

# Modelling of Survival Time Among Adult HIV/AIDS Patients Under Antiretroviral Therapy in Moi Teaching and Referral Hospital in Kenya

Mengich Kibichii Robert<sup>\*</sup>, Ann Mwangi, Gregory Kibet Kerich, Nyakundi Omwando Cornelious

Department of Mathematics, Physics and Computing, Moi University, Eldoret, Kenya

## Email address:

[mengichk@gmail.com](mailto:mengichk@gmail.com) (M. K. Robert)

<sup>\*</sup>Corresponding author

## To cite this article:

Mengich Kibichii Robert, Ann Mwangi, Gregory Kibet Kerich, Nyakundi Omwando Cornelious. Modelling of Survival Time Among Adult HIV/AIDS Patients Under Antiretroviral Therapy in Moi Teaching and Referral Hospital in Kenya. *International Journal of Statistical Distributions and Applications*. Vol. 7, No. 2, 2021, pp. 35-47. doi: 10.11648/j.ijstd.20210702.12

Received: May 17, 2021; Accepted: June 2, 2021; Published: June 15, 2021

---

**Abstract:** Survival modelling is a technique which exploits repeated measures of continuous covariates to predict explanatory variable's effects on the response factor. The survival modelling helps design interventions in the health sector, which has seen one of its applications in the management of Human Immune Virus/ Acquired Immune Deficiency Syndrome (HIV/AIDS). However, despite improvement in Anti-Retroviral Therapy (ART) interventions over the years, the observed disease effects (morbidity, progression and mortality) remain high and varies across geographical borders. This study utilizes survival models to determine the predictors of survival among adult HIV/AIDS patients on ART in Moi Teaching and Referral Hospital (MTRH) Kenya. This is achieved by fitting a Cox proportional hazard regression model to adult HIV/AIDS patients data and determine predictors of survival amongst the study subjects. A retrospective study design was adopted where a target population of 10,038 patients who were on ART and were enrolled between January 2005 and January 2007 were investigated for a ten years follow-up period. The Cox proportional hazard regression model (CPHRM) was fitted to the data using log partial likelihood function. The log rank test and 95% confidence Interval (C.I) were used to analyze the significance of the hazard ratios of each variable. The results showed that HIV severity with unadjusted Hazard Ratio [UHR=0.729,  $p=0.032$ ], level of education [lower UHR=0.952,  $p=0.019$ ], and perfect adherence of antiretroviral drugs (ARV) [UHR=0.668,  $p=0.004$ ] positively influenced patient survival time. Patient's gender [male UHR=1.633,  $p<0.001$ ] showed negative effect on patient survival time. The adjusted hazard ratios for multivariate Cox model were, HIV severity [AHR=1.18,  $p=0.735$ ] age category between 30-40 in reference to age less than 30 [AHR=0.459,  $p=0.178$ ] and age category above 40 years [AHR=0.644,  $p=0.447$ ], Body Mass Index (BMI) less than 18.5kg/m<sup>2</sup> in reference to between 18.5-<25kg/m<sup>2</sup> [AHR=1.65,  $p=0.847$ ] and BMI above 25 kg/m<sup>2</sup> [AHR=0.861,  $p=0.847$ ], level of education [lower AHR=0.931,  $p=0.209$ ], patients' gender [male AHR=1.884,  $p=0.19$ ] and ARV adherence [perfect AHR=1.393,  $p=0.498$ ]. In conclusion, HIV severity, level of education, ARV adherence and patients' gender were significant predictors of survival time. In addition, none of the patient's characteristics predicted survival time in the multivariate Cox model. Therefore, this study recommends to the government of Kenya to spearhead the development of policy framework for the provision of regular screening services for the male population to avoid late diagnosis and interventions of HIV/AIDS disease.

**Keywords:** Survival Modeling, Survival Time, Survival Analysis, Censoring Censored Observation

---

## 1. Introduction

Survival models are essential in research because of their unique attributes. The factors of interest comprise of survival time, defined by either death or censoring, and the presence

of explanatory variables. Binary outcomes are common in medical research, where "success" may indicate that the patient is alive after treatment while "failure" implies the death of a patient. Survival analysis is interested in the statistical study of such occurrence (time until event) in a

group of individuals. The follow-up period of these individual patients occurs within a definite period with attention on the time in which the outcome occurs, known as failure time, survival time, or event time. The event time can be measured in hours, days, weeks, months, quarters, semiannually, annually, or years. Examples include death occurrence or re-occurrence of the disease, marriage, and divorce.

The existence of partial data regarding the survival time of some individuals complicates the analysis of survival data. This partial data was majorly as result of the right censoring. Therefore, this study tried to find out the proportions of HIV/AIDS patients who survived past the study period, the hazard rate among the survivors who surpassed the study time and effects of specific variable on the probability of survival.

In order to estimate model parameters, survival models correctly incorporate information from both uncensored and censored data.

Modeling strategies for survival data falls into three categories comprising of nonparametric, parametric, and semi-parametric analysis [10]. The extent of parametric assumption reliance influences the classifications of these models. When the true distribution is not known or difficult to approximate, non-parametric models are used which does not require assumptions about survival. These include Life tables and Kaplan Meier estimation. When the interest is in the association between time to event and covariates, parametric and semi-parametric models are applicable. For parametric models, all the covariates in the model require a specification and full characterization of the hazard function and the Accelerated Failure Time (AFC) model is an example. The semi-parametric model, which contains both parametric and non-parametric components, is more flexible and make less assumption about the distribution of survival data. In addition to these attributes, the model still allows for estimation of relative hazard between covariates to evaluate the effects of explanatory variables with or without specification of the baseline distribution. An example is the Cox Proportional Hazard model [4].

The outcome of interest is composed of two important parts; the time until an event occurred and the event status. These guides in the estimation of survival probabilities and hazard ratios which are essential in describing the distribution of survival time and effects of specific variable on the survival of the patients. The survival function gives the probability of not experiencing the event and the hazard function measures potential risk of death given that an individual patient survives beyond the study period.

In regards to censoring, this research considered right censoring defined as the censoring, which occurs when an individual patient survives beyond the study time.

The underlying principle behind Cox proportional hazard regression model is that two or more individual patients have constant proportional hazard ratios which do not depend on time [13].

The covariates are classified either as fixed or time-

dependent. The hazard ratio depends only on baseline covariates such as age and sex. However, when it is necessary to examine if continuous covariates contributes to the risk of death the extended Cox model is useful [2].

In this study, the focus was on the Cox Proportion Hazard model [4]. The model considers the effects of numerous explanatory variables at a given time and investigate the association of survival distribution with the variables considered. Hazard function is the dependent variable at given time ( $t$ ). In order to approximate Parameter values of the CPHRM, the partial likelihood weights are maximized. In this model the log hazard of an individual patient is a linear function of their static covariates and population -level baseline hazard function that varies over time.

Thus, the adopted CPHRM was given by:

$$\lambda(t/x) = \beta_o(t) \exp(\sum_{i=1}^n \beta_i(x_i))$$

Where  $\beta_o(t)$  is the baseline hazard function and  $\beta_i$  is the regression coefficient of the corresponding  $x_i$ .

The mathematical sign that precedes the regression coefficients'  $\beta_i$ , is significant in drawing inference in the sense that a positive sign for  $\beta_i$  means the risk of occurrence of an event was higher. Conversely, a negative sign means that the risk of occurrence of an event (for this case death) was lower.

