



# Non-Linear Regression Models for Heart Attack Data – An Empirical Comparison

## KEYWORDS

Heart attack, Cox proportional hazard model, time dependent covariates, frailty model.

**V. Vallinayagam**

Department of Mathematics, St. Joseph's College of Engineering, Chennai, India

**S. Prathap**

Department of Mathematics, Jeppiaar Engineering College, Chennai, India.

**P. Venkatesan**

Department of Statistics, National Institute for Research in Tuberculosis, Chennai, India

## ABSTRACT

*Non linear regression is a popular statistical tool that has been used successfully in different areas of research. The Cox proportional hazard model has achieved widespread use in the analysis of time-to-event data with censoring and covariates. The covariates may change their values over time. A frailty model is a random effects model for time variables, where the random effect (the frailty) has a multiplicative effect on the hazard. It can be used for univariate (independent) failure times, i.e., to describe the influence of unobserved covariates in a proportional hazards model. The aim of this paper is to compare the performance of Cox proportional hazard model, Cox time dependent model and Frailty model using Heart attack data. The result shows that Cox time dependent model is better than other models.*

## 1. Introduction

Survival analysis is concerned with studying the time between entry to a study and a subsequent event and becomes one of the most important fields in statistics. The techniques developed in survival analysis are now applied in many fields, such as biology (survival time), engineering (failure time), medicine (treatment effect of drugs), quality control (lifetime of components), credit risk modeling in finance (default time of a firm).

The Cox proportional hazard model is now the most widely used for the analysis of survival data in the presence of covariates or prognostic factors. This is the most popular model for survival analysis because of its simplicity, and not being based on any assumptions about the survival distribution. The model assumes that the underlying hazard rate is a function of the independent covariates, but no assumptions are made about the nature or shape of the hazard function. In the last several years, the theoretical basis for the model has been solidified by connecting it to the study of counting processes and martingale theory, which was discussed in the books of Fleming and Harrington (1991) and of Andersen et al (1993). These developments have led to the introduction of several new extensions of the original model. However the Cox proportional hazard model may not be appropriate in many situations and other modifications such as stratified Cox model (Kleinbaum, 1996) or Cox model with time dependent variables (Collett, 2003) can be used for the analysis of survival data.

A frailty model is a random effects model for time variables, where the random effect (the frailty) has a multiplicative effect on the hazard. It can be used for univariate (independent) failure times, i.e., to describe the influence of unobserved covariates in a proportional hazards model. However, is to consider multivariate (dependent) failure times generated as conditionally independent times given the frailty. This approach can be used for survival times for individuals, like twins or family members, and for repeated events for the same individual. The standard assumption is to use a gamma distribution for the frailty, but this is a restriction that implies that the dependence is most important for late events (Hougaard, 1995). The word 'frailty' was introduced by Vaupel et al. (1979) for univariate data. These models are used to explain the deviant behavior of mortality rates at advanced ages (Vaupel and Yashin, 1985), to correct biased estimates of regression coefficients in Cox-type models of hazard rate (Chamberlain, 1985) and to separate compositional and bio-

logical effects in aging studies (Manton et al., 1986). Frailty models play an important role in the interpretation of the results of stress experiments (Yashin et al., 1996a) and in centenarian studies (Yashin et al., 1999). The use of survival data on related individuals opens a new avenue in frailty modeling. Genetic variation, heritability, and other properties of individual susceptibility to death can now be analyzed using correlated frailty models (Yashin and Lachine, 1995; Yashin and Lachine, 1997).

In Heart attack studies, the main outcome of interest is the time to occurrence of event like death, relapse etc. Coronary heart disease (CHD) is the leading cause of death world wide (Mackay & Mensah, 2004). Although men have higher rates than women at all ages, and coronary disease occurs up to 10 years later in women (Sharp, 1994), CHD is a major cause of death for both sexes: the World Health Organisation estimates that 3.8 million men and 3.4 million women around the world die from it each year (Mackay & Mensah, 2004). Despite recent improvements, the mortality rate in the UK remains amongst the highest in the world and coronary prevention is a priority (The Scottish Office, 1999; Department of Health, 1999).

The objective of this paper is to compare the performance of Cox proportional hazard model, Cox time dependent model and Frailty model using Heart attack data.

