Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229-5666
Vol. 11, Issue 2 (2018): 37-56

BAYESIAN SURVIVAL ANALYSIS OF DIABETES MELLITUS PATIENTS: A CASE STUDY IN TIKUR ANBESSA SPECIALIZED HOSPITAL, ADDIS ABABA, ETHIOPIA

Tigabu Hailu Kassa
Department of Applied Statistics, Samara University, Assosa University, Semera, Afar Region, Ethiopia
E Mail: tigabu.hailu2004@gmail.com

Received October 14, 2017
Modified June 11, 2018
Accepted December 05, 2018

Abstract

Diabetes is a complex, chronic illness that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Globally, 415 million (340-536 million) people have diabetes in 2015 with regional prevalence of 8.8% (7.2-11.4%) by 2040 this figure will expect rise to 642 million (521-829 million) with predicted prevalence rate of 10.4% (8.5-13.5%) and more than 22 million people in the African Region; by 2040 this figure will almost double (IDF, 2015).The statistical result of World health organization estimated that the number of cases of diabetics in Ethiopia to be about 796,000 in 2000, and projected that it would increase to about 1,820,000 by the year 2030 (WHO, Diabetes estimates and Projections, 2003).But,according to the report of international diabetes federation atlas in 2017 there were around 2,567,900 [1,094,000-3,795,400] million diabetes cases in Ethiopia in 2017 (IDF atlas, 2017). The general objective of the study is to identify the determinant risk factors for the survival of Diabetic mellitus patients. From 2474 patients a sample of 451 diabetes patients administered the treatment in Tikur Anbessa specialized Hospital between September 11/ 2008-9/5/2014 were included in the study. The data were analyzed using classical and Bayesian Accelerated failure time model because of the failure in proportional hazard assumption. Bayesian Accelerated failure time model was better model than Classical Accelerated failure time model because it contains smaller AIC. Descriptive statistics and the Kaplan-Meier survival curves were used to estimate and compare the survival time of diabetes patients among different categorical characteristics of the patients. From the result, the survival time until death is significantly related to the age category, BMI, types of diabetic disease, alcohol use, diabetic complication, blood pressure, cholesterol level, family history of diabetic, fasting blood sugar, comorbidity, density lipoprotein, triglyceride level and smoking habit. The patients should keep their normal body weight and change their life style such as smoking habit, alcohol consumption, and take care of on their lipid cholesterol level.

Key Words: TIKUR Anbessa, Diabetes Mellitus, Accelerated Failure Time, Bayesian Analysis, Winbugs.

1. Introduction

Insulin is a sugar regulator hormone produced in the pancreas that regulates blood sugar, or glucose (Alvin C., 2012). A person already known who is with diabetes does not absorb glucose in his or her body properly, because of this glucose remains

circulating in the blood which leads to damaging body tissues over time. This damage can lead to different disabling and plenty life threatening health complications such as heart attack, stroke, kidney failure, leg amputation, vision loss and nerve damage. Diabetes and its complications bring about substantial economic loss in people, health systems and national economies through direct medical costs and loss of work and wages (Alvin C., 2012). During the last decade, diabetes mellitus has emanating and emerged as an important clinical, public health and economic problem throughout the world including the developing counties.

2. Methodology

2.1 Study Area and Data

For this study, retrospective cohort data on adult diabetes patient's history card with follow up from Tikur Anbessa specialized hospital, Addis Ababa was collected which is a teaching hospital and Addis Ababa university of medicine and health. The researcher would use a secondary data and simple random sampling technique to select a sample data from patient's under the follow up period of from September 10/8/2008 up to August September 1/8/2014 (six year data). All patients with age greater than 18 were included in the study. But, the study could exclude those patients who were under 18 years old and diagnosed with Gestational diabetes mellitus.

2.2 Study Population

A total of 2474 diabetic patients were on active follow up. All diabetic patients greater than or equal to 18 years old and placed under treatments that have started the follow up between September 11/2008 and August 10/2014 with six years data with the exception of Gestational Diabetic mellitus were included. Patents' follow up time was one, two, three and six months gab according to the order of the doctor.

2.3 Sample Size Determination

According to the sample size determination formula (Cochran, W. 3rd Edt. 1977, page 86-89)

$$n = \frac{\frac{Z^2\alpha_{/2}pq}{d^2}}{1+\frac{1}{N}\left(\frac{Z^2\alpha_{/2}pq}{d^2}-1\right)} = \frac{n_0}{1+\frac{1}{N}(n_0-1)}$$

$\alpha = 5\%$is a level of significance, the estimated proportion of death due to DM disease was p=0.07 (According to Bekele, T., Hymete, A., Tadesse, M. and Mekonnen, Y. 2008), the researcher have been used a maximum error of 0.0213 and 5% significance level. The sample size would be 451.

2.4 Study Variable

2.4.1 Dependent Variable

The dependent variable also called outcome variable, of this study is the length of time in month of a given patient until the occurrence of event which is death happens T_i , Survival time to event (days since acceptance).

2.4.2 Independent Variable

There were many covariates that might produce variations on the survival of DM patients. Covariates may be quantitative or qualitative. Such as, Sex, Baseline age, Height, Initial weight, Diabetic type, Smoking habit, Region, Diabetic complications, Alcohol use, Body mass index, Blood pressure(systolic & diastolic blood pressure), Cholesterol level, Triglyceride, Protein level (high density lipoprotein and low density lipoprotein), baseline Fasting blood sugar, Family History, Comorbidity, etc.

2.5 Data and Model

Survival analysis is the phrase used to describe the analysis of data in the form of a well-defined time origin until the occurrence of some particular event or end point & for which the response variable of interest is the length of time until a certain event occurs. If the end point is the death of a patient or time to cure from a certain disease for a patients in given hospital, the resulting data are indicates and refers to survival times. To determine the survival time denoted by t_i, three basic elements are needed such as, Starting time (an unambiguous time origin or the beginning of the study), A measurement scale for the passage of time or a time scale like real time (days, years), An ending event of interest (a definition of when the endpoint occurs) (Lee, E. T., & Wang, J. 2003).

Most of the time, we cannot fully observe this random variableT_i but only observe some boundaries for this time. This is called censoring. The data that are collected over a finite period of time and consequently the "time to event" may not be observed for all the individuals in our study population or sample. This result in what is called Censored data. That is the "time to event" for those individuals who have not experiencing the event under study is censored (by the end of the study). It is also common that the amounts of follow-up for the individuals in the sample vary from subject to subject(Lee, E. T., & Wang, J. 2003).The combination of censoring and differential follow-up creates some unusual difficulties in the analysis of such data that cannot be handled properly by the standard statistical methods. Because of this, a new research area in statistics has emerged which is called Survival Analysis or Censored Survival Analysis (Lee, E. T., & Wang, J. 2003).The researcher considers only right censoring – events would necessarily take place after the follow-up period.

2.5.1 Semi-Parametric Survival Model

Cox's semi-parametric model is widely used in survival analysis to model the effect of covariates on hazard rates. The model assumes that the effect of explanatory variables is to multiply the hazard by some constant. In this model, the conditional hazard of an individual, given the covariates values,X_1, X_2, X_3,…., X_p is defined

$$\lambda(t|X) = \lambda_0(t)e^{\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p} = \lambda_0(t)e^{\beta' \mathbf{x}}$$

Where, $\lambda_0(t)$ is the baseline hazard which indicates the hazard of patients when the independent variables are zero.Independent variables in Cox proportional hazard model areentering the model linearly in the exponential scale. Many model assumption checking procedures are based on quantities known as residuals such as Schoenfeld residuals and scaled Schoenfeld residuals.

