Ee-JRIF
Ethiopian e-journal
For Research and Innovation Foresight

Vol 7, no 1, (2015): pp(20 -39)

Survival Analysis of Diabetes Mellitus Patients Using Parametric, Non-Parametric and Semi-Parametric Approaches: Addis Ababa, Ethiopia

Derbachew Asfaw Teni * , Fikre Enquoselassie ** and Cheru Atsmegiorgis ***

Abstract

Diabetes is a chronic illness that requires lifelong care to control blood glucose levels and prevent complications. The aim of this study is to model survival probability of diabetic patients who were under follow-up and identify significant risk factors for mortality and morbidity. Recorded hospital data were obtained for a cohort of 462 patients at Yekatit-II Hospital, Ethiopia. The patients have been under follow-up from September 2003 to August 2011. The Parametric, Non-Parametric and Semi-Parametric Survival Models are used to estimate the survival time as well as examine the association between the survival time with different demographic, health and risk behavior variables. The analysis shows that most factors significantly contribute to a shorter survival time of diabetes mellitus patients. These factors include being overweight, alcohol use, tobacco use, complication of diabetes mellitus, patients diagnosed with type I diabetes mellitus, uncontrollable blood pressure, high blood cholesterol level, excess amount of fasting blood sugar level and positive family history of diabetes mellitus. It is therefore recommended that people ought to be cognizant on the burden of these risk factors and well informed about the disease.

Key words: Diabetes Mellitus; Mortality and Morbidity; Parametric, Non-Parametric and Semi-Parametric Survival Models

* Statistics Department, Arba Minch University, Ethiopia. Email: dereapeace@gmail.com or dasfaw469@gmail.com; Tel: +251 94 655 5818

** School of Public Health, Addis Ababa University, Addis Ababa, Ethiopia

*** School of Mathematical and Statistical Science, Hawassa University, Hawassa, Ethiopia

1. Introduction

For centuries, communicable diseases were the main causes of death around the world. With advancement of medical research and improvements of life conditions in industrialised countries, non-communicable diseases started to take the place of communicable diseases. At the same time, owing to demographic and life style change, these diseases started to occur in developing countries, resulting in the double burden of communicable and non-communicable diseases (Abdesslam, B. and Saber, B., 2005).

Chronic diseases, such as diabetes mellitus (DM), cardiovascular diseases and cancer are increasing worldwide and they are associated with poor quality of life and increased economic burden; therefore, development of preventive measures against chronic diseases is imperative (Lahham, H.N. M., 2009). Developing countries are encountering a growing burden of chronic diseases, besides infectious diseases and nutritional problems. Although chronic diseases represent a considerable proportion of the disease burden in African countries (WHO, 2002), adequate efforts are not devoted to their prevention and control (WHO and AFRO, 2005). The main chronic diseases, cardiovascular disease, diabetes, and cancers, share a few common risk factors that are related to diet and lifestyle behavior. These include high blood pressure, high cholesterol, tobacco use, excessive alcohol use, inadequate intake of fruit and vegetables, and being overweight, obese or physically inactive, all of which are on the rise in many African countries (Steyn, K. et al, 2005)

According to the World Health Organization's (WHO) statistics, chronic diseases such as cardiovascular diseases (CVD), diabetes, cancers, obesity and respiratory diseases account for about 60% of 56.5 million deaths each year and almost half of the global burden of diseases (Abdesslam, B. and Saber, B., 2005).

Diabetes was considered as one of the main global health issues in the world and the trend of diabetic sufferers was currently showing a significant increase. According to WHO, the estimated number of people with diabetes will increase from 151 million in 2000 to about 221 million people in 2010. An increase of 70 million people was equivalent to an increase of 46% within 10 years. Prediction compiled by Dr. Hillary King of the WHO indicated that this figure will rise to 300 million by the year 2025 (Wong, Y., 2007).

Recent estimate indicate that 5 to 8% of the urban adult population in Dares-Salaam and in South African townships are affected with diabetes, while 20 to 33% have hypertension. In addition, these conditions tend to affect economically active adults, on whom young and old members of the population are often dependent (Unwin U., et al, 2001). The rising prevalence of diabetes, its increasing morbidity and mortality, its disproportionate effect on disadvantaged individuals, communities and nations, and its high human and economic costs clearly establish diabetes as a significant global public health problem (Vinicor, F., 1994).

Similarly to other developing countries, little is done to quantify the prevalence of chronic diseases and their risk factors in Ethiopia. Small-scale surveys of bank employees in Addis Ababa and Ethiopian medical patients at different times have revealed the existence of these diseases and their risk factors; besides an increasing trend of myocardial infarction admissions were also recorded from 1988 to 1997 (Frances, T. and Oli, K., 2006). A burden of disease analysis carried out in rural Ethiopia found that chronic diseases have contributed to 24% of disability-adjusted life year (DALY) lost compared to 72% for other health problems including communicable diseases. According to the Ethiopian Ministry of Health

report on health and health- related indicators, hypertension without mention of heart was the 9th cause of death nationwide in 2003/04 (Ministry of Health, 2003/2004).

In Ethiopia, no population based prevalence study exist but hospital based studies show that the prevalence of diabetes admission has increased from 1.9% in 1970 to 9.5% in 1999 of all medical admissions. It accounts for about 7% of all deaths over the age of 55 years in the medical wards of referral hospitals. According to WHO estimate, the number of diabetic cases in Ethiopia in 2000 was 800,000 and is expected to increase to 1.8 million by 2030 (Feleke, Y. and Enquselassie, F., 2005).

1.1 Statement of the Problem

The problem of chronic diseases is gaining increasing attention in many developing countries. The WHO has led the development of appropriate methods and techniques for surveillance of chronic disease risk factors, and assisted countries through training, and provision of financial and technical support, among others. These efforts are yielding crucial information on the burden of risk factors, such as overweight and obesity, high blood pressure, smoking, and physical inactivity. While many African countries have joined the force, Ethiopia lagged behind the regional initiative; a reflection of the low priority it has until recently accorded to the problem of chronic diseases (Fikru, T., 2008).

Diabetes is a chronic illness that requires lifelong care to control blood glucose levels and prevent complications. Individuals with diabetes are at risk for developing long-term complications that include: loss of vision, kidney disease, nerve damage, peripheral circulatory disorders and other complications. Diabetics are also at risk for stroke and heart disease. Hospitalisations with diabetic complications per 1,000 with a diabetes diagnosis increased to 24% from 1992 to 2000 (Perkins, L, 2004). In 2000, for every 1,000 hospitalisations with a diabetes diagnosis, approximately 192 of those hospitalised included a diabetes related complication diagnosis. In the same year, there were 31 patient deaths per 1,000 hospitalisations with a diabetes diagnosis. Diabetes related deaths increased from 13.8 per 1,000 in 1982 to 21.7 per 1,000 in 2000 (Perkins, L, 2004).

Diabetes can eventually cause a variety of disabling and life-threatening complications. Increasing public awareness of the seriousness of diabetes and its complications, as well as promoting good self-management and treatment among those diagnosed with the disease is key in combating the adverse health effects and economic burden to society associated with this disease (Perkins, L, 2004).

1.2 Objective of the Study

The main objective of this study is to model, estimate and compare survival probability of diabetic patients who were under follow-up and identify significant risk factors for mortality and morbidity.