## 2. Non-linear Regression Models

### 2.1 Cox Proportional Hazard Model

It is a mathematical modeling approach for estimating survival curve when considering several explanatory variables simultaneously. It is also called semi parametric model. The proportional hazard model describes the relationship between the hazard function of the risk of an event and a set of covariates. The Cox proportional hazard model is usually written in terms of the hazard model. It is given below as described by Cox (1972)

$$h(t, X) = h_0(t) \cdot \exp\left(\sum_{i=1}^p X_i \beta_i\right) \quad (1)$$

where  $h_0(t)$  is baseline hazard and  $\beta_i$  is parameter vector and  $X_i$  are independent variables. The above equation (1) reveals that the hazard at time  $t$  is the product of two quantities. The first of these,  $h_0(t)$  is called the baseline hazard function. The second quantity is the exponential expression. This model gives an expression for the hazard at time  $t$  for an individual with a

given specification of a set of explanatory variables denoted by  $X$ . An important feature of equation (1), which concerns the proportional hazard assumption, is that the baseline hazard is a function of  $t$ , but doesn't involve the  $X$ . The baseline hazard function is left unspecified so that the time to event random variable is not assumed to follow any particular distribution and this is one of the essential properties of proportional hazard model (Lee, 1992).

**2.2 Cox Time Dependent Covariate Model**

Time dependent covariates have been studied a number of authors (Crowley & Hu, 1977; Cox & Oakes, 1984; Andersen, 1986; Fisher & Lin 1999). A baseline Cox analysis ignores the change of updated covariate values usually yields smaller effect estimates than a time dependent analysis using all temporal information available (Aydemir et al., 1999). Also Altman and De Starola (1994) called this is the time decay of the effects of entry values. One of the earliest applications of the use of time varying covariates in a biomedical setting may be found in Crowley (1977).

Let  $X_i(t)$  denote the value of the covariate measured at time  $t$ . Let  $X_{ip}$  denote the value of the covariate for subject  $p$  at time  $t$

$$X_i(t_i) = [X_{i1}(t_i), X_{i2}(t_i), X_{i3}(t_i), \dots, X_{ip}(t_i)] \quad (2)$$

The notation in the above equation is completely general in the sense that, if a particular covariate, is fixed then and this has lead to use the time dependent notation in equation (1) exclusively. The generalization of the proportional hazard regression function to include possibly multiple time varying covariates is

$$h(t, X(t), \beta) = h_0(t) \cdot \exp(X' \beta(t)) \quad (3)$$

and the generalization of the partial likelihood function

$$l_p(\beta) = \prod_{i=1}^n \left[ \frac{e^{X_i \beta}}{\sum_{j \in R(t_i)} e^{X_j \beta}} \right] = \prod_{i=1}^n \left[ \frac{e^{X_i'(t_i) \beta}}{\sum_{i \in R(t_i)} e^{X_i'(t_i) \beta}} \right]^{c_i} \quad (4)$$

**2.3 Frailty Models**

Frailty is an unobserved random factor that modifies multiplicatively the hazard function of an individual or group or cluster of individuals. Vaupel et al. (1979) introduced univariate frailty model (with a gamma distribution) into survival analysis to account for unobserved heterogeneity or missing covariates in the study population. The idea is to suppose that different patients possess different frailties and patients more frail or prone tend to have the event earlier than those who are less frail. The model is represented by the following hazard given the frailty:

$$h(t | Z, X) = Z h(t | X) \quad (5)$$

can be equal to the baseline hazard function (in Cox regression model). The baseline hazard function can be chosen non-parametrically, or parametrically. An important point is that the frailty  $Z$  is an unobservable random variable varying over the sample which increases the individual risk if  $Z > 1$  or decreases if  $Z < 1$ . The model can also be represented by its conditional survivor function

$$S(t | X) = \exp\left(-z \int_0^t h(u | X) du\right) = \exp(-Z \Lambda(t | X)), \quad (6)$$

where  $\Lambda(t | X) = \int_0^t h(u | X) du$

The marginal survivor function can be calculated by

$$S(t | X) = \int S(t | Z, X) g(z) dz = E[S(t | Z, X)] = L(\Lambda(t | X)) \quad (7)$$

Elbers and Ridder (1982) proved that frailty model with mean is identifiable with univariate data, when covariates are included in the model. Many distributions can be chosen for the frailty, but the most common frailty distribution with gamma distribution. The Gamma distribution has been widely applied as a mixture distribution (Clayton, 1978; Hougaard, 2000; Oakes, 1982; Vaupel et al., 1979; Yashin et al., 1995). Other distributions which are sometimes applied for the frailty distribution are the well known normal, the log normal (McGilchrist and Aisbett, 1991), the three parameter distribution (PVF) (Hougaard, 1986), and inverse Gaussian distribution. The effect of different frailty distribution is investigated by Congdon (1995). If the value of the frailty is assumed to be constant within groups, the models are called shared frailty models.