2.6 Accelerated Failure Time Model

2.6.1 Classical Accelerated Failure Time Model

When the possibility of proportionality assumption of the Cox proportional hazard model is not valid, the model cannot be used in modeling length of time to the occurrence of an event rather some parametric survival models and approaches are appropriate. In a parametric survival model, the distribution of the response variable, length of time to the happening of an event is specified in term of a finite number of unknown parameters. One of the well-known parametric models is accelerated failure time (AFT) model.In which the length of time is to happening of the event is assumed to be a function of explanatory variables. In this model the covariates are acting multiplicatively effect on time and so affect the 'rate of passage' to the event. This model also assumes that the relationship of logarithm of length of time to the happening of the event (T) and x is linear related like as follows

$$\log(t_i) = \boldsymbol{X\beta} + \sigma\varepsilon_i$$

$$\varepsilon_i = \frac{\left(\log(t_i) - \beta_0 - \beta_1 \mathrm{x_{i1}} - \beta_2 \mathrm{x_{i2}} - \beta_3 \mathrm{x_{i3}} - \cdots - \beta_p \mathrm{x_{ip}}\right)}{\sigma}$$

Let $\varepsilon \sim \mathrm{F_\varepsilon}(.|\sigma)$, such that $\mathrm{F_\varepsilon}(.|\sigma)$ is the known cumulative probability distribution associated with probability density function $\mathrm{f_\varepsilon}(.|\sigma)$ with scale parameter σ. The survival and hazard function of ε are $\mathrm{S_\varepsilon}(.|\sigma) = 1 - \mathrm{F_\varepsilon}(.|\sigma)$ and $\mathrm{h_\varepsilon} = \frac{\mathrm{f_\varepsilon}}{\mathrm{S_\varepsilon}}$ respectively. Where,$\boldsymbol{\beta}_{\mathbf{px1}}$ are vectors of coefficients and ε_i is a residual error term with a specified probability distribution. The commonly used distributions for the random error termεare extreme, logistic and normal; these three distributions are, respectively, log-transformation of Weibull, the log-logistic and the log-normal distributions. These distributions are appropriate parametric distributions for analyzing the length of time to event data (Christensen R, 2011).

The survivor function

$$S(\mathrm{t_i}) = p(\mathrm{T_i} \geq \mathrm{t_i}) = p(log\mathrm{T_i} \geq log\mathrm{t_i}),\text{so}$$

$$S(\mathrm{t_i}/X_i) = p(\varepsilon_i \geq \left(\log(t_i) - \beta_0 - \beta_1 \mathrm{x_{i1}} - \beta_2 \mathrm{x_{i2}} - \beta_3 \mathrm{x_{i3}} - \cdots - \beta_p \mathrm{x_{ip}}\right)/\sigma)$$

As an example $\varepsilon_i$follows a logistic probability distribution with a density & survival function as follows

$$f(\varepsilon) = \frac{e^{\varepsilon}}{(1+e^{\varepsilon})^2} \text{ and } S(\varepsilon) = \frac{1}{(1+e^{\varepsilon})}$$

Then, substitute another expression of ε_i in the above survival function.

$$\mathrm{S(t_i|x_i)} = \frac{1}{\left(1 + e^{(\log(t_i) - \beta_0 - \beta_1 \mathrm{x_{i1}} - \beta_2 \mathrm{x_{i2}} - \beta_3 \mathrm{x_{i3}} - \cdots - \beta_p \mathrm{x_{ip}})/\sigma}\right)}$$

Suppose we observe"n" independent vectors of the form denoted by (T_i, δ_i, X_i), where T_i is the length of time to the occurrence of an event, δ_i is indicator variable telling us whether δ_i is uncensored (unobserved) or censored (observed) and X_i is the vector of explanatory variables that are expect to affect the survival of diabetes patients.

$$\delta_i = \begin{cases} 0\ censoring\ observation \\ 1\ event\ or\ dead\ (failure) \end{cases}$$

$$T_i \sim AFT(F_\varepsilon, \boldsymbol{\theta}|\mathrm{x_i})$$

Let the precision parameter be denoted by$\tau = \frac{1}{\sqrt{\sigma}}$, then

$$S(\mathrm{t_i}/X_i, \boldsymbol{\theta}) = 1 - F_\varepsilon\left[(log(t_i) - \boldsymbol{X_i\beta})\sqrt{\tau}\right]$$

$$S(\mathrm{t_i}/X_i,\boldsymbol{\theta}) = \frac{\sqrt{\tau}}{t_i} f_\varepsilon\left[(log(t_i) - \boldsymbol{X_i\beta})\sqrt{\tau}\right]$$

$$h(\mathrm{t_i}/X_i,\boldsymbol{\theta}) = \frac{\sqrt{\tau}}{t_i} h_\varepsilon\left[(log(t_i) - \boldsymbol{X_i\beta})\sqrt{\tau}\right]$$

Accelerated failure time model assumes that the effects of the explanatory variables are to decelerate or accelerate the life of patients. It is possible to show that the coefficients from Cox proportional hazard and accelerated failure time model are linked in the following manner i.e $\hat{\beta}_{AFT} = \frac{-\hat{\beta}_{PH}}{\hat{p}}$ and $\hat{\beta}_{PH} = \frac{-\hat{\beta}_{AFT}}{\hat{\sigma}}$. Thus, the estimated coefficients are equivalent up to a scale factor equal to $\hat{\sigma}$, where, $\hat{p} = \frac{1}{\hat{\sigma}}$. In accelerated failure time models, the sign of the regression coefficient shows how the explanatory variables affect the log survival times. Thus, a positivevalue ofregression coefficient increases the log survival time and, hence, the expected duration. A negative value of regression coefficient decreases the log survival time and, hence, the expected duration.

2.6.2 Bayesian Accelerated Failure Time Model

The researcher can use this type of inference to estimate the parameter by considered them as a random variable and by specifying their own probability distribution to describe the uncertainty of parameters called prior distribution.

The likelihood function for the set of unknown vectors of parameters denoted by,θ

- The likelihood function of the set of unknown parameters, $\boldsymbol{\theta}$ in the this types of censoring (right censoring) can be written as

$$l(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=0)} * S(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=1)}\right]$$

Log-likelihood would be as follows

$$l(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=0)} * S(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=1)}\right]$$

$$l(\boldsymbol{\theta}) = log\left\{\prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=0)} * S(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=1)}\right]\right\}$$

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} log\{\left[f(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=0)} + S(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=1)}\right]\}$$

Where, $f(t_i|x_i,\boldsymbol{\theta})$ and $S(t_i|x_i,\boldsymbol{\theta})$are the density and survival distributions respectively (Ganjali, M., & Baghfalaki, T. 2012, Ibrahim, J. G., Chen, M. H., and Sinha, D. 2001). In these models, when both of the regression coefficients β and scale parameters, σ are unknown, no joint conjugate prior is available. A typical joint prior specification can be expressed as a product of a multivariate normal (for parameter$\beta|\sigma^2$) and an inverse gamma prior (for σ^2), that is

$\beta|\sigma^2 \sim N_p(\mu_0, V_0\sigma^2{}_0)$, $\sigma^2 \sim IG(a,b)$

Let combining the likelihood function above with the prior probability distribution on (β, σ^2) and the full conditional probability distributions for these unknown parameters is as follows

$$\prod(\beta|\sigma^2,t,x) \propto \prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=0)} * S(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=1)}\right] * \prod(\beta|\sigma^2)$$

$$\prod(\sigma^2|\beta,t,x) \propto \prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=0)} * S(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=1)}\right] * \prod(\beta|\sigma^2) \ * \prod(\sigma^2)$$

The posterior probability distribution for the model specification above does not have closed form solution for the parameters. For these models, Markov chain Monte Carlo (MCMC)simulation type like the Gibbs sampler can be implemented using the WinBUGS software (Ganjali, M., & Baghfalaki, T. 2012, Ibrahim, J. G., Chen, M. H., and Sinha, D. 2001).

