SURVIVAL MODELLING OF UNDER-FIVE HIV INFECTED CHILDREN TAKING ANTIRETROVIRAL THERAPY AT CHIRADZULU DISTRICT HOSPITAL, MALAWI: AN APPLICATION OF COX AND PARAMETRIC REGRESSION MODELS

MSC (BIOSTATISTICS) THESIS

AUBREY JAZZA

UNIVERSITY OF MALAWI

MARCH, 2022

SURVIVAL MODELLING OF UNDER-FIVE HIV INFECTED CHILDREN TAKING ANTIRETROVIRAL THERAPY AT CHIRADZULU DISTRICT HOSPITAL, MALAWI: AN APPLICATION OF COX AND PARAMETRIC REGRESSION MODELS

MSc. (Biostatistics)

By

AUBREY JAZZA

BSc. (Statistics & Computing) – Malawi University of Business and Applied Sciences

Submitted to the Department of Mathematical Sciences, Faculty of Science, in partial fulfilment of the requirements for the degree of Master of Science (Biostatistics)

University Of Malawi

March, 2022

DECLARATION

I, the undersigned, hereby declare that this work is substantially my own original work which has not been submitted to any other institution for similar purposes. Where other people's work has been used, acknowledgements have been made.

Aubrey Jazza

Full Legal Name

_____

Signature

17/03/2022

_____

Date

CERTIFICATE OF APPROVAL

The undersigned certify that this thesis represents the student's own work and effort and has been submitted with our approval.

Signature: _____Date: _____

Professor Lawrence Kazembe, PhD

Main Supervisor

Signature: _____Date: _____

Patrick Sawerengera, (Lecturer)

Programme Coordinator (Master of Science in Biostatistics)

DEDICATION

I dedicate my dissertation work to my family, Director of Health and Social Services for Chiradzulu District Hospital, friends and relatives. Special thanks should go to my parents, whose words of encouragement ring in my ears and my wife Maggie, my son Blessings and Chiradzulu District Commissioner: Mrs. Reinghard Kaweta Chavula for their tirelessly support.

ACKNOWLEDGEMENTS

I would like to greatly thank my supervisor, Professor Lawrence Kazembe for his invaluable comments and support throughout the execution of this thesis. He was very patient with me and was always available whenever I had questions that needed his guidance. I have learnt a lot of things over the last couple of months and Professor Kazembe is responsible for such knowledge. I would not have come this far without him. Above all, I thank God for the strong health life, joy and new mercies for me over my completed work on this project.

To all the lecturers who taught me the modules, I say thank you for the knowledge I gained from you.

I do not forget Earnest H. Kaphuka, Evaristar N. Kudowa and Vincent J. Samuel for their motivation during my research process.

ABSTRACT

The hazard of a child dying before reaching 5 years is highest in sub-Saharan African countries including Malawi, with elevated risk in rural areas. Chiradzulu district has a higher under-five mortality rate as compared to a national average mortality rate for Malawi. The main aim of the study was to model and determine survival patterns of under-five children living with HIV who were on antiretroviral therapy. Using retrospective secondary data for 186 cases of under five children on anti-retroviral therapy, collected from July 2011 through July 2016, both Cox and parametric models were analyzed. The effect of the following covariates were investigated: Weight, height, sex, residence, mother's occupation, mother's education level and mother's marital status. Weibull, exponential, Gompertz, loglogistic and lognormal regression were performed as parametric models and Cox as a semi-parametric model. Akaike Information Criterion (AIC) was used to compare the efficiency of fitted models. Residence, mother's education level, mothers' occupation and weight factors influenced the survival of under-five children living with HIV in Chiradzulu District. During the study, 56 (30.0%) of the participants died. Gompertz model was found to be the best fit model for predicting survival of under-five children living with HIV on antiretroviral therapy in Chiradzulu district. It is recommended that the results should be substantiated by similar survival studies from other parts of the district to generalize the results to other individuals in the country. Researchers should check the underlying assumptions of Cox model before using it in order to use a proper model during analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

AHR	Adjusted Hazard Ratio
AIC	Akaike Information Criterion
AIDS	Acquired Immuno-Deficiency Syndrome
ART	Antiretroviral Therapy
ARV	Antiretroviral
BIC	Bayesian Information Criterion
CD4	Cluster of Differentiation 4
CDF	Cumulative Distribution Function
CI	Confidence Interval
DIC	Deviance Information Criteria
EGPAF	Elizabeth Glaser Pediatric AIDS Foundation
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immuno-Deficiency Virus
HR	Hazard Ratio
LMICs	Low and Middle Income Countries
LTFU	Lost to Follow Up
ICU	Intensive Care Unit
MDGs	Millennium Development Goals
MDHS	Malawi Demographic and Health Survey
MLE	Maximum Likelihood Estimation
MoH	Ministry of Health
MPHIA	Malawi Population based HIV Impact Assessment
MTCT	Mother to Child Transmission

NAC	National AIDS Commission
NSO	National Statistics Office
NVP	Nevirapine
PLWHA	People Living With HIV and AIDS
PMTCT	Prevention of mother to child transmission
PrEP	Pre-Exposure Prophylaxis
P-Value	Probability Value
SDGs	Sustainable Development Goals
SSA	Sub-Saharan African
TB	Tuberculosis
UN	United Nations
UNAIDS	United Nations Programme on HIV and AIDS
UNC	University of North Carolina
UNICEF	United Nations Children Education Fund
UNIGME	United Nations Inter-Agency Group for Child Mortality Estimation
USAID	United States Agency for International Development
VMMC	Voluntary Medical Male Circumcision
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Background and history

As of 2018, of the estimated nearly 38 million people worldwide living with HIV, approximately 1.7 million were children under 15 years of age. Since 2010, new HIV infections among children have declined by 41%, but only half (54%) of all children living with HIV are getting treatment and 100,000 children died of AIDS-related illnesses in 2018 (UNAIDS, 2019).

Globally, it was estimated that about 5.4 million children aged under-five die each year, with huge variations and trends across regions and nations (UNICEF, 2017). A 2017 UNICEF report on child mortality states that most of these children deaths happened in two particular regions: sub-Saharan Africa (SSA) (38%) and South Asia (39%). Over half of these deaths occurred in just five nations: the Democratic Republic of the Congo, India, Ethiopia, Nigeria and Pakistan. All things considered, 1 out of every 13 children born in sub-Saharan Africa dies before their fifth birthday as against 1 out of every 185 in high-income countries- meaning that the highest proportions of under-five mortality are concentrated in sub-Saharan Africa and South Asia (UNICEF, 2017).

The burden of mortality in children has remained a key area of concern for nations and organizations in the world. The year 2018 recorded approximately 5.3 million children and infant deaths worldwide. The risk of under-five mortality in the WHO Africa region

was 76 deaths per 1000 live births, which was eight-times higher than the WHO European region (WHO, 2018). This is far from ideal and is a worrying situation.

The Malawi Population Based HIV Impact Assessment (MPHIA) is the first survey in Malawi to measure National HIV incidence, pediatric HIV prevalence and viral load suppression. Malawi has an overall adult HIV prevalence of 10.6% in the general population and of the estimated one million people living with HIV, 10% are children with pediatric HIV prevalence of 1.6% (MPHIA, 2016). Globally, under-five mortality rate declined by 61 per cent, from 93 deaths per 1,000 live births in 1990 to 37 in 2020 (UNIGME, 2021; UNICEF, 2021). This is equivalent to 1 in 11 children dying before reaching age 5 in 1990, compared to 1 in 27 in 2020. With the end of the Millennium Development Goals (MDGs) era, the international community agreed on a new framework – the Sustainable Development Goals (SDGs) where the target is to end preventable deaths of new-borns and children under 5 years of age. The goal is for all countries to reduce under-five mortality to at least as low as 25 per 1000 live births. In 2019, child mortality rate for Malawi was 41.6 deaths per 1,000 live births. Child mortality rate of Malawi fell gradually from 341.3 deaths per 1,000 live births in 1970 to 41.6 deaths per 1,000 live births in 2019 (Malawi Child Mortality rate, 1960-2019, knoema.com).

Credible estimates of what this burden entails are essential for the establishment of informed health policies in order to implement effective health interventions. Although data exist on the rate of mortality and morbidity of children under the age of five years exposed to HIV in Africa, there is still little information regarding the area-specific infectious disease morbidity and mortality rates.

In 2011, Malawi started implementing Prevention of Mother to Child Transmission (PMTCT) *Option B+* policy, making life-long ART available for all HIV infected pregnant and breastfeeding women, regardless of clinical stage or cluster of differentiation 4 (CD4) count. This has resulted in a 66% reduction of vertical transmission within 3 years. This Malawi-pioneered strategy has since been included in global guidance by World Health organisation (WHO). As of February 2014, 12 other African countries were implementing Option B+ (WHO, 2018).

1.2 HIV prevalence in Chiradzulu

Chiradzulu district is located in the southern region of Malawi. The district has an estimated population of 356,875 (NSO Malawi, 2018). Chiradzulu district hospital started providing free antiretroviral (ARV) drugs in 2001. According to the 2015/2016 Malawi Demographic and Health Survey (MDHS), Chiradzulu District has an HIV prevalence of 9.2%, which is higher than the national average of 8.8% (MDHS, 2016), with a total number of 49,348 sexually active populations who have ever been tested for HIV and received results.

Supported by Médecins Sans Frontières, decentralization of ART provision and treatment to other 11 health facilities was completed in 2003. Task shifting allowing nurses to initiate ART started in 2006. Other HIV programs such as HIV counselling and testing (HTC) and prevention to mother transmission (PMTCT) were also scaled up and decentralised. The Elizabeth Glaser Paediatric AIDS Foundation (EGPAF) also provides paediatric HIV prevention, care and treatment services.