**3. Application to Heart attack data**

The data obtained from the John Wiley & Sons website, ftp://ftp.wiley.com/public/scitech\_med/survival. It may also be obtained from the website for statistical services at the University of Massachusetts at Amherst by going to the data sets link and then the section on survival data http://www.umass.edu/statdata/statdata. The data from the Worcester Heart Attack Study (WHAS) have been provided by Goldberg (2005) of the Department of Cardiology at the University of Massachusetts Medical School. Data have been collected during 13 one-year periods (1975-1988), on all myocardial infarction (MI) patients admitted to hospital in the Worcester, Massachusetts Standard Metropolitan Statistical Area. Event is coded as 1 and censoring is coded as 0. Table 1. describes the subsets of covariates used, with their codes and values.

**Table 1. Description of the covariates obtained from WHAS**

COVARIATES	DESCRIPTION	CODES/UNITS
AGE	Age	years
SEX	Gender	0 = Male, 1 = Female
CPK	Peak Cardiac Enzyme	Int. units
SHO	Cardiogenic shock complications	0 = No, 1 = Yes
CHF	Left Heart Failure Complications	0 = No, 1 = Yes
MIORD	MI Order	0 = First, 1 = Recurrent
YRGRP	Grouped Cohort Year	1 = 1975 & 1978 2 = 1981 & 1984 3 = 1986 & 1988

**4. Model Results**

**Table 2. Parameter Estimates under different Models**

COVARIATES	MODELS					
	COX PH		COX TIME DEPENDENT (AGE)		FRAILITY (GROUPED COHORT YEAR, YRGRP)	
	$\beta$	S.E.	$\beta$	S.E.	$\beta$	S.E.
AGE	.035*	.006	.033*	.006	.032*	.006
SEX	.045	.134	.008	.133	-.013	.133
SHO	1.860*	.217	1.906*	.215	1.822*	.207
CHF	.563*	.143	.564*	.144	.585*	.143
MIORD	.269*	.132	.211	.132	.226	.132

YRGRP	-.205*	.094	-.150	.092	-	-
DEVI- ANCE	2662.699	2659.617	2662.277			

\* P < 0.05 SIGNIFICANT

## 5. Results

The non-linear regression models were fitted using STATA 12 and the results are presented in Table 2. From the table we see that the three covariates namely AGE, SHO, CHF are significantly associated with the survival time under all the model. Among the models, Time dependent Cox model has the lowest level of deviance compared to all other models. It is further noted that all other models have significantly higher deviance compared to Time dependent Cox model.

## 6. Discussion and Conclusion

From the Table 2, it is found that the, SEX is not significant to the survivorship. Controlling for other covariates the risk of death increases by 3.6 % as a patient's age increases by 1 year. Controlling for other covariates, the risk of death of patients with recurrent heart attack is 1.3 times the risk of patients with first heart attack. Using Cox PH model for all covariates the result showed significant p-values except the covariate SEX. When Extended Cox PH model was used for the time dependent covariate, Peak Cardiac Enzyme (CPK). The result is based on CPK as a time dependent covariates, after adjusting all the covariates. We observed that the result of time dependent covariates is better than the results of Cox PH Model. The application of frailty model to the same data assuming individual heterogeneity is also considered. For Frailty model, we see that all the covariates are statistically significant except SEX and MIORD. This leads to the conclusion that the individual heterogeneity in SEX and MIORD differs significantly between the patients. Among the models, Time dependent Cox model has the lowest level of deviance compared to all other models.