To high light & derived final form of baseline models for Posterior probability distribution of the three parametric probability distribution in the above equations, as follows. A straightforward way to state the Weibull accelerated failure time models is to let $T_i \sim Weibull(\lambda_i, \alpha)$ and $\log(\lambda_i) = -x'\beta$. And a joint probability distribution specification is to take $\beta \sim N(\mu_0, V_0)$ and $\alpha \sim IG(a,b)$

IG $= \frac{b^a x^{-a-1} e^{\frac{-b}{x}}}{\gamma(a)}$

The posterior probability distribution for the model specification above does not have final solution or closed form solutions for the unknown parameters. To apply the Bayesian analysis, Markov chain Monte Carlo simulation techniques can be used to sample a data from the joint posterior distribution of these models. One of the special Markov chain Monte Carlosimulation typeis the Gibbs sampler, which need only the specification of the conditional posterior probability distribution for each parameter. In situations where those probability distributions are simple to sample from the approach is easily implemented. The final form by combining the likelihood function with the prior distribution on (β, σ^2) and the full conditional probability distributions for unknown parameters in log-logistic probability and log-normal probability models are given by:

i. Density and Survival distribution of log-logistic distribution

$$T \sim LL(X'\beta, \sqrt{\tau}), \quad \text{(Christensen, 2011)}$$

$$f(t|x,\beta,\tau) = \frac{\sqrt{\tau} e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}}{\left(1+e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}\right)^2}$$

$$S(t|x,\beta,\tau) = 1 - F_{\varepsilon}[(log(t) - X'\beta)\tau] = \frac{1}{1+e^{\{(log(t)-X'\beta)\sqrt{\tau}\}}}$$

$$= \left[1+e^{\{(log(t)-X'\beta)\sqrt{\tau}\}}\right]^{-1}$$

- Posterior Distribution for Log-logistic

Posterior

$$\pi(\beta|\sigma^2,t,x) = \frac{\prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{\delta_i} * S(t_i|x_i,\boldsymbol{\theta})^{1-\delta_i}\right]\pi(\beta|\sigma^2)}{\int_{-\infty}^{\infty}\prod_{i=1}^{n}[f(t_i|x_i,\boldsymbol{\theta})^{\delta_i} * S(t_i|x_i,\boldsymbol{\theta})^{1-\delta_i}]\pi(\beta|\sigma^2)d\,\beta}$$

$$\beta \sim \mathrm{N}(\mu_0, \sigma_0)$$

$$\pi(\beta|\sigma^2,\mathrm{t},\mathrm{x}) = \frac{\prod_{\mathrm{i}=1}^{\mathrm{n}}\left[\left(\frac{\sqrt{\tau}e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}}{\left(1+e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}\right)^2}\right)^{\delta_\mathrm{i}} * \left(\frac{1}{1+e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}}\right)^{1-\delta_\mathrm{i}}\right] * \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}\left(\frac{\beta-\mu_0}{\sigma_0}\right)}}{\int_{-\infty}^{\infty}\prod_{\mathrm{i}=1}^{\mathrm{n}}\left[\left(\frac{\sqrt{\tau}e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}}{\left(1+e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}\right)^2}\right)^{\delta_\mathrm{i}} * \left(\frac{1}{1+e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}}\right)^{1-\delta_\mathrm{i}}\right] * \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}\left(\frac{\beta-\mu_0}{\sigma_0}\right)}\mathrm{d}\beta}$$

Full conditional probability distributions for the vectors of unknown parameters in log-logistic probability

$$\prod(\beta|\sigma^2,t,x) \propto \prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=0)} * S(t_i|x_i,\boldsymbol{\theta})^{I(\delta_i=1)}\right] * \prod(\beta|\sigma^2)$$

$$\pi(\beta|\sigma^2,t,\mathrm{x}) \propto \prod_{\mathrm{i}=1}^{\mathrm{n}}\left[\left(\frac{\sqrt{\tau}e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}}{\left(1+e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}\right)^2}\right)^{\delta_\mathrm{i}} * \left(\frac{1}{1+e^{\{(\log(t)-X'\beta)\sqrt{\tau}\}}}\right)^{1-\delta_\mathrm{i}}\right] * \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}\left(\frac{\beta-\mu_0}{\sigma_0}\right)}$$

ii. Density and Survival distribution of lognormal distribution

$T \sim LN\left(X'\beta, \frac{1}{\tau}\right)$ (Christensen, 2011)

$$f(t|x,\beta,\tau) = \frac{1}{\sqrt{2\pi}}\frac{\sqrt{\tau}}{t}e^{\left\{-\frac{\tau}{2}[log(t)-X'\beta]^2\right\}}$$

$$S(t) = 1-\Phi\left(\frac{logt-X'\beta}{\sigma}\right) t>0$$

- Posterior distribution for lognormal distribution

$$X \sim IG = \frac{b^a x^{-a-1} e^{\frac{-b}{x}}}{\gamma(a)}$$

Then the prior for

$$\sigma^2 \sim \frac{b^a(\sigma^2)^{-a-1}e^{\frac{-b}{(\sigma^2)}}}{\gamma(a)}$$

$$Posterior = \pi(\beta|\sigma^2,t,x) = \frac{\prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{\delta_i} * S(t_i|x_i,\boldsymbol{\theta})^{1-\delta_i}\right]\pi(\beta|\sigma^2)\pi(\sigma^2)}{\int_{-\infty}^{\infty}\prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{\delta_i} * S(t_i|x_i,\boldsymbol{\theta})^{1-\delta_i}\right]\pi(\beta|\sigma^2)\pi(\sigma^2)d\beta d\sigma^2}$$

$$\pi(\beta|\sigma^2,\mathrm{t},\mathrm{x}) = (\mathrm{A}*\mathrm{B})/\int_{-\infty}^{+\infty} A*B\,d\beta.d\sigma^2$$

where

$$\mathrm{A} = \prod_{\mathrm{i}=1}^{\mathrm{n}}\left[\left(\frac{1}{\sqrt{2\pi}}\frac{\sqrt{\tau}}{t}e^{\left\{-\frac{\tau}{2}[\log(t)-X'\beta]^2\right\}}\right)^{\delta_\mathrm{i}} * (1-\Phi(\frac{logt-X'\beta}{\sigma}))^{1-\delta_\mathrm{i}}\right]$$

$$\mathrm{B} = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}\left(\frac{\beta-\mu_0}{\sigma_0}\right)} \frac{b^a(\sigma^2)^{-a-1}e^{\frac{-b}{(\sigma^2)}}}{\Upsilon(a)}$$

