

Demonstrating the Performance of Accelerated Failure Time Model over Cox-PH Model of Survival Data Analysis with Application to HIV-Infected Patients under HAART

Getnet Begashaw (✉ getnetbogale145@gmail.com)

Research article

Keywords: AFT Model, HAART, HIV/AIDS Data, Log-logistic Model, Survival Data Analysis

Posted Date: August 1st, 2019

DOI: <https://doi.org/10.21203/rs.2.12269/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at American Journal of Theoretical and Applied Statistics on January 1st, 2019. See the published version at <https://doi.org/10.11648/j.ajtas.20190806.11>.

Abstract

Background: Human Immunodeficiency Virus (HIV) is a virus that kills CD4 cells. These CD4 cells are white blood cells that fight infection. CD4 count is like a snapshot of how well our immune system is functioning. Studying the way of CD4+ count over time provides an insight to the disease evolution.

Methods: This study was considering the data of HIV/AIDS patients who were undergoing Antiretroviral Therapy in the ART clinic of Menellik II Referral Hospital, Addis Ababa, Ethiopia, during the period 1st January 2014 to 31st December 2017. The data was analyzed in separate survival models i.e non parametric, semi parametric (Cox PH) and parametric survival model (AFT models). For the purpose of model diagnosis cox-snail residual analysis were incorporated. Results: For separate survival model log-logistic model is more appropriate for the survival data than other parametric models. Therefore; functional status and regimen class are significant covariates in determining the hazard function patients. . In the Log-logistic model, among the covariates we have included in the survival model: functional status (working subgroup) and regimen class (all subgroup) were significant at 5% level of significance. But, sex, age, baseCD4, marital status and WHO-clinical stage are not significance at 5% significance level. Using cox-snail residual shows proportionality not satisfied for these WHO stage, regimen class and marital status. Conclusions: Log rank and Wilcoxon tests showed that the significant difference in survival situation among the categorical variables selected for this study sex, marital status, functional status, WHO-clinical stages and regimen class subgroups. But, there was no significant difference in the time-to-event between subgroups of sex, Marital Status and WHO clinical Stage, while, Regimen Class and Functional Status there was a significant difference in the time-to-event between subgroups.

Background

Human Immuno deficiency Virus (HIV) is a family of lentivirus virus that causes for HIV infection. Since the first cases of acquired immuno deficiency syndrome (AIDS) were reported in 1981, infection with human immuno deficiency virus (HIV) has grown to pandemic proportions, resulting in more than 78 million infections and 39 million deaths [17].

HIV infection has four distinct stages [18]. The first stage is known as primary HIV infection, and lasts for two to four weeks. Infected individuals may have no visible symptoms at all, although it is common to develop a brief flue like illness. The symptoms may include headache, fever, loss of appetite, sweating, swelling of lymph nodes, sore throat. The second stage is the asymptomatic stage and lasts for an average of three years. The infected persons do not show any signs or symptoms of the disease. The third stage is the stage of symptomatic HIV infection, where a lot of symptoms such as diarrhea, heavy weight loss, fever, cough, fatigue, shortness of breath, etc are started to show up. The fourth stage is progression from HIV to AIDS class. This is the final stage when the immune system is extremely weakened. The gradual decrease in CD4 T lymphocytes cells (i.e., becomes below 200 cells/mm³) ultimately results in loss of control over the immune response and various opportunistic infections start appearing. This is terminal stage of HIV infection and called AIDS [12].

Sub-Saharan Africa carries a disproportionate burden of HIV, accounting for more than 70% of the global burden of infection. In 2013, sub-Saharan Africa was home for approximately 24.7 million people living with HIV, accounting for 71% of the global total with around 1.5 million new HIV infections and 1.1 million AIDS related deaths as cited in [3]. In sub-Saharan African countries, the prevalence rate among both female and male sex workers was still high (13%) [8].

Ethiopia is the second most populous nation in Africa, next to Nigeria, and one of the seriously affected countries in sub-Saharan Africa. Since first evidence of the HIV epidemic was detected in Ethiopia in 1984, AIDS has claimed the lives of millions. It has been estimated that approximately 45,200 deaths were related to AIDS and that 793,700 people were living with HIV in 2013 as cited in [7].

In order to reduce the number of deaths caused by HIV, it is better to supply free Antiretroviral (ARV) drugs for patients in HIV. These drugs are taken in combination with three or more drugs; this kind of approach is commonly known as Highly Active Antiretroviral Therapy (HAART). After the introduction of ART the death of HIV patients decreases tremendously. Antiretroviral drugs (ARVs) are medications that slow or block the replication of Human Immuno Deficiency Virus (HIV), help to boost the body's ability to fight of HIV infection (immune system) and decrease the viral load in the body.

This therapy positively influences the quality of life and the survival of seropositive HIV carriers [7]. In Ethiopia, there were more than 222,000 patients on ART at the end of 2010 and ART has improved the survival of patients with HIV/AIDS and the quality of life [9]. It is quite common in clinical research that both the repeated measurement data and time-to-event data are simultaneously observed. For example, in HIV/AIDS studies, the trajectories of CD4 counts and time-to-death are collected. In such studies, the interest often lies in understanding the relationships between the longitudinal history of a process and its effect on the risk of an event [4].

The main objective of this study was to analyze the survival endpoints of HIV infected patients under HAART based on a data records in Menilik II Referral Hospital, Addis Ababa, Ethiopia. Specifically this study was aimed to

1. Empirically investigate the major factors that determine the survival time among those patients.
2. To assess which, male or female, patients are on average tend to have more survival time or to recognize whether or not survival time of the male and female patients is different
3. To demonstrate the performance of Accelerated Failure Time Model analysis techniques than Non-parametric and Semi-parametric model analysis of survival data.

