Research & Development
2022; 3(4): 215-223
http://www.sciencepublishinggroup.com/j/rd
doi: 10.11648/j.rd.20220304.13

Modelling HIV-Free Survival Among Children Under Five-Year-Old from the Uganda Population-Based HIV Impact Assessment (UPHIA) 2016-2017 Survey

Muriuki Francis Maina*, Waititu Anthony, Wanjoya Anthony

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

mwakanai6@gmail.com (Muriuki Francis Maina), agwaititu@gmail.com (Waititu Anthony), awanjoya@gmail.com (Wanjoya Anthony)

To cite this article:

Muriuki Francis Maina, Waititu Anthony, Wanjoya Anthony. Modelling HIV-Free Survival Among Children Under Five-Year-Old from the Uganda Population-Based HIV Impact Assessment (UPHIA) 2016-2017 Survey. *Research & Development*. Vol. 3, No. 4, 2022, pp. 215-223. doi: 10.11648/j.rd.20220304.13

Received: September 22, 2022; **Accepted**: October 21, 2022; **Published**: October 29, 2022

Abstract: The Human Immunodeficiency Virus (HIV) pandemic is currently the most challenging public health matter that faces third world countries, especially those in Sub-Saharan Africa. Uganda, in East Africa, with a generalised and highly heterogeneous epidemic, is no exception, with HIV/AIDS affecting most sectors of the economy. The barrier HIV-stigma presents to the HIV treatment cascade is increasingly documented; however, no research has been done on the effects of HIV free survival of children under five-year olds in the presence of clustering. This study present models for analysing the effects of covariates on HIV free survival of children in Uganda in which tests for the association among children in different enumeration areas (EA) are carried out. The main objective of this study is to test for association between covariates and HIV free survival among under five-year old children born to HIV positive mothers in Uganda. This study uses the shared frailty model from the Uganda Population-Based HIV Impact Assessment (UPHIA 2016-2017) dataset. The study reviews the Cox proportional hazard model to include clustering within children from different EAs. The study first checks for proportional hazard assumption in the model then extend the Cox PH model to include clustering of children in different EA in Uganda using the shared frailty model. The basis for testing proportional hazard assumption in Cox regression analysis is to assess whether the effects of the covariates change overtime. When observations are clustered within groups or multiple event times are clustered within individuals, dependence between event times in a cluster is of interest. The study then uses the shared frailty model which is a random effect model which helps explain the unaccounted heterogeneity in the data.

Keywords: Cox Proportional Hazard Model, Frailty Model, Mother-To-Child Transmission, HIV-Free Survival

1. Introduction

1.1. Background of Study

Globally, an estimated 35.3 (32.2–38.8) million people were living with HIV in 2012. An increase from previous years as more people are receiving the life-saving antiretroviral therapy. There were 2.3 (1.9–2.7) million new HIV infections globally [15]. Many prevention strategies are being employed to minimize new infections and to improve the living standards of HIV-infected persons. An estimate of over two thirds of all persons living with HIV are in sub-Saharan Africa [15]. Furthermore, this region also has the highest prevalence and HIV infection incidence in the entire world [15].

The launch of "Start Free Stay Free" AIDS Free framework in 2016 by the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the United States President's Emergency Plan for AIDS Relief (PEPFAR) was to build on the achievements of the Global Plan towards the Elimination of New HIV Infections among Children by 2015 and Keeping their Mothers Alive, which ended in 2014. The "Start Free" component of this programme aims to secure free beginning of every child by ending new HIV infections among children by, among other actions, reaching 95% of pregnant women living with HIV and sustaining them on lifelong antiretroviral therapy (ART) by 2018 [6]. Even though mother-to-child transmission (MTCT) of HIV is

preventable through antiretroviral treatment (ART) during pregnancy and postpartum, the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimates that 160,000 new HIV infections occurred among children in 2018 [6].

Each day an estimated 1600 children born to HIV-infected mothers become infected, 1500 of whom are in sub-Saharan Africa. Mother to child transmission (MTCT) during pregnancy, delivery and breastfeeding is by far the most common route of HIV infection in children. The estimated risk of infection is 5-10% during pregnancy, 10 - 20% during labour and 10 - 20% during breastfeeding [11]. Risk factors for MTCT include maternal viral load levels, vaginal delivery, prematurity and breastfeeding.

The Cox proportional hazard model is one of the most common methods used in analysis of time to event data. The idea of the model is to define hazard level as a dependent variable which is being explained by the time-related component (so called baseline hazard) and covariates-related component. Model is based on several restrictive assumptions which need to be carefully verified before interpretation of parameters estimates. One of them is the assumption of proportional hazard which results directly from the model formula and means that hazard ratio needs to be constant over time. However, if this assumption is violated, it does not necessarily prevent analyst from using Cox model.