Full conditional probability distributions for unknown parameters in log-normal probability distribution

$$\pi(\sigma^2|\beta,t,x) \propto \prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{\delta_i} * S(t_i|x_i,\boldsymbol{\theta})^{1-\delta_i}\right] * \pi(\beta|\sigma^2) * \pi(\sigma^2)$$

$$\pi(\sigma^2|\beta,\mathrm{t},\mathrm{x}) \propto \prod_{\mathrm{i}=1}^{\mathrm{n}}\left[\left(\frac{1}{\sqrt{2\pi}}\frac{\sqrt{\tau}}{t}e^{\left\{-\frac{\tau}{2}[\log(t)-X'\beta]^2\right\}}\right)^{\delta_i} * \left(1-\Phi\left(\frac{logt-X'\beta}{\sigma}\right)\right)^{1-\delta_i}\right]$$

$$* \frac{1}{\sqrt{2\pi\sigma}}e^{\frac{-1}{2}\left(\frac{\beta-\mu_0}{\sigma_0}\right)}\frac{b^a(\sigma^2)^{-a-1}e^{\frac{-b}{(\sigma^2)}}}{\Upsilon(a)}$$

Where, $\pi(\beta|\sigma^2)$ and $\pi(\sigma^2)$ are prior distributions for β and σ^2 in the log-logistic probability and lognormal probability model. Also, the full conditional probability distributions for Weibull probability model are given in the same notations, where σ^2 replaced by α and $\pi(\beta)$ and $\pi(\alpha)$ independent prior distributions **(**Ganjali, M. and Baghfalaki, T. 2012, Christensen et al. 2011).

iii. Density and Survival distribution of Weibull distribution

$$T\sim Weib(\sqrt{\tau}, e^{-X'\beta\sqrt{\tau}}), \text{ (Christensen R., 2011)}$$

$$S(t) = \exp\left(-\exp\left[(\log(t)-X'\beta)\sqrt{\tau}\right]\right)$$

$$f(t|x,\beta,\tau) = \sqrt{\tau}e^{\left[(log(t)-X'\beta)\sqrt{\tau}\right]}exp\left(-exp\left[(log(t)-X'\beta)\sqrt{\tau}\right]\right)$$

Where, $\sigma = \frac{1}{\alpha}$ and $\lambda = e^{-\frac{\mu}{\sigma}}$

$$X\sim IG = \frac{b^a x^{-a-1}e^{\frac{-b}{x}}}{\gamma(a)}$$

Then the prior for

$$\alpha\sim\frac{b^a(\alpha)^{-a-1}e^{\frac{-b}{(\alpha)}}}{\gamma(a)}$$

- Posterior distribution for Weibull distribution

$$Posterior = \pi(\beta|\sigma^2,t,x) = \frac{\prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{\delta_i} * S(t_i|x_i,\boldsymbol{\theta})^{1-\delta_i}\right]\pi(\beta|\alpha)\pi(\alpha)}{\int_{-\infty}^{\infty}\prod_{i=1}^{n}\left[f(t_i|x_i,\boldsymbol{\theta})^{\delta_i} * S(t_i|x_i,\boldsymbol{\theta})^{1-\delta_i}\right]\pi(\beta|\alpha)\pi(\alpha)d\beta d\alpha}$$

$$\pi(\beta|\alpha,\mathrm{t},\mathrm{x}) = \frac{A}{B} \qquad where$$

$$\mathrm{A} = \prod_{\mathrm{i}=1}^{\mathrm{n}}\left[\left(\sqrt{\tau}e^{\left[(\log(t)-X'\beta)\sqrt{\tau}\right]}\exp\left(-\exp\left[(\log(t)-X'\beta)\sqrt{\tau}\right]\right)\right)^{\delta_i}\right.$$

$$\left. * \left(\exp\left(-\exp\left[(\log(t)-X'\beta)\sqrt{\tau}\right]\right)\right)^{1-\delta_i}\right].$$

$$\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{-1}{2}\left(\frac{\beta-\mu_0}{\sigma_0}\right)}\frac{b^a(\alpha)^{-a-1}e^{\frac{-b}{(\alpha)}}}{\gamma(a)}$$

and

$$B = \int_{-\infty}^{\infty} \prod_{i=1}^{n} \prod_{i=1}^{n} \Big[\left(\sqrt{\tau} e^{\,[(\log(t) - X'\beta)\sqrt{\tau}]}\right) \exp\left(-\exp\left[(\log(t) - X'\beta)\sqrt{\tau}\right]\right)^{\delta_i} * \left(\exp\left(-\exp\left[(\log(t) - X'\beta)\sqrt{\tau}\right]\right)\right)^{1-\delta_i} \Big] . \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}\left(\frac{\beta-\mu_0}{\sigma_0}\right)} \frac{b^a(\alpha)^{-a-1} e^{\frac{-b}{(\alpha)}}}{\gamma(a)} d\,\beta d\alpha$$

The denominator of the above three posterior probability distributions is a constant called normalizing constant it makes to integrating to one to the posterior probability. Full conditional distributions for unknown parameters in Weibull distribution

$$\pi(\beta|\alpha, t, x) \propto \prod_{i=1}^{n} \Big[\left(\sqrt{\tau} e^{\,[(\log(t) - X'\beta)\sqrt{\tau}]}\right) \exp\left(-\exp\left[(\log(t) - X'\beta)\sqrt{\tau}\right]\right)^{\delta_i} * \left(\exp\left(-\exp\left[(\log(t) - X'\beta)\sqrt{\tau}\right]\right)\right)^{1-\delta_i} \Big] \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}\left(\frac{\beta-\mu_0}{\sigma_0}\right)} \frac{b^a(\alpha)^{-a-1} e^{\frac{-b}{(\alpha)}}}{\gamma(a)}$$

Even when an analytical solution exists, it is often easier to use simulations (Christensen et al, 2011). For these models, MCMC type like the Gibbs sampler simulation can be apply by using the WinBUGS software (Spiegelhalter et al., 2003, Ganjali, M. & Baghfalaki, T., 2012, Christensen et al, 2011).

2.6.2.1 Simulation Convergence Test Statistic

The Markov Chain Monte Carlo simulation convergence diagnostics tests are widely used to determine how many initial "burn-in" iterations should be discarded from the output of a MCMC sampler in the hope that the remaining samples are representative of the target posterior probability distribution of interest. The best way to do this is choosing the number of burn-in iterations "*r"* by applying convergence diagnostics to one or more plot chains, and then basing estimation and inference on a separate long chain from which the first "r" iterations have been discarded (Ntzoufras I., 2009).

Tests used for checking convergence of a Bayesian analyses were as follows: Time series plot: Time series plot of the different independent initial values of the chains should be mix together or overlapped. Kernel Density plot: The plots for the parameters of predicator variables should be resemble the curves of normal distribution if so the simulated parameter values will be converged. Gelman-Rubin Statistic, Bayesian Gelman - Rubin diagnostic compares the values of within-chain variability and the between-chain variability of the MCMC simulated values and if the ratio (converges approximately to one or if lines for each chain on the Bayesian Gelman-Rubin are nearly together, this is an indication for the convergence of statistics (Spiegelhalter, 2004).

Gelman-Rubin Statistics = $\sqrt{R} = \sqrt{\frac{Var(\hat{\beta})}{W}}$

V ($\hat{\beta}$) = overestimates variance (variations between the chains), W = underestimates total posterior variance (variation within the chains) The ratio R of W and V ($\hat{\beta}$) close to 1 to be called convergence of Gelman-Rubin statistic was achieved.

☞ **Autocorrelation plot**

The simulated values of Markov Chain Monte Carlo expect that the "k^{th}" lag autocorrelation values must smaller as "k" increases, which mean that the "2^{nd}" and "50^{th}" draws should be less correlated than our "2^{nd}" and "4^{th}" draws. If autocorrelation function is still relatively high for higher values of "k", this shows that there is a high degree of correlation between our draws and slow mixing. The degree of autocorrelation can be quantified using the autocorrelation function as follows (Christensen R, 2011).