In the absence of facilities for CD4 T lymphocytes count, all patients with WHO stages 3 and 4 disease should start ART. Those with WHO stages 1 and 2 disease need supervision and should be monitored carefully, with a minimum of three-monthly clinical reviews and at any time if new symptoms develop. Initiation of ART is recommended for all patients with pulmonary TB or severe bacterial infections and CD4 counts < 350 cells/mm³. Initiation of ART is also recommended for all pregnant women with any stage 3 disease and a CD4 count < 350 cells/mm³. In response to successful ART, the CD4 T

lymphocytes count typically increases by > 50 cells/ μL within weeks after viral suppression, and then increases 50–100 cells/ μL per year thereafter until a threshold is reached [20].

Antiretroviral therapy ART increases the CD4 lymphocyte counts and reduce risk of opportunistic infections and improve survival of HIV-infected people [Santos ACOD, Almeida AMR 2013] as cited in [9]. HIV affects the CD4 cell count in the human body, so it can be employed to make appropriate decisions for the initiation of HAART and proper management of the progression of the infection [19]. In addition to HIV, demographic variables are other factors affecting CD4 cell count changes. Such as age (older ages are predictors of lower count response to HAART [13]; [10], sex (females experience better CD4 count response to HAART compared to males [1], and residence area (rural patients who start ART with a deteriorated CD4 cell count at the initiation of HAART poorly respond to HAART) [9]. ART cannot cure HIV, but it can help the patient live a longer, healthier life and reduce their risk of HIV transmission [14].

The effectiveness of ART has been assessed by clinical observations, CD4 cell counts and determination of plasma viral load [Awoke T., et al., 2016]. In the national treatment guideline of Ethiopia 2010, the first-line ART contains four NRTIs backbone which are Stavudine (d4t), zidovudine (AZT), Abacavir (ABC) and Tenofovir (TDF) plus lamivudine (3TC) with additional two NNRTI drugs (efavirenz (EFV) or nevirapine (NVP) [16]. The combination regimens which have been used most frequently in Ethiopia are d4t–3TC-EFV, d4t–3TC-NVP, AZT–3TC-EFV, AZT–3TC-NVP, TDF–3TC-EFV, or TDF–3TC-NVP [WHO, 2004]. Whereas, patients switch to second-line regimen when the first-line regimen failed due to different reasons. A failure in treatment is measured in three ways: (1) clinical-when new or recurrent WHO stage 4 condition, (2) immunological-when persistent CD4 level below 100 or 50% fall from on-treatment peak value and (3) virological-when plasma viral load above 10,000 copies/ml in duplicates after six months on ART [2].

Methods

The method of estimators in this analysis: Kaplan-Meier estimator, Cox's proportional hazards regression model (Cox PH) and accelerated failure time (AFT) model will be discussed. The model comparison and model diagnosis techniques used in this data analysis are also included. Survival data analysis involves the analysis of time-to-event data that have a principal endpoint-the time until an event occurs (time-to-event data). Survival data are censored in the sense that they did not provide complete information for a variety of reasons; subjects may not have experienced the event of interest. The existence of variables that change over time is also a distinguishing feature in survival analysis.

Descriptive analysis of survival data utilizes non-parametric methods to compare the survival functions of two or more groups. Kaplan-Meier estimator (product-limit-estimator) of the survival function used for this purpose. The log-rank test is utilized to test whether observed differences in survival experience between/among the groups is significant or not. The multivariable model used is the semi-parametric regression model known as the proportional hazards regression (PHR) model. When a study involves multiple characteristics, appropriate statistical techniques must be used to select variables that have significant effects on survival and which are judged to be clinically meaningful for inclusion in a PHR

model. The following sections present the method of non-parametric, semi-parametric and fully-parametric estimation statistical techniques.

2.1 Non-parametric Survival Model

Kaplan-Meier (KM) Estimator: the standard non-parametric technique to estimate the survival function is proposed by Kaplan and Meier (1958). Some said this product-limit estimator. Kaplan and Meier commonly called as Kaplan-Meier estimator. Let $t_1 < t_2 < \dots < t_D$ represent the observed failure times in a sample of size n and y_i be the number of individuals at risk prior to time t_i . This estimator is defined as:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

for all t in the range, where d_i means the number of events at time t_i .

The Product-limit estimator is a right continuous step function with jumps at the observed event times and it provides an efficient means of estimating the survival function for right-censored data. An alternative non-parametric estimator is suggested by Nelson (1972). It has better small-sample-size performance estimator based on the Product-limit estimator.

2.2 Cox's Proportional Hazard Regression Model

The Cox-PH regression model [Cox DR, 1972] is widely used in epidemiological research to examine the association between an exposure and a health outcome. In a typical approach to the analysis of epidemiological data with a continuous exposure variable, the exposure is transformed to an ordinal or nominal variable and relative risk (RR) is modeled as a step function of the exposure. The Cox-PH model is used to analyze censored data. Suppose the observed data are the triples (t_i, z_i, c_i) where t_i is the possibly censored survival time, z_i the scalar predictor variable, and c_i the event indicator, taking values of 1 if the event of interest occurred and 0 if it did not. Survival analysis model that has fully parametric regression structure, but leaves its dependence on time unspecified is known as semi-parametric regression model. Cox PH model is a semi-parametric regression model. Then, the Cox's PH model takes the form:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where $h_0(\cdot)$ is an arbitrary unspecified hazard function and β is the regression coefficient.

In other instances, when the covariate $Z(t)$ may be thought of as a stress factor affecting the individuals under study at time t . With such time-dependent covariates, the Cox's proportional hazard model is of the form:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where $h_0(t)$ reflects how hazard function changes with survival time, and $Z(t)'\beta$ characterizes how hazard function changes with covariates. Cox has proposed exponential function the function of covariate x , and parameter β , say it $f(x, \beta)$, and the hazard function formulated as:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

when x changes from x_0 to x_1 , the hazard ratio is computed as

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

this model in literature is termed as Cox proportional hazard model Cox DR [6]. In this model researchers are mainly interested on the parameter β , interpreted as changing rate of hazard when the covariate changed by unit value of $(x_1 - x_0)$. Remembering that the given baseline hazard function $h_0(t)$ remains known, because of this is that the model is called semi-parametric model [11].