Frailty models are very useful for analysing correlated survival data, when observations are clustered into groups or for recurrent events. Frailty models [5] are extensions of the Cox proportional hazards model [1] which is the most popular model in survival analysis. It does so by adding random effects which act multiplicatively on the hazard function.

1.2. Statement of the Problem

The fact that HIV/AIDS has greatly ravaged children, the young and the old generations globally, this has caused alarm, being that these are the most productive of any population. Ever since the discovery of HIV/AIDS much information has been disseminated, but despite these, levels of awareness are still low among some sections of the population. Mother-to-child transmission is the most known mode of acquisition of HIV infection in children. There have been concerns about development of resistance to ARV drugs especially to young mother hence increasing the risk of MTCT in HIV positive women and in infants who nevertheless become infected.

A steady increase in HIV infection especially among young girls has been witnessed, thus confounding past assertions that increase in awareness leads to a decrease in HIV infections. The number of new HIV infections among adolescent girls and young women, many of whom become mothers, remain to be too high. WHO recommendations for HIV-infected women on infant feeding are not adhered to. It is not clear whether HIV prevention programs imposed have influenced the target populations. Fewer children newly infected with HIV and improved health for mothers living with HIV stand out as achievements of the global AIDS response in recent years.

Therefore, modelling HIV free survival among children born to HIV positive mothers and taking the association between the event times in a cluster into account is an integral part of the global response to the AIDS epidemic which should be considered an essential component of an effective MTCT control and prevention strategy. Here, the study focuses on the influence of different covariates on clustered multivariate survival variable in which the different response event times are grouped into clusters (Enumeration areas).

1.3. Justification of the Study

It has been argued that the most common reason for high HIV infections among under-five year old children is the transmission from mother to child during the pregnancy period, birth and breastfeeding period. Determining the effects of covariates on HIV free survival among children born to HIV positive mothers in the presence of clustering at this -time of HIV epidemic, is very important for planning and programming. Currently, there is lack of empirical evidence on the HIV free survival rates and factors that influence HIV free survival among under five-year-old children from different EAs in Uganda. This study is important because it will provide information which shall be used in reducing MTCT since HIV-free survival is a standard measure of progress toward eliminating MTCT. The basis for testing the proportional hazard assumption in a Cox regression analysis helps in assessing whether the effects of the covariates change overtime. In many clinical applications, the study population needs to be considered as a heterogeneous sample or as a cluster of homogeneous groups of individuals such as families or geographical areas. Sometimes, due to lack of knowledge or for economic reasons, some covariates related to the event of interest are not measured. The frailty approach is a statistical modelling method which aims to account for the heterogeneity caused by unmeasured covariates. Since the data consist of children coming from different enumeration areas in Uganda, which is clustering within EA, the best way to model is to use the frailty model.

1.4. Objectives

1.4.1. General Objective of the Study

The main objective of the study is to estimate HIV free survival among children born to HIV positive mothers from the UPHIA 2016-2017 dataset.

1.4.2. Specific Objectives of the Study

The specific objectives of the study are as follows;

1) To develop time to child being HIV negative and alive model for children born to HIV positive mothers.
2) To estimate the survival model parameters.
3) To construct the hazard function of the survival time of HIV patients in the presence of clustering.

2. Literature Review

Different models have been used by many researchers in studying the HIV outcome. The Cox regression model and the frailty model have been used to model HIV outcome. The frailty model is one of the most widely used model to model clustered and correlated survival times. In this section, the study briefly describes some important literature on modelling HIV data with the Cox regression model and the frailty model for modelling clustered survival data.

Garc´ıa de Olalla et al. [3] used the Cox regression with time dependent variable to assess the effect of antiretroviral therapy (ART) adherence on survival in HIV- infected patients. In their study, they performed a univariate analysis in which survival curves were obtained using an extended Kaplan-Meier method by estimating and testing associated relative hazards by means of proportional hazard regression with late entries and therapy was introduced as time-dependent variable. In the multivariate analysis, Cox proportional risk models were used to calculate relative risks of dying, and 95% confidence intervals were calculated.

A study on modelling Viral Suppression, Viral Rebound and State-Specific Duration of HIV Patients with CD4 count adjustment: parametric multistate frailty model approach was done in South Africa [2]. In this study, a full-parametric and semi-parametric Markov multistate models were applied to assess the effects of covariates. Viral rebound was found to be significantly associated with many sex partners, higher eosinophils count, younger age, lower educational level, higher monocyte counts, having abnormal neutrophils count, and higher liver enzyme abnormality. The analysis result also showed that patients with a stable sex partner, higher educational levels, higher QoL scores, lower eosinophils count, lower monocyte counts, and higher RBC indices were more likely to spend more time in undetectable viral load state.