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{\sqrt{var(X_t)}\sqrt{var(X_{t+k})}} = \frac{E[(X_t-\theta)(X_{t+k}-\theta)]}{\sqrt{E[(X_t-\theta)^2]}\sqrt{E[(X_{t+k}-\theta)^2]}}$$

3. Result of the Study

3.1 Results of the Descriptive Statistics

Out of 451 patients included in the study 380 (84.3%) and 71 (15.7%) were right-censored (dropout, transferred, loss and alive till the study period)and dead patients respectively. From 451 patients in the study 237(52.5%) of the patients were females and 214(47.5%) were males. Out of 451 patients included in the study 380 (84.3%) and 71 (15.7%) were censored and dead patients respectively. From 237 female patients 192 (81%) and 45(19%) of which were censored and death respectively and out of 214 male patients 188 (87.9%) and 26(12.1%) were censored and death respectively.A summary of the data for each level of variables is provided intable 1 below.

Variable	Category	Status of patients		
		Censored (%)	Death (%)	Number in the sample (%)
Age	=<30 30<Age=<45 45<Age=<60 >60	48(96%) 102(94.4%) 175(88.4%) 55(57.9%)	2(4%) 6(5.6%) 23(11.6%) 40(42.1%)	50(11.1%) 108(23.9%) 198(43.9%) 95(21.1%)
Sex	Female Male	192(81%) 188(87.9%)	45(19%) 26(12.1%)	237(52.5%) 214(47.5%)

BMI	Normal(18.5-24.5) Under weight(<18.5) Over W (24.5-29.9) Obese(>=30)	181(93.3%) 15(83.3%) 131(85.6%) 53(61.6%)	6(6.7%) 3(16.7%) 22(14.4%) 33(38.4%)	194(43%) 18(4%) 153(33.9%) 86(19.1%)
Types of DM	Type 1 Type 2	10(95.2%) 380(82.5%)	53(4.8%) 8(17.5%)	63(14%) 388(86%)
Alcohol Use	Yes No	15(26.3%) 365(92.6%)	42(73.7%) 29(7.45)	57(12.6%) 394(87.4%)
DM compl.	No complication D_Kidney failure D_eye disease Neuropathy Heart (CVD) Diabetic foot ulcer More than 2 comp.	241(98.4% 31(70.5%) 33(82.5%) 21(70%) 32(62.7%) 15(65.2%) 7(38.9%)	4(1.6%) 13(29.5%) 7(17.5%) 9(30%) 19(37.3%) 8(34.8%) 11(61.1%)	245(54.3%) 44(9.8%) 40(8.9%) 30(6.7%) 51(11.3%) 23(5.1%) 18(4%)
SBP	Normal (<120) Pre-HTN(120-139) Stage 1 (140-159) Stage 2 (>=160)	77(96.2%) 179(90.4%) 69(75%) 55(67.9%)	3(3.8%) 19(9.6%) 23(25%) 26(32.1%)	80(17.7%) 198(43.9%) 92(20.4%) 81(18%)
Cholesterol level	Normal <200 High >=200	339(95%) 41(43.6%)	18(5%) 53(56.4%)	357(79.25) 94(20.8%)
Family history	Positive Negative	15(23.4%) 365(94.3%)	49(76.6%) 22(5.7%)	64(14.2%) 387(85.8%)
Comorbidity	No HTN Dyslipidemia Obesity Others	184(95.8%) 69(22.5%) 32(84.2%) 22(56.4%) 73(77.7%)	8(4.2%) 20(22.5%) 6(15.8%) 17(43.6%) 21(22.3%)	192(42.6%) 89(19.7%) 38(8.4%) 23(8.6%) 94(20.8%)
HDL levels	Normal = above 60 Borderline 40-59.9 Undesirable < 39.9	93(91.2%) 163(84.9%) 124(79%)	9(8.8%) 29(15.1%) 33(21%)	102(22.6%) 192(42.6%) 157(34.8%)
LDL level	Normal <100 Above normal 100-12.9 Borderline 130-159.9 High >=160	115(95%) 103(96.3%) 103(82.4%) 59(62.8%)	6(8%) 8(7.2%) 22(17.6%) 35(37.2%)	121(26.8%) 111(24.8%) 125(27.7%) 94(20.8%)
Triglyceride	Normal < 150 High >=150	184(90.6%) 211(79%)	19(9.4%) 52(21%)	203(45%) 281(55%)
Smoking habit	Non-smoker Smoker	367(92.9%) 13(23.2%)	28(7.1%) 43(76.8%)	395(87.6%) 56(12.4%)

Table 1: Descriptive statistics for each level of the covariates

Comparisons of Survival Curves

In order to investigate if there is significant difference between the survivals of a patient between categories of covariates, Kaplan-Meier survivor estimates all significant covariates in the log-rank test.

		Log-Rank			Breslow		
Variable	Mean survival	Chi-sq	df	Sig	Chi-sq	Df	Sig
Age C		125.843	3	0.000	115.393	3	0.000
=<30	78.689						
30<age=<45	77.927						
45<age=<60	70.524						
>60	44.199						
Sex		3.17	1	0.075	0.995	1	0.318
Female	68.70						
Male	71.236						
BMI		52.094	3	0.000	44.899	3	0.000
Normal	74.825						
Underweight	66.747						
Overweight	71.231						
Obese	55.27						
Types_ DM		10.712	1	0.001	10.98	1	0.001
Type 1	68.781						
Type 2	78.781						
Alcohol		205.91	1	0.000	156.29	1	0.000
Yes	60.25						
No	75.24						
DM_com		120.068	5	0.000	105.19	5	0.000
No com	79.20						
D_Neph	58.15						
D_retino	67.59						
D-Hear	52.15						
D_Neurophat.	52.79						
Others	49.96						
SBP		44.73	3	0.000	36.054	3	0.000
Normal	78.23						
Pre-HTN	74.24						
Stage 1 HTN	63.15						
Stage 2 HTN	57.07						
DBP		79.82	3	0.000	60.31	3	0.000
Normal	77.15						
Pre-HTN	74.29						
Stage 1HTN	64.06						
Stage 2 HTN	58.27						
Cholesterol		169.59	1	0.000	128.48	1	0.000
Normal	76.06						
High	49.24						
F_His		244.59	1	0.000	174.27	1	0.000
Positive	46.24						
Negative	74.16						
Comorbidity		54.71	4	0.000	47.01	4	0.000
No com	78.14						
HTN	65.14						
Dyslipidemia	68.22						

Obesity Others	61.26 63.29						
HDL_ca Normal Borderline Undesirable	 73.05 72.10 64.10	8.13	2	0.037	6.567	2	0.039
LDL_C Normal Above normal Borderline High	 77.24 76.18 66.25 56.12	64.73	3	0.000	73.613	3	0.000
Triglyceride Normal High	 75.04 66.33	13.52	1	0.000	19.13	1	0.000
Smoking Hab Non-smoker Smoker	 74.12 49.26	368.61	1	0.000	344.25	1	0.000

Table 2: Comparisons of Survival Curves on different level of covariates using numerical hypothesis testing

The researcher used the graph of Kaplan-Meier survival function and the numerical comparison like log-rank (mantel-cox) and Breslow (generalized Wilcoxon) to compare the survival experience of the diabetes patients from the different groups of the covariates. From the groups of the covariates included in the study with the above curve had a highest survival than the category with lower Kaplan-Meier curve. Because of this patients with age category less than or equal to 30 had more survival time than the other age group of patients.