Similarly, the magnitude of the regression coefficient is useful in making inference and drawing conclusions in that a value of the hazard ratio equating to one (1.00) has no effect on the risk of failure whereas less than one (1.00) decreases the hazard and greater than one (1.00) increases the failure rate.

The assumptions considered in CPHRM applied in this study include: - multiplicative effects of the covariates on the hazard ratio, the hazard ratios of two subjects' remains the same at all time and the end times (death) of the patients are autonomous of each other.

For successful implementation of ART care among HIV/AIDS patients, there is need for such quantitative statistical analysis in appropriate settings to be able establish determinants of survival time among AIDS patients on antiretroviral therapy in MTRH, Kenya.

## 2. Methods

### 2.1. Study Design, Area and Population

A longitudinal study design was adopted. The target population were HIV/AIDS patients who were 18years and above during the start of ART in MTRH (Eldoret, Kenya). The patients were enrolled in January 2005 and were followed up to January 2007 for a period of ten years.

### 2.2. Data Collection Instruments

The required data set was extracted from an existing research database at MTRH. The data obtained was de-identified before sharing to the researcher to enhance the study subject's privacy and confidentiality. The research data were retrospectively collected from MTRH antiretroviral

database management information system, in Ampath, Eldoret Kenya.

Therefore, patients attending Ampath programme had a unique ART number and this was a requirement for each patient who voluntarily wished to start ART according to the national treatment guidelines [12].

### 2.3. Study Variables

The variables which were considered in this research were

classified into dependent and independent variables

#### 2.3.1. Dependent Variable

The dependent variable was survival time of HIV/ AIDS patients'

#### 2.3.2. Independent Variables

The predictor variables and their levels were recorded as follows:

**Table 1.** Independent variables considered and the levels.

Variable	Descriptions	Coding	Data type
Gender	The gender of the respondents	0-Female 1-Male	Binary
Age group (Years)	This describes the age group of the respondents	0- 18-29 1- 30-39 2 $\geq$ 40	Ordinal
Marital status	The marital status of the patients	0- Never married 1- Married 2- Divorced/Separated 3-Widowed	Nominal
Education Status	The education of the patients under study	0-Primary 1-Secondary 2- Tertiary	Ordinal
Employment status	The employment status of the respondents	0- None 1- Paid employee 2- Peasant 3-Other	Ordinal
Body Mass Index (BMI)	The ratio of body weight to the square of patient's height of the respondents in kg/m <sup>2</sup>	0- $< 18.5$ 1- $> 18.5 - < 25$ 2- $25 - < 30$ 3- $> 30$	Ordinal
WHO stage	The WHO staging level of the respondents	0- Stage II 1- Stage III 2- Stage IV	Ordinal
Drug Adherence	Drug adherence status of the patients	0- perfect 1- non- perfect	Binary
House holds	The number of people in the household	0=1-2 1=3-4 2- $> 4$	Ordinal
CD4 count (HIV severity)	The Cell Differential four count of the patients	0- $< 350$ 1- $\geq 350$	Binary
Opportunistic infections	Patient having opportunistic infections	0- yes 1- no	Binary
Functional status	The body functional status of the patient	0- Bedridden 1- Ambulatory 2- working	Nominal
Drug abuse	The use of alcohol and other drugs abuse	0- yes 1- no	Binary
Condom use	The use of condoms	0- never 1- sometimes 2- mostly 3- always	Nominal

### 2.4. Data Analysis Techniques

Data management was done using Statistical Analysis system (SAS) version 13, this means that data was cleaned with the aim of removing outliers hence remaining with the desired patients' characteristics before being analyzed.

Data analysis was done at two levels using R- software,

first, a descriptive baseline summary of the HIV/AIDS patient's demographic and clinical characteristics was made using frequency distributions and summary statistics where appropriate.

Second, an analysis of disparities in survival times of demographic and clinical characteristics was done using the log rank for both unadjusted and adjusted Cox models. The

purpose of the analysis was to establish predictor variables of survival in HIV/AIDS patients on antiretroviral treatment.

## 2.5. The Cox Proportional Hazard Model

### 2.5.1. The Survival Time Predictors for Adult HIV/AIDS Patients on ART in Kenya

The CPHRM is the basic regression model applied on survival data. This model was used to investigate the determinant factors of survival time in HIV/AIDS patients on ART in MTRH. This model was introduced by [4] and was named after him. The CPHRM is multiple regression approach to analyze the connection between survival times and one or more predictor variables or covariates. These predictor variables may dependent or independent of time. The Cox proportional hazard regression model is function of a set of explanatory variables ( $x$ ) and baseline hazard function. This model characterizes how the hazard function changes as a function of survival time and  $r \{(X', \beta), \beta\}$  as a function of subject covariates.

The data structure was in the following form:  $\{(t_1, \delta_1, x_1), \dots, (t_n, \delta_n, x_n)\}$  where;  $t_i$  is the observed survival time for  $i^{\text{th}}$  individual,  $\delta_i$  is an indicator of censoring, and  $x_i$  are explanatory variables.

Let  $X' = (x_1, x_2, \dots, x_p)$  and  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$

For  $(X', \beta) = \exp(X' \cdot \beta)$  then the hazard ratio is:

$$\lambda(t, X', \beta) = \lambda_0(t) \exp(X' \cdot \beta) \quad (1)$$

The model in equation (1) is known as the Cox Proportional Hazard Model.

Where:  $X'$  are  $P \times 1$  vector of predictor variables.

$\beta$  Is a vector of  $1 \times P$  of parameters.

Where  $X'_i$ 's is the vector of values of the explanatory variables for the  $i^{\text{th}}$  individual at time  $t$  and  $\beta$  is the vector of unknown regression parameters that are assumed to be the same for all individuals in the study, which measures the influence of the covariate on the survival experience. The Cox model formula has the property that if  $X'$ 's is entirely equal to zero, the formula reduces to the baseline hazard function. That is, the exponential part of the formula  $e^0$ , which is 1. This property of the Cox model is the reason why  $\lambda_0(t)$  it is called the baseline function.

Note that the cumulative hazard function at time  $t$  for a subject is given by:

$$\Lambda(t, x, \beta) = \int_0^t \lambda(u, x, \beta) du = r(X', \beta) \int_0^t \lambda_0(u) du = r(X', \beta, \Lambda_0(t)) \quad (2)$$

The survival function for the semi parametric hazard model is

$$S(t, X', \beta) = [S_0(T) \exp(X' \cdot \beta)] \quad (3)$$

And in the equation,  $S_0(T)$  is the baseline survival function.

The CPHRM is widely and commonly used on survival data and can handle covariate of interest that are both time-dependent and independent. R statistical software was used

to analyze patients' data.

### 2.5.2. Fitting the Cox Proportional Hazard Regression Model

To obtain the maximum likelihood (ML) estimates of the CPHRM parameters the likelihood function ( $L$ ) is maximized. 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 ( $\beta$ 's) in the model being considered. The  $L$  is sometimes written notational as  $L(\beta)$  where  $\beta$  denotes the collection of unknown parameters.