Goethals et al. [4] examined the parametric proportional hazards frailty model for time to first insemination in dairy heifer cows based on the marginal likelihood maximization. The authors begin by working out the hessian matrix for the frailty model with Weibull baseline and one covariate; the effect of parity on time to insemination and then maximizing the marginal log likelihood with respect to the parameters. The results are then expressed in terms of median first insemination time. They extend the marginal likelihood maximization approach to the accelerated failure time model in the case of a Weibull baseline hazard. The parameter estimates for the accelerated failure time model are obtained by simple transformation from the parameter estimates for the proportional hazards model. The parameter estimates for the baseline hazard and the frailty term remained the same in the accelerated failure time model.

According to Muttai et al. [10] in his study on development and validation of a sociodemographic and behavioural characteristics‐based risk‐score algorithm for targeting HIV testing among adults in Kenya. In their study, it was found that the overall final algorithm comprised the following variables: age category 35–39/40–44 years; occupation manual/domestic or trade/sales/service); marital status (polygamous marriage, separated/divorced or widowed); in the prior 12 months having ≥ 2 sexual partners or reporting treatment for an STI; and having never been tested for HIV or having a negative HIV test result > 12 months prior. The study demonstrated that a HIV predictive risk-score algorithm, derived from a set of sociodemographic and behavioural characteristics, can be used to identify sub-populations who have higher risk of HIV infection to whom HIV testing could be targeted.

Mbougua et al. [9] performed a nonlinear multiple imputation for continuous covariate within semiparametric Cox model: application to HIV data in Senegal. They evaluated the proposed method on the motivated data set collected in HIV-infected patients enrolled in an observational cohort study in Senegal, which contains several incomplete variables. It was found that their method performs well to estimate hazard ratio compared with the linear imputation methods, when data are missing completely at random, or missing at random.

Lancaster and Intrator [8] analysed the hospitalization experience of a panel of HIV-positive patients. In their study, they model the joint distribution of the inpatient episodes and the survival times of a panel of patients over 15 months using the shared frailty model. The model induces correlation between hospitalization and death via an unmeasured, person-specific, frailty term, and it allows rates of hospitalization and of death each to be affected by time-invariant and time-varying covariates. The study concluded that the model without covariates that the model describes rather well several important features of the joint distribution of the observed hospitalization and mortality processes. It was also noted that their approach can be generalized in many ways, some of which were attempted. For example, it is possible to have the frailty effect in the rate of hospitalization different from that in the mortality rate.

3. Methodology

This section gives the research design and process that will be used including identifying the population sample and the methods that will be employed in analysing the data.

3.1. Data

3.1.1. Study Area

Uganda Population-based HIV Impact Assessment (UPHIA) 2016-2017 survey is a National household HIV survey that is part of a multicounty PHIA survey assessment being conducted in 13 other sub-Saharan countries in Africa. UPHIA, was conducted between August 2016 and March 2017 to measure the status of Uganda's national HIV response. UPHIA provides key information for health policy makers and implementers on the size of the HIV epidemic, the impact that recent and ongoing HIV programming are having on the epidemic and a gap-assessment of key areas requiring further focus in the fight against HIV.

3.1.2. Study Subjects

The survey collected a representative sample of 12,812 eligible households selected from the 10 regions of Uganda. The data was collected from August 2016 to March 2017. The survey targeted 41,374 persons of whom 17,217 were eligible women and 13,364 were eligible men aged 15 to 64 and 10,793 eligible children aged 0 to 14 years. It gives estimates for persons aged 0-64 years for the National HIV Incidence, National & region (10 regions) HIV Prevalence.

3.2. Methods

3.2.1. Sample Size

The sample size was calculated to provide a representative national estimate of HIV incidence among adults. This survey was based on a nationally representative sample of over 12,800 households throughout Uganda. The sample size was 33,243 for adults, and 10,956 for children, sampling of the two paediatric age groups, ages 0-4 years and ages 5-14 years, was done at different rates to meet specified precision targets. UPHIA described demographic characteristics of respondents, and obtained data on uptake and coverage of services including prevention of mother-to-child-transmission (PMTCT) of HIV, male circumcision, HIV testing and awareness of HIV status, HIV care and treatment, tuberculosis (TB), syphilis, Hepatitis B, and intimate partner violence. UPHIA also collected information about HIV testing and treatment in children. Through blood tests, UPHIA measured indicators of the national impact of the HIV programme including HIV incidence and HIV viral load suppression (VLS) in adults, and HIV burden estimates based on HIV prevalence among adults and children.

3.2.2. Sampling Strategy

UPHIA used a two-stage, stratified cluster sample design. The sampling frame comprised all households in the country, based on the 2014 National Housing and Population census, which includes 80,000 EAs, containing an estimated 7,800,000 households [15]. The first stage selected 520 EAs (clusters) using a probability proportional to size method. The EAs were stratified by ten regions: Central 1, Central 2, Kampala, East-Central, Mid-Eastern, North-East, West Nile, Mid-North, Mid-West, and South-West. During the second stage, a sample of households was randomly selected within each EA, or cluster, using an equal probability method.