1.3 Study Design and Data

The study is a retrospective study; all the events - exposure had already occurred in the past; and reviews the patients' cards and information sheets. The researcher merely collect the data and investigate the risk factors associated with the survival of patients with DM diseases. The data consist of patients that visited the Yekatit-II Hospital (Addis Abeba, Ethiopia) with a case of DM. The sample selection mechanism that the researcher used was simple random sampling method in which each of the patients had an equal chance of being selected to be part of the study. A sample size of 462 patients was drawn out of a total population of 2,477.

2. Non-Parametric Survival Model

2.1 Kaplan-Meier Estimator

The Kaplan-Meier estimator of the survivorship function (or survival probability) $S(t) = P(T \geq t)$ is defined as:

$$\hat{S}(t) = \prod_{t_j<t} \left(\frac{n_j - d_j}{n_j} \right) \qquad (1)$$

where d_j is the number of individuals who experience the event at time ⟦ t⟧ _j, and n_j is the number of individuals who have not yet experienced the event at that time (Hosmer, D. and Lemeshow, S., 1998).

2.2 Comparison of Survivorship Functions

When comparing groups of subjects, it is always a good idea to begin with a graphical display of the data in each group. The figure in general shows the pattern of one survivorship. If the function lies above another then it means the group defined by the upper curve lived longer, or had a more favorable survival experience, than the group defined by the lower curve. The statistical question is whether the observed difference seen in the figure is significant.

The general form of this test statistic is given by:

$$Q = \frac{[\sum_{i=1}^{m} w_i(d_{li} - \hat{e}_{li})]^2}{\sum_{i=1}^{m} w_i{}^2 \hat{v}_{li}} \qquad (2)$$

where $\hat{e}_{li} = \frac{n_{li} d_i}{n_i}$, $\hat{v}_{li} = \frac{n_{1i} n_{oi}(n_i - d_i)}{n_i^2 (n_i - 1)}$, n_{oi} is the number at risk at observed survival time $t_{(i)}$ in group 0, n_{1i} is the number at risk at observed survival time $t_{(i)}$ in the group 1, d_{oi} is the number of observed deaths in group 0, d_{1i} is the number of observed deaths in group 1, n_i is the total number of individuals or risk before time $t_{(i)}$ and d_i is the total number of deaths at $t_{(i)}$. If weights (w_i) equal to 1, i.e. $w_i = 1$, the Cochran-Mantel-Haenszel Log Rank test (Q_LR) is used; otherwise if w_i = ⟦ n⟧ _i, the Generalized Wilcoxon test (Q_GWt) is used.

$$Q_{LR} = \frac{[\sum_{i=1}^{m}(d_{li} - \hat{e}_{li})]^2}{\sum_{i=1}^{m} \hat{v}_{li}}$$

$$Q_{GWt} = \frac{[\sum_{i=1}^{m} n_i(d_{li} - \hat{e}_{li})]^2}{\sum_{i=1}^{m} n_i{}^2 \hat{v}_{li}} \qquad (3)$$

3. Semi-Parametric Survival Model

3.1 The Cox Proportional Hazard Model

It is usually written in terms of the hazard model formula. 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 and it is generally given by:

$$h(t, \boldsymbol{X_i}, \boldsymbol{\beta}) = h_o(t) \exp(\boldsymbol{\beta}' \boldsymbol{X_i}) \qquad (4)$$

where $h_o(t)$ is the baseline hazard function that characterizes how the hazard function changes as a function of survival time, $\boldsymbol{X_i}$ is the vector of values of the explanatory variables for the i^{th} individual at time t and $\boldsymbol{\beta}$ is the vector of unknown regression parameters that are assumed to be the same for all individuals in the study, which measure the influence of the covariate on the survival experience.

The cumulative hazard function is given by: $H(t) = H_o(t)\, exp(\boldsymbol{\beta}' \boldsymbol{X})$.

The survivor function is: $S(t, \boldsymbol{X}, \boldsymbol{\beta}) = S_o(t) exp(\boldsymbol{\beta}' \boldsymbol{X})$. where $S_o(t)$, is the baseline survival function.

3.2 Fitting Proportional Hazard Model

The Maximum Likelihood estimates of the Cox model parameters are derived by maximizing a likelihood function usually denoted as *L*. The likelihood function is a mathematical expression which describes the joint probability of obtaining the data actually observed on the subjects in the study as a function of the unknown parameters (the $\boldsymbol{\beta}$'s) in the model being considered. *L* is sometimes denoted as $L(\boldsymbol{\beta})$ where $\boldsymbol{\beta}$ denotes the collection of unknown parameters.

The partial likelihood can be written as the product of several likelihoods, one for each of, say *k* failure times. Thus, at the j^{th} failure time, l_j denotes the likelihood of failing at this time, given survival up to this time. Note that the set of individuals at risk at the j^{th} failure time is called the "risk set," $R(t_{(j)})$, and this set may changes and actually gets smaller in size as the failure time increases.

$$(\beta) = \prod_{j=1}^{k} l_j \qquad (5)$$

In a general sense, the partial likelihood is given by the expression

$$l_p(\beta) = \prod_{i=1}^{m} \left[\frac{e^{X_i\beta}}{\sum_{j \in R(t_i)} e^{X_j\beta}} \right]^{c_i} \qquad (6)$$

where the summation in the denominator is over all subjects in the risk set at time t_i denoted by $R(t_i)$, when there are no tied times, and it is often modified to exclude terms when $c_i = 0$, yielding

$$l_p(\beta) = \prod_{i=1}^{m} \left[\frac{e^{X_i\beta}}{\sum_{j \in R(t_i)} e^{X_j\beta}} \right], \log\left(l_p(\beta)\right)$$
$$= \sum_{i=1}^{m} \left\{ X_{(i)}\beta - \ln\left[\sum_{j \in R(t_{(j)})} e^{X_j\beta} \right] \right\} \qquad (7)$$

where the product is over the m distinct ordered survival time and $X_{(i)}$ denotes the value of the covariance for the subject with ordered survival time $t_{(i)}$.

3.3 Extensions of the Proportional Hazard Model

To accommodate the non-proportionality assumption, one can apply stratified proportional hazard models in which the stratification in most cases is done by using a covariate fixed by design. Suppose we have $s = 1,2,\ldots,S$ strata, and then allow the baseline unspecified hazard function to vary among the strata. The hazard function for stratum, S is:

$$h_s(t, X_i, \beta) = h_{so}(t)\exp(\beta' X_i) \qquad (8)$$

The form of the partial likelihood for the s^{th} stratum is identical to the partial likelihood used in proportional hazard models, but it includes an additional subscript, s indicating the stratum. The contribution to the partial likelihood for the s^{th} stratum is:

$$L_{sp}(\beta) = \prod_{i=1}^{n_s} h(t_i, x_{si}, \beta)^{\delta_{si}} S(t_{si}, x_{si}, \beta) \qquad (9)$$

Where $n_{si}\ is$ the number of observations in the s^{th} stratum, t_{si} is the i^{th} observed value of time in s^{th} stratum, δ_{si} is the value of the censoring indicator associated with t_{si}, $R\ (t_{si})$

, the risk set for subjects in stratum s at time t_{si}, X_{si} is the vector of p covariates for subject i in stratum s.

The full stratified partial likelihood is obtained by multiplying the contributions to the likelihood, namely:

$$L_{Sp}(\beta) = \prod_{s=1}^{S} L_{sp}(\beta) \qquad (10)$$

The maximum stratified partial likelihood estimator of the parameter vector, β is obtained by solving the p equations obtained by differentiating the $\log L_{Sp}(\beta)$ with respect to the p unknown parameters and setting the derivatives equal to zero. Finally model building and model assessment is the same as that of proportional hazard models.

