applied Cox's proportional hazard model to assess the influence of two covariates, viz., age at the time of first MI (years) and the number of MI's experienced by the patient, on the survival time of CVD patients. It is found that age at the time of first MI is significantly related to the survival time of the patient in the presence of the effect due to the number of MI's. The estimated coefficient B for number of MI's (2) has been found to be not significantly related to the survival time of the patient. But, since the contribution of any

one variable is determined in the context of the contribution of all other variables in the model [4], if other variables are included in the model, number of MI's (2) might be found to make a significant contribution.

Acknowledgement

References

[1]. Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes : A large sample study. *The annals of statistics*, 10 (3), 1100-1120.

[2]. Andersen, P.K., Borch-Johnsen, K., Deckert, T., Green, A., Hougaard, P., Keiding, N. and Kreiner, S. (1985). A Cox regression model for the relative mortality and its application to diabetes mellitus survival data. *Biometrics*, 41, 921-932.

[3]. Barai, U. and Teoh, N. (1997). Multiple statistics for multiple events, with application to repeated infections in the growth factor studies. *Statistics in medicine*, 16, 941-949.

[4]. Collett, D. (2003). *Modeling survival data in medical research (2nd edition).* London: Chapman and Hall/CRC.

[5]. Cook, R.J. and Lawless, J.F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine*, 16, 911-924.

[6]. Chiang, C.L. (1968). *Introduction to Stochastic Processes in Bio-Statistics*. New York: J. Wiley.

[7]. Gross, J.A. and Clark, A.V. (1975). *Survival distributions : Reliability applications in the biomedical sciences.* , New York, London, Sydney, Toronto: John Wiley and Sons.

[8]. Kakade, S.V., Tyagi, N.K., Kadam, R.N. (2006). Application of logistic regression to estimate prognosis in acute myocardial infarction. *Indian journal of community medicine*, 31 (2), 69-72.

[9]. Lambert, P., Collett, D., Kimber, A. and Johnson, R. (2004). Parametric Accelerated Failure Time Models with Random Effects and an Application to Kidney Transplant Survival. *Statistics in Medicine*, 23, 3177-3192.

[10]. Lee, E.T., Wang, W.Y. (2003). *Statistical Methods For Survival Data Analysis, Third Edition*. New Jersey: John Wiley & Sons.

[11]. Li, Q.H., Lagakos, S.W. (1997). Use of the Wei-Lin-Weissfeld Method for the Analysis of a Recurring and a Terminating Event. *Statistics in Medicine*, 16, 925-940.

[12]. Lin, D.Y., Wei, L.J. and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika*, 85 (3), 605-618.

[13]. Mahe, C., Chevret, S. (2001). Analysis of Recurrent Failure Times Data: Should the Baseline Hazard be Stratified?. *Statistics in Medicine*, 20, 3807-3815.

[14]. Metcalfe, C., Thompson, S.G. (2006). The importance of varying the event generation process in simulation studies of statistical methods for recurrent events. *Statistics in medicine*, 25, 165-179.

[15]. Moschopoulos, P.G. (1985). The distribution of the sum of independent gamma random variables. *Ann. Inst. Statist. Math.*, 37, Part A, 541-544.

[16]. Murray CJ. Lopez AD. (1997). Global Mortality, Disability, and the Contribution of Risk Factors; Global Burden of Disease Study. *Lancet*, 349, 1436-1442.

[17]. Reddy KS.and Yusuf S. (1998). Emerging epidemic of Cardiovascular Disease in Developing Countries. *Circulation*, 97, 596-601.