3.2.3. Ethical Clearance

Ethical approval was received from Uganda virus Research Institute (UVRI), Columbia University and the United States Centres for Disease Control and Prevention (CDC) institutional review boards (IRB).

3.3. Model Specification

3.3.1. Cox Proportional Hazard

The proportional hazards model, proposed by Cox [1], has been primarily used in medical testing analysis, to model the effect of covariates on survival. Cox proportional hazard model is one of the most common methods used in analysis of time to event data. The idea of the model is to define the hazard level as a dependent variable which is being explained by the time-related component (baseline hazard) and covariates-related component.

Denote the observed data for individual i by (t_i, δ_i, x_i), $i = 1,...,n$, where t_i is the observed event or censoring time, $\delta_i = 1$ if t_i is uncensored and $\delta_i = 0$ otherwise, and x_i is a vector of covariates.

The model is then defined as follows:

$$\lambda(t,x) = \lambda_0(t)\exp(\beta x)$$

Where;

$\lambda(t,x)$ – hazard function that depends on timepoint t and vector of covariates **x**,

$\lambda_0(t)$ – baseline hazard function that depends on time only,

$\exp(\beta x)$ – covariates-related component.

The model is based on several restrictive assumptions which need to be carefully verified before interpretation of parameters estimates. One of them is the assumption of proportional hazard which results directly from the model formula and means that hazard ratio needs to be constant over time.

Estimation of the Cox proportional hazards model parameters:

The regression coefficients will be estimated using Maximum Likelihood statistical method [13]. Parameter estimates in the Cox PH model are obtained by maximizing the partial likelihood as opposed to the likelihood. Cox and others have shown that this partial log-likelihood can be treated as an ordinary log-likelihood to derive valid (partial) Maximum Likelihood Estimates (MLEs) of β. Therefore, hazard ratios and confidence intervals can be estimated using maximum likelihood.

The likelihood contribution of individual i is $p_i f(t_i \mid Y=1; z_i)$ for $\delta_i = 1$ and $(1-p_i) + p_i S(t_i \mid Y=1; z_i)$ for $\delta_i = 0$ where $p_i = pr(Y_i = 1; x_i)$. The observed full likelihood is then;

$$L(b,\beta,\wedge_0) = \prod_{i=1}^{n}\left\{p_i\lambda_0(t_i \mid Y=1)e^{z_i'\beta}e^{-\wedge_0(t_i|Y=1)\exp(z_i'\beta)}\right\}^{\delta_i} \times\left\{(1-p_i) + p_i e^{-\wedge_0(t_i|Y=1)\exp(z_i'\beta)}\right\}^{1-\delta_i}$$

Denote the complete data by $(t_i, \delta_i, z_i, y_i)$, $i = 1,...,n$, which includes the observed data and the unobserved y_i's. The complete-data full likelihood is

$$L(b,\beta,\wedge_0;y)=\prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}\prod_{i=1}^{n}\left\{\lambda_0(t_i \mid Y=1)\exp(z_i'\beta)\right\}^{\delta_i y_i} \times e^{-y_i \wedge_0(t_i|Y=1)\exp(z_i'\beta)} = L_1(b;y)L_2(\beta,\wedge_0;y)$$

Where y is the vector of y_i values. The estimates $\hat{b}$ and $\hat{\beta}$ are chosen such that they maximizes $L(b,\beta,\wedge_0)$.

3.3.2. The Frailty Model

When observations are clustered within groups or multiple event times are clustered within individuals, dependence between event times in a cluster is of interest. The study uses the shared frailty model which is a random effect model which helps explain the unaccounted heterogeneity in the data.

The aim for multivariate analysis is to account for the dependence in clustered event times. A natural way to model dependence of clustered event times is through the introduction of a cluster-specific random effect - the frailty. This random effect explains the dependence in the sense that if frailty was known then the events would be independent. This approach can be used for survival times of related individuals like family members, parent-child, twins) or recurrent observations on the same person.

In the shared frailty model, the cox proportional hazard model proposed by Cox [1] is extended to include a multiplicative individual frailty term Z_i assuming that Individuals in the same cluster have a common frailty Z_i.

The hazard function for the j^{th} individual in cluster i conditional on the frailty Z_i is defined by the model;

$$h(t/Z_i)=Z_i h_0(t)\exp\left(X_{ij}^t\beta\right)$$

Where;

$h_0(t)$ – baseline hazard function.

X_{ij}^t -is the vector of covariates associated with the regression coefficient β

Z_i - is the random effect called the frailty for the i^{th} cluster.

The frailties Z_i are usually assumed to be independent and identically distributed with a distribution referred to as the *frailty distribution*. Individuals with a higher frailty can be thought of as being frailer and therefore expected to die sooner than other individuals with the same measured covariates. If the event of interest is a positive outcome, like pregnancy or recovery, subjects with a higher "frailty" are expected to experience the positive outcome sooner than others with the same covariates.