4. Parametric Regression Modeling

In previous topics we focused entirely on the use of semi-parametric model and proportional hazard Cox regression model, in the analysis and prediction of the survival time of patients with diabetes mellitus. The basis of this method was to avoid having to specify the hazard function completely. However, there may be settings in which the distribution of the survival time is in specific parametric distribution that justifies the use of a fully parametric model to better address the goal of the analysis.

4.1 Weibull Regression Model

Survival time t is a positive random variable with Weibull probability density function can be expressed as:

$$f(t;\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} exp\left(\left(-\frac{t}{\mu}\right)^{\alpha}\right) \qquad (11)$$

where, $\mu > 0$ and $\alpha > 0$ and the baseline hazard function of the distribution becomes:

$$h(t;\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \qquad (12)$$

This yield the following survivorship functions: $S(t) = \exp\left[-\left(\frac{t}{\mu}\right)^{\alpha}\right]$ and the cumulative hazard function becomes: $H(t) = \left(\frac{t}{\mu}\right)^{\alpha}$

Depending on the value of α, the hazard function can increase or decrease with increasing survival time. Hence the Weibull model can yield an accelerated failure time model. Independent observations $(t_i, \delta_i), i = 1, \dots, n$ with survival time t_i and censoring indicator δ_i which has value of one if i^{th} observation is not censored and zero when the i^{th} observation is censored and let β be the unknown parameter. The likelihood function is:

$$L(\beta) = \prod_{i=1}^{n}\left\{f(t_i)^{\delta_i}\left(s(t_i)\right)^{1-\delta_i}\right\}$$
$$= \prod_{i=1}^{n}\left\{\left(\frac{f(t_i)}{s(t_i)}\right)^{\delta_i} s(t_i)\right\}$$
$$= \prod_{i=1}^{n}\left\{\{h(t_i)\}^{\delta_i} s(t_i)\right\}$$

$$= \prod_{i=1}^{n}\left\{\left(\frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1}\right)^{\delta_i} \exp\left[-\left(\frac{t}{\mu}\right)^{\alpha}\right]\right\} \qquad (13)$$

Re-parameterising the Weibull distribution using $\lambda = \mu^{-\alpha}$ then $h_0(t) = \lambda\alpha t^{\alpha-1}$ will be the baseline hazard function. Now incorporate covariates X in the hazard function, the Weibull regression models become:

$$h(t;X,\beta) = \lambda\alpha t^{\alpha-1}\exp(X\beta) \qquad (14)$$

The model assumes that individual *i* and *j* with covariates x_i and x_j have proportional hazard

function of the form: $\frac{h(t;x_i)}{h(t;x_j)} = \frac{exp(x_i\beta)}{exp(x_j\beta)} = exp\big((x_i - x_j)'\beta\big)$

A different parameterisation is used with intercept v and scale parameter σ and covariate effects γ_j having relationship with original parameterization as $\beta_j = \frac{-\gamma_j}{\sigma}, \alpha = \sigma^{-1}$ and $\mu = \exp(v)$.

4.2 The Exponential Regression Model

The time data is skewed to the right with exponential distribution, the time of survival for a set of covariates $\boldsymbol{X}$, which is called, accelerated failure time is expressed as:

$$T = exp(\beta' X + \varepsilon) \qquad (15)$$

where, ε^* is the error component

The exponential model $(t \sim Exp(\alpha))$ is the simplest parametric model and assumes a constant risk or hazard over time, which reflects the property of the distribution appropriately called 'lack of memory' because the hazard function $h(t) = \alpha$ does not depend on time. Hence the probability of failure in a time interval $[t, t + \delta t]$ does not depend on previous interval. The survivorship function may be obtained by expressing in terms of time as: $S(t, X, \beta) = exp\big(-te^{-\beta' X}\big)$ and the hazard function of the exponential regression model is $h(t, X, \beta) = e^{-(\beta' X)}$. For the exponential regression survival model, the hazard ratio for the dichotomous covariate is $HR(x = 1, x = 0) = e^{-\beta_1}$.

4.3 The Log-Logistic Regression Model

Multiple covariate log-logistic accelerated failure time may be expressed as:

$$\ln(T) = \beta' X + \sigma\varepsilon \qquad (16)$$

where σ is the scale parameter and ε is the residual (unexplained) variation in the transformed survival times (Collette, D. , 1994) The survivorship function for the model (16) is: $\mathrm{s(t, X, \beta, \sigma)} = [1 + \exp(z)]^{-1}$ Where z is the standardized log-time outcome variable, that is; $z = \frac{(y - \beta_0 - \beta_i X)}{\sigma}$ and $y = \ln(t)$.

The odds of a survival time of at least t are, $OR = \frac{\mathrm{s(t,x,\beta,\sigma)}}{1 - \mathrm{s(t,x,\beta,\sigma)}} = \exp(-z)$, assumes that the covariate is dichotomous and coded 0 or 1. The odds- ratio at time t from the ratio the odds of a survival time evaluated at x= 0 and x= 1 is:

$$OR(\mathrm{x} = 1, \mathrm{x} = 0) = \frac{\exp\left[\frac{-(y - \beta_0 - \beta_1 \times 1)}{\sigma}\right]}{\exp\left(\frac{-(y - \beta_0 - \beta_1 \times 0)}{\sigma}\right)} = exp\left(\frac{\beta_1}{\sigma}\right)$$

This is independent of time.

4.4 The Lognormal Regression Model

The log-normal model may take censored time dependent variable that allows the hazard rate to increase and decrease. The log-normal model assumes that ε ~N (0, 1). Let $h(t)$ be the hazard function of T for (16) when $\beta = 0$ *i*.e. $\beta_0 = \beta_1 = \ldots = \beta_p = 0$. Then, it can be shown that $h(t)$ has the following functional form:

$$h(t) = \frac{\phi\left(\frac{\log(t)}{\sigma}\right)}{\left[1 - \Phi\left(\frac{\log(t)}{\sigma}\right)\right]\sigma t} \qquad (17)$$

where, $\phi(\mathrm{t}) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\mathrm{t}^2}{2}\right)$ is the probability density function, and

$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) du$ is the cumulative distribution function of the standard normal distribution. Obviously we no longer have a proportional hazard model. If the baseline hazard function is desired, it can be obtained from equation (17) by setting x = 0. The survival function $s(t/X)$ at any covariate x can be expressed as:

$$s(t/X) = \Phi\left[\beta_0^* + \beta_1^* x_1 + \cdots + \beta_p^* x_p - \alpha \log(t)\right] \quad (18)$$

where, $\alpha = \frac{1}{\sigma}$, $\beta_j^* = \frac{\beta_j}{\sigma}$ for $j = 0, 1, \ldots, p$. This is the final survival model with intercept depending with t.

5. Model Selection

To select the model that can predict the survival time of diabetic patients, we have two methods. The first is graphical approach (Schoenfeld, D., 1984). For this method the Cox-Snell residual plot is the common one. It is used to determine how well a specific distribution fits to the observed data. This plot will be approximately linear if the specified theoretical distribution is the correct model. Easy fit displays the reference diagonal line along which the graph points should fall along with the goodness of fit tests; the distribution plots can be helpful to determine the best fitting model. The fundamental difference of this approach is that it is quite subjective to come to conclusion while the goodness of fit tests are "exact" in the sense that the results do not depend on the researcher (provided that the tests are performed correctly). Using the plot method is considered to be a more empirical approach in model selection.