In cox proportional hazard regression model there are basic concepts we have to know, these are: (i) the baseline hazard $\lambda_0(t)$ depends on t but not on the covariates x_1, x_2, \dots, x_n ; (ii) the hazard ratio $\exp(\beta X)$, depends on the covariates x_1, x_2, \dots, x_n but not on time t ; and also, (iii) the covariates x_1, x_2, \dots, x_n do not depend on the time t . These concepts are assumptions of Cox PH model. To check these Cox PH assumptions we can develop diagnostic plots and formal tests. We have been developed diagnostic plots of selected graphical representations and used formal tests to test time-varying coefficients using proportionality test by including a covariate and time interaction terms (time-dependent covariates) associated with the graphical ones.

Allison *et.al* 1995) and Hosmer, Lemeshow (1999) presented two different things about the PH model assumptions. Allison *et.al* 1995) was presented that violation of PH assumptions is not that much serious. On the contrary, Hosmer, Lemeshow (1999) put their ideas as violation of PH assumptions should be taken into account and appropriate modification of the model should be used for the concise interpretation of the model and covariates. Partial likelihood used for the cox model allows using time-dependent covariates. A covariate is time-dependent mean that if the covariate value is change over-time for an individual.

2.3 Accelerated Failure Time (AFT) Model

The AFT model is an alternative model when the proportional hazards assumption does not fulfill. For the past twenty years the Cox proportional hazards model has been used widely to study the covariate effects on the hazard function for the failure time variable. On the other hand, the AFT model, which

simply regresses the logarithm of the survival time over the covariates, has rarely been utilized in the analysis of censored survival data. The AFT model has an intuitive physical interpretation and would be an alternative method to the Cox model in survival analysis.

The AFT model treats the logarithm of survival time as the response variable and includes an error term that is assumed to follow a particular some distribution. Equation below shows the log-linear representation of the AFT model for the i^{th} individual, where $\log(T_i)$ is the log-transformed survival time, T_i survival-time, X_1, X_2, \dots, X_p are explanatory variables with coefficients $\beta_1, \beta_2, \dots, \beta_p$, respectively, and ϵ_i represents residual or unexplained variation in the log-transformed survival times, while β_0 and σ are intercept and scale parameters, respectively [5]. The main reason we use logarithm of T_i is that to consider the truth behind the survival times are always positive with probability of 1.

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

In the absence of censored data, we can estimate the data by using ordinary least square (OLS) estimators. By assigning a new variable, $Y = \log(T)$, using the linear regression model with Y as the response variable. But, survival data have at least some censored observations and these are difficult to manage in OLS. Instead we have to use the MLE with different assumptions on ϵ . For each the distribution of ϵ , there should be a corresponding distribution of survival time, T .

In the presence of only a single explanatory variable, X_1 was defined as a 0–1 indicator variable, and the deceleration factor (c^{\wedge}) was calculated from the coefficient estimate associated with X_1 (i.e. $c^{\wedge} = \exp(\beta^{\wedge}_1)$). When there are many covariates were available, additional terms $\beta_2 X_2 \dots \beta_p X_p$ were included in the model, where the added variables represented factors such as gender, age, marital status, weight, WHO Stage, functional status, regimen class.

Results

Based on the patients marital status 142 (79.78%), 135 (79.41%), 148 (77.49%) and 162 (78.26%); single, married, widowed and divorced responds were censored; respectively, the remaining responds 36 (20.22%), 35 (20.59%), 43 (22.51%), and 45 (21.74%) were event occurred. Also in functional status of HIV-positive ART followers working 255 (77.98%), ambulatory 196 (76.56%) and bedridden 157 (82.63%) responds censored and the remaining 72 (22.02%), 60 (23.77%) and 33 (17.37%) responds respectively as event occurred.

WHO-clinical stages of HIV-positive patients such as stage-I, stage-II, stage-III and stage-IV were 112 (74.17%), 175 (77.09%), 162 (79.02%) and 159 (83.68%) responds censored, the remaining responds 39 (25.83%), 52 (22.91%), 43 (20.98%) and 31 (16.32%) event occurred; respectively. Baseline ART regimen class AZT+3TC+NVP or EFV 147 (74.24%), TDF+3TC+NVP or EFV 152 (75.62%), d4t+3TC+NVP or EFV

150 (78.53%) and others 159 (86.89%) responds censored, the remaining percentage value responds event occurred.

In survival interest response variable that the length of time from HAART start date until the date of death or censor (measured in months). 608 (78.70%) and 165 (21.30%) of patients were alive on HAART (Censored) and died due to HIV/AIDS related death, respectively.

Exploring Univariate Survival Data Analysis

In any data analysis process it is always better to do some uni-variate analysis before directly goes to the more complicated models. In survival analysis it is recommended that visualizing the Kaplan-Meier curves for all the categorical predictors in the dataset. This was show insight to the shape of the survival function for each categorical group and gives an idea whether or not the groups are proportional (i.e., the survival functions are approximately parallel). We have also done the tests of equality across strata to explore whether or not the predictor included in the final model, which is a non-parametric test.

The Kaplan-Meier estimator plots, which used to estimate the survival curves for categorical variables is presented. Plot of the Kaplan-Meier estimates for only four selected categorical variables; sex, functional status, marital status and Regimen Class was displayed (Figure 1). In Figure 1: (i) reveals that both male and female seems have equal survival time. But in some extent female are more survived than male. In functional status of patients in ambulatory are more survived than patients working and bedridden.

According to marital status of patients from plot (iii) also one can see that all strata have equal survival time, no more difference between them. Similarly, in plot (iv) patients who take another regimens other than the listed have some more survival time than patients who allocated in the three class. As a general, all the survival time in all variables are very delayed, no rapid decrements.

The results in the Log rank and Wilcoxon tests showed that the significant difference in survival situation among the categorical variables selected for this study sex, marital status, functional status, WHO-clinical stages and regimen class subgroups (Table 3). But, there was no significant difference in the time-to-event between subgroups of sex, Marital Status and WHO clinical Stage, while, Regimen Class and Functional Status there was a significant difference in the time-to-event between subgroups. The log-rank test of equality across strata for these predictors are insignificance, indicate that no need to include them as potential candidate for the final model. From the graph also we have seen the subgroups in each variable were not parallel and overlap for some part of the graph. This lack of parallelism could gain a problem when we include these predictors in the Cox PH model since one of the assumptions is proportionality of the predictors.