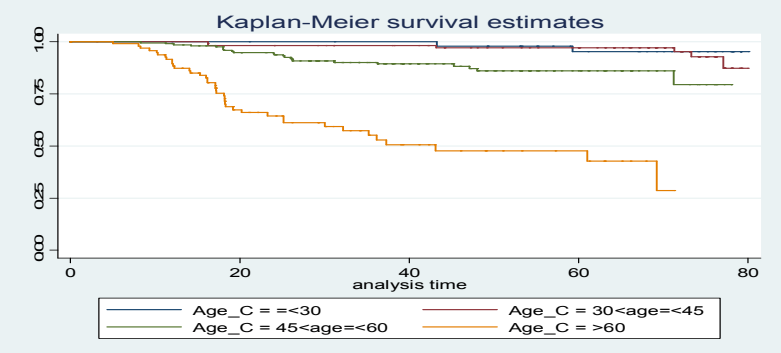

Figure 1:Kaplan-Meier Curve for age Category

The difference in survival were supported by numerical Statistical hypothesis tests, since log-rank (p-value=0.000 and Chi-square=125.84, 3 df) test and Breslow (p-value=0.000 & Chi-square 115.4, 3 df) output revealed in Table 2 shows that there is significant difference between survival time of the different age groups of patients. The survival variation of patients on the other level covariates is present in Table 2.

3.2 Results of the Proportional Model

3.2.1 Proportional Hazard Model Assumption Checking

The proportionality of Cox proportional hazard model can be tested using rho statistic, p-value and Scaled Schoenfeld residuals. The large value of rho showed that strong correlation between residuals and time because of this there is the existence of systematic patter on the graph this showed that proportional hazard assumption is not satisfied. The p-value of rho statistic is less than 5% for a given covariate indicates the rejection of null hypothesis of the proportionality of cox proportional hazard model.

Variable	Rho	Chi2	Df	Prob>chi2
SBP	0.22124	4.23	1	0.0397
Comorbidity	-0.25565	4.39	1	0.0318
LDL_ca	-0.26563	4.72	1	0.0298
Trig	-0.34923	8.99	1	0.0027
Global test		43.78	18	0.0001

Table 3: Test of proportional-hazards for covariates violate the assumption only

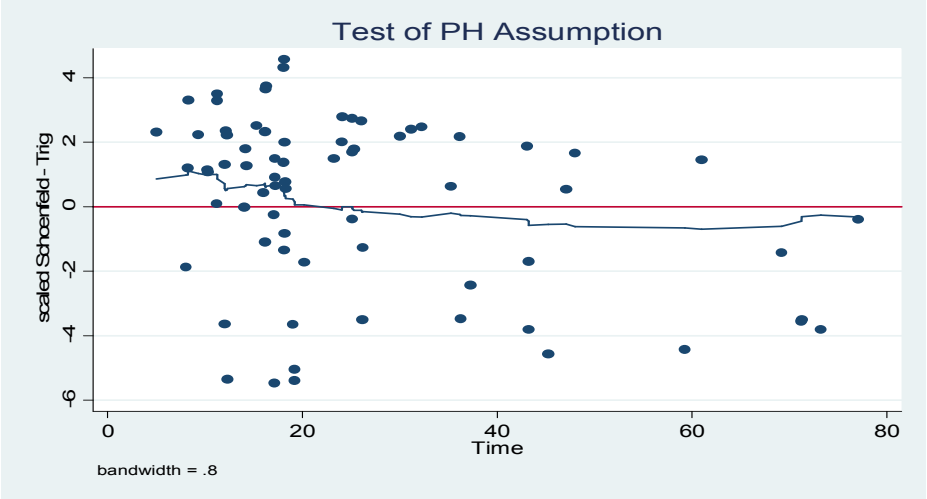

Figure 2: Scaled Schoenfeld residuals of covariate triglyceride

The reference line and Lowess curve never cross to each other and is approximately horizontal (parallel) if the assumption fulfilled. This shows that there was some evidence of a departure from the proportional hazards assumption for the covariates that are included in the model like triglyceride, LDL, SBP and comorbidity.
The effect of "ties" in this model is leads to make the parameter estimates biased towards zero. As a rule of thumb if more than 10% of the observation are tied we have recommend to deal with them. Breslow's approximation, Efron's approximation etc, but these approximations told us there were no effect of tie values on our estimators and lie survival times are not more than 10% of the observation.

3.3 Bayesian Accelerated Failure Time Model

3.3.1 MCMC Estimation methods

The researcher use non-informative normal prior distribution with mean zero and variance 1000 variance and inverse gamma distribution with scale =0.01, shape=0.01 parameters. In this simulation study of Bayesian inference using MCMC the researcher used 40,000 Markov Chain samples by fixing the burn-in state at 15,000 using WinBugs & SAS software. This implies the parameters of the covariates were estimated by 25,001 Markov chain sample values, simply using the Markov Chain samples after the burn-in state. After this simulation study additional covariates that are not significant in classical AFT model, like sex, are statistical significant.

Distribution	AIC	AICC	BIC	DIC	pD
Weibull	203.655	213.877	388.671	210.122	37.305
Lognormal	210.527	220.749	395.543	218.987	41.500
Loglogistic	207.943	218.165	392.959	220.434	42.674

Table 4: Bayesian Model comparison for probability models (SAS output)

From the above table, Weibull distribution has smallest AIC, AICC, BIC, and DIC because of this Bayesian Weibull accelerated failure time model is preferable to analyze the data in Bayesian Paradigm.

Convergence diagnosis

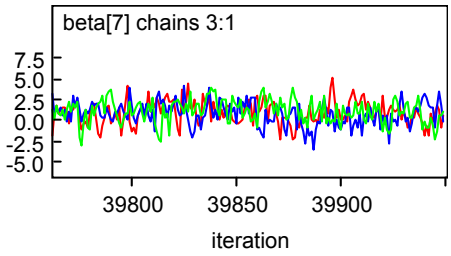

Figure 3:Trace plot of the parameter weight

The plot indicates the convergence of the Markov chain samples.

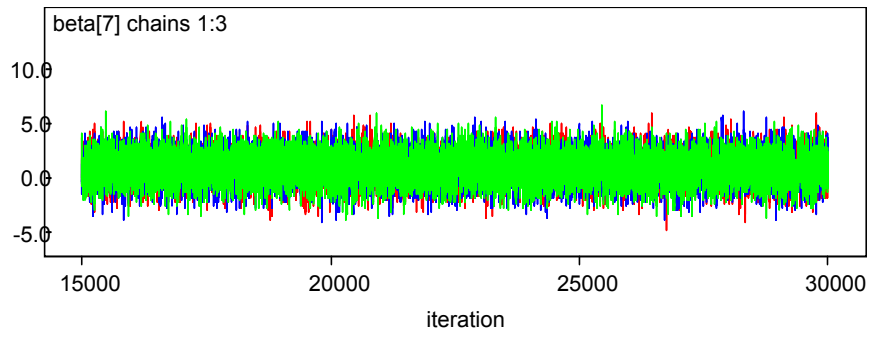

Figure 4: Time series plot of covariates weight

- Time series plot of covariate weight indicates a good convergence and three independently generated chains will mix together or overlapped.