Suppose the survival data is represented by  $(t_i, \delta_i, X_i)$  for  $i=1, 2, \dots, n$  where  $t_i$  the length of time a subject is observed (survival time),  $\delta_i$  an indicator of censoring for the  $i^{\text{th}}$  individual and  $X_i$  a vector of covariates for the  $i^{\text{th}}$  individual. The likelihood for right-censored data includes both the survival and hazard functions given by equation (4)

$$L(\beta) = \prod_{i=1}^m [\lambda(t_i, X_i, \beta)]^{\delta_i} S(t_i, X_i, \beta) \quad (4)$$

The proposed partial likelihood function avoids specification of the baseline hazard function, treating it as a parameter of nuisance and excluding it from the equation of estimation due to its assumption that there are no tied values among the observed survival times. Suppose we have  $m$  distinct failure times and let  $X_{(i)}$  is the vector of covariates at ordered failure time  $t(i)$ . We define the Partial Likelihood by the following expression:

$$L_P(\beta) = \prod_i^m \left[ \frac{e^{X_i \cdot \beta}}{\sum_{j \in R(t_i)} e^{X_j \cdot \beta}} \right]^{d_i} \quad (5)$$

Where  $d_i$  is the number of deaths, we assume there are no deaths so excluded for  $d_i=0$ . And,  $R(t_i)$  the set of subjects at risk at the time just before  $t_i$  ( $t_i-0$ ). And the summation in the denominator is over all subjects in the risk set at time  $t_i$  denoted by  $R(t_i)$

The log partial likelihood function is given as:

$$L_P(\beta) = \sum_{i=1}^m \{X_i \cdot \beta - \ln \left[ \sum_{j \in R(t_i)} e^{X_j \cdot \beta} \right]\} \quad (6)$$

We obtain the maximum partial likelihood estimates by differentiating the above equation with respect to  $\beta$ , setting the derivative equal to zero and solving for the unknown parameter.

Interpretation of the fitted CPHRM was based on the hazard function i.e.,  $e^{\hat{\beta}}$  is the maximum partial likelihood estimator of  $\beta$ .

The  $(1-\alpha) * 100\%$  confidence interval (C.I) for the parameter estimates is given by equation (3.16) below.

$$\hat{\beta} \pm Z_{\alpha/2} \text{ S.e}(\hat{\beta}) \quad (7)$$

The hazard ratios confidence intervals are obtained by

exponentiation of the parameter estimates.

$$\exp \hat{\beta} \pm Z_{\alpha/2} \text{S.e}(\hat{\beta}) \quad (8)$$

## 2.6. Ethical Clearance, Data Safety and Confidentiality

The ethical clearances were obtained from MTRH and Institutional Research Ethics Committee (Moi University). However, because the study was conducted using secondary data individual pediatric patients' consent was not required. The retrieved research database used only patients' identification numbers that were uniquely linked to the patient. However, the patient's identification numbers were changed to serial numbers by the assigned supervisor responsible for ART data in Ampath before to the researcher encrypted laptop for security and confidentiality.

## 2.7. Limitations to the Study

The limitations included use of data from only one hospital. Consequently, the findings were to be interpreted carefully if anyone intends to draw inference at the national level. In addition, there was insufficient published literatures in Kenya related to the study. Finally, loss to follow up resulted to missing the data for some patients who dropped from the study.

# 3. Results

## 3.1. Fitting the Cox Proportional Hazard Model

In order to study the relationship between survival time and predictor variables, a regression modeling approach to survival data using Cox proportional hazard regression model was used with the aim of estimating the regression coefficients, performing statistical test, construction of confidence intervals and drawing inference based on the hazard function. Model development and its adequacy assessment was done before drawing inference from the results of the fitted model. The main objective of model development was to attain a model which describes the data optimally.

### 3.1.1. Model Development

The first procedure in Model development process was to select variable which were important for the study. The selection process was achieved by considering clinical importance of the variables, statistical significance and adjustment for confounding.

Once the variables were selected the univariate Cox proportional hazard regression model was fitted to the selected covariate data using log likelihood and partial log likelihood functions from equations (4) and (6) respectively. This result to unadjusted Cox proportional model in table 4. The confidence intervals for the univariate model were constructed using equation (8).

Purposeful selection of the covariates was the most useful method for selecting variables to the initial multivariable model. This began with a model that contained all variables which were significant in univariate at 20-25% level in relation to survival time among HIV/AIDS patients enrolled

at Ampath center in MTRH.

The six variables found statistically significant in univariate analysis were the candidates for inclusion in the initial multivariate Cox model. These variables included; HIV severity, age, BIM, education level, gender and ARV adherence. HIV severity variable was created by merging the CD4 count and WHO staging to avoid multi-collinearity of the variables. However, marital status, number of households, employment status, presence of opportunistic infections, patient functional status, drug abuse and condom use were not significant at 20-25% level of significance and hence were excluded from initial multivariate Cox model.

Continuous covariates i.e., age, CD4 count and BMI were categorized in basis of quartile and median values. Their significance was studied in two schemes, in categorical and continuous form as seen in table 1. These continuous covariates were significant in both coding schemes at 5% significance level hence selected for inclusion in the initial adjusted Cox multivariate model in continuous form.

By method of likelihood ratio tests, HIV severity, age, BMI, education level, gender and ARV regimens adherence were found statistically significant variables.

After fitting the initial multivariate model using equations (4), (6) and (8). The second procedure in modelling process was to check assumption of linearity for the continuous covariates in the model i.e., age, BMI and CD4 count using smoothed plots of Martingale residuals. The plots demonstrated linearity i.e., the plots were found to be random, showing no systematic patterns and approximated a straight line for each continuous covariate. As a result, they were linear in the model.

The final step in model development process was the consideration of interaction term for inclusion in the model for purpose of improving inference and obtain a more realistic model. This step began with creation of a list of plausible interactions formed between the six variables of the initial Cox model and were found to be statistically significant at 5% level in prediction of the hazard rates.

Now, each significant interaction was added back to the initial Cox multivariate model and fitted again using forward likelihood function.

Therefore, the Cox model in table 4 remains as the "final model" but interpretation based on this model should not be made until its fit and adherence to the model assumption are checked.

### 3.1.2. Model Assessment

In order to evaluate how well the fitted Cox regression model describes the data set, model adequacy was assessed on the preliminary final model which fulfilled the model development stages. The prerequisites for model assessment included testing the assumption of the proportional hazards, checking the presence of leverage and measuring the overall goodness of fit of the model.

The basic assumption of the proportional hazard model was that the hazard ratios were constant over time. This implies that the risk of failure was the same no matter how

long the subjects were followed up. The proportional hazard assumption was tested through creation of an interaction between variables and logarithm of time (time dependent covariates) and checking their significance in predicting hazard ratios. The results in table 4 demonstrates that all time dependent covariates were not significant, this assures that there was 'proportional hazard.' Furthermore, to test for the assumption above for each covariate, the function *cox.zph()* in the *r* codes was used to correlate the corresponding sets of scaled Schoenfeld residuals with time and performed global test for the model as whole as displayed by table 3 below

**Table 2.** Test for proportional hazard assumption for each covariate and global test.

Variable	Chi-square (chisq.)	Degrees of freedom (d.f)	p-value
HIV Severity	0.124	1	0.725
Age category	2.023	2	0.364
BMI category	2.689	2	0.261
Level of education (lower)	2.286	1	0.131
Gender (male)	3.426	1	0.064
ARV perfect adherence	0.624	1	0.429
Global test	10.187	8	0.252

From the above output, the test for independence between scaled Schoenfeld residuals and time was not statistically significant. In addition, global test is not statistically significant. This confirms that the final model met proportional hazard assumption.