Akaikie (1974) proposed an information criterion (AIC) statistic to compare different models and/or models with different numbers of parameters. For each model the value is computed as:

$$AIC = -2\, log\,(likelihood) + 2\,(p + 1 + s) \quad (19)$$

where p denotes the number of covariates in the model without including the constant term and s is the number of parameters minus one i.e. $s = 0$ for the exponential regression model and $s = 1$ for Weibull, Log logistic and Lognormal regression models. According to the criterion, a model with small AIC value will be considered a better fit.

6. Results and Discussion

The medical cards of 462 diabetic patients were reviewed, 56.9% were females and 43.1% were males. Among these patients 14.1% were dead while 85.9% censored. The proportion of deaths among females (11.4%) is lower than males (17.6%).

6.1 Comparison of Survival Experiences

Among different diabetic categories, type 1 DM patients had the lowest survival time and it was also statistically significant (p<.000). These two variables log-rank test and Breslow (generalized Wilcoxon) test for survival difference were all highly significant. Patients who had diabetic nephropathy had lowest survival time, followed by diabetic retinopathy, diabetic neuropathy and cardiovascular disease (CVD). Statistical tests using log-rank and Breslow test also show that there were significant differences among the diabetic complication groups. Similarly the results depict that patients with poor health indicators like drinking alcohol, smoking tobacco, high blood pressure, high blood cholesterol level, pre-existing health problem and positive family history of diabetes mellitus had small survival time and all were highly significant (p<.000).

Table 1: Summary results of diabetes mellitus death events by different demographic, health and risk behavior variables

Covariates	Status		Total 462 (%)
	Number censored (%)	Number of deaths (%)	
Sex			
Female	233(88.6%)	30(11.4%)	263(56.9%)
Male	164(82.4%)	35(17.6%)	199(43.1%)
Place of residence			
Rural	118(86.1%)	19(13.9%)	137(29.7%)
Urban	279(85.8%)	46(14.2%)	325(70.3%)
Body mass index (BMI)			
Underweight	52(88.1%)	7(11.9%)	59(12.8%)
Healthy	140(97.9%)	3(2.1%)	143(31.0%)
Overweight	205(78.8%)	55(21.2%)	260(56.2%)
Alcohol use			
No	305(90.0%)	34(10.0%)	339(73.4%)
Yes	92(74.8%)	31(25.2%)	123(26.6%)
Tobacco use			
No	292(91.8%)	26(8.2%)	318(68.8%)
Yes	105(72.9%)	39(27.1%)	144(31.2%)
Types of diabetic disease diagnosed			
Type 1	157(72.7%)	59(27.3%)	216(46.8%)
Type 2	240(97.6%)	6(2.4%)	246(53.2%)
Family history of diabetes mellitus			
Negative	286(91.4%)	27(8.6%)	313(67.7%)
Positive	111(74.5%)	38(25.5%)	149(32.3%)

Table 2: Comparison of survival experience on diabetic patients using demographic, health and risk behavior variables

Test of Equality over Groups							
Variable	**Mean survival time (in year)**	**Log Rank (Mantel-Cox)**			**Breslow (Generalized Wilcoxon)**		
		Chi-Square	**Df**	**Pr>Chi-Square**	**Chi-Square**	**Df**	**Pr>Chi-Square**
Body mass index							
Underweight	8.337						
Healthy	8.879	25.930	2	0.000	21.622	2	0.000
Overweight	7.941						
Alcohol use							
No	8.557	17.879	1	0.000	22.249	1	0.000
Yes	7.520						
Tobacco use							
No	8.671	33.080	1	0.000	38.745	1	0.000
Yes	7.412						
Types of diabetic disease diagnosed							
Type 1	7.632	58.151	1	0.000	44.873	1	0.000
Type 2	8.839						
Diabetic complications							
None	8.709	48.073	5	0.000	38.179	5	0.000
D_Nephropathy	7.318						
D_Retinopathy	7.022						
D_Neuropathy	8.235						
CVD	8.601						
Others	7.762						
Blood pressure							
Normal	8.462	45.552	2	0.000	30.516	2	0.000
High	8.453						
Uncontrollable	6.998						
Cholesterol level							
Normal	8.668	59.555	1	0.000	49.067	1	0.000
High	7.330						
Pre-existing health condition							
None	8.534	39.646	2	0.000	37.716	2	0.000
Hypertension	8.504						
dyslipidemia	6.987						
Family history of diabetes mellitus							
Negative	8.579	22.395	1	0.000	23.596	1	0.000
Positive	7.647						

6.2 Multiple Covariates Analysis of Stratified Proportional Hazard Model

After adjusting other covariates, the risks of death of patients having abnormal blood pressure, has been increased (adjusted HR=0.605, 95% CI=0.265-1.382). Similarly, the hazard rates of those patients who had uncontrollable blood pressure was raised (adjusted HR=4.269, 95% CI=1.870-9.749).

In addition, after adjusting other covariates, patients who had normal BMI were found to be associated with high survival time, whose hazard rates were 0.522 times that of underweight patients (adjusted HR=0.522, 95% CI=0.123-2.207) which means the survival time of patients who had normal BMI was increased by 47.8% and the increment could be as low as 87.7% and as high as 120.7%. Similarly, the hazard rates of patients who were overweight was 9.325 times that of patients who had been underweight (adjusted HR= 9.325, 95% CI=3.002-28.967). Looking at the effect of alcohol use, after adjusting other confounding variables, the hazard rates of those patients who took alcohol was 1.870 times the hazard rates of those who didn't take alcohol (adjusted HR=1.870, 95% CI=0.975-3.586) indicating that the survival time was reduced by 87%. On the other hand, the hazard rates of patients who use tobacco were about 1.901 times higher than patients who didn't use tobacco (adjusted HR=1.901, 95% CI=1.009-3.580).

Similarly, the hazard rates of patients who had diabetic nephropathy, diabetic retinopathy, diabetic neuropathy, CVD and other type of complications were respectively 5.051, 5.539, 6.027, 3.655 and 8.121 times greater than patients with no diabetic complication (adjusted HR=5.051, 95% CI=1.688-15.110 for diabetic nephropathy, adjusted HR=5.539, 95% CI=1.718-17.864 for diabetic retinopathy, adjusted HR=6.027, 95% CI=1.920-18.925 for neuropathy, adjusted HR=3.655, 95% CI=1.307, 10.221 for CVD and adjusted HR=8.121, 95% CI=2.922-22.568 for others). Moreover, after adjusting other confounding variables, the hazard rates of patients having high cholesterol was 2.191 times the hazard rates of those having a normal cholesterol level (adjusted HR=2.191, 95% CI=1.118-4.292).

The estimated coefficient for the amount of fasting blood sugar level being $\hat{\beta}$ = .003 for continuous risk factor implies the hazard ratio was $exp(\hat{\beta})$ = 1.003. This indicates the change of hazard rates for every one unit (in mg/dl) increase in the amount of fasting blood sugar level (adjusted HR=1.003, 95% CI=0.999-1.007). For a better understanding, one can consider that a 10 unit (in mg/dl) increase in the amount of fasting blood sugar level (FBS) would result in the estimated hazard ratio for the survival of diabetic patients to be exp (10 *.003) = 1.03. This can be interpreted as "for a patient whose FBS increased by 10 units (in mg/dl), the survival time decreased by 3%". The 95% confidence interval suggests that an increase in the hazard rates may be as high as 1.007 or even a decrease rate of 0.999 is consistent with the data. On the other hand, initial weight decreases the hazard time of the patients by 4.1% (adjusted HR=.959, 95% CI=0.927-.992), that is, for every 1 kilogram increase in the initial weights of patients, the hazard rates decreased by 4.1% after controlling the effects of all other covariates in the model.