In addition to what the non-governmental organizations are doing, the hospital has also resources which are being used to address the HIV prevalence like drugs for the

prevention of mother to child transmission of HIV. It does also provide information about HIV and AIDS, civic education and communications through posters among others. Despite these efforts, lack of some drugs like flaconazole, vincristine as well as cotrimoxazole preventive therapy for treating opportunistic infections continue to be a major factor contributing to the mortality of people living with HIV and AIDS and this poses challenge for the district hospital (Chiradzulu District Council, 2017).

1.3 Problem statement

Pediatric HIV significantly contributes to overall child mortality and morbidity especially in high-burden countries. Malawi government has put in place strategies to meet the target of reducing by two thirds the mortality of under-five children. The Malawi Growth Development Strategy (MGDS III), Key Priority Area 5 (Health and Population), that is linked to Sustainable Development Goal (SDG) number 3, *"Ensure healthy lives and promote well-being for all at all ages"*, targets reduction of under-five mortality at least as low as 25 per 1000 live births. Currently, under-five mortality rate for Chiradzulu district is at 65 deaths per 1000 live births (Chiradzulu District Council, 2017).

Since the initiation of ART and Pre ART follow up program in 2001 at Chiradzulu District hospital, there has been no formal analysis to report on the survival patterns of under-five children living with HIV who are enrolled on HIV care. A study of children initiating ART younger than 5 years of age in rural Zambia between 2008 and 2015 responded well to treatment (Jessica et al., 2021). Hence this study aimed at reporting on survival of under-five children living with HIV and determinants for survival by comparing Cox and parametric models in Malawi, Chiradzulu.

There are two major regression models used for right censored data: proportional hazards model (Cox) as a semi parametric method (Cox, 1972) and parametric model. Many of the standard parametric models such as Weibull, Exponential, Loglogistics, Gompertz and Lognormal are accelerated failure time models. However, Cox regression is the most widely employed model in survival analysis while parametric models lead to some benefits (Lawless, 2011). Researchers in medical sciences often tend to prefer semi parametric instead of parametric because of its less assumptions but some comments recommended that under certain circumstances, parametric models estimate the parameter more efficient than Cox (Oakes, 1977).

Ata and Sozer (2007) argued that Cox regression model relies on the hazards being proportional, i.e. on a given covariate, its coefficient not changing over time. If this assumption is violated, the general Cox regression model is not suitable, and more appropriate analyses such as the stratified Cox regression model or the extended Cox regression model, including parametric duration models can be applied.

1.4 Study objectives

General objective

The main aim of the study was to model and determine survival patterns of under-five children living with HIV enrolled on antiretroviral therapy.

Specific objectives

The study aimed at addressing the following specific objectives:

i) To develop the best statistical model for the survival of under-five children living with HIV in Chiradzulu;
ii) To compare statistically between Cox proportional hazard and parametric hazard models for analyzing data of under-five children living with HIV on antiretroviral therapy in Chiradzulu;
iii) To determine factors affecting the survival of under-five children living with HIV on antiretroviral therapy in Chiradzulu.

1.5 Significance of the study

Apprehension of the survival patterns of under-five children living with HIV on ART and determinants for their survival rate is important to the development and implementation of HIV programme for under five Children. The absence of adequate progress in reducing the rates of under-five mortality by many of the developing countries has resulted in the newly adopted Sustainable Development Goals (SDGs), with the target of reducing under-five mortality to 25 per 1,000 or less by 2030 (United Nations, 2015).

The Malawi Growth and Development Strategy (MGDS) III emphasises on improved quality of health services, patient safety and people centered services in order to reduce incidence and prevalence of diseases. Its strategy is to strengthen prevention and management of infectious diseases such as HIV/AIDS and sexually transmitted diseases. In Malawi, the following HIV prevention programmes are implemented: condom availability and use, HIV education and approach to sex education, prevention from mother to child transmission, voluntary medical male circumcision (VMMC) and Pre-exposure prophylaxis (PrEP). Therefore, it is expected that the results from this study will assist in improving HIV and AIDS programme for under five children.

Furthermore, the results of this study could lead to improvement of district social economic profile in addressing the issue of infant mortality.

1.6 Definition of terms

1) Hazard Ratio (HR): is a measure of the relative survival experience of two groups.
2) Confidence Interval (CI): is a range of values around an estimate.
3) P-value: is the probability of obtaining results at least as extreme as the observed results of statistical hypothesis test, assuming that null hypothesis is correct.
4) The 90-90-90 targets: refer to the pathway by which a person is tested, linked and retained in HIV care, and initiates and adheres to antiretroviral drugs.

1.7 Structure of this thesis

The remainder of this thesis has been organized as follows: Chapter two discusses the literature review with combined statistical theory for AIDS and survival models. Chapter three discusses the methods employed in analysing the data for this study. Chapter four presents the results and discussion. Finally, chapter five presents the conclusion and recommendations for this study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter discusses the relevant literature for the theory of analysis of time to event data. The first part of the chapter briefly discusses the HIV and AIDS studies in children conducted in other countries plus ART provision in Malawi.

There are several models which researchers fit when dealing with survival data. The most used model is the Cox model. This is usually used because, it does not require any distribution to represent the survival time, and it is used to study the relationship between survival rate and covariates in the model.

2.2 HIV and AIDS in East and Southern Africa

Eastern and Southern Africa is the region hardest hit by HIV. It is home to more than 60 percent of children and adolescents worldwide living with HIV. In 2018, an estimated 1.8 million children and adolescents aged 0-19 years in Eastern and Southern Africa were living with HIV (UNICEF, 2019).

South Africa accounted for more than a quarter (240,000) of the region's new infections in 2018. Seven other countries accounted for more than 50% of new infections: Mozambique (150,000), Tanzania (72,000), Uganda (53,000), Zambia (48,000), Kenya (46,000), Malawi (38,000) and Zimbabwe (38,000) (UNAIDS, 2019). Overall, new infections in the region have declined by 28% since 2010 (UNAIDS, 2019). Around 310,000 people died of AIDS-related illnesses in the region in 2018, although the

number of deaths has fallen by 44% since 2010 (UNAIDS, 2019). Despite the continuing severity of the epidemic, huge strides have been made towards meeting the UNAIDS 90-90-90 targets. The 90-90-90 targets refer to the pathway by which a person is tested, linked and retained in HIV care, and initiates and adheres to antiretroviral drugs (ARVs).

Despite substantial improvements in accessibility of ART and improved program implementation, death and loss to follow-up (LTFU) have been a prevailing challenge among people living with HIV and AIDS (PLWHA) of all ages. However, attrition is much more pronounced in pediatric cases. There have been various studies conducted in Sub-Saharan African (SSA) countries to determine the rate of mortality among pediatric ART patients. A systematic review conducted by Fox et al. (2015), estimated attrition (death and LTFU) of pediatric ART patients in low and middle-income countries (LMICs) based on studies from 2008 to 2013. However, the pooled magnitude of mortality at different ART follow-up periods have not been separately analysed and reported. The aim of this review was to determine the pooled magnitude of mortality at different follow-up period among pediatric patients who were on first-line ART in SSA countries based on studies published since 2014. This timeframe was selected to include new studies that were not covered in the systematic review conducted in LMICs (Fox et al., 2015). The lessons from such studies can guide pediatric HIV program implementation in SSA and help policy makers and program managers to make informed decisions to prevent deaths among pediatric ART patients.

2.3 HIV and ART provision in Malawi

Malawi's HIV prevalence is one of the highest in the world, with 9.2% of the adult population (aged 15-49) living with HIV (UNAIDS, 2019). In 2018, an estimated one million Malawians were living with HIV and 13,000 Malawians died from AIDS-related illnesses (UNAIDS, 2019). The Malawian HIV epidemic plays a critical role in the country's life expectancy of 61 years for men and 67 years for women (WHO, 2021).

Over the last decade, impressive efforts to reduce the HIV epidemic have been made at both national and local levels. In 2018, 90% of people living with HIV in Malawi were aware of their status, of whom 87% were on treatment. Of these people, 89% were virally suppressed, meaning the country was very close to reach the UNAIDS 90-90-90 targets. This equates to 78% of all people living with HIV in Malawi on antiretroviral treatment (ART) and 69% of all people living with HIV virally suppressed (UNAIDS, 2019). Among children (0-14 years) treatment coverage is lower at only 61% of HIV-positive children accessing ART (UNAIDS, 2019).

New infections have dramatically declined from 66,000 new infections in 2005 to 38,000 in 2018 (UNAIDS, 2019). An impressive prevention of mother-to-child transmission (PMTCT) programme in Malawi has also driven down new HIV infections among children (ages 0-14). In 2018 there were 3,500 new pediatric infections, compared with 15,000 in 2010 (UNAIDS, 2019).

In 2015, the World Health Organization (WHO) announced new universal treatment guidelines for antiretroviral treatment (ART), which supports initiation of ART for all

individuals living with HIV, independent of their immunologic or clinical status (WHO, 2018). Since then, countries throughout sub-Saharan Africa have adopted the "Universal Test-and-Treat" (Test and Treat) strategy. The strategy is expected to contribute to improved client outcomes and attaining UNAIDS 90-90-90 treatment targets, specifically the ART coverage target (UNAIDS, 2014).

2.4 Scaling-up antiretroviral therapy in Malawi

Before the scale-up, an estimated 930 000 people in Malawi were HIV-infected, with 170,000 in immediate need of ART. About 3000 patients were on ART in nine clinics (Andreas et al., 2016). Relevant changes by December 2015, cumulatively 872,567 patients had been started on ART from 716 clinics, following national treatment protocols and using the standard monitoring system (Andreas et al., 2016).

2.5 Theory of analysis for survival data

The survival patterns following HIV infections in African population in the era before antiretroviral therapy form an important baseline for measuring future success of treatment programmes. The knowledge of the survival times of patients with HIV and variables that influence survival is important for increasing understanding of the patho-physiology of the disease, clinical decision making and planning health services interventions (Isingo et al., 2007). The survival of patients with AIDS may depend on a variety of factors including hosts, the patterns of diseases present, access to health care, diagnostic routines and therapeutic interventions (Robert et al., 1995).