Graphically checking of proportional hazard was made by plotting scaled Schoenfeld residuals of the continuous variables with the corresponding survival time through lowess plots. The graph showed that each plot was random, smooth and had approximately zero slopes. The plots had parallel shape pattern, implying that there was no time dependent covariate, hence no violation of proportional hazard assumption.

The next important aspect of model evaluation was thorough examination of regression diagnostic statistics to identify which subjects had unusual configuration of covariates i.e., to identify which subjects exerts undue influence on the estimates of the parameters and had undue influence on the fit of the model. An outlier is an extreme observation and by examining Scaled Score residuals and normal probability plots are helpful in identifying outliers. The effects of outliers on the proportional hazard regression were easily checked by dropping these points and refitting the regression equation. In this study outliers were examined by Scaled Score residuals. The final step in assessment of the model adequacy was the measure the overall goodness of fit. All measures depend on the proportion of values that are censored. A perfect adequate model has low  $R^2$  due to presence of censored data [9].

### 3.2. Investigating Predictors of Survival Time Among HIV Patients on ART at MTRH

This section introduces how to investigate the final multivariate Cox proportional hazard model that fulfilled the

above procedures. In this study, the Cox model had HIV severity, age, Body Mass Index (BMI), level of education and ARV perfect adherence as its explanatory or predictor variables and time to the death as the dependent variable also known as survival time. The Cox regression parameters ( $\beta$ ) were obtained by taking natural logarithm ( $\ln$ ) of the respective proportional hazard values of the variables however, in cases where the variables had categories, the parameter was treated first as stratum and in order to obtain its value, the hazard values were multiplied for all categories within the variable and then takes natural logarithm of the product to obtain the parameter.

**Table 3.** Variables and parameter estimates for both unadjusted and adjusted Cox proportional model.

Covariate of interest	Symbol	Unadjusted parameter( $\beta$ )	Adjusted parameter ( $\beta$ )
HIV severity (yes)	$x_1$	-0.32	+ 0.17
Age	$x_2$	-0.09	-1.22
BMI	$x_3$	-0.231	+0.35
Level of education (lower)	$x_4$	-0.05	-0.07
Gender (sex) male	$x_5$	+ 0.49	+ 0.63
ARV perfect adherence (yes)	$x_6$	-0.40	+0.33

#### 3.2.1. Unadjusted Cox Proportional Hazard Model

From above parameters' estimates, the unadjusted Cox proportion hazard regression models can thus be written as: -

$$\begin{aligned} \ln \{\lambda_t\} &= -0.32X_1 & \beta_1 &= -0.32 \\ \ln \{\lambda_t\} &= -0.09X_2 & \beta_2 &= -0.09 \\ \ln \{\lambda_t\} &= -0.23X_3 & \beta_3 &= -0.23 \\ \ln \{\lambda_t\} &= -0.05X_4 & \beta_4 &= -0.05 \\ \ln \{\lambda_t\} &= +0.49X_5 & \beta_5 &= +0.49 \\ \ln \{\lambda_t\} &= -0.40X_6 & \beta_6 &= -0.40 \end{aligned}$$

Only gender (male,  $X_5$ ) of the patients' indicated worse prognosis i.e., it was associated with increased risk of death or decreased survival because its parameter had positive coefficient whereas patients' severity ( $X_1$ ), age ( $X_2$ ), BMI ( $X_3$ ), lower level of education ( $X_4$ ) and perfect adherence to antiretroviral drugs ( $X_6$ ) variables had protective effect (i.e., associated with improved survival), since their coefficients are negative.

#### 3.2.2. Adjusted Cox Proportional Hazard Model

$$\ln \{\lambda_t\} = 0.17X_1 - 1.22X_2 + 0.35X_3 - 0.07X_4 + 0.63X_5 + 0.33X_6$$

In the adjusted model above, HIV severity ( $X_1$ ), BMI ( $X_3$ ), gender ( $X_5$ ), and ARV perfect adherence ( $X_6$ ) had positive parameter coefficients which indicated that these predictor variables are associated with worse prognosis or decreased survival among HIV/AIDS patients on ART in MTRH. While age ( $X_2$ ) and lower-level education ( $X_4$ ) had negative parameter coefficient indicating that it had protecting effect or associated with improved survival or decreased risk of the death among HIV/ AIDS patients.

The change in parameter coefficients estimates sign (from negative to positive) for HIV severity, BMI and ARV perfect adherence was due to the multicollinearity between the above variables and male gender variable.

**Table 4.** Unadjusted/Adjusted Hazard Ratios of factors associated with longer survival time among HIV patients on ART in MTRH.

Covariates of interest	Unadjusted HR (95%CI)	P-value	Adjusted HR (95%CI)	P-value
HIV Severity: Yes	0.729 (0.547, 0.972)	0.032	1.18 (0.451, 3.085)	0.735
Age Cat-<30	Ref		Ref	
Age Cat-30-40	0.824 (0.565, 1.201)	0.314	0.459 (0.148, 1.424)	0.178
Age Cat-40+	1.114 (0.77, 1.609)	0.567	0.644 (0.207, 2.003)	0.447
BMI Cat-18.5-<25	Ref		Ref	
BMI Cat-<18.5	1.383 (0.564, 3.393)	0.478	1.65 (0.634, 4.294)	0.305
BMI Cat-25+	0.574 (0.131, 2.51)	0.461	0.861 (0.188, 3.948)	0.847
Level of education (lower).	0.952 (0.914, 0.992)	0.019	0.931 (0.833, 1.041)	0.209
Sex: Male	1.633 (1.239, 2.152)	<.0001	1.884 (0.731, 4.855)	0.19
ARV_perfect_adherence: Yes	0.668 (0.506, 0.882)	0.004	1.393 (0.534, 3.633)	0.498

Ref=Reference category.

### 3.2.3. The Unadjusted Hazard Ratios

From the above Cox model (table 3), the unadjusted hazard ratio, confidence interval and p-value for each explanatory variable considered in this study were discussed as follows:

Results of the unadjusted Cox model showed that patients who were HIV/AIDS severe had less risk of death or associated with improved survival since the UHR was less than 1.00 i.e. [UHR=0.729, 95% C.I=0.547, 0.972 p-value=0.032]. This association was statistically significant since its 95% C.I does not contain the null value (1) and the p-value was less than 0.05. However, patients who were HIV/AIDS severe had 72.9% risk of death compared to 27.2% among non-severe patients.

HIV/AIDS patients with age category between 30 to 40 years had decreased risk of death or associated with improved survival compared to the reference age category less or equal to 30 years, with [UHR=0.824, 95%CI=0.565-1.201, p-value=0.314] whereas patients with age category 40 years and above had 11% higher risk of death compared to the same reference age category with [UHR=1.11, 95%CI=0.77-1.609, p-value=0.567]. Both associations were not statistically significant since their 95% C.I contained the null value (1) and their p-values were greater than 0.05.

Patients with BMI less than 18.5kg/m<sup>2</sup> had 38% higher risk of death or decreased survival rate compared to the reference BMI category of 18.5-<25KG/M<sup>2</sup>, with [UHR=1.383, 95%CI=0.564-3.393, p-value=0.478], whereas patients with B.M.I category ≥ 25kg/m<sup>2</sup> had increased chances of survival or less risk of death as compared with the same reference category with [UHR=0.574, 95%CI=0.131-2.51, p-value=0.461]. Both associations were not statistically significant since their 95% C.I contained the null value (1) and their p-values were greater than 0.05.