Finally, family history of DM was another predictor variable related with risk of death of patients. The hazard rates of patients who had a family history of DM were found to be 2.044 times the hazard rates of those who do not have any history of DM (adjusted HR=2.044, CI=1.160-3.600).

Table 3: Results for the final proportional hazard model stratified by type of diabetic disease diagnosed

Variables	$\widehat{\beta}$	SE	Wald	df	Sig.	Exp($\widehat{\beta}$)	95% CI for Exp($\widehat{\beta}$)
Weight	-0.042	0.018	5.718	1	0.017	0.959	(0.927, 0.992)
Body mass index							
Underweight (R)			25.126	2	0.000		
Normal	-0.650	0.736	0.782	1	0.377	0.522	(0.123, 2.207)
Overweight	2.233	0.578	14.904	1	0.000	9.325	(3.002, 28.967)
Alcohol use							
No (R)	0					1	
Yes	0.626	0.332	3.547	1	0.040	1.870	(0.975, 3.586)
Tobacco use							
No (R)	0					1	
Yes	0.642	0.323	3.954	1	0.047	1.901	(1.009, 3.580)
Diabetic complications							
None (R)			17.667	5	0.003		
D_Nephropathy	1.620	0.559	8.390	1	0.004	5.051	(1.688, 15.110)
D_Retinopathy	1.712	0.597	8.212	1	0.004	5.539	(1.718, 17.864)
D_Neuropathy	1.796	0.584	9.469	1	0.002	6.027	(1.920, 18.925)
CVD	1.296	0.525	6.103	1	0.013	3.655	(1.307, 10.221)
Others	2.094	0.522	16.129	1	0.000	8.121	(2.922, 22.568)
Blood pressure							
Normal (R)			22.488	2	0.000		
High	-0.503	0.422	1.424	1	0.233	0.605	(0.265, 1.382)
Uncontrollable	1.451	0.421	11.872	1	0.001	4.269	(1.870, 9.749)
Cholesterol level							
Normal (R)	0					1	
High	0.784	0.343	5.224	1	0.022	2.191	(1.118, 4.292)
Fasting blood sugar level	0.003	0.002	1.982	1	0.049	1.003	(0.999, 1.007)
Family history of diabetes mellitus							
Negative (R)	0					1	
Positive	0.715	0.289	6.117	1	0.013	2.044	(1.160, 3.600)

6.3 Multivariate Analysis of Weibull Regression Model

Looking at the covariates diabetic complications, the hazard rate increases for patients who had diabetic nephropathy ($exp(\hat{\beta}) = 10.34$) followed by diabetic neuropathy ($exp(\hat{\beta}) = 8.39$), diabetic retinopathy ($exp(\hat{\beta}) = 6.84$), and CVD ($exp(\hat{\beta}) = 4.13$) respectively. Moreover, by letting other covariates constant, the hazard rates of patients who were diagnosed with type 1 DM had been increased by 87.9%. On the other hand, the hazard rates of patients who had high and uncontrollable BP were increased and the rates of increment were ($exp(\hat{\beta}) = 1.94$) and $(exp(\hat{\beta}) = 5.05)$, respectively. Similarly, the hazard rates of patients who had high blood cholesterol were 2.65 times more than those who had normal blood cholesterol level. Finally keeping other covariates constant, the hazard of patients who had positive family history of DM was 2.13 times more than patients who had no diabetics in their family history.

Using the regression model of equation (14) and with the parameters found, the survival time of diabetic patients have Weibull distribution, which can be expressed as $t \sim weibull(\alpha, \mu)$, with parameters $\mu = \exp(v) = 5.9944$ and $\alpha = \frac{1}{\delta} = 2.03332787$, $time \sim weibull(2.03332787, 5.9944)$. By substituting the parameters in the final Weibull model with substitution of $\lambda = \mu^{-\alpha} = 5.9944^{-2.03332787} = 0.026217299$ then $h_0(t) = \lambda\alpha t^{\alpha-1} = 0.053308365 \times t^{1.03332787}$ the Weibull hazard regression model that predicts the survival time of patients with DM with identical data settings were:

$$h(t; X, \beta) = 0.053308365 \times t^{1.03332787} \times \exp(X\beta) \qquad (20)$$

In parametric settings, except for exponential regression models the baseline function is not proportional for all subjects as a case of Cox regression model. For the Weibull regression model the baseline hazard vary with $h_0(t) = \lambda\alpha t^{\alpha-1}$, so the base line hazard function of diabetic patients in every increase in time measured in years:

$$h_0(t) = \lambda\alpha t^{\alpha-1} = 0.053308365 \times t^{1.03332787} \qquad (21)$$

Table 4: Parameter estimates, standard errors and the hazard ratios in the final Weibull regression model

Covariate	Covariate effects(γ_j)	$\widehat{\beta}$	SE	Wald	p-value	$Exp(\widehat{\beta})$	95% CI for $Exp(\widehat{\beta})$
WEIGHT	0.0284	-0.0578	0.0197	8.64	0.003	0.944	(0.901, 0.981)
BMI							
Under weight (R)	0	0				1	
Normal	0.2123	-0.432	0.757	0.325	0.568	0.649	(0.14, 2.861)
Over weight	-1.2328	2.507	0.620	16.32	0.000	12.27	(3.64,41.3)
ALC_USE							
No (R)	0	0				1	
Yes	-0.4197	0.853	0.359	5.617	0.018	2.35	(1.16, 4.75)
TOB_USE						1	
No (R)	0	0					
Yes	-0.3287	0.668	0.327	4.162	0.041	1.95	(1.03, 3.70)
DIB_COM							
None (R)	0	0				1	
D_Nephropathy	-1.1491	2.336	0.677	19.90	0.001	10.34	(2.74, 39.0)
D_Retinopathy	-0.9454	1.922	0.763	6.35	0.012	6.84	(1.53, 30.5)
D_Neuropathy	-1.0460	2.127	0.691	9.486	0.002	8.39	(2.17, 32.5)
CVD	-0.6998	1.419	0.620	5.244	0.022	4.13	(1.23, 13.9)
Others	-1.2848	2.612	0.654	15.92	0.000	13.63	(3.78, 49.1)
TY_DIB_DD							
Type 1(R)	0	0				1	
Type 2	1.0403	-2.115	0.638	11.02	0.001	.121	(0.035, 0.421)
BP							
Normal (R)	0	0				1	
High	-0.3259	0.663	0.656	1.02	0.313	1.94	(.54, 7.019)
Uncontrollable		1.619	0.497	10.56	0.001	5.05	(1.90, 13.4)
BLD_CHOL							
Normal (R)	0	0				1	
High	-0.4790	0.974	0.519	3.497	0.061	2.65	(0.96, 7.337)
FBS	-0.0012	0.0025	0.002	1.21	0.023	1.0025	(0.998, 1.01)
FAM_HIST							
Negative (R)	0	0				1	
Positive	0.3711	0.755	0.304	6.20	0.013	2.13	(1.17, 3.86)

6.4 Discussions

Results obtained from this study were found to be analogous with literature on the topic. The first factor that affects survival time of diabetic patients was their weights. As it was indicated both in stratified proportional hazard models and Weibull regression models the hazard rates of patients who had unhealthy weight was about 4.1% and 5.6% higher than patients who had normal weight respectively. This result is in accordance with the studies by (Hu, F.B, et al, 2001). Other studies by Vinicor, F., (1994) and Harder, T et al, (2007) indicated that there exists a relation between birth weight and later life risk of non-insulin dependent diabetes mellitus (NIDDM) which is not linearly inverse but U-shaped.