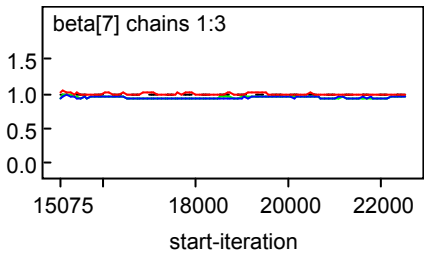

Figure 5: Gelman Rubin test for covariate weight

- For covariate weight R ratio of W and $V(\hat{\beta})$ close to 1.This implies convergence of Gelman-Rubin statistic was achieved.

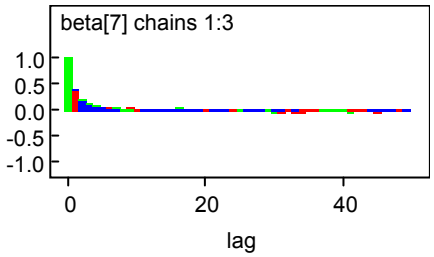

Figure 6: Autocorrelation of the estimated covariate weight

The plot of Autocorrelation indicates that the sampled values of the Markov are independent since the Autocorrelation are diminishing before lag 20.

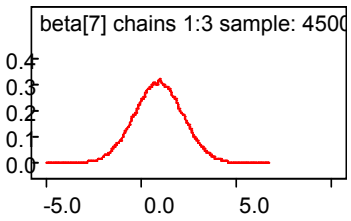

Figure 7: Kernel density of estimated parameter for covariate weight

The plot for the predicator variables weight in the Figure 7 is resembles the curve of normal distribution. This clearly indicated that the coefficient has normal distribution. Hence the simulated parameter values were converged.

node	Mean	S.D	5%*S.D	MC error	2.5%	median	97.5%	Variable
beta[1]*	0.2133	0.1473	0.007365	0.00331	0.179	1.589	2.998	Constant
beta[2]	-0.7709	0.2588	0.01294	0.00316	-0.8389	-0.912	-1.1	Age-c1
beta[3]*	-0.1692	0.1364	0.00682	0.00229	-0.6968	-0.4205	-0.1442	Age-c2
beta[4]*	-0.2661	0.09001	0.004501	0.00208	-0.445	-0.233	-.017	Age-c3
beta[5]*	-0.0721	0.04985	0.002493	0.00141	-0.1649	-0.06801	-0.0064	Trt type
beta[6]*	0.1142	0.134	0.0067	0.00409	0.009	0.22525	0.4415	Sex
beta[7]*	-0.0131	2.68E-03	0.000134	2.41E-05	-0.0485	-0.0415	-0.0045	Weight
beta[8]	-0.597	0.48	0.024	0.01332	-1.556	-0.951	-0.345	BMI1
beta[9]*	-0.143	0.05993	0.002997	0.00131	-1.276	-0.646	-0.016	BMI2
beta[10]*	-0.183	0.2062	0.01031	0.00865	-1.138	-1.789	-0.102	BMI3
beta[11]	-0.9654	0.6608	0.03304	0.0059	-1.098	-0.9554	0.0537	Gl_typ_hyper
beta[12]*	0.1406	0.522	0.0261	0.01101	0.0275	0.66625	1.305	Alco-no
beta[13]*	-0.2575	0.09927	0.004964	0.00246	-0.2678	-0.1829	-0.0979	D-neph
beta[14]*	-0.1754	0.8926	0.04463	0.03533	-0.797	-0.0734	-0.049	D-retino
beta[15]*	-0.234	0.6629	0.033145	0.01472	-1.450	-0.7545	-0.059	D-hear
beta[16]*	-0.1965	0.7147	0.035735	0.01646	-1.6533	-0.8569	-0.0605	D-Neuro
beta[17]*	-0.2695	0.06451	0.003226	0.001569	-0.672	-0.3555	-0.0389	Others
beta[18]	-0.1487	0.15431	0.007716	0.00473	-0.6085	-0.4358	0.179	Sys. Pre-HTN
beta[19]*	-0.1373	0.2132	0.01066	0.01019	-1.130	-0.5831	-0.0362	Sys. Stage 1
beta[20]*	-0.1989	0.5809	0.029045	0.01632	-1.594	-0.844	-0.094	Sys. Stage2
beta[21]*	-0.0327	0.434	0.0217	0.01206	-0.9713	-0.7848	-0.0274	FBS
beta[22]	-0.993	0.3426	0.01713	0.01289	-1.284	-0.6875	0.091	Dia. Pre-HTN
beta[23]*	-0.1133	0.4784	0.02392	0.01613	-2.033	-1.912	-0.087	Dia. Stage1
beta[24]*	-0.1491	0.3334	0.01667	0.01199	-2.47	-1.2813	-.0925	Dia. Stage 2
beta[25]*	-0.2283	5.79E-04	0.000029	5.21E-06	-0.6283	-0.1083	-0.0814	Chole-high
beta[26]*	0.175	0.14155	0.007078	0.00364	0.091	1.317	2.543	F_his_Nega
beta[27]*	-0.187	0.19992	0.009996	0.00591	-0.3032	-0.1877	-0.0690	Com_HTN
beta[28]	-0.844	0.2932	0.01466	0.0106	-1.218	-0.6633	0.1085	Com-Dysli
beta[29]*	-0.061	0.2995	0.014975	0.02709	-0.252	-0.134	-0.015	Com_Obesity
beta[30]*	-0.2623	0.09126	0.004563	0.00451	-0.9177	-0.557	-0.196	Com-Others
beta[31]	-0.5032	0.07367	0.003684	0.00655	-0.9178	-0.506	0.0933	HDL-c-border
beta[32]*	-0.097	0.298	0.0149	0.01154	-0.885	-0.4565	-.0279	HDL_C-undes
beta[33]	-0.0585	0.05075	0.002538	0.00137	-0.1405	-0.0608	0.04258	LDL-1-abo-no
beta[34]	-0.2086	0.09781	0.004891	0.00444	-0.2979	-0.1531	0.0057	LDL_2-border
beta[35]*	-0.2435	0.19993	0.009997	0.00533	-0.484	-0.324	-0.212	LDL_C3-high
beta[36]*	-0.2047	0.19583	0.009792	0.0058	-1.666	-0.8905	-0.115	Trig_high
beta[37]*	-0.2466	0.1001	0.005005	0.00189	-1.3552	-0.744	-0.233	Typ_ dm
beta[38]*	-0.161	0.49883	0.024942	0.02465	-1.152	-0.6145	-0.077	Smoker
beta[39]	0.171	0.2968	0.01484	0.01239	-0.358	3.333	2.37	Height
beta[40]	-0.197	0.18188	0.009094	0.0073	-0.7561	-0.6171	0.025	Oromia
beta[41]	-0.2705	0.2831	0.014155	0.01293	-0.4461	-0.3279	0.062	Amara
beta[42]	0.066	0.2001	0.010005	0.0089	-0.1302	0.6667	0.7885	SNNPRS
beta[43]	0.1672	0.5982	0.02991	0.02673	-0.616	0.189	0.6095	Others
Sigma	0.14897	0.1473	0.007365	0.00331	0.00216	0.00396	0.01455	

Table 7: Bayesian posterior summary for parameter Estimate for covariates with their corresponding credible interval (* indicates statistically significant)

In the above value the MC error (SEM) indicates that how much uncertainty there is about the true posterior mean via the sampled mean. As a rule of thumb, the Markov Chain Monte Carlo simulation should be run until the Monte Carlo error for each parameter of interest is less than about 5% of the sample standard deviation (JESA,2011).Table 7 above shows the Monte Carlo error (MC-error), sample standard deviation (SD), 5% of the SD and the 5% credible intervals for all parameters. It can be seen that for all parameter estimates the Monte Carlo error (MC-error) is less than 5% of standard deviation. So we can use this parameter estimate for inferential purpose. Using the relationship between parameter estimates of accelerated failure time and proportional hazard model interpret the final model using hazard ratio or directly using time ratio.