## REFERENCE

- [1] Andersen, P.K. (1986). Time dependent covariates and Markov processes In: Modern statistical methods in chronic disease epidemiology. Moolgavkar SH & Prentice RL(eds), Wiley, NY, pp:82-103. | [2] Andersen, P.K. (1991). Survival analysis 1982-1991: The second decade of the proportional hazard regression model. *Stat.med.*10, 1931-1941. | [3] Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N. (1993). Statistical methods based on Counting Processes. Springer, NY. | [4] Aydemir, O., Aydemir, S. and Dirschedl, P. (1999). Analysis of time-dependent covariates in failure time data. *Stat. Med.* 18, 2123 - 2134. | [5] Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika.* 65, 141-151. | [6] Chamberlain, G. (1985). Heterogeneity, Omitted variable Bias, and Duration Dependence. *Longitudinal Analysis of Labour Market Data.* Eds. J. Heckman and B. Singer, Cambridge University Press, pp.3-38. | [7] Collett, D. (2003). Modeling Survival data in Medical Research. Chapman and Hall, London. | [8] Cox, D.R. (1972). Regression model and life tables (with discussion). *J. Royal Stat. Soc. (B)*, 34, 187-220. | [9] Cox, D.R. (1975). Partial Likelihood. *Biometrika.* 62, 269 - 276. | [10] Cox, D.R., and Oakes, D. (1984). Analysis of Survival data, Chapman and Hall, London. | [11] Crowley, J. and Hu, M. (1977). Covariate analysis of heart transplant survival data. *J. Am. Stat. Ass.* 78, 27-36. | [12] Fleming, T.R. and Harrington, D.P. (1991). Counting Processes and Survival Analysis, John Wiley & Sons, NY. | [13] Fisher, L.D. and Lin, D.Y. (1999). Time dependent covariates in the Cox-proportional-hazards regression mode. *Ann. Rev. Pub. Health,* 20, 145-57. | [14] Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis,* 1995; 1: 255-73. | [15] Kalbfleisch, J.D. and Prentice, R.L. (1980). The statistical analysis of failure time data, John Wiley & Sons, NY. | [16] Kleinbaum, D.G. (1996). Survival analysis: A self learning text, Springer-Verlag, NY. | [17] Lee, E.T. (1992). Statistical methods for survival data analysis, 2nd edition, John Wiley & Sons, NY. | [18] Manton, K.G., Stallard, E. and Vaupel, J.W. (1986). Alternative Models of Heterogeneity in Mortality Risks Among the Aged. *J. Am. Stat. Ass.* 81: 635-644. | [19] Mack, J. and Mensah, G. (2004). Atlas of Heart Disease and Stroke, Geneva: WHO | [20] Ponnuraja, C. and Venkatesan, P. (2010). Survival models for exploring tuberculosis clinical trial data- an empirical comparison. *IJST, Vol.* 3, 755-758. | [21] Ponnuraja, C. and Venkatesan, P. (2010). Correlated frailty model: an advantageous approach for covariate analysis of tuberculosis data. *IJST, Vol.* 3, 151-155. | [22] Sharp, I. (1994). Coronary heart disease: are women special? London: National Forum for CHD prevention. | [23] The Scottish Office. (1999). Towards a Healthier Scotland, HMSO, Edinburgh. | [24] Vaupel, J.W., Manton, K. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography.* 16, 439-454. | [25] Vaupel, J.W. and Yashin, A.I. (1985). Heterogeneity's Ruses: Some Surprising Effects of selection on population Dynamics. *American Statistician,* 39: 176-185. | [26] Vallinayagam, V., Prathap, S., and Venkatesan, P. (2011). Comparison of Reliability Models for Life Time data". Recent Trends in Statistics and Computer Application, Eds. Senthamarai Kannan K. and Loganathan A., Manonmaniam Sundaranar University, Tirunelveli, India, pp. 163-170. | [27] Vallinayagam, V., Prathap, S., and Venkatesan, P. (2014). Parametric Regression Models in the Analysis of Breast Cancer Survival Data. *International Journal of Science and Technology Vol.*3, 163-167. | [28] Yashin, A.I. and Iachine, I.A. (1995). How long can human live? Lower bound for biological limit of human longevity calculated from Danish twin data using correlated frailty model. *Mechanism of Ageing and Development* 80: 147-159. | [29] Yashin, A.I. and Iachine, I.A. (1995). Genetic analysis of durations: Correlated frailty model applied to survival of Danish Twins. *Genetic Epidemiology* 12: 529-538. | [30] Yashin, A.I. and Manton, K.G. (1997). Effect of Unobserved and Partially Observed Covariate Processes on System Failure: A Review of Models and Estimation Strategies. *Statistical Science,* 12, 20-34. | [31] Yashin, A.I., Begun, A.Z. and Iachine, I.A. (1999). Genetic Factors of susceptibility to Death: Comparative Analysis of Bivariate Survival Models. *Journal of Epidemiology and Biostatistics, Vol.*4, 53-60. |