The BMI of patients was a prognostic factor that significantly predicts the survival time of diabetic patients. The hazard rates of the obese (overweight) patients was much higher. The result is comparable with earlier studies by Joseph et al, (2010) and Vazquez, G et al, (2007).

Smoking cigarette was an important predictor of survival of patients. This study revealed that the hazard rates of patients who smoke cigarette was higher than nonsmokers. The present result concords with earlier results in Hu, F.B et al, (2001) showing that current smoking is associated with a 44% increased risk of diabetes. Similarly, alcohol was the stronger predictor of survival of diabetic patients. The hazard rates of alcohol users were 1.870 and 1.951 times greater than those who didn't took alcohol in both methods.

The types of diabetic disease diagnosed were also a prognostic factor that significantly predicts the survival time of diabetic patients. The result obtained from this study indicates that the hazard rates of type 2 diabetic patients were about 87.9% lower than type 1 diabetic patient. This result is in accordance with the studies from Canada (Talbot P., 2011) showing that the survival time for type 1 population was shorter than type 2 populations.

Blood pressure has been found to be significant predictor of death due to diabetic disease. According to the study by Talbot, P. (2011) hypertension is consistently and independently associated with the risk of morbidity and mortality from DM. On the other hand, blood cholesterol level also has a great impact on the survival of diabetic patients. The result of this study depict that patients who had high blood cholesterol had higher risk as compared to the others who had a regular blood cholesterol level. The finding is confirmed by previous study, Joseph et al, (2010) in Tromsø. In addition to those variables, the family history of diabetic patients also had a significance effect on their survival time. The finding illustrates that, the risk of death due to diabetes mellitus disease is higher for patients who had positive family history of diabetics than those who had negative family history of DM. The result is analogous with an earlier study by Nsamba, S. , (2011) in Uganda.

7. Conclusions and Recommendations

7.1 Conclusions

The Cox regression analysis showed that the major factors that affect the survival of diabetic patients are initial weight, body mass index, alcohol use, tobacco use, diabetic complications, blood pressure, blood cholesterol level, fasting blood sugar level and family history of diabetes mellitus. Patients involved in risky behaviors such as taking alcohol and smoking cigarettes have higher death rates. Similarly, patients with poor health indicators like being overweight, high blood pressure, high blood cholesterol level, diabetic complications, high amount of fasting blood

sugar level and positive family history of diabetics, were less likely to survive. The result of this study also indicated that survival probability of a patient is not statistically different among groups classified by sex, age, place of residence, height, region and pre-existing health conditions.

To predict and model the survival time of diabetic patients, the Exponential, Weibull, Log logistic and Lognormal parametric regression models were applied. Among these, the Weibull regression survival model is best fitted to predict the survival time of diabetic patients.

7.2 Recommendations

Key prevention through health education in primary and secondary care settings on life style factors are necessary to prevent smoking, alcohol use, reduce obesity and early detection of lipids and high blood pressure. Patients with type-1 diabetes mellitus have high hazard of death. Since, type-1 diabetics are insulin dependent, regular checkup of blood glucose level and proper use of insulin is imperative. Early screening of those who have a family history of diabetes needs to be introduced. In addition, different educational programs in prevention of diabetes mellitus disease need to be presented to the public.

According to the results of this study the main predictive factors for the survival time of diabetic patients are more of clinical variables. So, health workers should be cautious when a patient has uncontrollable BP, is overweight, has high blood cholesterol level, has diabetic complications and high amounts of FBS.

The Weibull regression model provide better predictions to the survival probability of diabetic patients. So, future researchers could make use of this model. Future studies also need to assess the level of awareness, treatment and control of these risk factors. The economic and social consequences of diabetes mellitus and other chronic diseases should also receive due attention in future research, as these diseases involve lifelong medical care and social support with significant socioeconomic burden to the individual and the society at large.

REFERENCES

Abdesslam, B. and Saber, B. (2005). The Burden of Non-communicable Diseases in Developing Countries. International J. of Equity in Health, Vol. 4: P 3-9.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19: pp. 716-723

Collette, D. (1994). Modeling Survival Data in Medical Research. Chapman and Hall, London

Feleke, Y. and Enquselassie, F. (2005). An Assessment of the Health Care System for Diabetes in Addis Ababa. Ethiopian Journal of Health Development, Vol 19: P 203-210.

Fikru, T. (2008). Epidemiology of Cardiovascular Disease Risk Factors in Ethiopia. The Rural-urban Gradient. Umea University, Sweden.

Frances, T. and Oli, K. (2006). Chronic Non Infectious Diseases of Adults. The Ecology of Health and Disease in Ethiopia, Shama Books, Vol 21: P 702-712.

Harder, T., Rodekamp, E., Schellong, K. and Dudenhausen, J.W. (2007). Birth Weight and Subsequent Risk of Type 2 Diabetes: A Meta-Analysis. *American Journal of Epidemiology*, Vol 165: P 849–857.

Hosmer, D. and Lemeshow, S. (1998). Applied Survival Analysis: Regression Modeling of Time to Event Data. Canada.

Hu, F.B., Manson, J.E., Stampfer, M.J. and Colditz, G. (2001). Diet, Lifestyle and the Risk of Type 2 Diabetes Mellitus in Women. *New England Medical Journal of Medicine,* Vol 345(11): P 790-7.

Josepha, J., Svartberg J., Njølstad I., and Schimer H., (2010). Incidence of and Risk Factors for Type 2 Diabetes in a General Population. The Tromsø Study. Scand Journal of Public Health, Vol 38 (7): P 768-775.

Lahham, H.N.M., (2009). Cardiovascular Diseases and Risk Factors among Diabetic Patients. Nablus District, West Bank, Palestine

Ministry of Health, Ethiopia (2003/2004). Health and Health related Indicators in Ethiopia.

Nsamba, S. (2011). Factors Associated with Survival of In-Patients with Diabetes at Mulago Hospital, Uganda.

Perkins, L. (2004). The Factors That Impact the Incidence of Diabetes Mellitus and Mortality in the Seven Different Regions of Florida. Florida State University. March, 17.

Schoenfeld, D. (1982). Chi-squared goodness of fit tests for the proportional Hazards regression model. Biometrika, Vol 7: P 145-153.

Steyn, K., Sliwa, K., Hawken, S., Commerford, P., Damasceno, A., Ounpuu, S. and Yusuf, S. (2005). Risk factors Associated with Myocardial Infarction in Africa. The Inter heart Africa Study. Circulation, Vol 112: P 3554-3561.

Talbot, P., (2011). Factors Associated With Survival for A Cohort of Clinically Confirmed Diabetes Cases in Nova Scotia. Canada.

Unwin, N., Setel, P., and Rashid, S. (2001). Non Communicable Disease in Sub Saharan Africa. International Journal of public health, Vol 79 (10): P 947-953

Vazquez, G., Duval, S., Jacobs, D.R. and Silventoinen, K. (2007). Comparison of Body Mass Index, Waist Circumference, and Waist／Hip Ratio in Predicting.

Vinicor, F. (1994). Is Diabetes A Public Health Disorder? Diabetes Care, Vol 1: P 22-7.

WHO (2002). The Cost of Diabetes. Fact Sheet No. 236.

WHO and AFRO (2005). Cardiovascular Diseases in African Region: Current Situation and Perspectives Report of the Regional Director. Fifty Fifth Sessions. Maputo, Mozambique.