The assumption of Cox models is that child survival is dependent on a baseline survival and certain risk factors, however, this is not often true in reality as survival data are

dependent when clusters or locations are considered. This dependency introduces unobserved random effects (frailties) present at various levels, and suggests the presence of community level characteristics that influence health outcomes (Cambridge, 2015). The existence of these effects caused either by a location, or a presence in certain population clusters leads to the use of spatial survival models to capture these unobserved effects, especially if they are geographical.

The Ethiopian Demographic and Health Survey data was used for the study of survival analysis of under-five children and its associated risk factors in Ethiopia. In this study, it was attempted to find out the impact of socioeconomic, demographic, environmental, health related and nutritional factors in under- five mortality of child. Firstly, data was analysed using Kaplan-Meier, non-parametric method of estimation of survival function and compared the survival time of different categories of region and other covariates that influence the child survival. Cox proportional hazard model and stratified Cox proportional model were also used to compare the hazard of under-five mortality of child for different covariates comparison to the reference categories. The potential covariates which influence under five mortality were region, mothers' education level, sex, mothers' age at first birth, preceding birth interval, contraceptive use, breast feeding, place of delivery, number of antenatal visits during pregnancy and father occupation. The study recommends that under-five mortality of child among regions is significant. This is an indication that the risk of under-five mortality of child varies from one region to another. Thus, in order to have a bearing on policy recommendations, future studies should focus on identifying risk factors of under-five mortality of child for each region of Ethiopia separately in high mortality area (Getachew, 2016).

A retrospective cohort study was conducted to determine the main factors that affect under-five mortality in West Sulawesi using Intercensal Population Survey 2015. There were 2549 cases in West Sulawesi. In this study, the impact of mother's education, age of first delivery, previous birth interval, birth type, the gender of the under-five child, and paramedics help during labour were investigated using the Cox proportional hazard regression. All variables impacted mostly to the survival rate of under-five children. Female under-five children had a lower hazard (risk) of death compared to the males. Twins had a 3 times higher hazard of death as compared to single born children. In addition, higher mother's education tends to have a lower hazard than those with lower education (Nurmalasari et al., 2019).

The effect of antiretroviral therapy on survival of HIV and Tuberculosis (TB) infected patients in Ukraine was assessed in prospective cohort study. The Kaplan-Meier method was used to determine the survival of 80 patients and the effect of Highly Active Antiretroviral Therapy (HAART) and survival was evaluated using Cox proportional hazard models. From the results of the analysis, it was found that patients with CD4 cell count<100µl had 5-fold higher risk of mortality and those with pulmonary tuberculosis with a 2-fold increased risk. Another assessment as to whether highly active ART was associated with improved survival in critically ill HIV-infected patients was conducted at the Saint–Louis teaching hospital in Paris, France. In this study, multivariable logistic regression was used to identify risk factors for death. It was found that five factors were independently associated with increased intensive care unit mortality: delayed intensive care unit (ICU) admission, acute renal failure, hepatic cirrhosis, admission for coma and severe sepsis (Mykhailo & Dmytro, 2013).

The adherence to ART in Benin city and identification of the contributing factors in a prospective study conducted on 125 out patients were assessed at the University of Benin teaching hospital. The authors used logistic regression models to determine the predictors of ART adherence relative to socio-demographic and clinical treatment variables. The proportional data were compared using Chi-square test or fischer exact at statistical significance of 95% confidence interval. The results from the findings showed that poor financial status, medication adverse effects, lack of confidentiality, occupational factors and stigmatisation were the major reasons given for ART non-adherence (Ayalu et al., 2012).

The investigation on the relationship between tuberculosis infection and death in people living with HIV and AIDS was conducted on 1575 subjects residing in both rural and urban areas of Yala province in Thailand and were followed between January 1992 and April 2010. Cox proportional hazard model was used to analyse the relationship and the model reported statistically significant relationship in people living with HIV and AIDS with tuberculosis and patients without tuberculosis. The people living with HIV and AIDS with tuberculosis were more likely to live shorter compared to those patients without tuberculosis after accounting for demographic factors.

The prospective study aimed at estimating the short-term disease progression among people living with HIV was conducted in Asia and Pacific region. In this study, the authors used Cox proportional hazard model to assess the predictors of disease progression and prognostic models were developed using Weibull models. It was found out in the analysis that the patients' not on treatment had higher rate of disease progression with 17.6 per 100 person-years against 8.1 per 100 person-years in the

patients receiving antiretroviral treatment. The results showed that the baseline CD4 count was the strongest predictor of disease progression. The authors reported that prognostic models were successful at identifying patients at high risk of short-term disease progression. The study was conducted on the outcome of antiretroviral treatment in rural public hospital in South nations, nationalities and people region in Ethiopia. The study used historical retrospective cohort study for patients visiting from January 1, 2005 to January 31, 2009. In the study, the authors used Kaplan-Meier models to estimate mortality and Cox proportion hazard models to identify predictors of mortality. It was found that the hazard of death was higher in males patients with WHO stage IV at baseline compared to WHO stage I (Zhou & Kumarasamy, 2005).

The survival rate of people living with HIV and AIDS after receiving free antiretroviral treatment was determined in Dehong Prefecture, Yunnan Province, China. A retrospective cohort analysis was conducted on all the people living with HIV and AIDS aged over 16 years who had started ART during January 2007 throughout December 2009 in Dehong Prefecture.

Assessing survival is not always an easy task. The choice of method is guided by the type of survival data (e.g. collecting age at death or monitoring live individuals with perfect or imperfect detection), the species life-history (e.g. single or numerous stages or ages) and the environment it experiences (e.g. controlled conditions versus variable environments). While it is relatively straightforward to gather survival data and assess survival using simple models under controlled lab conditions (Klein, 2016), monitoring survival in free-ranging populations often require more sophisticated capture-mark-recapture (CMR) techniques to deal with imperfect detection of individuals (Williams

et al., 2002). Indeed, survival data are often 'incomplete' in free-ranging populations, timing and cause of death can be hard to assess and multiple environmental factors are at play in influencing survival. Humans are an exception, with the existence of several consequent databases with perfect knowledge of age and cause of death for several human populations.

Survival analysis estimates and interprets survival functions as well as hazard functions from time to event data. In addition, survival analysis can be used to compare survival and hazard functions. It also helps to identify and asses the relationship of explanatory variables to survival time. This helps practitioners to concentrate on areas that can help improve people's welfare thereby, improving their survival times.

2.5.1 Survival terminology

By definition, survival analysis involves the application of statistical procedures for analysing data for which the outcome variable of interest is time until a study unit experiences an event (Kleinbaum & Klein, 2010).

The survival function is the probability that the survival time is greater than *t* (Kalbfleisch and Prentice, 2011).

Let T denote a random variable that describes survival time from ART initiation into the study to death.

That is,

$$S(t) = P(T > t), \quad (1)$$

$$S(t) = 1- F(t),$$

where F(t) is the cumulative density function of the random variable T.

$$F(t) = P(T \leq t). \quad (2)$$

where $F(t) = P(T \leq t)$ denotes the Cumulative Distribution Function (cdf) denoted as F(t) informs the probability that length of time T is less than or equal to any given value of *t*.

The probability density function is the slope of the cdf (failure function),

$$f(t) = \lim_{\Delta t \to 0} \left(\frac{p(t \leq T \leq t + \Delta t)}{\Delta t} \right) = \frac{\partial F(t)}{\partial t}. \quad (3)$$

Hence $\frac{\partial S(t)}{\partial t}$ is the probability of an individual dying in the interval (t, t+Δt). The survival function *S(t)* and the failure function *F(t)* are and hence have the properties of probabilities. It can be observed that, in particular, the survivor function lies between zero and one and strictly decreasing function of *t.*

The survivor function is equal to one at the start of the follow up *(t=0)* and zero at infinity. This implies that

$$0 \leq S(t) \leq 1, \quad (4)$$

$$S(0) = 1,$$

and that,

$$\lim_{t \to 0} S(t) = 0. \quad (5)$$

Further, $\frac{\partial S}{\partial t} < 0$, and the density function is non-negative

$$f(t) \geq 0. \quad (6)$$

2.5.2 The hazard rate

The hazard rate gives the instantaneous failure rate at time given that an individual has survived up to time *t* (Kalbfleisch & Prentice, 2011). The continuous time, hazard rate is defined as,

$$\theta(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{S(t)}. \quad (7)$$

It can also be demonstrated that there is a clear relationship between hazard and survival functions given as follows:

$$\theta(t) \quad = \frac{f(t)}{1-F(t)}, \quad (8)$$

$$= \frac{-\partial[1-F(t)]}{\partial(t)} / 1 - F(t), \quad (9)$$

$$= \frac{-\partial\{-ln[S(t)]\}}{\partial(t)}. \quad (10)$$

Integrating both sides gives the following:

$$\int_0^t \theta(u)du = ln[1 - F(t)] \Big|_0^t. \quad (11)$$

since $F(0) = 0$ and $ln(1) = 0$ then,

$$ln[1 - F(t)] = ln[S(t)] = -\int_0^t \theta(u)du,$$

so that,

$$S(t) = \exp\left(-\int_0^t \theta(u)du\right),$$

which becomes

$$S(t) = \exp[-H(t)]. \quad (12)$$

Then it follows that:

$$H(t) = \int_0^t \theta(u)du,$$

which is cumulative hazard function,

$$H(t) = -\ln[S(t)]. \quad (13)$$

From this it can be observed that,

$$H(t) \geq 0,$$

and $$\frac{\partial H(t)}{\partial t} = \theta(t). \quad (14)$$

2.5.3 The probability density function of survival time

The probability density function can be written as follows:

$$f(t) = \theta(t)\, exp\left(-\int_0^t \theta(u)du\right), t \geq 0. \quad (15)$$

The three functions outlined above are equivalent specifications of the distributions of the survival time. The survival function is useful for comparing survival progress of two or more groups. Among the functions of the survival analysis, the hazard function provides useful description of the risk of failure at any time point.