Figure 1: Kaplan-Meier Survival Estimates

3.1 Cox's Proportional Hazard Model Analysis

In order to determine the hazard function in survival data analysis, it is better to fit Cox's PH Regression Model. Therefore; to undertake the hazard function, we have fitted the full model that contains all main effect variables. These may help to identify the effects of each main effects of the study interest. Therefore; the full cox regression model is examined using these seven main effect variables: Sex, functional status, age, regimen class, marital status, WHO-clinical stage and baseline CD4 that were under study. But before building the full model we have to check model with covariate or without covariate (null model) is effective.

From the Table 4: the value of -2LOG L is less, in model with covariates than model without covariates which indicate that the Log likelihood is larger in model with covariates. But the value of AIC and BIC contradict with concept. According to AIC and BIC value no need to include covariates in the Cox PH model. When testing proportional assumption in Cox PH model there are a lot of techniques, these makes difficult and they may contradict one to other. In this case we have to proof using further test. So, another test is global null hypothesis test.

The global null hypothesis test from Table 4 also reveals that all three tests (basically Likelihood Ratio, Score and Wald test) are significantly different from zero and yield similar conclusions, that is, the model without explanatory variables was more effective than the full model. This support the AIC and BIC conclusion what we seen in Table 4. Even the K-M plot in Figure 1: the covariates were intersected with in groups and this is an indication that the PH assumption violated for that variable. None of the variables is globally important in explaining in changes the survival time or hazard rate of the patients.

The cox PH regression model presented as:

$$\lambda_i(t) = \lambda_0(t) \{ \beta_{Am}(Ambulatory2_j) + \beta_{AZ}(AZT + 3TC + NVP\text{or}EFV1_j) + \beta_{TD}(TDF + 3TC + NVP\text{or}EFV2_j) + \beta_{d4}(d4t + 3TC + NVP\text{or}EFV3_j) \}$$

From the type III tests of the PH regression model Wald test p-value is 0.045, 0.047 for functional status and regimen class are seems statistically significant at 5% level of significance. In the cox PH regression model, model parameters interpreted in the form of exponential function of the parameter multiplied by the covariate and the results indicates the ratio of hazards.

HIV patients in Ambulatory status have the probability of event (death) to decrease by 34% than bedridden HIV patients. Likewise, HIV patients who are taking AZT+3TC+NVP or EFV, TDF+3TC+NVP or EFV and d4t+3TC+NVP or EFV increase the death opportunity 97%, doubling and 61% respectively.

Testing Proportionality: To check the validity of the PH model we have to test by examining the interaction between time-dependent covariates with time. The cox PH regression model just we have been conducted was assumed that the risks are proportional, which means the proportion is constant over time. To check this assumption of risks proportionality, we have used time-dependent covariates or variables and test whether they are significant or not. If they are not significant, it shows us time did not

affect the relative risk, and we can conclude that the risks in our model are proportional. Here; we have done proportionality test for time-dependent variables.

The value of the proportionality test found was not significant (p-value = 0.3491). Therefore; none of two (age and CD4+) time-dependent variables individually or jointly were significant, when makes interaction with time. So, the risk is constant overtime. This result taken from Cox PH is satisfactory for detail conclusion and recommendations we have to invite AFT models.

3.2 Accelerated Failure Time (AFT) Model Analysis

All AFT models are named for the distribution of T rather than the distribution of ε or $\log T$. The reason for allowing different distribution assumptions is that they have different implications for the shapes of hazard function. To fit the AFT model we have applied six parametric survival models (exponential, weibull, normal, log-logistic, Gamma and log-normal) to the effects of covariates to HIV/AIDS patients on ART on patient's survival, the data under study. To compare the models, we have used AIC, BIC and -2 Log Likelihood smaller is better.

From the fitted six parametric survival models, Log-logistic seems fit the data better than the others. With the minimum AIC (961.107), BIC (1029.222) and -2 Log likelihood (931.107), Log-logistic model suggests appropriate for the data under study. Therefore; we have fitted our model for parametric survival analysis using the assumed Log-logistic distribution survival function for survival time (T).

Subsequent analysis of the survival data are based on this Log-logistic model. In the Log-logistic model, among the covariates we have included in the survival model: functional status (working subgroup) and regimen class (all subgroup) were significant at 5% level of significance. But, sex, age, baseCD4, marital status and WHO-clinical stage are not significance at 5% significance level. However; the AFT model do not undergoing on the significance of the p-value for these predictors. Instead the significance of the study variables is tested parametrically by the selected Log-logistic distribution.

From the AFT model in Table 8, the survival time is interpreted as exponential of the parameter estimates. According to the Log-logistic model, the survival time for HIV-positive patients under ART increased by ($\exp(2.0737) = 7.9541$) almost 8% by keeping all variables in the study.

Likewise, survival time for males have slightly better than females after adjusting other covariates constant. The average survival time for males is $\exp(0.0365) = 1.04$ than that of males. Males have 4% increases in survival time with respect to females. HIV-positive patients who are in working class have better survival time than Bedridden adjusting other covariates constant. Average survival time of working groups is $\exp(0.2641) = 1.30$ than that of the two subgroups. They have 30% increases in survival time with respect to Bedridden.

Among the class of ART treatment patients who are allocated in AZT+3TC+NVP or EFV and d4t+3TC+NVP or EFV treatment have ($\exp(-0.2461) = 0.78$) 22% decreased in survival time than the

patients allocated other than the listed treatment class. Also patients in TDF+3TC+NVP or EFV drug have (exp $(-0.2853) = 0.75$) 25% decreased in survival time than the patients who take another combination drugs.