Estimation of the gamma frailty parameters:

To estimate the unknown parameters in a frailty model, the study uses the maximum likelihood method for the gamma frailty model as used by Klein [7]. The flexibility of the gamma frailty distribution makes it possible to detect a complex frailty distribution structure which may otherwise be missed. It is combined with the methods for Cox's proportional hazards model and provides the maximum likelihood estimate of β and a non-parametric estimate of $h_0(t)$.

We obtain the maximum likelihood estimates of the unknown parameters in the generalized gamma frailty model based on an approximated likelihood function. Given the observed data $(t_{ij},\delta_{ij},x_{ij})\ i=1,...,n\ j=1,...,m_i$ the unconditional likelihood function of the generalized gamma frailty model is

$$L(q,\sigma,\beta,\gamma)=\prod_{i=1}^{n}\int_0^{\infty}\left(\prod_{j=1}^{m_i}\left[y_i h_0(t_{ij},\gamma)e^{\beta' x_{ij}}\right]^{\delta_{ij}} S_0(t_{ij},\gamma)^{y_i\exp(\beta' x_{ij})}\right)g(y_i,q,\sigma)dy_i$$

$$=\prod_{i=1}^{n}\left(\prod_{j=1}^{m_i}\left[h_0(t_{ij},\gamma)\exp(\beta' x_{ij})\right]^{\delta_{ij}}\right)\int_0^{\infty} y_i^{\delta_i}\exp\left[-y_i A_i(\beta,\gamma)\right]g(y_i,q,\sigma)dy_i$$

where $S_0(t,\gamma)=\exp\left[-\int_0^t h_0(u,\gamma)du\right]$ and $H_0(t,\gamma)=-\log S_0(t,\gamma)$ are the corresponding baseline survival and cumulative hazard functions, respectively, $\delta_i=\sum_{j=1}^{m_i}\delta_{ij}$ and $A_i(\beta,\gamma)=\sum_{j=1}^{m_i}H_0(t_{ij},\gamma)\exp(\beta' x_{ij})$.

3.4. Model Diagnostics

To build the model, the study first includes a classical cox survival model, group level variables and test significance at 5% level in Kaplan Meier Analysis; thereafter, the random effects are added. Thereafter, the likelihood ratio (LR) test is used test to compare the Cox model and the Frailty model.

The best model will be selected using Bayesian Information Criterion (BIC) which measures the goodness of fit and the complexity of the model.

The BIC value is given by;

$$BIC=-2l+\log(d)df$$

Where L is the maximum likelihood value and d is the number events.

In this study, the model with the minimum value of BIC will be judged as the best model.

4. Data Analysis and Results

4.1. Descriptive Statistics

In the UPHIA data set, children were classified according to maternal report of HIV status during pregnancy. In this study, data for mother's status before and during pregnancy was analyzed. Out of 6814 eligible mothers, response on HIV status before and during pregnancy resulted to 3011 (44.19%) HIV negative mothers, 119 (1.75%) HIV positive, 87 (1.28%) did not receive their HIV test results, 13 (0.19%) did not know their HIV status, 3 (0.04%) were uncertain about their HIV status and 3581 (52.55%) did not respond. The outcome of interest in this study is the time until a child born to HIV positive mother is alive and negative. Therefore, the event in this case is a child being HIV negative and alive at the time of the survey. Children born to HIV positive mothers were censored at their age at HIV diagnosis, or their age at the time of survey. Under five-year old children were considered in the analysis. There were high numbers among the uncensored children (96.6%) compared to the censored (3.4%).