Wong, Y. (2007). Survival Analysis of Diabetic Patients in Queen Elizabeth Hospital, Kota Kinabalu. HS, Vol 1: P 2004-1813.

Appendix 1: Checking for the Linearity of Continuous Covariates in the Model

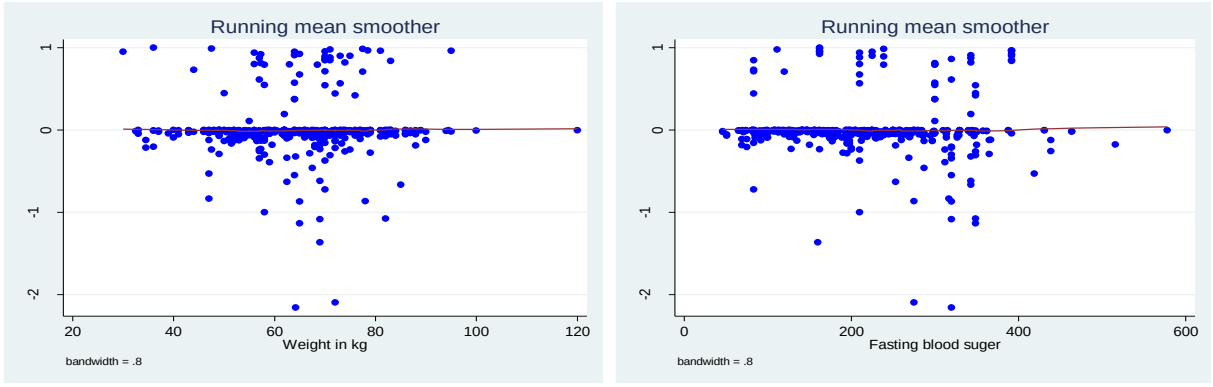


Figure 1: Plots of the Martingale residuals against the covariate initial weight and fasting blood sugar.

Appendix 2: Identification of influential and poorly fit subjects

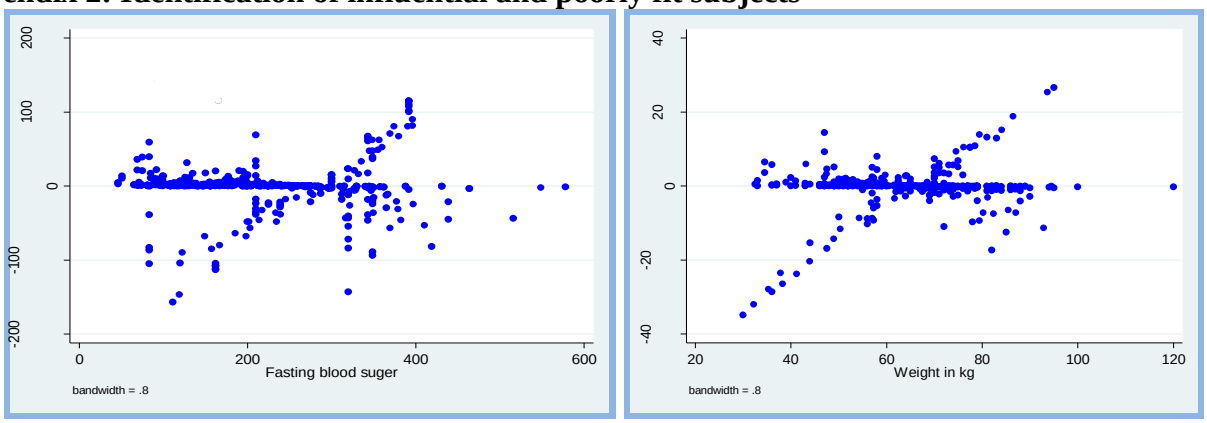
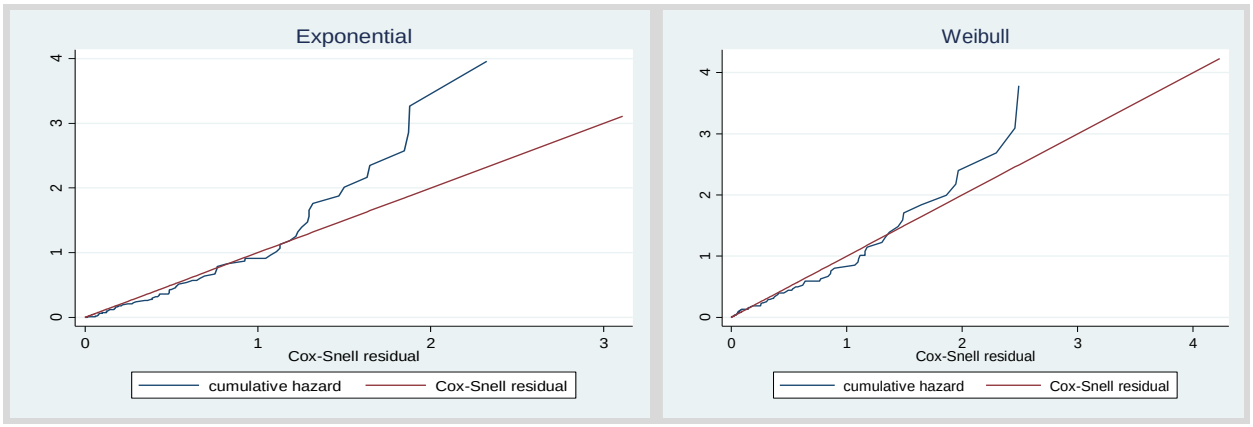


Figure 2: Plots of score residuals for initial weight and fasting blood sugar to detect the existence of influential observation in Cox proportional hazard model

Appendix 3: Cox-Snell Residual plots for model assessment



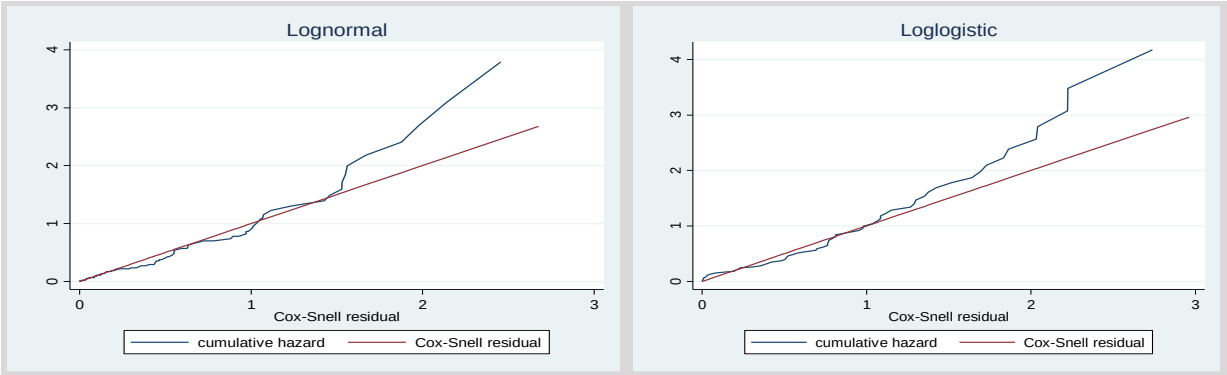


Figure 3: Plots of parametric survival models to examine models that fit the data better