3.3 Separate Survival Model Selection

The global null hypothesis test from *Table 4* in Cox PH semi-parametric survival model have the LRT ($\chi^2 = 18.3855$ and $\text{Pr}>\text{ChiSq} = 0.1434$), Score ($\chi^2 = 18.09$ and $\text{Pr}>\text{ChiSq} = 0.1541$) and Wald test ($\chi^2 = 17.5766$ and $\text{Pr}>\text{ChiSq} = 0.1743$) were statistically insignificant. But in this Cox PH model functional status (Ambulatory) and regimen class (AZT+3TC+NVP or EFV, TDF+3TC+NVP or EFV and d4t+3TC+NVP or EFV) were seems statistically significant. Likewise, *Table 7* from the fitted six parametric survival models (exponential, weibull, normal, log-logistic, Gamma and Log-normal), Log-logistic fit the data better than the others with the minimum AIC (961.107), BIC (1029.222) and -2 Log likelihood (931.107). In this AFT model with Log-logistic distribution functional status (working), regimen class (AZT+3TC+NVP or EFV, TDF+3TC+NVP or EFV and d4t+3TC+NVP or EFV) and WHO clinical stage (stage I, stage II and stage III) were statistically significance at 5% level significance.

3.4 Separate Survival Model Diagnostics

0000 *Cox-Snell Residuals*: Parallel lines indicated that hazards between groups were proportional and assumption of proportionality was satisfied. But in Figure 2: (i, ii and iii) bellow shows proportionality not satisfied for these WHO stage, regimen class and marital status respectively.

Figure 2: Cox-Snell Residuals

Conclusions

4.1 Conclusions

Separate analysis of time-to-event outcomes reveals 165 (21.30%) were event occurred and the remaining 608 (78.70%) are censored from the ART as well as the estimated mean survival is 32.27 (\pm SD 8.40) months. From a Log-logistic parametric model functional status (Working) and regimen class (AZT + 3TC + NVP or EFV, TDF + 3TC + NVP or EFV and d4t + 3TC + NVP or EFV) found to be a significant factor on time to death. This lead us to conclude HIV positive patients who can able to work day-today activities have an immunity in resisting the disease burden and delaying the disease event with treatment. Also the regimens that comprise AZT+3TC+NVP or EFV, TDF+3TC+NVP or EFV and d4t+3TC+NVP or EFV have a significant role in increasing the number of CD4+ for HIV-positive patients, which helps to live long than other regimen class ordered by clinician for patients.

As recommendations: This thesis shall be recommend that further studies of this nature include other important socioeconomic and clinical covariates that were not included in this study like opportunistic infections, body mass index, condom use, drug addiction, education level and many others which are recorded for long duration. Beside this health officers and data clerks working with recording HIV-positive patients data either in chart or database need training to improve quality of medical data especially in keeping dataset from missingness.

Declaration

I hereby, declare that the paper entitled as “Demonstrating the Performance of Accelerated Failure Time Model over Cox-PH Model of Survival Data Analysis with Application to HIV-Infected Patients under HAART” is the research report of my work, which is carried out by me. This project paper performed in the way of research metrology on the course of “Survival Data Analysis” in degree of MSc Bio-Statistics during the second semester (Jun, 2018/19). This research paper not previously formed the on basis of any degree, diploma, fellowship or other similar titles of universities or institutions. I authenticated that this project paper is definitely my own hand work via my signature as follows.

Ethics Approval and Consent to Participate

I obtained a formal ethical clearance from Debre Berhan University Department of Statistics (Ref no. Stat42/9/18). Personal information will have to be kept confidentially without disclosing to others during data collection from patient cards. Afterward I got ethics approval from Menellik hospital in orally with oath. Since I am the only study participant stated in ethics committee I have no peer consents.

Consent for publication

Not Applicable

Availability of data and material

The data that I take from hospital is full secured because of ethical oath. If anyone can access this data set he/she have to license from hospital. The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Competing interests

Non-financial competing interests

Acknowledgments:

I am deeply indebted to my supervisor Dr. Haftu Gebrehiwot (Asst.Prof.) for his academic and professional guidance, prompt reviews of my drafts and the future directions also for his strongest moral advise that makes me be strong. I am equally grateful to my co-advisor Ayele Admasu (MSc). Finally I would like to thank more Menelik Hospital front desk workers for their permission to use their data sets to my research paper. Thank you everyone that I could not mention here, who have encouraged and helped me in different ways. I never miss your kindness from my heart.

References

1. Asfaw A, et al. (2015) CD4 cell count trends after commencement of antiretroviral therapy among HIV-infected patients in Tigray, Northern Ethiopia: a retrospective cross-sectional study. PLoS ONE.10(3):e0122583
2. Awoke T, Worku A, Kebede Y, Kasim A, Birlie B, Braekers R, et al. (2016) Modeling Outcomes of First-Line Antiretroviral Therapy and Rate of CD4 Counts Change among a Cohort of HIV/AIDS Patients in Ethiopia: A Retrospective Cohort Study. PLoS ONE 11(12): e0168323. doi:10.1371/journal.pone.0168323
3. Ayesha B. M. Kharsany and Quarraisha A. Karim; 2016, 10, 34–48, HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities, DOI:10.2174/1874613601610010034, The Open AIDS Journal
4. C. Brombin, (2016). Evaluating treatment effect within a multivariate stochastic ordering framework: Nonparametric combination methodology applied to a study on multiple sclerosis. Statistical Methods in Medical Research, 2016, Vo l. 25(1) 366384, DOI:10.1177/0962280212454203
5. Collett, D. (2004), Modeling Survival Data in Medical Research, 2nd edition, Chapman and Hall. Cox DR. Regression models and life tables. Journal of the Royal Statistical Society Series BStatistical Methodology. 1972; 34:187–220.
6. Cox, D. R. (1972). Regression models and Life Tables (with Discussion). Journal of the Royal Statistical Society. Series B, 34, 187–220
7. Dinberu S., et al., (2017). Risk Factors for Mortality among Adult HIV/AIDS Patients Following Antiretroviral Therapy in Southwestern Ethiopia: An Assessment through Survival Models, International Journal of Environmental Research and Public Health, vol (14), 296; doi:10.3390/ijerph14030296.
8. East S, Africa S-S. (2010). Towards universal access: scaling up priority HIV/AIDS interventions in the health sector. Europe.85:000
9. Feleke DG, Yemanebhane N, Gebretsadik D (2017). Nutritional Status and CD4 Cell Counts in HIV/AIDS Patients under Highly Active Antiretroviral Therapy in Addis Ababa, Ethiopia. J AIDS Clin Res 8: 688. doi:10.4172/2155–6113.1000688