4.1.1. Distribution of HIV Free Survival by Covariates

Table 1. *Distribution of HIV free survival by Covariates.*

Covariates	HIV+ N (%)	HIV- N (%)	Total (%)	Covariates	HIV+ N (%)	HIV- N (%)	Total
Gender				*Child on ART*			
Female	6 (0.2)	3401 (99.8)	3407 (50)	HIV -ve	2 (0.03)	6801 (99.97)	6803 (99.8)
Male	7 (0.2)	3400 (99.8)	3407 (50)	On ART	11 (100)	0 (0.0)	11 (0.2)
Region				*Mother on ARV during pregnancy*			
Central 1	1 (0.2)	541 (99.8)	542 (8.0)	No ARV	2 (22.2)	7 (77.8)	9 (7.6)
Central 2	0 (0.0)	515 (100)	515 (7.6)	Newly intd	1 (3.1)	31 (96.9)	32 (26.9)
East Central	1 (0.1)	808 (99.9)	809 (11.9)	ARV 1st ANC	1 (1.3)	77 (98.7)	78 (65.5)
Kampala	3 (0.8)	361 (99.2)	364 (5.3)	*Mother on ARV during breastfeeding*			
Mid-East	2 (0.2)	1003 (99.8)	1005 (14.7)	Continued ARV	2 (2.0)	100 (98.0)	102 (2.9)
Mid-North	3 (0.6)	538 (99.4)	539 (7.9)	Didn't cont	0 (0.0)	1 (100)	1 (0.03)
Mid-West	1 (0.2)	625 (99.8)	626 (9.2)	Miss status	3 (0.1)	3416 (99.9)	3419 (96.47)
North East	2 (0.2)	990 (99.8)	992 (14.6)	Never brstfd	0 (0.0)	22 (100)	22 (0.6)
South West	0 (0.0)	434 (100)	434 (6.4)	*Mother HIV status before pregnancy*			
West Nile	0 (0.0)	988 (100)	988 (14.5)	No	2 (0.1)	2452 (99.9)	2454 (96.5)
				Yes	2 (2.2)	88 (97.8)	90 (3.5)
Place of residence				*Mother HIV status during pregnancy*			
Rural	8 (0.2)	5288 (99.8)	5296 (77.7)	Didn't rc reslt	0 (0.0)	87 (100)	87 (2.7)
Urban	5 (0.3)	1513 (99.7)	1518 (22.3)	Don't know	0 (0.0)	13 (100)	13 (0.4)
Place of delivery				Negative	1 (0.0)	3010 (3.6)	3011 (93.1)
Health facility	4 (0.1)	2736 (99.9)	2740 (40.2)	Positive	4 (3.4)	115 (96.6)	119 (3.7)
Home	0 (0.0)	695 (100)	695 (10.2)	Uncertain	0 (0.0)	3 (100)	3 (0.1)
In transit	0 (0.0)	36 (100)	36 (0.5)	*Mother's age*			
Other	9 (0.3)	3334 (99.7)	3343 (49.1)	15-19	0 (0.0)	328 (100)	328 (5.5)
Child's age time of survey				20-24	2 (0.1)	1704 (99.9)	1706 (28.6)
Below 1 year	2 (0.2)	1284 (99.8)	1286 (18.9)	25-29	1 (0.1)	1613 (99.9)	1614 (27.1)
1 year	1 (0.1)	1327 (99.9)	1328 (19.5)	30-34	2 (0.2)	1145 (99.8)	1147 (19.2)
2 years	3 (0.2)	1290 (99.8)	1293 (19.0)	35-39	4 (0.5)	746 (99.5)	750 (12.6)
3 years	1 (0.1)	1440 (99.9)	1441 (21.1)	40-44	0 (0.0)	288 (100)	288 (4.8)
4-5 years	6 (0.4)	1460 (99.6)	1466 (21.5)	Above 45	0 (0.0)	133 (100)	133 (2.2)

4.1.2. Diagnostics for Cox-PH Model

Table 2. *Testing for Proportional Hazard assumption.*

Covariates	Chi square	DF	P Value
Child's gender	0.76	1	0.383
Place of residence	1.13	1	0.288
Mother's ARV status bstfd	7.48	3	0.058
Mother's ARV status pregnancy	0.89	2	0.641
Place of delivery	5.43	3	0.143
Child on ART	7.63e-08	1	1.000
GLOBAL	13.4	11	0.267

In order to establish which covariates satisfy the Cox PH assumption a Cox PH model is fitted to the data. It is important to determine whether the fitted Cox regression model adequately represents the data. Table 2 represents results for the covariates that satisfy the PH assumption. The results show that the covariates presented were non-significant as is the global test for proportional hazards, thus the proportionality of hazards assumption holds for these covariates.

For the covariates that violated the PH assumption, a Schoenfeld residuals plot is used to assess proportionality of hazards. The results are presented in Figure 1. The p values show that the two covariates that is region of residence and HIV exposed baby are statistically significant. Also the smooth pointwise confidence band deviates from zero for both covariates. This clearly shows the two covariates interacts significantly with time.

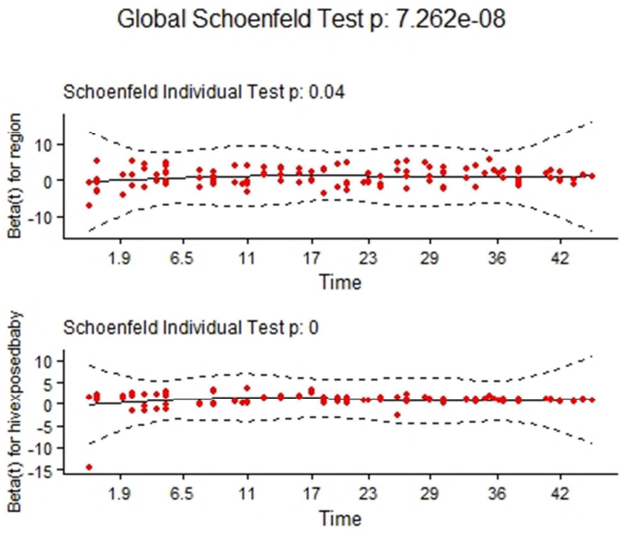

Figure 1. *Plots of scaled Schoenfeld residuals for covariates.*

4.2. Cox PH and Frailty Models

To test for association between covariates and HIV free survival among under five-year old children born to HIV positive mothers in Uganda, the study used the shared frailty model. Firstly, the study checks correlation between children in the same Enumeration area. This study used the *frailtypack* [12] and the *survival* [14] packages for this analysis.