2.5.4. Censoring

It is difficult to ascertain exact survival times for study participants that did not experience an event of interest. Exact survival times can only be calculated for patients with outcomes during or by the end of the study. In this case, patients without an outcome are censored. Events can be left censored or right censored. In practice, most survival data are right censored (Kleinbum & Klein, 2010). Therefore, analysis of such data requires statistical methods that must consider a key analytical problem of censoring, if survival functions of individuals are to be meaningful.

Where there is no censoring, the survival function can be estimated as

$$\hat{S}(t) = \frac{N^*}{N},$$

where N* denotes the number of individuals with survival times > t and

N denotes the number of individuals in the data set.

2.5.5 Cox regression model

Cox regression model is a semi parametric model which is popular in survival data proposed by Cox (1972). The Cox model is defined as

$$\mathrm{h(t|x)} = h_0(t)\exp(\beta_1 x_1 + \cdots \beta_p x_p), \quad (16)$$

where $h(t|x)$ is the hazard function at time t for a subject with explanatory variables $X=(x_1, x_2 \ldots x_k)$, $h_0(t)$ is the baseline hazard function, that is the hazard function when all covariates equal to zero and β_i is the regression coefficient for the i^{th} covariate. The baseline can take any form (unspecified) but not negative. The Cox model assumes that the hazard functions for the two different levels of a covariate are proportional for all values of it and is given by:

$$\frac{h(t_k|X_k)}{h(t_j|X_j)} = \frac{h_0(t)\exp\{\beta_1 X_k 1 + \cdots \beta_p X_k p\}}{h_0(t)\exp\{\beta_1 X_j + \cdots \beta_p X_j p\}},$$

$$= \exp\{\beta_1(X_{k1} - X_{j1}) + \beta_p(X_{kp} - X_{jp})\}. \quad (17)$$

Hence $h_0(t)$ cancels out and this means that the ratio is the same at all-time points.

2.5.5.1 Assumption of the Cox model

The Cox model makes the following assumptions:

a) The structure of the model is assumed correct. That is for example, model is multiplicative and all relevant covariates have been met.
b) The continuous covariates have a linear form.
c) The proportional hazards assumptions are satisfied.

2.5.5.2 Cox model popularity

The Cox model has the following key properties:

a) It is robust hence a safe choice of model in many situations.
b) Estimated hazard are always non-negative.
c) Even though $h_0(t)$ is unspecified, β_i's can be estimated and thus compute the hazard ratio.
d) The $h_0(t|x)$ and $S(t|x)$ can be estimated for a Cox model using a minimum of assumption.

The Cox proportional hazard model can fit by maximizing the likelihood function and this procedure estimates the $h_0(t)$ and β. The popular approach is proposed by Cox (1975) in which a partial likelihood also called Cox likelihood function that does not rely on $h_0(t)$ is realized for β. The partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameter $h_0(t)$ in the Cox PH model. Assume K different failure times $t_{(1)}, t_{(2)}...t_{(k)}$ such that there is exactly one failure at each $t_{(i)}$, i=1,...,k . Let [i] denote the subject with an event time $t_{(i)}$ and $R(t)$ the risk set at time t, then the partial likelihood is given as:

$$\mathrm{L}(\beta) = \frac{\prod_{j=1}^{k} exp(\sum_{i=1}^{p} \beta_i X_{ji})}{\sum_{i \epsilon R(t|j)} \exp(\sum_{i=1}^{p} \beta_i X_{ji})}. \quad (18)$$

The likelihood considers probabilities for subjects who fail and does not consider probabilities for censored subject explicitly. The censored subjects are taken into account in the risk set. The estimates of β *'s* is denoted by $\hat{\beta}_i$'s. The $\hat{\beta}_i$'s solves:

$$\frac{\partial logL}{\partial \beta_i} = 0, i = 1, ...p. \quad (19)$$

Therefore, $\hat{\beta}_i$'s maximise the Cox likelihood.

2.6 Testing proportional hazard assumptions

The proportional hazard is the core assumption of the Cox model. There are a number of procedures for ensuring that a model satisfies the assumption of proportionality before the model results can be safely applied. The proportional hazard means that the survival for two subjects have hazard functions that are proportional overtime (constant relative). The proportional hazards (PH) assumption tests can be checked using statistical tests and graphical diagnostics based on the scaled Schoenfeld residuals.

If a variable violates PH assumption, Machin et al. (2006) recommends that a stratified Cox PH regression model be fitted. This is because stratification controls the effect of such a variable in a Cox model without making the PH assumption.

2.6.1 Statistical tests

2.6.1.1 Logrank method

Logrank test is used to find out whether the true survival curves differ from group to group with hypothesis testing:

H_0: No differences between survival times curves

H_1: There is a difference between survival times

It consists of observed versus expected events. For example, letting $t_{(1)} < \ldots < t_r$ be r distinct deaths times for each group. At time (j), let $d_{(1j)}$ and $d_{(2j)}$ be the number of deaths in group I and II respectively, $d_{(1j)}$ and $d_{(2j)}$ be the number of persons at risk prior to the time $t_{(j)}$. Then the log rank test statistic is:

$$\chi^2 = \frac{(\sum_{j=1}^{r} d_{ij} - Ex_{ij}))^2}{Var} \sim \chi^2(1)\,, \tag{20}$$

where $Ex_{ij} = \frac{n_1 d_j}{n_j}$.

The mean of the hypergeometric random variable and the variance of d_{ij} are given as,

$$Var(d_{ij}) = \frac{n_{1j} n_{2j} d_j (n_{1j} - d_j)}{n^2 j (n_j - 1)},$$

where $Var = \sum_{i=1}^{r} var(d_{1j})$. (Mantel, 1966)

2.6.1.2 Global test for the stratified

The global test for the stratified Cox model is an extension of the global test for the ordinary Cox model (Goeman et al., 2005).

Let us assume that *n* observations of q predictors are organized in a data matrix $X \epsilon R^{(n*q)}$with elements X_{ij}, further define $R_x = XX'$. The stratified Cox model, the hazard function of individual *i* at time *t* is,

$$h_i(t) = h^{(si)}(t)\exp(r_i),$$

where $h^{(i)}(.), \dots, h^{(m)}(.)$ are the unknown baseline hazards of strata 1, ..., m and

$$\mathrm{r}_i = \sum_{l=1}^{q}(\beta_l)x_{il}, \tag{21}$$

is the linear effect of the predictors.

Observing a sample of size *n* consisting of the predictor matrix *X*, follow-up times,

$$t = (t_1, \dots, t_n),$$

and status indicators

$$\boldsymbol{d} = (d_1, \dots, d_n).$$

We are interested in testing the null hypothesis that the predictors are not associated with survival, i.e. the hypothesis testing that:

$$H_0: \beta_1 = \cdots = \beta_q = 0.$$

2.6.2 Graphical diagnostics

The Cox PH survival function can be obtained by the relationship between hazard function and survival function,

$$S(t|X) = S_0(t)\exp\left(\sum_{i=1}^{k}\beta_1 X_k\right). \tag{22}$$

where $X = (x_1, x_2, \dots, x_k)$ is the value of the vector of predictor variables for a particular individual. Taking the logarithm twice, we have

$$\ln[-\ln S(t|\ X)] = \sum_{i=1}^{k}\beta_1 X_k + \ln[-\ln S_0(t)].$$

It can be noted that the difference in log-log curves corresponding to two different individuals with variables $X_1 = (x_{11}, x_{12}\ \dots x_{1k})$ and $X_2\ = (x_{21}, x_{22}\ \dots x_{2k})$ is given by,

$$\ln [-\ln S (t| X_1) - \ln S (t| X_2] = \sum_{i=1}^{k} \beta_i (X_{1i} - X_{2i}), \quad (23)$$

which does not rely on *t*. By plotting estimated log (-log (survival) versus survival time for the two groups, parallel curves would be realized if the hazards are proportional. However, this method does not work well for continuous predictors or categorical predictors that have many levels because the graph becomes "cluttered". Moreover, the curves are sparse when there are few time points and it may be difficult to gauge how close to parallel is close enough.

2.6.2.1 Kaplan-Meier method

The Kaplan-Meier estimator, also known as the product limit, is an estimator for estimating the survival function from lifetime data. It measures the fraction of patients living for a certain amount of time after treatment (Kaplan & Meier, 2018). A plot of the Kaplan-Meier estimate of the survival function is a series of horizontal steps of declining magnitude which, when a large enough sample is taken, approaches the true survival function for that population. The value of the survival function between successive distinct sampled observations is assumed to be constant. The method is defined as: let $X_1, X_2, \dots, X_n$ be independently identically distributed survival times having distribution function F(*x*) and let *G*(*c*) be the distribution of independently identically distributed censoring times $C_1, C_2, \dots, C_n$. Let $t_i = min\{X_i, C_i\}$ is the observed survival time and $\delta_i = I(X_i \leq C_i)$ indicate whether the survival time is censored (0 = alive) or event (1 = dead). Let the number of individuals who are alive just before time t_i including those who are about to die at this time, be n_i and d_i denote the number who die at this time. The Kaplan-Meier estimator is defined as:

$$\frac{\prod_{i=1}^{n}((n_i - d_i))^{\delta i}}{n_i}. \quad (24)$$

The assumption of the Kaplan-Meier survival function is that the distribution of censoring times is independent of exact survival times**.**

2.6.2.2 Deviance residuals

The deviance residuals (Therneau, Grambsch and Fleming, 1990) is defined by

$$\mathrm{r}D_i = sign(rx_i[-2\{rx_i + \delta_i \log(\delta_i - rx_i\}]^{\frac{1}{2}}, \quad (25)$$

where the function sign (.) is the sign function which takes the values 1 if rx_i is positive and -1 if rx_i if negative.

$$rx_i = \delta_i - rc_i,$$

is the martingale residuals for the i^{th} individual, $\delta_i = 1$ for uncensored observation and $\delta_i = 0$ for censored observation. The deviance residuals are normalized transformation of the martingale residuals (Therneau et al., 1990). They have a mean zero but are approximately symmetrically distributed about zero when fitted model is appropriate. Very large or small value can indicate potential outliers.