Lower level of education had protective effect on the risk of death (improved survival) since the UHR was less than (1.00) i.e., its unadjusted hazard ratio was [UHR=0.952, 95%CI=0.914-0.992, p-value=0.019]. This association was statistically significant in that the 95% C.I does not contain the null value (1) and its p-value was less than 0.05. However, patients with lower level of education had higher risk of death (95.2%) compared to those who had higher levels of education, with 4.8% risk of death.

Male patients had 63% higher risk of death compared to female patients on ART in MTRH with [UHR=1.633, 95% C.I=1.239-2.152, p-value < 0.0001]. The association was highly statistically significant in that the 95% C.I does not contain the null value (1) and its p-value was approximately 0.00.

Patients who perfectly adhered to ARV had decreased risk of death or had enhanced survival compared to those who did not adhere, with unadjusted hazard rate ratio [UHR=0.668, 95%CI=0.506-0.882, p-value 0.004]. The association was highly statistically significant in that the 95% C.I does not contain the null value (1) and its p-value was approximately 0.00.

### 3.3. Adjusted Hazard Ratios

From Cox model (table 4), the adjusted hazard ratio, confidence interval and respective p-values for each explanatory variable considered while adjusting for the other variables were as follows:

HIV/AIDS patients' with severe cases had 18% higher risk of death as compared with non-severe patients' when age, BMI, school years completed, gender and ARV Adherence was adjusted for with adjusted hazard ratio [AHR=1.18, 95%CI=0.451-3.085, p-value=0.735]. Though the association was statistically insignificant since the 95% C.I for AHR contained the null value (1) and the p-value was greater than 0.05.

Patients' with age category of between 30 to 40 years had less risk of death compared to reference age category less or equal to 30 years when other variables were adjusted for with [AHR=0.459, 95%CI=0.148-1.424), p-value 0.178]. However, HIV/AIDS patients' with age category 40 years above had higher risk of death compared to the same reference age category with [AHR=0.644, 95% C I=0.207-2.003, p-value=0.447]. Both associations were not statistically significant in that their 95% C.I contained the null value (1) and their p-values were greater than 0.05.

The patient's with BMI less than 18.5kg/m<sup>2</sup> had 65% higher risk of death compared to the reference BMI category of 18.5-<25 when other covariates in the study were adjusted for in the model, with [AHR=1.65, 95%CI=0.634-4.294, p-value=0.305]. However, for patients with B.M.I ≥ 25kg/m<sup>2</sup> had increased chances of survival as compared to the same

BMI reference category with [AHR=0.861, 95%CI=0.188-3.948, p-value 0.847]. However, both associations were not statistically significant since their 95% C.I contained the null value (1) and their p-values were greater than 0.05.

Lower level of education had protective effect on the risk of death (improved survival) when other variables considered in the model were adjusted for, since the AHR was less than (1.00) i.e., its adjusted hazard ratio was [AHR=0.931, 95%CI 0.833-1.041, p-value 0.209]. However, Patients' who had lower level of education had higher risk of death (decreased survival) i.e., 93.1% compared to those who had completed higher levels of education when other covariates were adjusted for, although the relationship was insignificant since the 95% C.I contained the null value (1) and its p-values was greater than 0.05.

When HIV severity, age, BMI, school years completed and perfect adherence to ARV were adjusted for in the model, the male patients had 88% higher risk of death (decreased chances of survival) compared to the female counterpart with [AHR=1.884, 95%CI 0.731- 4.855, p-value=0.19]. However, the association was statistically insignificant since the 95% C.I contained the null value (1) and their p-value was greater than 0.05.

Patients' who perfectly adhered to ARV had 39.3% higher risk of death compared to those who did not perfectly adhered to the ARV drugs, with adjusted hazard ratio [AHR=1.393, 95%CI=0.534- 3.633], p-value 0.498] when other covariates are adjusted for in the Cox model. However, the association was statistically insignificant since the 95% C.I contained the null value (1) and its p-value was greater than 0.05.

## 4. Discussion

In this study, Cox proportional hazard model was applied to estimate the predictor variables of survival time among HIV/AIDS patients under ART in Moi Teaching and Referral Hospital (MTRH). The ten years retrospective adult study of HIV/AIDS patients' in MTRH (ART) Center gave an insight into survival time patterns of patients and its determinants in a hospital setting in Kenya. The Cox proportional hazard regression model was fitted to the data. HIV severity, age, BMI, level of education, gender and perfect adherence to the ARV were explanatory variables and survival time was dependent variable in the study. This was similar to all past studies that considered fitting the Cox model to medical human data [8] & [14]. However, for this study Patients' data were classified into severe and non-severe cases of the HIV/AIDS disease based on the level of CD4 cell count and W.H.O staging, where patients with CD4 count less than 350cells/mm<sup>3</sup> (stage III and IV) have been classified to be severe and those with CD4 count equal or more than 350 cells/mm<sup>3</sup> (stage I and II) classified as non-severe cases of HIV disease. The reason for this classification is to avoid the multi-collinearity effects of the two variables i.e., WHO staging and CD4 count clustering.

HIV/AIDS severity was statistically significant predictor

of survival time when other factors were held constant i. e. It was associated with improved survival (protective variable) with [UHR=0.729, 95%CI=0.547- 0.972, p-value=0.032]. This means that patients with severe cases of HIV/AIDS would die at the rate of 0.729 times or would be 1.371 years death free from HIV/AIDS related cases, and statistically significant in the sense that its confidence interval does not contain the null value i.e. (1) and its p-value less than 0.05. The reason for the above observation was that patients with severe cases would tend to strictly adhere to the ARV drugs and ART guidelines compared to the non-severe cases in order for them to improve their disease conditions.

However, adjusting for other covariates included in the model, increases the risk of death by 18% among patients with severe cases as compared to those with non-severe disease cases or decreases the years of death free time to 0.85 years. Although the association was not statistically significant due to interrelationships of the considered covariates in the model. This results was consistent with most of the past studies in Kenya [15], who established that WHO stage IV was found the main predictor of death i.e. patients' in stage IV had lesser chance of survival this was supported by most of the African nations studies [8, 14]. Age of the patients in this study was statistically insignificant predictor of survival time in both univariate and multivariate Cox models since the C.I for both unadjusted and adjusted hazard ratios contain the null value (1) and their p-values were greater than 0.05 i.e., UHR for age category 30-40 and above 40 years were [UHR=0.824, 95%CI=0.565-1.201, p-value=0.314] and [UHR=1.11, 95%CI=0.77-1.609, p-value=0.567] respectively. Also, AHR for the respective age category were [AHR=0.459, 95%CI=0.148-1.424), p-value 0.178] and [AHR=0.644, 95% C I=0.207- 2.003, p-value=0.447]. Nevertheless, when survivorship of categorized aged group was done it was established that the rate of failure increases with increase in the age of the patients for both unadjusted and adjusted Cox models, table (4), i.e. the failure rate for age brackets 30-40 years and 40 years were 82.4% and 111.4% for unadjusted and 45.9% and 64.4% for the adjusted models respectively. The value above 100% signifies decreased survival time or increased rate of failure in reference to  $\leq 30$  years age category. The above results could also be interpreted to mean, patients who were 30-40 age category would be 1.21 years and 2.18 years death free from HIV/AIDS defining illness for unadjusted and adjusted Cox models respectively. Conversely patients with age bracket  $\geq 40$  would be 0.90 years and 1.55 years death free from AIDS defining illness for unadjusted and adjusted models respectively. These results indicated that there was significant difference in the survival rates within age category. This was consistent with the studies done in Ethiopia [14] who found out that HIV/AIDS patients who were above 60 years had shortest survival time as compared to other age categories below that age group. This was attributed to decrease in body physiological functions for instance decreases in cell differentiation rendering the body to have incompetent immunity status and thus would be at higher risk

of complications and respond poorly to ART.