Table 3 represents results from the Cox PH and the frailty models. The variance of the frailty term (Theta=0.698, p value=0.024) is significantly different from zero, hence there is heterogeneity among children from different enumeration areas. This means HIV free survival of children born to HIV positive mothers varies due to unobserved covariates shared within an enumeration area. Furthermore, the frailty model results in an increase in effect of some covariates; that is mother on ARV during breastfeeding and mother on ARV during pregnancy.

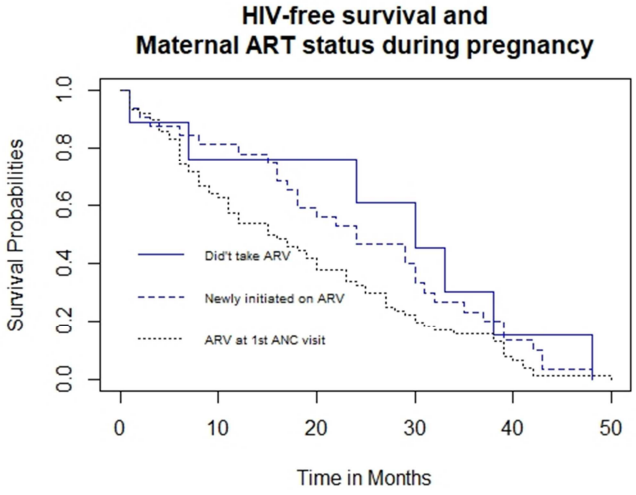

Figure 2. *Kaplan Meir survival curves for HIV-free survival and maternal ART status during pregnancy.*

Table 3. *Estimated hazard ratios obtained from both Cox proportional hazards model and frailty model for HIV free survival in children of HIV positive mothers.*

Covariates	Cox PH Model		Frailty Model		
	Beta (p value)	HR	Beta (p value)	HR	Global Chi square (p value)
Gender					
Female	Ref				
Male	-0.0103 (0.961)	0.989	0.246 (0.437)	1.279	
Place of Residence					
Rural	Ref				
Urban	-0.170 (0.423)	0.844	-0.397 (0.256)	0.672	
Mother on ARV during breastfeeding					7.122 (0.068)
Continued taking ARV	Ref				
Didn't continue	-0.500 (0.632)	0.607	-1.923 (0.209)	0.146	
Missing ARV status	0.735 (0.340)	2.086	2.909 (0.0116)	18.338	
Never breastfed	0.528 (0.328)	1.695	0.359 (0.589)	1.433	
Mother on ARV during pregnancy					11.667 (0.0029)
Didn't take ARV	Ref				
Newly initiated on ARV	0.512 (0.565)	1.669	3.144 (0.007)	23.206	
On ARV at 1st ANC visit	0.996 (0.285)	2.706	4.157 (0.001)	63.857	
Place of delivery					0.628 (0.8900)
Health Facility	Ref				
Home	0.004 (0.993)	1.004	0.301 (0.656)	1.351	
In transit	-0.187 (0.821)	0.829	-0.242 (0.819)	0.785	
Other	0.270 (0.714)	1.31	0.611 (0.523)	1.842	
Mother's age					8.435 (0.208)
15-19	Ref				
20-24	-0.467 (0.466)	0.627	0.427 (0.645)	1.532	
25-29	-0.337 (0.574)	0.714	0.122 (0.888)	1.13	
30-34	-0.697 (0.248)	0.498	-0.393 (0.652)	0.675	
35-39	-0.466 (0.453)	0.627	-0.085 (0.925)	0.919	
40-44	-1.330 (0.039)	0.265	-1.474 (0.122)	0.229	
Above 45	-0.771 (0.446)	0.463	-0.356 (0.794)	0.701	

Covariates	Cox PH Model		Frailty Model		
	Beta (p value)	HR	Beta (p value)	HR	Global Chi square (p value)
Wald Test	14.44 on 16 df, p=0.6				
Likelihood Ratio test	15.88 on 16 df, p=0.5				
Penalized marginal Log Likelihood			-427.71		
Score (log rank) test	14.95 on 16 df, p=0.6				
Frailty Parameter (Theta)			0.698		
Frailty P value			0.024		

5. Conclusions

With HIV/AIDS still being a major epidemic, especially in the sub-Saharan Africa, more research needs to be done to reduce the number of new HIV infections, in particular mother to child transmission. This study aimed at assessing the effects of HIV-free survival children among five-year old born to HIV positive mothers in Uganda. Unlike Cox-PH model that doesn't account for random effects, this study adopted Frailty model to test for association between covariates of interest and HIV-free survival children in different enumeration areas. Some of the covariates that were considered for this study included gender, region, place of residence, place of delivery, and mother's HIV status during and after pregnancy and during and after breastfeeding.