2.6.2.3 Schoenfeld residuals

Schoenfeld residuals are computed with one per observation per covariate. It is only defined at observed event times for the i^{th} subject and k^{th} covariate. The estimated Schoenfeld residual $\hat{r}_{ik}$, (the covariate value for the individual that failed minus its expected value) is given by,

$$\hat{r}_{ik} = X_{ik} - \hat{\bar{\mathrm{X}}}_{wik},$$

where X_{ik} is the value of the k^{th} covariate for individual *i* and $\hat{\bar{\mathrm{X}}}_{wik}$ is the weighted mean of covariate values for those in the risk set at the given event time. Positive value of risk shows X value that is higher than expected at that death time. The Schoenfeld residuals sum to zero.

2.7 Adding time-dependent covariates in the Cox model

This is done by creating interactions of the predictors and a function of survival time. Let X_i be the predictor of interest, and creating $X_j(t)$ as a time–dependent covariate, then

$$X_j = X_i * g(t),$$

where g(t) is a function of time. The model assessing PH assumption for X_j adjusted for other covariates is:

$$h(t,(t)) = h_0(\mathrm{t})\exp[b_1X_1 + b_2X_2 + \ldots\beta_jXj + \cdots + \beta_pXp + \delta X_j * \mathrm{g(t)}], \quad (26)$$

where $X(t) = (x_1, x_2, x_p, x_j(t))$ is the value of predictor variables for a particular individual. The null hypothesis to check proportionality is that, $\delta = 0$ where δ is time varying coefficient for X_j covariate. The test statistic can be done using either a Wald test or a likelihood ratio test. These statistics have chi-square distribution with one degree of freedom under the null hypothesis. If the time–dependent covariate is significant, then the predictor is not proportional.

2.8 Parametric survival models

Parametric survival models are regression models in which the distribution of the response is chosen to be consistent with what one would see if the response is time to-failure (Gutierrez**,** 2010). The parametric models are fitted to the survival data using maximum likelihood method, the procedure is described as follows:

Suggesting that the survival times $t_1, t_2 \ldots t_n$ are observed and q of the n individuals die at times $t_{(1)} < t_{(2)} \ldots < t_{(q)}$ and that the survival times of the remaining n-q (q<n) individuals are censored. If $f(t)$ denotes the probability density function of the survival time t and $S(t)$ be the survival function then, the likelihood is given by,

$$\prod_{i=1}^{n}\{f(t_i)\}^{ci}\{S(t_i)\}^{1-ci}, \quad (27)$$

where, c is an indicator variable, taking value 0 when the survival time is censored and 1 for the uncensored survival time.

2.8.1 Weibull model

The Weibull distribution is the generalized version of the exponential distribution. It is preferred for performing survival data analysis in industrial engineering (Rinne, 2020). However, when implementations in the discipline of medicines are examined, one may see that it is an important distribution model. It is a flexible distribution that allows monotonous increasing and decreasing of mortality ratio in patients' groups. In a study carried by Viscomi et al. (2006), the distribution of the survival period of childhood leukemia patients was analyzed using the Weibull distribution. In a study conducted in Italy on the national wide estimations of the cancer patients, some estimations were made for defining the parameters of Weibull distribution. The Weibull distribution has the following functions:

$$h(t\,|x) = \lambda, \tag{28}$$

is the hazard rate, and the hazard function is given by,

$$h(t) = \lambda t,\ \lambda > 0,\ t > 0. \tag{29}$$

The survival function is defined as:

$$S(t) = \exp(-\lambda t). \tag{30}$$

The probability density function for the exponential regression model is given by,

$$f(t) = \lambda \exp(-\lambda t). \tag{31}$$

The exponential model assumes that the baseline hazard is constant (Lawless, 2011). Therefore, the hazard is given by,

$$h(t|x_i) = \exp(\beta_0 + \beta_1 x_i)\ . \tag{32}$$

The survival function is given by,

$$S(t|x_i)=\exp\{-\exp(\beta_0+\beta_i x_i)\}. \quad (33)$$

Then the hazard function for a particular person with the explanatory variables $(x_1, x_2 \ldots x_p)$ is given by:

$$h(t|x) = \lambda p(t)^{p-1} \exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots \beta_p x_p\right) = \lambda p(t)^{p-1} \exp(\beta` x). \quad (34)$$

2.8.2 Gompertz model

Gompertz model is used frequently by medical researchers and biologists in modeling mortality ratio data (Wilson, 1994). The model was formulated by Gompertz. It has these functions:

Hazard function:

$$h(t) = \lambda \exp(\gamma t)\ , \quad (35)$$

for $0 \geq t < \infty$ where λ is positive value and is the scale parameter and γ is the shape parameter. When g = 0 survival times have an exponential distribution, where $\gamma > 0$ the hazard increases monotonically with time and when $\gamma < 0$ the hazard decreases with time.

$$h_0(t) = \lambda \exp(\gamma t)\ \exp(\beta_0)\ , \quad (36)$$

the model now becomes:

$$h(t|x_i) = h_0(t)\exp(x_i \beta x)\ t=\exp(\gamma t)\ \beta x\ . \quad (37)$$

The survival function is given by:

$$S(t)=\exp\left(\frac{\lambda}{\gamma}\right)(1-\exp(\gamma t). \quad (38)$$

The distribution is characterized by the fact that the log of hazard is linear in it. The hazard function for a particular person using the Gompertz distribution is given by:

$$h(t|x) = \lambda \exp(\gamma t)\ \exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots \beta_p x_p\right) =$$

$\lambda \exp(\beta^T x) \exp(\gamma t)$.(39**)**

2.8.3 Log-logistic model

The log logistic distribution is continuous for the random variable which is not negative in probability and statistics. The mortality ratio in a life analysis slowly decreases after it reaches to the maximum point over a finite period and it is suitable to use a non-monotonic failure rate distribution model on the life and lost (Viscomi et al., 2006).

Zhou et al. (2007) conducted study in which he emphasized that the maximum likelihood estimation was the most suitable method in estimating the parameters when performing analyses using log logistic distribution on grouped data such as half censored data.

These are the distributions functions for the log logistic:

$$h(t) = \frac{\lambda p t^{p-1} nx}{1+\lambda t^{p}}, \tag{40}$$

is the hazard function, *h(t*) increases then decreases if *p*>1, monotonically decreases when *p*=1, λ gives information on the covariate,

$$\lambda_i = \exp(x_i \beta)$$

and the following function,

$$h(t) = \frac{\lambda p t^{p}}{1+\lambda p t^{p}}, \tag{41}$$

is the survival function.

The Accelerated Failure time for the log logistic regression:

$$\lambda t_i = \exp(-x_i \beta x) t_i, \tag{42}$$

with $t_i \sim Loglogistic\ (\beta_0, \gamma)$

This has the cumulative distribution function,

$$F(t) = 1 - [1 + \{\exp(-\beta_0)\, t\}^{\frac{1}{\gamma}}]^{-1} \tag{43}$$

$$\ln(t_i) = x_i \beta x + \ln(\tau_i) = (\beta_0 + x_i \beta x + \mu_i) \tag{44}$$

where μ_i follows a logistic distribution with mean 0 and the standard deviation,$\frac{\pi\gamma}{\sqrt{3}}$. This follows**:**

$$E\{\ln(t_i|x_i)\} = [\beta_0 + x_i\beta x]. \tag{45}$$

The base line survivor functions of t_i is given by,

$$S_0(t) = 1 + [1 + \{\exp(-\beta_0)\, t_i\}^{\frac{1}{\gamma}}]^{-1}\,. \tag{46}$$

Hence the effect of the covariates is to accelerate time by the factor of $exp(-x_i\beta x)$. Then the Accelerated Failure time model is given by,

$$S(t_i|x_i) = S_0\{\exp(-x_i\beta x)\, t_i\}\,, \tag{47}$$

$$= [1 + \{\exp(-\beta_0)\exp(-x_i\beta x)t_i\}^{\frac{1}{\gamma}}]^{-1},$$

$$= [1 + \{\exp(-\beta_0 - x_i\beta x)\, t_i\}^{\frac{1}{\gamma}}]^{-1}. \tag{48}$$

2.8.4 Lognormal model

The lognormal is a skewed distribution where the average values are low, variances are high and the values are not negative. The survival distributions of Hodgkin's disease, chronic leukemia were analyzed via lognormal distribution, indicated positively skewed and with survival period distributed normally (Lee & Wang, 2003). In the lognormal, the hazard function increases from 0 to reach maximum and then decreases monotonically approaching 0 as *t*→ ∞. The survival function is given by:

$$h(t) = 1 - \Phi\left\{\frac{\ln(t)-\mu}{\sigma}\right\}, \tag{49}$$

where Φ is the standard normal cumulative density function and $\mu = \beta x$. The hazard rate is given by:

$$\frac{1}{t\sigma\sqrt{2\pi}}\exp\,[-\frac{1}{2\sigma^2}\{\frac{\ln(t)-\mu^2}{1-\Phi\{\frac{\ln(t)-\mu}{\sigma}}\}], \tag{50}$$

and the hazard rate rises first then falls. The density function is given by:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\{\ln(t) - \mu^2\}\right]. \quad (51)$$

However, the lognormal has no proportional hazard interpretation. Hence, its interpretation is in the AFT metric (Cleves, 2010). It assumes that t_i ~lognormal ($\beta_{0,}\sigma$) and it has the cumulative distribution as given by:

$$F(t) = \Phi\left[\frac{\ln(t)-\beta_0}{\sigma}\right], \quad (52)$$

is the cumulative distribution function for the standard Gaussian(normal) distribution hence,

$$\ln(t_i) = x_i\beta x + \ln(\tau_i). \quad (53)$$

The lognormal model transforms time into ln(*time*) and converts the problem into simple linear regression:

$$E\{\ln(t_i)\,|x_i\} = \beta_0 + x_i\beta_x. \quad (54)$$

The baseline survivor function is realized as:

$$S_0(t) = 1 - \Phi\{\frac{\ln(t)-\beta_0}{\sigma}\}. \quad (55)$$

2.8.5 Exponential model

This is the simplest parametric model. It assumes that the risk is constant over time. If X is a random variable, the exponential distribution is defined as:

$$X \sim Exp(\lambda), \lambda > 0. \quad (57)$$

The exponential distribution has a memoryless property which can be expressed as

$$P(x \leq X \leq x + \delta | X \geq x) = P(X < \delta),$$

for a positive δ. The probability to experience an event depends only on the length of the interval. The probability density function is then

$f(x) = \lambda e^{-\lambda x}$.