Furthermore, there was decreased risk of death within age categories when other factors were adjusted for in the model. This was due to multicollinearity among protective univariate variables (HIV severity, level of education and perfect adherence to ARV) with age of the patients.

Body Mass Index was statistically insignificant predictor of survival in both univariate and multivariate Cox models in (table 4) for the reason that their 95% C.I contained the null value (1) and their p-values were greater than 0.05. This was marked by the UHR of [UHR=1.383, 95%CI=0.564-3.393, p-value=0.478] for BMI < 18.5KG/M<sup>2</sup> and [UHR=0.574, 95%CI=0.131-2.51, p-value=0.461] for BMI ≥ 25KG/M<sup>2</sup> in addition the adjusted hazard ratios of [AHR=1.65, 95%CI=0.634-4.294, p-value=0.305] for BMI < 18.5KG/M<sup>2</sup> and [AHR=0.861, 95%CI=0.188-3.948, p-value 0.847] for BMI ≥ 25KG/M<sup>2</sup>. This was not in agreement with many other studies which found out that BMI was statistically significant predictor of survival. However, some of the studies uses weight as a proxy measure of BMI because most hospitals do not record heights of the patients. One study Similar to this was done in Tanzania yielded similar results that patients with lower body mass index had higher mortality in comparison to those with higher BMI. This was evident by the failure rate of 138.3% and 57.4% for BMI < 18.5kg/m<sup>2</sup> and ≥ 25kg/m<sup>2</sup> in reference to BMI category of ≥ 18.5-≤ 25 kg/m<sup>2</sup> respectively in univariate Cox models (table 4). Similarly, there was decreased risk of death with increase in BMI among HIV/AIDS positive patients on ART when other factors considered in the multivariate Cox model table (4) were adjusted This observation was validated by failure rate of 165% and 86.1% for BMI < 18.5kg/m<sup>2</sup> and ≥ 25kg/m<sup>2</sup> Respectively.

Alternative inference on the results was that patients with BMI < 18.5kg/m<sup>2</sup> were 0.72- and 0.61-years death free for unadjusted and adjusted models respectively. Moreover, patients with BMI ≥ 25KG/M<sup>2</sup> were 1.74- and 1.16-years death free for unadjusted and adjusted models respectively from AIDS defining illnesses.

From above results it was observed that adjusting for other covariates in multivariate model increase the risk of death within categories of BMI, due to interrelationship between HIV severities, male gender, non-adherence to ARV and BMI.

BMI contribute to drug metabolism and would therefore have an effect to the efficacy of HAART. Patients with lower BMI tend to have high viral loads than those with higher BMI category due to emaciated CD4, CD8 and T lymphocyte cells caused by malnutrition hence increases the risk of death.

Lower level of education (primary or no school) was significant predictor of improved survival time in univariate Cox model with unadjusted hazard ratio of [UHR=0.952, 95%CI=0.914-0.992, p-value=0.019] since the 95% C.I does not contain the null value (1) and its p-value was less than 0.05. However, the same level of education was not statistically significant in the multivariate model when other covariates were adjusted. i.e. The adjusted hazard ratio was [AHR=0.931, 95%CI 0.833-1.041, p-value 0.209] contained the null value (1) and its p-value was greater than 0.05.

However, patients with lower level of education had 95.2% risk of death compared to those with higher levels (tertiary level) patients with 4.8% risk of death. These analogous results would also be interpreted to mean, patients with lower level of education would be 1.05 years death free from AIDS defining illness compared to 2.08 years of death free among patients with higher level of education. The results were statistically significant in univariate Cox model which was consistent in to other comparable studies [16]. The observation was attributed to better know-how by patients with higher level of education about the importance of drug compliance, their ability to afford balance diet, their understanding on the need for earlier presentation for hospitalization (ART) and rightful medication procedures. In addition, higher education reduces stigmatization attitude.

Adjusting for other covariates in the study decreases the risk of death among patients with lower level of education from 95.2% to 93.1% or increases the time of death free from 1.05 years to 1.07 years this was due to interrelationship between variables with protective effects (non-severity and perfect adherence to ARV) and level of education.

Gender (male) was the most statistically significant predictor of death in univariate Cox model with UHR of [UHR=1.633, 95% C.I=1.239-2.152, p-value < 0.0001] reason being its 95% C.I does not contain the null value (1) and the p- value was less than 0.05 This output was consistent to most of the early studies for instance [5], established that most men were on higher risk of death in comparison to women patients.

Conversely gender (male) was not statistically significant predictor of survival time in multivariate Cox model with [AHR=1.884, 95%CI 0.731- 4.855, p-value=0.19], the reason was that its 95% C.I contained the null value (1) and its p-value was greater than 0.05.

From table 4 male patients had 63.3% and 88.4% higher risk of death compared to the female patients in univariate and multivariate Cox models respectively. This discrepancy was accounted for by late diagnosis among male patients i.e., living in denial for a long time, while female patients undergo compulsory testing during both pre-natal and post-natal clinics visits.

The above results (gender variable) were also interpreted to mean; male patients were 0.61- and 0.53-years death free from AIDS defining illnesses in univariate and multivariate models respectively.

Adjusting for other covariates in multivariate model increases the risk of death among the male patients from 63.3% to 88.4% due to multicollinearity between variables associated with worst prognosis (HIV Severe cases, lower BMI, lower level of education, old age and non-adherence to ARV drugs) with gender variable.

Adherence of ART pertains intake of all ARV pills in correctly prescribed doses at the right time in the right way observing any dietary restrictions [14]. Successful antiretroviral therapy is dependent on sustaining high rate of adherence. The minimum level of adherence required for ARV to work effectively is 95% [5].

In this study Patients 'perfect adherence to ART was statistically significant predictor of survival time i.e., was associated with improved survival time in the univariate Cox model with [UHR=0.668, 95%CI=0.506-0.882, p-value 0.004] since its 95% C.I does not contain the null value (1) and the p- value was less than 0.05. This result was in tandem to similar comparable studies in Ethiopia, [3] and [6]. However, analogous comparable study in South Africa [11] showed that estimated survivorship was statistically significant different. Also a Kampala study in Uganda, [1] established that non-adherence was found a significant predictor of survival time.

Patients' perfect adherence to ART was not statistically significant predictor of survival time in multivariate Cox model since the 95%CI contained the null value and its p-value was greater than 0.05 given by [AHR=1.393, 95%CI=0.534- 3.633], p-value 0.498].

Similarly, this result would be interpreted to mean patients with perfect adherence to ART were 1.50 years and 0.72 years death free from HIV/AIDS defining illness in unadjusted and adjusted models respectively.

Adjusting for other covariates in model (table 4) increases the risk of death by 39.3% among patients who had perfectly adhered to ART compared to the non-adherence due to the confounding effects of variables considered jointly in the multivariate model.

Non adherence to antiretroviral drugs results in treatment failure by increasing the chances of mutation that could lead to drug resistant virus and finally death.