The study found out that there is heterogeneity among children from different enumeration areas since the frailty term variance was significant. This indicates that the HIV-free survival of children born to HIV-positive mothers changed as a result of unobserved factors that were shared within an enumeration region. The findings of this research could be used to improve programmes focused on HIV positive females in the child-bearing age and mothers, and consequently reducing significantly the number new HIV infections of new born children. Future research could also adopt hybrid frailty models such as mixed frailty copula model with Gompertz marginal and Bayesian multivariate joint frailty model.

Declaration

A Research Report Submitted in partial fulfilment of the requirements for the Degree of Master of Science in Applied Statistics of Jomo Kenyatta University of Agriculture and Technology.

This research project is my original work and has never been submitted elsewhere for a degree award. No part of this report may be reproduced without prior permission of the author and/or the University.

Declaration by Supervisors

This research project has been submitted for examination with our approval as supervisors.

1. *Prof. Anthony Waititu*, Statistics and Actuarial Sciences Department, JKUAT.
2. *Dr. Anthony Wanjoya*, Statistics and Actuarial Sciences Department, JKUAT.

List of Abbreviations and Acronyms

AIDS Acquired Immunodeficiency Syndrome
ART Antiretroviral Treatment
HIV Human Immunodeficiency Virus
UPHIA Uganda Population-Based HIV Impact Assessment
MLE Maximum likelihood Estimator
MTCT -Mother-To-Child Transmission
NASSEP IV National Sample Survey and Evaluation Programme IV
PEPFAR President's Emergency Plan for AIDS Relief
STIs Sexually Transmitted Infections

References

[1] Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34 (2), 187-202.

[2] Dessie, Z. G., Zewotir, T., Mwambi, H., & North, D. (2020). Modeling viral suppression, viral rebound and state-specific duration of HIV patients with CD4 count adjustment: parametric multistate frailty model approach. Infectious diseases and therapy, 9 (2), 367-388.

[3] García de Olalla, P., Knobel, H., Carmona, A., Guelar, A., López-Colomés, J. L., & Caylà, J. A. (2002). Impact of adherence and highly active antiretroviral therapy on survival in HIV-infected patients. Journal of acquired immune deficiency syndromes (1999), 30 (1), 105-110.

[4] Goethals, K., Janssen, P., & Duchateau, L. (2008). Frailty models and copulas: similarities and differences. Journal of Applied Statistics, 35 (9), 1071-1079.

[5] Hougaard, P. (2012). Analysis of multivariate survival data. Springer Science & Business Media.

[6] Joint United Nations Programme on HIV/AIDS (UNAIDS). (2019). Start free Stay free AIDS free-2019 report. Geneva: UNAIDS.

[7] Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. Biometrics, 795-806.

[8] Lancaster, T., & Intrator, O. (1998). Panel data with survival: hospitalization of HIV-positive patients. Journal of the American statistical association, 93 (441), 46-53.

[9] Mbougua, J. B. T., Laurent, C., Ndoye, I., Delaporte, E., Gwet, H., & Molinari, N. (2013). Nonlinear multiple imputation for continuous covariate within semiparametric Cox model: application to HIV data in Senegal. Statistics in medicine, 32 (26), 4651-4665.

[10] Muttai, H., Guyah, B., Musingila, P., Achia, T., Miruka, F., Wanjohi, S.,... & Zielinski-Gutierrez, E. (2021). Development and Validation of a Sociodemographic and Behavioral Characteristics-Based Risk-Score Algorithm for Targeting HIV Testing Among Adults in Kenya. AIDS and Behavior, 25 (2), 297-310.

[11] Newell, M. L. (2003). Antenatal and perinatal strategies to prevent mother-to-child transmission of HIV infection. Transactions of the Royal Society of Tropical Medicine and Hygiene, 97 (1), 22-24.

[12] Rondeau, V., Gonzalez, J. R., Yassin Mazroui, A., Mauguen, A. D., Laurent, A., Lopez, M.,... & Rondeau, M. V. (2021). Package 'frailtypack'.

[13] Sy, J. P., & Taylor, J. M. (2000). Estimation in a Cox proportional hazards cure model. Biometrics, 56 (1), 227-236.

[14] Therneau, T. M., & Lumley, T. (2015). Package 'survival'. R Top Doc, 128 (10), 28-33.

[15] UBOS (2014). 2014 census - uganda bureau of statistics. https://www.ubos.org/2014-census/. (Accessed on 10/19/2022).