The survival function is:

$$S(x) = \int_x^{\infty} \lambda e^{-\lambda x}\, dx,$$

$$= -e^{-\lambda x},$$

$$= 0 - -e^{-\lambda x}$$

$$= e^{-\lambda x}. \quad (58)$$

2.8.6 Accelerate failure time model

The Accelerated Failure Time model is a linear regression model in which the response variable is the logarithm or known monotone transformation of a failure time (Lee & Wang, 2003). The accelerated failure time model describes a relationship between the survivor function of any two individuals. Taking T_i to be a random variable denoting the failure time for the i^{th} subject, and let $X_{i1}, X_{i2}, \ldots, X_{ip}$ be the values of p covariates of the subject. The model is then given by,

$$\log T = \{\beta_0 + \beta_1 X_{i1} + \beta_p X_{ip} + \delta\varepsilon_i\}, \quad (59)$$

where $\varepsilon_i \sim h_0(t)$, ε_i is a random disturbance term, β_0,..., β_p and σ are parameters to be estimated, $h_0(t)$ is a known baseline survival, T_i is actual survival times sometimes observed is a scale parameter and X_i is a fixed P*1 vector of covariates and X_i is assumed to affect *log T* linearly and no interactions. Moreover, σ is assumed to be constant and independent of X_i. The parametric accelerated failure time distribution is also assumed to be correctly specified. The survival function for the parametric baseline accelerated failure model is given as,

$$\mathrm{S}(t_i) = S_0\{exp\left(-{X_i}^t\beta\right) t_i, \quad (60)$$

where $S_0(t)$ is the baseline. These accelerated failure time models are named for the distribution of *T* rather than the distribution of *e* or *log T*. This is so because different

distributions have different implications for the shapes or hazard function (Cox & Oakes, 1996).

2.9 Assessment of the model fitness

When the model has been fitted, the adequacy of it needs to be assessed. There are a number of ways to check the adequacy, like using Cox-Snell, deviance among others.

2.9.1 Cox –Snell residuals

The Cox-Snell residuals is given by Cox and Snell (Cox & Oakes, 1984). The residuals for the i^{th} individual with the observed survival time t_i is given as follows,

$$r_{ci} = \exp(\beta X_i)\, H_0(t_i) = H_i(t) = -log\hat{S}_i(t_i). \quad (61)$$

Given that $H_0(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i and it was derived by Kalbfleish and Pretence (1973).

Letting *T* be the continuous survival distribution S(*t*) with the cumulative hazard,

$$H(t) = -\log(S(t)). \quad (62)$$

Then it follows that,

$$S_T(t)=\exp(-H(t)).$$

Taking $Y = H(t)$ be the transformation of *T* based on cumulative hazard function. It follows that the survival function for Y is now given as:

$$S_y = p(P > y) = p\{H(t > y)\}, \quad (63)$$

$P\left(T > H_T^{-1}(y)\right) = S_T(H_T^{-1}(y)) = \exp\left(-H_T(H_T^{-1}(y)\right) = \exp(-y).$

The new $Y = H(t)$ has an exponential distribution with unit one. If the model is well fitted, the actual value $\hat{S}_i(t_i)$ would have similar properties to those of $S_i(t_i)$. Therefore $r_{ci} = log\hat{S}_i(t_i)$ will have a unit exponential distribution with $f_R = \exp(-r)$.

Letting $S_R(\mathrm{r})$ denote the survival function of Cox-Snell residuals r_{ci}, then,

$$S_R = \int_r^{\infty} f_R \exp(-x)\, dx = \exp(-r), \tag{64}$$

and it follows that,

$$H_R(r) = -logS_R(r) = -\log(\exp)(-r) = r. \tag{65}$$

Hence a plot of $H(r_{ci})$ versus r_{ci} is used to check the fit of the model. This gives a straight line with a unit slope and zero intercept if the fitted model is correct.

2.10 Checking for model goodness of fit

There are a number of methods which are employed to check if a parametric distribution fits the observed data. The Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) a statistical criterion used for comparing models and residuals plots can be used to check the goodness of fit for the models.

2.10.1 Bayesian Information Criterion (BIC)

Bayesian Information Criterion (BIC) is a criterion for model selection among a finite set of models i.e. the model with the lowest BIC is selected. Can be calculated using the following formula:

$$BIC = -2ln\,(L) + ln\,(n)*K,$$

where n= sample size, K=number of parameters and L=Log-likelihood.

2. 10.2 Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) proposed in Akaike (1974) is a measure of the goodness of fit of an estimated statistical model which compares the models which have been fitted. The AIC is an operation way of trading off the complex of an estimated model against how well the model fits the data. The AIC is defined by:

$$AIC = -2ln\,(L) + 2k, \tag{66}$$

where L is the log likelihood, k is the number of covariates in the model. Lower values of the AIC suggest a better model. However, there is a difficulty in using AIC in the sense that there is no formal test statistically to compare different AIC values when two or more models have similar AIC values.

2. 10.3 Residual plots

The residual plots can be used to check the goodness of fit of the model. Among the useful plots is based on comparing the distribution of the Cox-Snell residuals with the unit exponential distribution. The Cox-Snell residual for i^{th} individual with observed time, t_i is defined as:

$$r_{ci} = \widehat{H}(t_i|x_i) = -\log[\hat{S}_i(t_i)x_i], \tag{67}$$

where t_i is the observed survival time for individual i, x_i is the vector covariate values for individual i and $\hat{S}_i(t_i)$ is the estimated survival function on the fitted model. The estimated survival function for the i^{th} individual is given by,

$$\hat{S}_i(t_i) = S_{\varepsilon i}(\frac{logt-\hat{\mu}-\hat{\alpha}x_i}{\hat{\sigma}}), \tag{68}$$

where $\hat{\alpha}, \hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimator of a, m and s respectively. $S_{\varepsilon i}(\varepsilon)$ is the survival function of ε in the Accelerated Failure Time model given by,

$$S(t_i|x_i) = S_0\{\exp(-x_i\beta x)\, t_i\},$$

and

$$\left(\frac{logt-\hat{\mu}-\hat{\alpha}x_i}{\hat{\sigma}}\right) = r_{\varepsilon i}, \tag{69}$$

is referred to as standard residuals.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

3.1 Study design

The study was a retrospective cohort. Data for under-five children living with HIV who were on ART at Chiradzulu district hospital from July 2011 to July 2016 was analysed. Events were censored by 31st, July 2016.

3.2 Study setting

The study used data collected at Chiradzulu District Hospital, Boma ART clinic in Chiradzulu District. Chiradzulu has an estimated population of 356, 875 (NSO Malawi, 2018), with HIV prevalence of 9.2% (MDHS, 2017). Chiradzulu district was the site of the first antiretroviral therapy (ART) program implemented in public facilities in Malawi, early 2001. By the end of June 2021, ART register indicated that there were over 28,500 patients at Boma ART Clinic.

3.3 Study population

Participants in this analysis were children under the age of five years who started antiretroviral therapy from July 2011 to July 2016.

3.4 Data collection

The data analysed in this study were sourced from under five ART patient cards. Authorization to use the data set was obtained through the Chiradzulu District Hospital Research and Ethics Committee. Two clerical statisticians gathered the data from the

patients' ART cards (see appendix 1 and appendix 2) with the help from ART clinical nurse. Guardians of under five children living with HIV were also traced and asked on their level of education, occupation, marital status and confirmation on the place of residence during clinic visits. Under-five children who were lost to follow up and alive by the end of study were censored.

3.5 Variables description

3.5.1 Dependent variable

The primary outcome of the study was survival time measured in months from ART initiation till death. The event of interest was death otherwise, right censored.

3.5.2 Independent variables

The study used pre-selected covariates based on the existing literature on the significant determinants of child mortality. The covariates ranged from weight of a child in kilograms (kg), height of a child in centimetres (cm), residence, sex for of a child, mothers' education level, mothers' occupation status and mothers' marital status. Weight and height were considered because they help to determine whether a child is having nutritious diet or not.

3.6 Statistical analyses

Statistical analyses were performed using the Stata version 14 software for windows. The effects of the pre-selected variables were estimated using a Cox proportional hazards model. The survival analysis to estimate the under-five mortality rate in Chiradzulu was performed using parametric models and the study compared two

survival analysis models, Cox and parametric models through the Akaike Information Criterion (AIC). Kaplain Meier as well as logrank methods were also used.

3.6.1 Kaplan-Meier method

Kaplan meier method was used to produce graphs of the Kaplan-Meier estimates of the survival function of under-five children for all preselected categorical variables: residence, sex for child, mother's education level, occupation status and marital status. The Kaplan-Meier estimate is done on categorical variables only and its assumption is that the distribution of censoring times is independent of exact survival times**.**

3.6.2 Logrank method

The logrank test was used to find out whether the true survival curves differ from group to group through this hypothesis testing.

H_0: No differences between survival times curves

H_1: There is a difference between survival times

The test was applied to pre-selected categorical variables only to determine statistical differences between groups.