## 5. Conclusions

After examination of the final estimated univariate Cox proportional regression models, it was established that HIV severity (yes), level of education (higher) and adherence to ARV (perfect) were significant predictors of longer survival time in MTRH. The reason being their estimated relative hazard ratio are all below 1.0 meaning they have protective effect (associated with improved survival time) hence longer survival, also excluded the null value (1) in their confidence interval and all had the p-value less than 0.05. In gender, male was a significant predictor of shorter survival time in the same hospital with a relative hazard ratio greater than 1.0, which was associated with increased risk of death (the 95% confidence interval exclude the null value and its p- value is less than 0.05). In the same crude models, age and BMI were not significant predictors of the survival time. None of the predictor variables in the adjusted multivariate model was found significant. This outcome was due to the interrelationship (confounding effects) of the considered covariates.

Patients with severe cases of HIV/AIDS had higher risk to death as compared to the non-severe cases. The female patients on ART had higher survival compared to male counterparts.

The risk of death increased with increase in age categories i.e., patient who occupied lower category of age (30 years

and below) had higher chance of survival as compared with those on higher category (40 years and above). Conversely an increase in body mass index decreases the risk of death, which increased survival time for patients.

The rate of survival increases with years of school completed. This means the higher school years completed the greater chance of living longer.

Perfect adherence of ARV at later stage of the disease (terminal stage) increases the risk of death as compared to non- perfect adherence.

Interaction effect of variables was assessed and there was no significant outcome of their association. However, within variable levels, evaluation of effect across categories indicated statistically significant difference with changes in patient survival time.

## 6. Recommendations

The main objective of establishment of ART service was to improve health of HIV infected patients hence prolonging their lives and reducing HIV related mortality. However, there is continued mortality within the ART. Based on the findings of the study, some of the feasible approaches of improving HIV/AIDS patient survival include: the government of kenya to spearhead development of policy framework for provision of regular compulsory screening services for male population to avoid late diagnosis and intervention of the HIV/AIDS disease and The intervention points to include; school opening days for primary, secondary and college children and male adults coming to seek treatment should begin with the HIV test first. Additionally, the MTRH to begin Patient counseling and visit training on importance of perfect adherence to ARV regimens at initial stage of HIV/AIDS infection.

## Appendix: SAS Codes

```
libname men 'D:\Mengich\Data';
```

```
PROC IMPORT OUT=WORK.men
  DATAFILE="D:\Mengich\Data\cleaned data.xlsx"
  DBMS=XLSX REPLACE;
  GETNAMES=YES;
  DATAROW=2;
RUN;
```

```
PROC IMPORT OUT=WORK.kib
  DATAFILE="D:\Mengich\Data\kib data.csv"
  DBMS=CSV REPLACE;
  GETNAMES=YES;
  DATAROW=2;
RUN;
```

```
*Check for duplicates;
proc sql; select count (distinct patient_identifier) into: num
from men;quit;
*Format dates;
```

```
/*data men0;set men;format
Date_of_last_clinic_encounter date9.;run;*/
```

```
*Explore dataset;
proc freq data=men;tables male_gender
ARV_perfect_adherence Death WHOSTage__enrolment
/missing;run;
```

```
proc means data=men n nmiss min max mean std median
p25 p75 maxdec=2;
var School_Years_Completed CD4__enrolment
age_at_therapeutic_ARV_start age_at_enrollment
Weight__enrolment;
run;
```

```
proc sql; select count (patient_identifier) into: num from
men where School_Years_Completed=0;quit;
```

```
proc sql; select count (patient_identifier) into: num from
men where WHOSTage__enrolment gt 3 | 0 <
CD4__enrolment <200;quit;
```

```
proc sql; select count (patient_identifier) into: num from
men where WHOSTage__enrolment=4 ;quit;
```

```
proc sql; select count (patient_identifier) into: num from
men where 0 < CD4__enrolment <200 ;quit;
```

```
proc print data=men (obs=10);var patient_identifier
WHOSTage__enrolment CD4__enrolment;where
WHOSTage__enrolment gt 3 | 0 < CD4__enrolment
<200;run;
```

```
*Derive analysis variables;
```

```
data men0;
```

```
set men;
```

```
*Create age categories (@ art therapeutic);
```

```
if age_at_therapeutic_ARV_start ge 18 and
age_at_therapeutic_ARV_start lt 30 then agecat_therap=1;
```

```
if age_at_therapeutic_ARV_start ge 30 and
age_at_therapeutic_ARV_start lt 40 then agecat_therap=2;
```

```
if age_at_therapeutic_ARV_start ge 40 and
age_at_therapeutic_ARV_start lt 55 then agecat_therap=3;
```

```
if age_at_therapeutic_ARV_start ge 55 then
agecat_therap=4;
```

```
*Create age categories (@ enrollment);
```

```
if age_at_enrollment ge 18 and age_at_enrollment lt 30
then agecat_enroll=1;
```

```
if age_at_enrollment ge 30 and age_at_enrollment lt 40
then agecat_enroll=2;
```

```
if age_at_enrollment ge 40 and age_at_enrollment lt 50
then agecat_enroll=3;
```

```
if age_at_enrollment ge 50 then agecat_enroll=4;
```

```
if age_at_enrollment ge 18 and age_at_enrollment lt 30
then agecat_enroll0=1;
```

```
if age_at_enrollment ge 30 and age_at_enrollment lt 45
then agecat_enroll0=2;
```

```
if age_at_enrollment ge 45 then agecat_enroll0=3;
```

```
*exlcude patients with WHOSTage greater than 3 or cd4
less than 200;
```

```
if WHOSTage__enrolment eq 4 or (CD4__enrolment ne .
& CD4__enrolment lt 200) then delete;*339 with WHO=4
& 4218 cd4 total=4303(overlap);
```

```
*create the HIV severity variable;
```

```
if WHOSTage__enrolment in (2,3) | (0 <
CD4__enrolment <=350) then H=1;
```

```
if WHOSTage__enrolment=1 | CD4__enrolment gt 350
then H=0;
```

```
*create education level variables;
```

```
if 0 < School_Years_Completed <=8 then educ_level=1;
```

```
if 8 < School_Years_Completed <=12 then educ_level=2;
```

```
if School_Years_Completed gt 12 then educ_level=3;
```

```
run;
```

```
*Explore dataset;
```

```
proc freq data=men0;tables male_gender
ARV_perfect_adherence Death WHOSTage__enrolment
/missing;run;
```

```
proc means data=men0 n nmiss min max mean std median
p25 p75 maxdec=2;
```

```
var School_Years_Completed CD4__enrolment
age_at_therapeutic_ARV_start age_at_enrollment
```

```
Weight__enrolment __days_prior_to_therapeutic_ARV
__days_post_therapeutic_ARV_star;
```

```
run;
```

```
proc freq data=men0;tables agecat_therap agecat_enroll
agecat_enroll0 H/missing;run;
```

```
*Create bmi at enrollment dataset;
```

```
*check visit date status,min and max - 01JAN2005 to
09FEB2018;
```

```
proc tabulate data=kib;var encounter_date;table
encounter_date,n nmiss (min max median)*f=date9.
range;run;
```

```
proc sort data=kib;by Patient_Identifier
encounter_date;run;
```

```
data bmi(rename=Body_Mass_Index=bmi_b
drop=encounter_date);
```

```
set kib (keep=Patient_Identifier encounter_date
Body_Mass_Index);
```

```
by Patient_Identifier encounter_date;
```

```
if Body_Mass_Index ne .;
```

```
if first.Patient_Identifier;
```

```
*create bmi_cat;
```

```
if Body_Mass_Index ne . & (Body_Mass_Index lt 15 or
Body_Mass_Index gt 40) then delete;
```

```
if Body_Mass_Index ne . & Body_Mass_Index lt 18.5
then bmi_cat=2;
```

```

if Body_Mass_Index ge 18.5 and Body_Mass_Index lt 25
then bmi_cat=1;
if Body_Mass_Index ge 25 and Body_Mass_Index lt 30
then bmi_cat=3;
if Body_Mass_Index ge 30 then bmi_cat=4;
run;