Appendix 4: Residual plots for model assessment

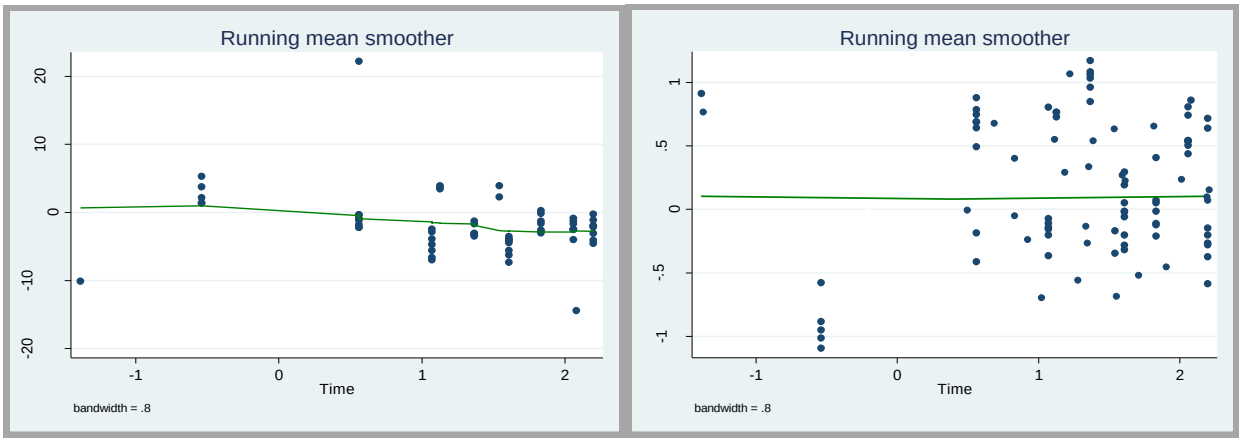
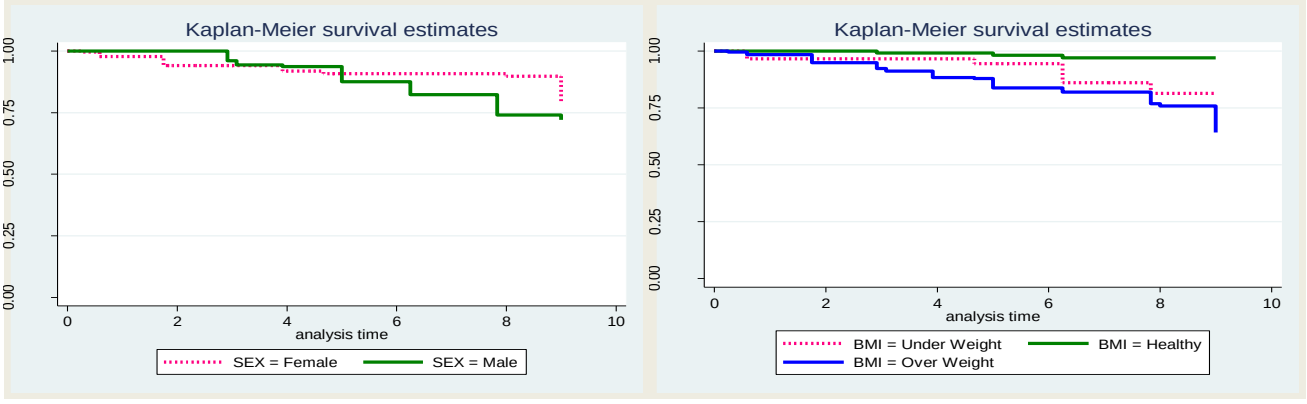


Figure 4: Plots of Scaled Schoenfeld Residuals against transformed time for each covariate in Cox proportional hazard model

Appendix 5: The Kaplan-Meier survival function estimates



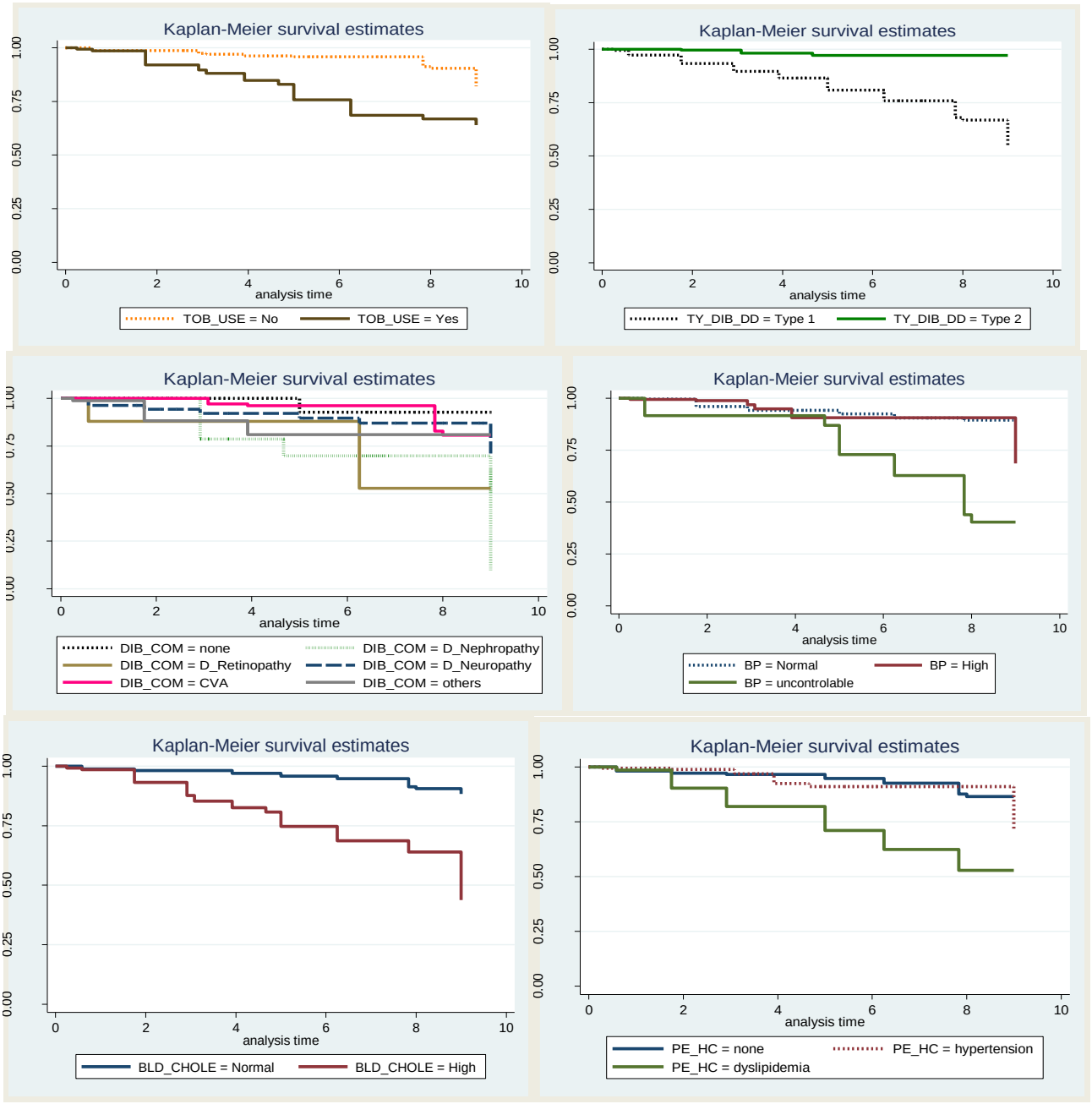


Figure 5: Kaplan Meier survivor estimates for categorical variables