3.6.3 Modelling strategies

3.6.3.1 Cox regression model

In this study, Cox regression model was used with the following covariates: X= (x_1= sex, x_2 =residence, x_3 = height, x_4 = weight, x_5 = mother education, x_6 = mother occupation, x_7 = marital status). Let T denote a random variable that describes survival time from ART initiation into the study to death where $h_0(t)$ is the baseline hazard function. The Cox proportional hazard model fitted was:

$$h(t) = h_0(t)e^{\beta_{sex}*Sex+\beta_{residence}*Residence+\beta_{height}*Height+\beta_{weight}}$$

$$*Weight + \beta_{mother\ Education} * Mother\ Education + \beta_{mother\ occupation} *$$

$$Mother\ Occupation + \beta_{marital\ status} * Marital\ Status.$$

3.6.3.2 Cox proportional hazard assumptions testing

Scaled Schoenfeld residuals was used to test the proportional hazards assumption for multivariable Cox regression model. A non-zero slope is an indication of violation of the proportion hazard assumption.

3.6.3. 3 Cox regression model goodness of fit test

The Cox model fit was evaluated using Cox –Snell residuals. If the hazard rate follows the 45-degree line, it suggests that its approximation has an exponential distribution with a hazard rate of one, and then the model fits the data well.

3.6.3.4 Parametric models

In the study, Weibull, Exponential, Gompertz, Loglogistics and Lognormal parametric models were fitted. These were also compared with Cox proportional hazards model to determine better model based on the decision of having a minimum Akaike Information Criterion (AIC) value.

3. 6.3.4.1 Weibull model

In this study, to estimate the effect size of covariates, Weibull model was fitted in the following manner:

$$h(t|x) = \lambda p(t)^{p-1} \exp\big(\beta_{sex} * Sex + \beta_{residence} * Residence + \beta_{height} *$$

$$Height + \beta_{weight} * Weight + \beta_{residence} * Residence + \beta_{mother\ education} *$$

$$Mother\ Education + \beta_{mother\ occupation} *$$

$$Mother\ occupation + \beta_{marital\ status} * Marital\ status).$$

3. 6.3.4.2 Gompertz model

In this study, hazard function for an under-five child using the Gompertz distribution was estimated by:

$$h(t|x) = \lambda \exp(\gamma t)\ \exp(\beta_{sex} * Sex + \beta_{residence} * Residence + \beta_{height} * Height + \beta_{weight} * Weight + \beta_{mother\ education} * Mother\ Education + \beta_{mother\ occupation} * Mother\ occupation + \beta_{marital\ status} * Marital\ status).$$

3. 6.3.4.3 Log-logistic model

An estimation of the effect size of covariates by log logistic model in this study was fitted using the following formula:

$$h(t) = \frac{1}{(1+\lambda t^{x})},$$

$$for\ x_i:\ x_1 = Sex, x_2 = Residence, x_3 = Height, x_4 = Weight, x_5 = Mother\ education, x_6 = Mother\ occupation\ and\ x_7 = Marital\ status.$$

3. 6.3.4.4 Lognormal

The survival function for lognormal model in this study was fitted by:

$$h(t) = 1 - \Phi\left\{\frac{\ln(t) - \mu}{\sigma}\right\},$$

where Φ = standard normal cumulative density function and

$$\mu = \beta x_i\ \ for\ x_1 = Sex,\ \ x_2 = Residence, x_3 = Height, x_4 = Weight, x_5 = Mother\ education, x_6 = Mother\ occupation\ and\ x_7 = Marital\ status\ .$$

The lognormal model transforms time into *ln*(time) and converts the problem into simple linear regression:

$$E\{\ln(t_i)\,|x_i\} = \beta_0 + \beta_x x_i.$$

3. 6.3.4.5 Exponential

In the study, the hazard function for the exponential model was fitted in the following manner:

$$h(t|x) = \lambda \exp\big(\beta_{sex} * Sex + \beta_{residence} * Residence + \beta_{height} * Height + \beta_{weight} * Weight + \beta_{residence} * Residence + \beta_{mother\ education} * Mother\ Education + \beta_{mother\ occupation} * Mother\ occupation + \beta_{marital\ status} * Marital\ status\big).$$

3.7 Model comparison

In this study the Akaike Information Criterion (AIC), a statistical criterion was used to compare models.

3.7.1 Akaike Information Criterion (AIC)

This Akaike Information Criterion (AIC) was used in this study. Thus, the lower values of the AIC suggested a better model.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Exploratory Analysis

A total number of 186 under-five children living with HIV entered to this study. The median age at ART enrolment was 3years. The median time to death for the under five children living with HIV on ART was 16 months. Out of 186 under five children living with HIV on ART, 56 (30.0%) died during the study while 130 (70.0%) were censored. Out of 130 censored participants, 30 (16.15%) were lost to follow up and 100 (53.85%) were alive by the end of the study. Table 1 shows a summary of descriptive statistics for of under five children living with HIV on ART at Chiradzulu District Hospital, Boma ART Clinic.

Table 1: Descriptive summary of variables

				Interquartile Range	
Continuous Variables	**Frequency**	**Median**	**Std. Dev**	**25th**	**75th**
Age (years)		3.00	0.91	2.00	4.00
Weight (kg)		10.00	8.16	8.90	12.00
Height (cm)		96.00	16.04	79.00	104.00
Time to Death (Months)		16.00	4.14	12.00	19.20
Categorical variables	**n (%)**				
Sex					
Male	84(45.16)				
Female	102(54.84)				
Residence					
Semi Urban	77(41.4)				
Rural	109(56.8)				
Mother Education					
No education	34(18.28)				
Primary school	73(39.25)				
Secondary school	64(34.41)				
Tertiary	15(8.06)				
Mother Occupation					
Not working	98(52.69)				
Working	88(47.31)				
Marital Status					
Single	72(38.71)				
Married	37(19.89)				
Divorced	77(40.40)				
Outcome Variables					
Dead	56(30.00)				
Censored					
Lost to follow up	30(16.15)				
Alive	100(53.85)				

4.2 Log rank tests for categorical variables

The results of log rank tests showed that residence, mother education and mother occupation were statistically significant. There were median survival differences in the groups of these categorical variables. The logrank tests were applied to categorical variables only.

4.3 Kaplan Meier survival curves for categorical variables

The Kaplan-Meier survival estimates results from Figure 1 have shown that under-five children living with HIV whose mothers work have a better median survival rate as compared to under-five children living with HIV on ART whose mothers do not work. Similar results were obtained from residence (survival rate for under-five children was better in urban than semi-urban) and mother's education level (survival rate better in under-five children with educated mothers as compared to under-five children with un-educated mothers). The different median survival rates were observed in the categories.

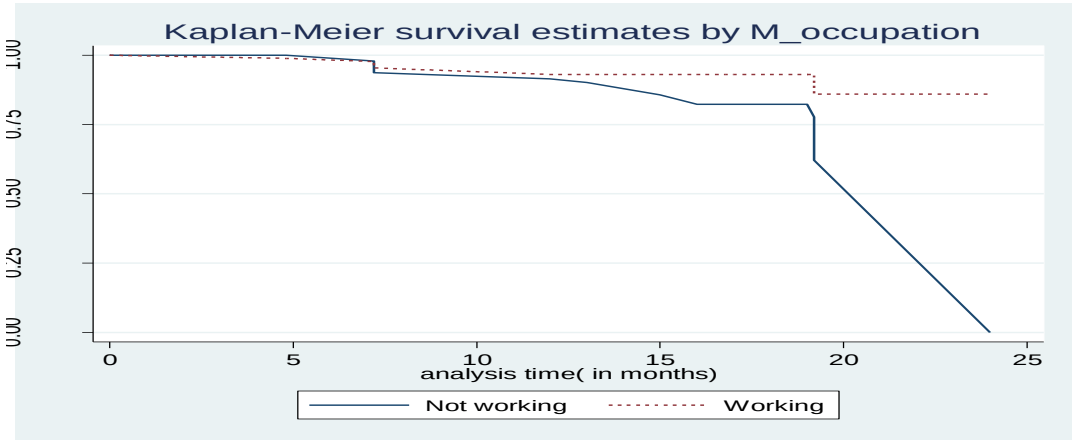


Figure 1: Kaplan Meier survival curves for mothers' occupation

4.4 Cox proportional hazard model assumption tests

4.4.1 Proportional hazard test by scaled Schoenfeld residuals

In all the plots of scaled Schoenfeld residuals for all pre-selected covariates (weight, height, residence, sex, mothers' education level, occupation status and marital status), there were almost flat lines being realised as illustrated in Figure 2 for mother's education levels. This is an indication that there was no violation of hazard proportionality by all the variables.