```

```

proc freq data=men0; tables agecat_therap agecat_enroll
H/missing; run;
proc means data=bmi n nmiss min max mean std median
p25 p75 maxdec=2; var bmi_b ; run;
proc freq data=bmi; tables bmi_cat/missing; run;

```

\*Merge data with the rest of the datasets and keep the analysis variables;

```

data men1
(rename=(__days_post_therapeutic_ARV_star=t_days
WHOSTage__enrolment=WHOSTage_b
CD4__enrolment=CD4_b));
merge men0(keep=Patient_Identifier male_gender
ARV_perfect_adherence agecat_therap agecat_enroll H
__days_post_therapeutic_ARV_star
WHOSTage__enrolment CD4__enrolment
School_Years_Completed Death agecat_enroll0 educ_level
age_at_therapeutic_ARV_start age_at_enrollment in=a) bmi;
by Patient_Identifier;
if a;

```

```

*create time object in years;
if __days_post_therapeutic_ARV_star eq . then delete;
t_yrs=__days_post_therapeutic_ARV_star/365.25;
if __days_post_therapeutic_ARV_star=0 then delete;*27
patients;

```

\*Create CD4 category;

```

if 200 <=CD4__enrolment <=350 then cd4_cat=0;
if CD4__enrolment gt 350 then cd4_cat=1;
run;

```

\*Check for duplicates;

```

proc sql; select count (distinct patient_identifier) into: num
from men1; quit;
proc means data=men1 n nmiss min max mean std median
p25 p75 maxdec=2; var School_Years_Completed t_days
t_yrs ; run;
proc sql; select count (patient_identifier) into: num from
men1 where t_yrs=0; quit;
proc sql; select count (patient_identifier) into: num from
men1 where CD4_b=200; quit;
proc sql; select count (patient_identifier) into: num from
men1 where School_Years_Completed=0; quit;
proc freq data=men1; tables Death H male_gender
educ_level cd4_cat/missing; run;

```

\*Export Analysis Dataset;

```

PROC EXPORT DATA=work.men1
OUTFILE="D:\Mengich\Data\men.csv"

```

```

DBMS=CSV REPLACE;
PUTNAMES=YES;
RUN;

```

## References

- [1] Abaasa, A. M., Todd, J., Ekoru, K., Kalyango, J. N., Levin, J., Odeke, E., & Karamagi, C. A. S. (2008). Good adherence to HAART and improved survival in a community HIV/AIDS treatment and care programme: The experience of the AIDS Support Organization (TASO), Kampala, Uganda. *BMC Health Services Research*, 8. <https://doi.org/10.1186/1472-6963-8-241>.
- [2] Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J. M., & Lesaffre, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*, 33 (18), 3167–3178. <https://doi.org/10.1002/sim.6158>.
- [3] Birtukan, T. (2010). *Predictors of Early Death in HIV Positive Individuals Enrolled for Chronic HIV Care in Jimma Zone, South Western Ethiopia*. Addis Abeba, Ethiopia: University.
- [4] Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34 (2), 187–202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>.
- [5] Dombrowski, J. C., Kitahata, M. M., Van Rompaey, S. E., Crane, H. M., Mugavero, M. J., Eron, J. J., Boswell, S. L., Rodriguez, B., Mathews, W. C., Martin, J. N., Moore, R. D., & Golden, M. R. (2013). High levels of antiretroviral use and viral suppression among persons in HIV care in the United States, 2010. *Journal of Acquired Immune Deficiency Syndromes*, 63 (3), 299–306. <https://doi.org/10.1097/QAI.0b013e3182945bc7>.
- [6] Ibrahim, N. (2007). *Education of factor affecting the chance of survival/death status among hiv positive people under the antiretroviral treatment program: The Case of Adama Referral Hospital (Doctoral dissertation, M. Sc. Thesis)*. Addis Ababa, Ethiopia: Addis Ababa University.
- [7] Levy, B., Correia, H. E., Chivore, F., Ronoh, M., Abebe, A., Kgosimore, M., ... & White, K. J. (2021). modelling the effect of HIV/AIDS stigma on HIV infection Dynamics in kenya. *Bulletin of mathematical biology*, 83 (5), 1-25.
- [8] Maposa, I. (2016). *Survival Modelling and Analysis of Hiv/Aids Patients on Hiv Care And Antiretroviral Treatment to Determine Longevity Prognostic Factors*. Cape Town, South Africa: University of the Western Cape. <http://hdl.handle.net/11394/5444>.
- [9] May, S., & Hosmer, D. W. (1998). A simplified method of calculating an overall goodness-of-fit test for the Cox proportional hazards model. *Lifetime Data Analysis*, 4 (2), 109-120. <https://link.springer.com/article/10.1023/A:1009612305785>.
- [10] Mecha, J. O., Kubo, E. N., Nganga, L. W., Muiruri, P. N., Njagi, L. N., Mutisya, I. N., Odionyi, J. J., Ilovi, S. C., Wambui, M., Githu, C., Ngethe, R., Obimbo, E. M., & Ngumi, Z. W. (2016). Trends in clinical characteristics and outcomes of Pre-ART care at a large HIV clinic in Nairobi, Kenya: a retrospective cohort study. *AIDS Research and Therapy*, 13 (1), 38. <https://doi.org/10.1186/s12981-016-0122-y>.

- [11] Mzileni, M. O., Longo-Mbenza, B., & Chephe, T. J. (2008). Mortality and causes of death in HIV-positive patients receiving antiretroviral therapy at Tshepang Clinic in Doctor George Mukhari Hospital. *Pol Arch Med Wewn*, 118 (10), 548-554.
- [12] National AIDS and STIs Control Programme (2016). *Kenya AIDS response progress report: Progress report 2016*.
- [13] Piulachs Lozada-Benavente, X., Alemany Leira, R., & Guillén, M. (2014). *UB Riskcenter Working Paper Series A joint longitudinal and survival model with health care usage for insured elderly A joint longitudinal and survival model with health care usage for insured elderly*. Barcelona, Spain: University of Barcelona. <http://hdl.handle.net/2445/98450>.
- [14] Shebeshi, D. (2011). Survival analysis of adult HIV/aids patients and stochastic modelling of aids disease progression: a case study of jimma. Addis Ababa, Ethiopia: jimma University Specialized Hospital.
- [15] Sika, A. M., Wools-Kaloustian, K., Mwangi, A. W., Kimaiyo, S. N., Diero, L. O., Ayuo, P. O.,... & Musick, B. (2010). Risk factors for death in HIV-infected adult african patients recieving anti-retroviral therapy. *East African Medical Journa*, 87 (11), 443-451.
- [16] Smith, A. D., Kimani, J., Kabuti, R., Weatherburn, P., Fearron, E., and Bourne, A. (2021) HIV burden and correlates of infection among transfeminine people and cisgender men who have sex with men in nairobi, kenya: an observational study. *The lancet HIV*, 8 (5), e274-e283.