10. Florence E, et al. (2003). Factors associated with a reduced CD4 lymphocyte count response to HAART despite full viral suppression in the EuroSIDA study. *HIV Med.* 4(3):255–62.
11. Hosmer DW JR, Lemenshow S, MAY S., (2008), *Applied survival analysis: regression modeling of time-to-event data*, second edition. Hoboken: WileyInterscience
12. Hu D. J., Dondero T. J., Rayfield M. A., George R., Schochetman G., (2016). The emerging genetic diversity of HIV,the importance of global surveillance for diagnostic research and prevention. *JAMA*,275(3), 210–216.
13. Kaufmann GR, et al. (2003). CD4 T-lymphocyte recovery in individuals with advanced HIV–1 infection receiving potent antiretroviral therapy for 4 years: the Swiss HIV Cohort Study. *Arch Intern Med.* 163(18):218795.
14. Keadu T., (2016). Modeling CD4 + Cell Counts of HIV-Positive Patients Following Antiretroviral Therapy (ART): A Case of Yekatit 12 Hospital, Addis Ababa.
15. Santos ACOD, Almeida AMR (2013) Nutritional status and CD4 cell counts in patients with HIV/AIDS receiving antiretroviral therapy. *Revista da Sociedade Brasileira de Medicina Tropical* 46: 698–703.
16. Smith CJ, et al. (2004). Factors inflencing increases in CD4 cell counts of HIV positive persons receiving long-term highly active antiretroviral therapy. *J Infect Dis.* 190(10):1860–8
17. UNAIDS. Fact sheet 2014: Global statistics.
18. UNAIDS: Fact Sheet (2016). In. Edited by Communications and Global Advocacy. Geneva, Switzerland.
19. van Leth F, et al. (2004). Comparison of fist-line antiretroviral therapy with regimens including nevirapine, efavirenz, or both drugs, plus stavudine and lamivudine: a randomised open-label trial, the 2NN Study. *Lancet.* 363(9417):1253–63.
20. WHO, (2007). Laboratory Guidelines for enumerating CD4 T Lymphocytes in the context of HIV/AIDS

Tables

Table 1: Frequency Distribution for Baseline Independent Variables together with the Censored Observations of Time-to-event

Characteristics	Category	Censored		Event (Death)		Total	
		Count	%	Count	%	Count	%
Sex	Male	305	77.81	87	22.19	392	50.71
	Female*	303	79.53	78	20.47	381	381
Marital Status	Single	142	79.78	36	20.22	178	23.03
	Married	135	79.41	35	20.59	170	21.99
	Widowed	148	77.49	43	22.51	191	24.71
	Divorced*	162	78.26	45	21.74	207	26.78
Functional Status	Working	255	77.98	72	22.02	327	42.30
	Ambulatory	196	76.56	60	23.77	256	33.12
	Bedridden*	157	82.63	33	17.37	190	24.58
WHO Clinical Stage	Stage I	112	74.17	39	25.83	151	19.53
	Stage II	175	77.09	52	22.91	227	29.37
	Stage III	162	79.02	43	20.98	205	26.52
	Stage IV	159	83.68	31	16.32	190	24.58
Regimen Class	AZT+3TC+NVP or EFV	147	74.24	51	25.76	198	25.61
	TDF+3TC+NVP or EFV	152	75.62	49	24.38	201	26.00
	d4t+3TC+NVP or EFV	150	78.53	41	21.47	191	24.71
	Others*	159	86.89	24	13.11	182	23.67

* Indicates the reference group for each patient characteristic

Table 2: Summary Statistics for Continuous Variables

Continuous Variables	N	Minimum	Maximum	Mean	Std. Deviation	Missing
CD4+	4615	140	455	303.59	52.30	23
Time	4638	1	6	3.50	1.71	0
Survival Time	773	6	36	32.27	8.40	0
Age of Patients	4422	18	48	32.88	8.21	216
Summary	Number of death = 165 Number of censored = 608			Percentage of death = 21.3% Percentage of censored = 78.7%		

Table 3: Logrank and Wilcoxon tests of Survival Differences for Categorical Covariates

Variable	Logrank Test			Wilcoxon Test		
	Test Statistics	DF	p-value	Test Statistics	DF	p-value
Sex	0.4060	1	0.5240	0.4788	1	0.4890
Marital Status	0.3817	3	0.9440	0.4343	3	0.9331
WHO Clinical Stage	5.3671	2	0.0683	4.0947	2	0.1291
Regimen Class	10.4995	3	0.0148*	10.1772	3	0.0171*
Functional Status	6.3752	2	0.0413*	7.1214	2	0.0284*

Table 4: Model Statistics for Cox's PH Model

Criterion	Without Covariates	With Covariates
-2 LOG L	1181.318	1162.933
AIC	1181.318	1188.933
BIC	1181.318	1228.071

Table 5: Testing Global Null Hypothesis for Cox's PH Model

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.3855	13	0.1434
Score	18.0900	13	0.1541
Wald	17.5766	13	0.1743

Table 6: Parameter Estimates for Cox's Proportional Hazard Model

Parameter	PE	SE	Chi-Squa	Pr>ChiSq	HR	95% HR CL	
						LCL	UCL
Age	0.00364	0.01035	0.1239	0.7248	1.004	0.983	1.024
Baseline CD4	-0.00101	0.00222	0.2054	0.6504	0.999	0.995	1.003
Male	0.05441	0.17077	0.1015	0.7500	1.056	0.756	1.476
Female (Ref)
Working	0.02246	0.23949	0.0088	0.9253	1.023	0.640	1.635
Ambulatory	-0.41283	0.18888	4.7770	0.0288*	0.662	0.457	0.958
Bedridden (Ref)
Single	-0.02097	0.23415	0.0080	0.9286	0.979	0.619	1.550
Married	-0.09055	0.24253	0.1394	0.7089	0.913	0.568	1.469
Widowed	0.02977	0.23085	0.0166	0.8974	1.030	0.655	1.620
Divorced (Ref)
AZT+3TC+NVP or EFV	0.67604	0.27327	6.1202	0.0134*	1.966	1.151	3.359
TDF+3TC+NVP or EFV	0.69908	0.27227	6.5924	0.0102*	2.012	1.180	3.431
d4t+3TC+NVP or EFV	0.47787	0.28115	2.8890	0.044*	1.613	1.293	2.798
Others (Ref)
Stage II	0.56879	0.35093	2.6270	0.1051	1.766	0.888	3.513
Stage III	0.12815	0.17677	0.5255	0.4685	1.137	0.804	1.607
Stage IV (Ref)