4. Discussion and Conclusion on the Results

The variables like triglyceride level, low density lipoprotein level, comorbidity and systolic blood pressure did not satisfy or fulfill the cox proportional hazard assumption. Because of this the researcher used Bayesian accelerated failure time model rather than to concern with stratify (time dependent covariates) the variable which not satisfy proportional hazards model assumption.

4.1 Discussion

The first factor which affects the survival of diabetes patients was age, the hazard of patients for the age group 45 to 60 year age group and age group greater than 60 was higher compared with the 18 to 30 year age group. This result is similar with other finding obtained by like Gurjeet S., (2009). The hazard of older age groups had higher hazard rate than younger age groups similar result with Mbanya V. (2008).

The hazard of patients with overweight and obese was higher compared with those having normal body weight this result also coincide with the result obtained by Gurjeet S,(2009), Josepha J, (2010), Fatimatou K. (2013). Blood cholesterol level also has a great impact on the survival of diabetes patients. The result revealed that patients who have high blood cholesterol had higher risk as compared to the others who had a normal blood cholesterol level. The find is similar with pervious study, by Josepha et al. (2010). The survival of patients without family history of diabetic had more survival time than patients with positive family history; it is also in line with the result obtained by Rajiv T., (2012). Density lipoprotein is one of the factors which affect the survival of diabetic patients, the result revealed that patients with low level of HDL had higher hazard rate than patients with high level of HDL, this result also coincides with the study done by Josepha J, (2010).

Smoking cigarette is an important predictor of survival of patients. This study revealed that the hazard rate of patients who smoke cigarette is higher than nonsmokers. The present result concord with earlier results in Josepha J. et al. (2010), similarly, alcohol is the stronger predictor of survival of diabetic patients.

Blood pressure has been found to be significant factor which influence the survival of diabetic patients. According to the study of Josepha J. et al. (2010), hypertension is consistently and independently associated with the risk of mortality from DM. and had higher hazard rate than patients without hypertension on the other hand, LDL level also has a great impact on the survival of diabetic patients. The result

depict that patients who have high LDL had higher risk as compared to the others who had a regular blood LDL. The finding is confirmed by pervious study, Josepha et al. (2010). In addition to those variables, type of diabetic also has a significance effect on the survival time of patients. The finding illustrate that, the risk of death due to type 1 diabetes mellitus disease is higher than for patients who had type 2 diabetes mellitus than those who had negative family history of DM. The result is analogous with earlier study Josepha J. et al. (2010).

Patients with type 1 diabetes have higher hazard than patients with type 2, patients without comorbidity (like HTN) have higher survival than patients' comorbidity this result is similar with the result obtained by Josepha J. et al. (2010). Patients with higher values of triglyceride had more hazard than patients with normal triglyceride Josepha J. et al. (2010). Finally patients with higher values of low density lipoprotein level had higher hazard rate than patients with normal LDL.

4.2 Conclusion on the Results

Bayesian accelerated failure time model showed that the major factors that affect the survival of diabetes patients are age group, BMI, weight, types of diabetic mellitus, alcohol use, diabetic complications, blood pressure, family history, fasting sugar level, cholesterol level, comorbidity, density lipoprotein, triglyceride and smoking habit.

Patients elder than 45 in general have higher hazard rate than that of younger than 45. Patients associated with obese and over body weight had higher hazard and lower survival time than patients with normal body weight. Similarly alcohol users, patients who had a habit of smoking cigarette have higher death rate and lower survival time. Similarly, patients with poor health indicators like stage 1 and stage 2 blood pressures, high blood cholesterol level, diabetic complications, and high amount of fasting blood sugar level, lower level of high density lipoprotein, high value of low density lipoprotein, positive family history and higher level of triglyceride, were unlikely to survive and had higher hazard rate. The result of this study also indicated that survival probability of a patient is not statistically different among groups classified by types of treatment, sex, region and glycemic type.

To analyze and model the survival time of diabetes patients, various parametric accelerated failure time regression models were applied. Among which the Weibull accelerated failure time regression survival model is better fits to analyze the survival time of diabetes patients of this data than log-logistic and log-normal parametric models. Bayesian accelerated failure time model used to survival analysis of diabetes patients because it's AIC (203.655) is smaller than its classical counterpart (AIC= 204.4898).

References

1. Alvin, C. (2012). Diabetes Mellitus, Harrison¨s Principles of Internal Medicine, 18th Edition.
2. Bekele, T., Hymete, A., Tadesse, M. and Mekonnen, Y. (2008). Antidiabetic activity and phytochemical screening of crude extracts of Stevia rebaudiana Bertoni and Ajugaremota Benth grown in Ethiopia on alloxan-induced

diabetic mice. Department of Pharmaceutical Chemistry, School of Pharmacy, Addis Ababa University April.

3. Christensen. R., Johnson. W., Branscum A. and Hanson. T. E. (2011). Bayesian Ideas and Data Analysis, An Introduction for Scientists and Statisticians, Chapman & Hall / CRC.
4. Cochran, W. G. (1977). Sampling Techniques, John Wiley and Sons, New York.
5. Diabetes estimates and Projections. World Health Organization. diabetes@who.int Accessed on Oct. 2003. Global Prevalence of Diabetes Estimates for the year 2000 and projections for 2030.
6. Fatimatou K. (2013) The Escalating Diabetes Epidemic: Determinants of Prevalence Disparity between Country Income Groups, Vol. 4, (2009), p. 24-59
7. Ganjali, M., and Baghfalaki, T. (2012). Bayesian analysis of unemployment duration data in presence of right and interval censoring, Journal of Reliability and Statistical Studies, 5(1), p. 17-32.
8. Gurjeet S.(2009)Diabetes Prevalence and Association Risk Factors among Canadians of South Asian Origin: Estimates from a National Survey,
9. Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). Bayesian Survival Analysis. Springer-Verlag, New York
10. International Diabetic Federation Atlas 2015
11. International Diabetic Federation Atlas 8th Edition 2017
12. Joseph, J., Svartberg, J., Njølstad, I. and Schirmer, H. (2010). Incidence of and risk factors for type-2 diabetes in a general population: the Tromsø Study, Scandinavian Journal Of Public Health, 38(7), p. 768-775.
13. Lee, E. T. and Wang, J. (2003). Statistical Methods For Survival Data Analysis (Vol. 476), John Wiley & Sons.
14. Mbanya, V. (2008). Type 2 diabetes and its association with life style factors (Master's thesis).
15. Ntzoufras, I. (2009). Bayesian modeling using WinBUGS (Vol. 698), John Wiley & Sons.
16. Rajiv T. E., David J. S., Mogamat S H., Ernesto B.-B., Zelda V., Andre P K. and Tandi E M. (2012), High Prevalence of Diabetes Mellitus and Metabolic Syndrome in a South African Coloured Population: Baseline Data of a Study in Bellville, Cape Town, Vol. 102, No. 11.
    Resource:http://www.who.int/diabetes/facts/world_figures/en/index.html
17. Spiegelhalter David J., Keith R. Abrams, Jonathan P. Myles. (2004). Bayesian Approaches to Clinical Trials and Health-Care Evaluation, John Wiley
18. Spiegelhalter, D.J., Thomas, A., Best, N. and Lunn, D. (2003). WinBUGS Examples, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK, available at http : //www.mrc bsu.cam.ac.uk/bugs .