Table 7: Test Proportionality for Time Dependent Covariates in Cox's PH Model

Label	Wald Chi-Square	DF	Pr > ChiSq
Test Proportionality	0.4932	2	0.7815

Table 8: Accelerated Failure Time (AFT) Model

Parametric Survival Model	AIC	BIC	-2 Log LL
Exponential	1811.174	1874.748	1783.174
Weibull	963.715	1031.831	933.715
Normal	1867.346	1935.462	1837.346
Log-logistic	961.107	1029.222	931.107
Gamma	967.670	1040.327	935.670
Log-normal	1785.867	1853.982	1755.867

Table 9: *AFT Parametric Survival Model (Log-logistic distribution) Parameter Estimates*

Parameter	Category	Estimate	SE	95% CL		Chi-Squ	P-value
Intercept	.	2.0737	0.3427	1.4020	2.7455	36.61	<.0001*
Sex	Male	0.0365	0.0744	-0.1093	0.1822	0.24	0.6239
	Female
FuncSt	Working	0.2641	0.1071	0.0542	0.4740	6.08	0.0136*
	Ambulatory	0.1103	0.0819	-0.0502	0.2707	1.81	0.1780
	Bedridden
MarSt	Single	-0.0035	0.1014	-0.2023	0.1953	0.00	0.9722
	Married	0.0824	0.1048	-0.1230	0.2878	0.62	0.4318
	Widowed	0.0385	0.1008	-0.1591	0.2361	0.15	0.7025
	Divorced
RegCl	AZT+3TC+NVP or EFV	-0.2461	0.1145	-0.4704	-0.0217	4.62	0.0316*
	TDF+3TC+NVP or EFV	-0.2853	0.1141	-0.5088	-0.0617	6.26	0.0124*
	d4t+3TC+NVP or EFV	-0.2380	0.1180	-0.4693	-0.0067	4.07	0.0437*
	Others
WHOST	Stage II	-0.0980	0.1655	-0.4223	0.2264	0.35	0.5539
	Stage III	-0.1020	0.0758	-0.2506	0.0466	1.81	0.1785
	Stage IV
Age	.	-0.0026	0.0046	-0.0116	0.0064	0.32	0.5741
baseCD4	.	-0.0006	0.0010	-0.0025	0.0014	0.32	0.5721
Scale	.	0.6478	0.0375	0.5782	0.7257	.	.

Remark: - * significance at 0.05 level of significance, confidence limits (CL)

Figures

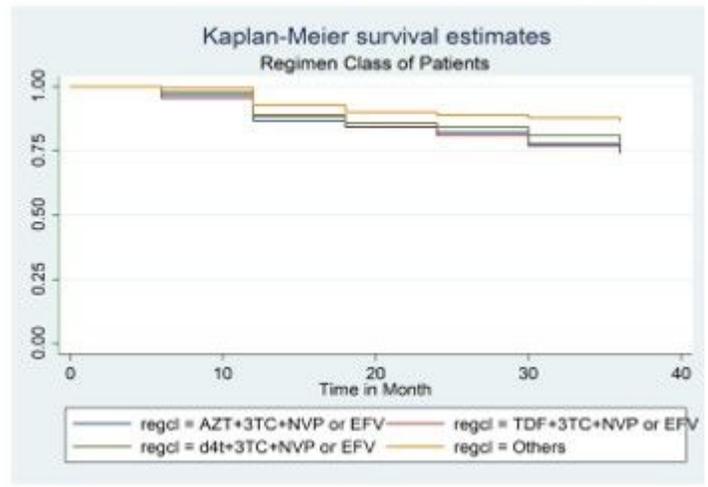
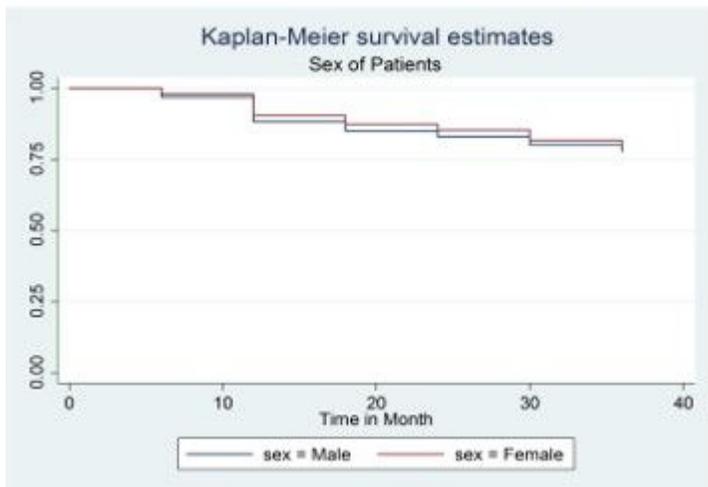
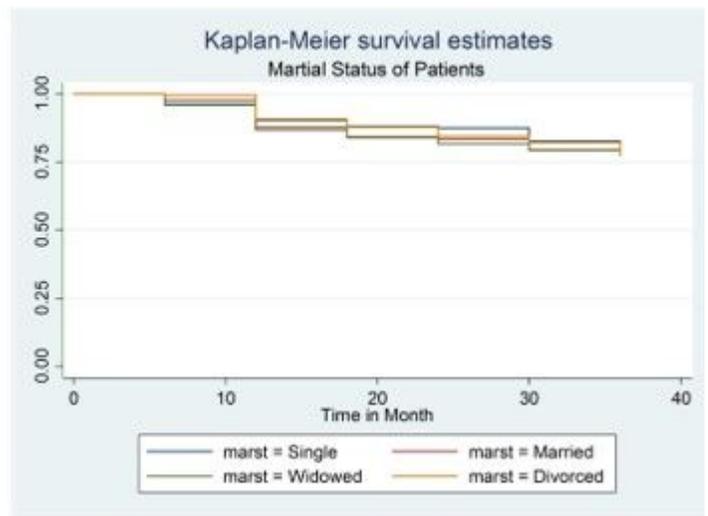
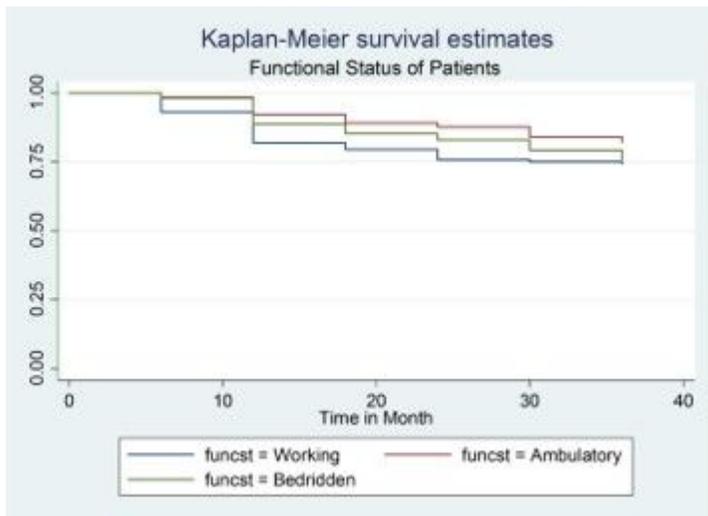


Figure 1

Kaplan-Meier Survival Estimates

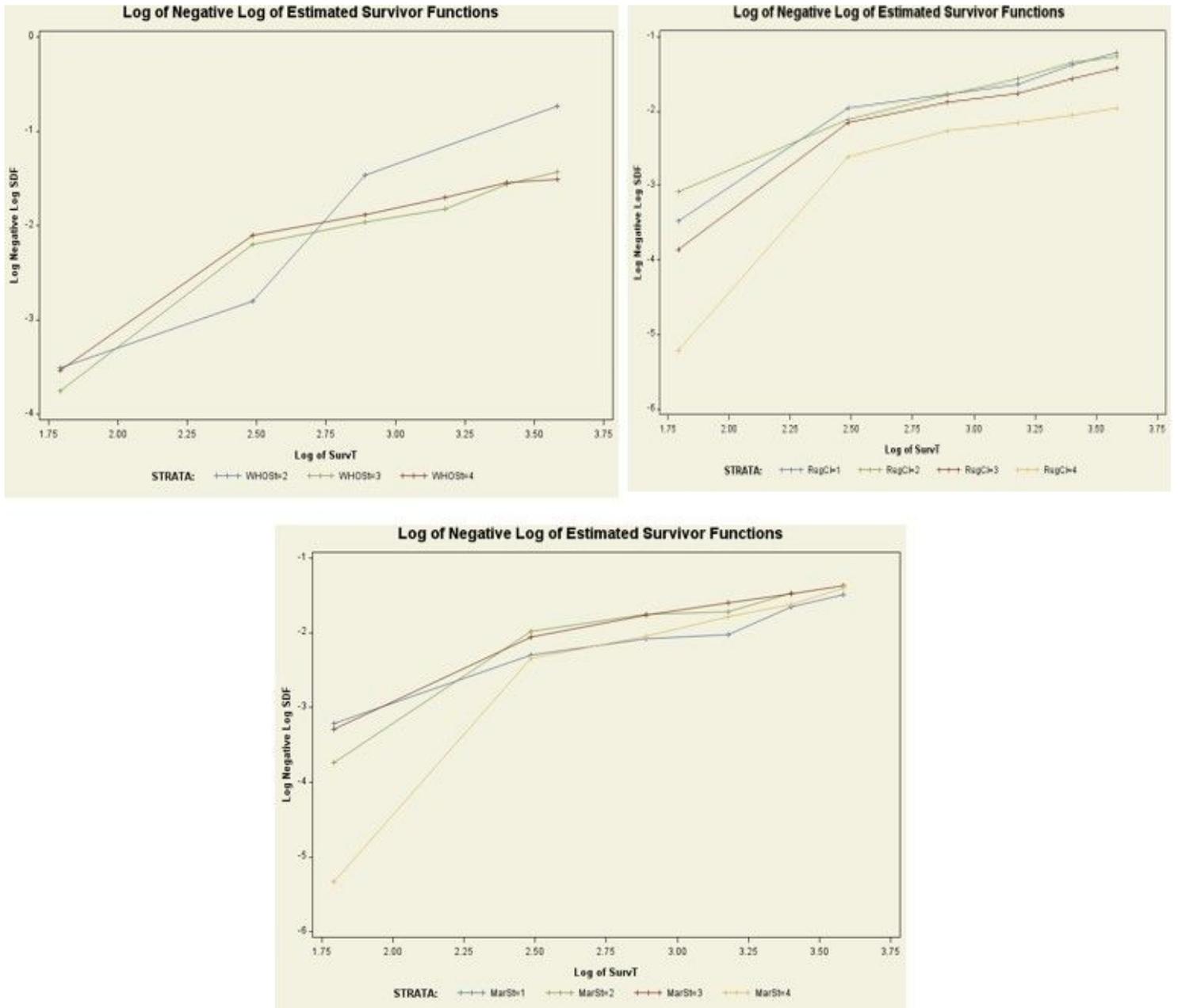


Figure 2

Cox-Snell Residuals

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [eq4.jpg](#)
- [eq5.jpg](#)
- [eq6.jpg](#)
- [eq1.jpg](#)

- eq2.jpg
- eq3.jpg