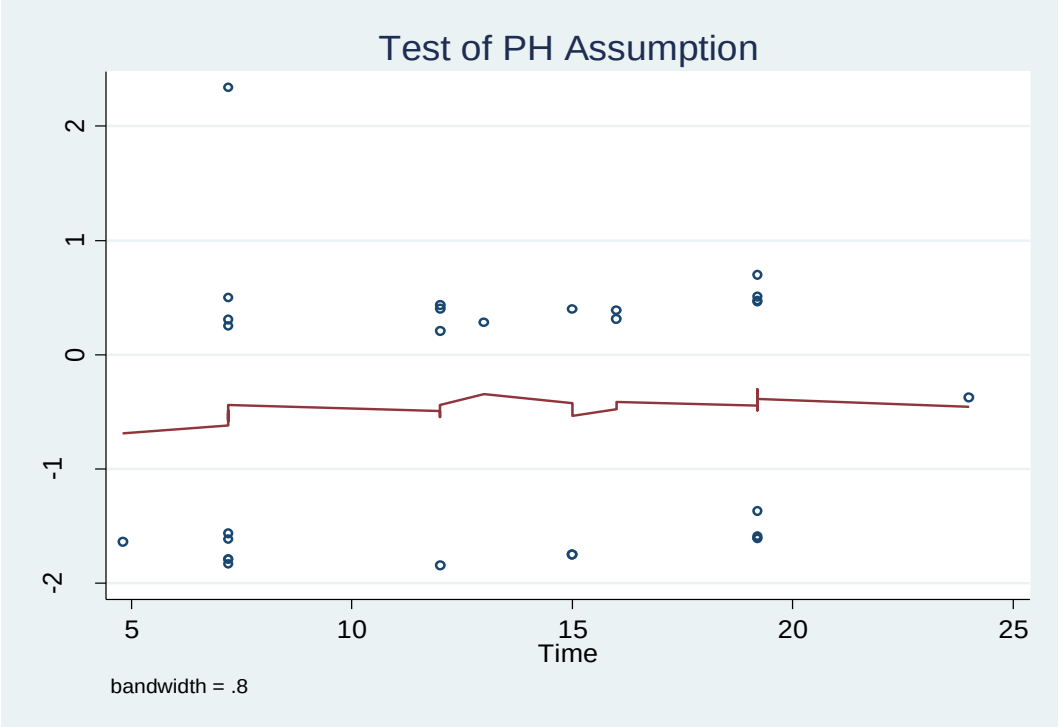


Figure 2: Graph of Scaled Schoenfeld residuals for Mothers' Education levels

The rest of the variables yielded similar graph results as of scaled Schoenfeld residuals for mother's education levels.

4.4.2 Cox Model bivariate and multivariate analysis

Table 2 indicates risk factors associated with survival rate of under-five children on antiretroviral therapy at Chiradzulu district hospital.

Table 2: Cox proportional hazards model: risk factors of under-five mortality

BIVARIATE			MULTIVARIATE	
COVARIATE	**HR**	**95% CI**	**AHR**	**95% CI**
Sex				
Male	(Reference)			
Female	1.40	[0.67, 2.94]		
Residence				
Semi Urban	(Reference)		Reference	
Rural	2.23	[1.95, 5.23]	1.88	[1.79, 4.50]
Mother Education				
No Education	(Reference)		Reference	
Primary	0.90	[0.39, 0.99]	0.93	[0.40, 0.95]
Secondary	0.34	[0.11, 0.65]	0.37	[0.10, 0.85]
Tertiary	0.00	- -	0.00	-
Mother Occupation				
Not working	(Reference)		Reference	
Working	0.37	[0.16, 0.83]	0.92	[0.32, 0.99]
Marital Status				
Single	(Reference)			
Married	0.89	[0.32, 2.35]		
Divorced	0.75	[0.33, 1.72]		
Weight	0.85	[0.71, 0.91]	0.86	[0.72, 0.98]
Height	0.99	[0.97, 1.01]		

Under five children living with HIV from rural areas were significantly associated with higher hazard of mortality than the hazard of under five children living with HIV from semi-urban areas in both bivariate analysis (HR: 2.23; 95% CI: 1.95, 5.23) and multivariate analysis (AHR: 1.88; 95% CI: 1.79, 4.50).

Under-five children living with HIV whose mothers have primary education were significantly associated with lower hazard of death than under-five children living with HIV whose mothers are not educated in both bivariate analysis (HR=0.34; 95% CI= 0.39, 0.99) and multivariable analysis (AHR: 0.93; 95% CI: 0.40, 0.95). Under-five children living with HIV whose mothers have secondary education are significantly associated with lower hazard of mortality than under-five children living with HIV whose mothers are not educated in both bivariate analysis (HR=0.34; 95% CI= 0.11, 0.65) and multivariable analysis (AHR: 0.37; 95% CI= 0.10, 0.85). Under-five children living with HIV whose mothers have tertiary education had 100% lower hazard of mortality than under-five children living with HIV whose mothers are not educated in both bivariate analysis (HR=0.0)and multivariable analysis (AHR: 0.00).

The hazard of death was less for under-five children living with HIV whose mothers work by a factor of (HR: 0.37; 95% CI: 0.16, 0.83) and (AHR= 0.92; 95% CI: 0.32, 0.99) for bivariate and multivariate respectively as compared to under-five children living with HIV whose mothers do not work.

Finally, weight was significant in both univariate analysis (HR = 0.85; 95% CI= 0.71, 0.91) and multivariable analysis (AHR = 0.86; 95% CI= 0.72, 0.98). For each additional kg of weight, the hazard of death in under-five children living with HIV reduces by 9% and 14% in bivariate and multivariable analyses respectively.

The estimated multivariable Cox Proportion hazard model became:

$$h(t|x) = h_0(t)\exp\left(\beta_1 x_{1,1} + \beta_2 x_{2,2} + \beta_3 x_{3,2} + \beta_3 x_{3,3} + \beta_5 x_{3,4} + \beta_6 x_{4,2}\right),$$

where

$h_0(t) = Baseline\ hazard, x_{1,1} = Weight\ ,$ $x_{2,2} = Rural\ ,$ $x_{3,2} = Primary\ Education,\ x_{3,3} = Secondary\ Education,\ x_{3,4} = Tertiary\ Education$ & $x_{4,2} = Working$

4.4.3 The goodness of fit for the Cox model

As illustrated in figure 3, the hazard rate of the Cox –Snell residuals for under five children on ART follows the 45-degree line. This suggests that the model fitted the data well.

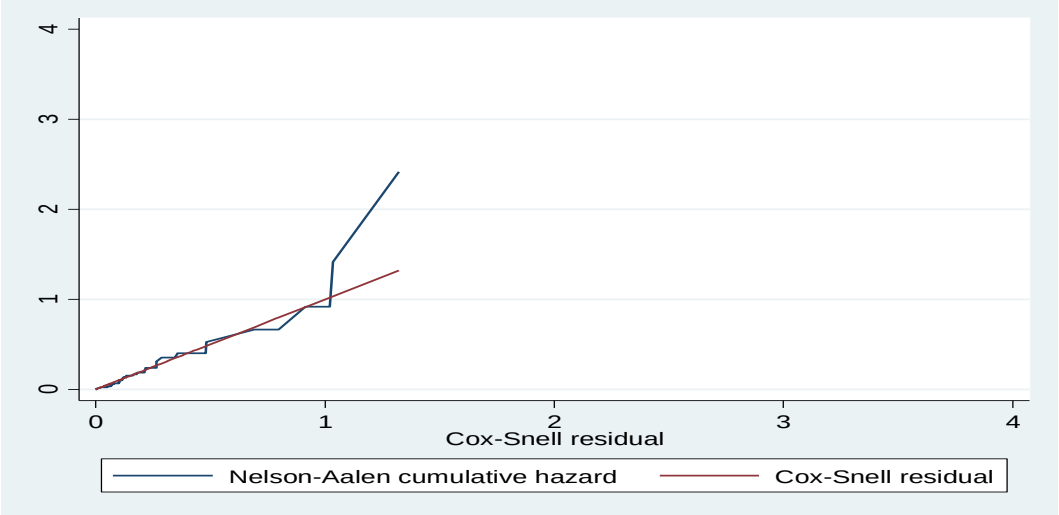


Figure 3: The goodness of model fit for Cox model

Overall, the final model fitted the data very well as illustrated in Figure 3.

4.5 Parametric models fitting

Tables 3 and 4 illustrates the hazard ratio estimates for parametric models in the analysis of bivariate and multivariable regression models. Table 5 illustrates the effect sizes of covariates in the parametric regression model of the multivariable analyses while Table 6 shows estimates for parametric model parameters.

Table 3: Bivariate Parametric regression models with hazard ratio estimates

	Weibull		Exponential		Gompertz		Loglogistic		Lognormal	
Covariate	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**Coeff**	**95% CI**	**Coeff**	**95% CI**
Weight	0.83	[0.68, 0.98]	0.80	[0.66, 0.94]	0.83	[0.70, 0.99]	0.07	[0.01, 0.15]	0.08	[0.01, 0.16]
Residence										
Semi-Urban	Reference		Reference		Reference		Reference		Reference	
Rural	2.22	[1.95, 5.20]	2.22	[1.95, 5.21]	2.24	[1.96, 5.24]	-0.30	[-0.64, -0.04]	-0.32	[-0.68, -0.04]
Education										
No	Reference		Reference		Reference		Reference		Reference	
Primary	0.91	[0.40, 0.99]	0.81	[0.36, 0.99]	0.96	[0.42, 0.98]	0.04	[0.01, 0.40]	0.04	[0.01, 0.55]
Secondary	0.55	[0.36, 0.86]	0.51	[0.33, 0.81]	0.57	[0.37, 0.87]	0.24	[0.04, 0.44]	0.28	[0.06, 0.49]
Tertiary	0.00	[0.00, 0.00]	0.00	[0.00, 0.00]	0.00	[0.00, 0.00]	5.898	[1.98, 8.7]	3.49	[2.16, 5.13]
Working Status										
Not Working	Reference		Reference		Reference		Reference		Reference	
Working	0.39	[0.17, 0.87]	0.41	[0.18, 0.92]	0.38	[0.17, 0.86]	0.34	[0.02, 0.66]	0.32	[0.23, 0.67]

The results from the fitted bivariate regression models in Table 3 have indicated that the continuous variable, weight, categorical variables residence (semi-urban and rural), mothers' education level (no education, primary, Secondary, tertiary) and mother's occupation (not working and working) significantly affect the survival rate of under-five children on antiretroviral therapy.

Table 4: Multivariable Parametric regression models with hazard ratio

	Weibull		Exponential		Gompertz		Loglogistic		Lognormal	
Covariate	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**Coef**	**95% CI**	**Coef**	**95% CI**
Weight	0.84	[0.71, 0.99]	0.81	[0.69, 0.96]	0.84	[0.71, 0.92]	0.07	[0.01, 0.15]	0.07	[0.01, 0.15]
Residence										
Semi-Urban	Reference		Reference		Reference		Reference		Reference	
Rural	1.94	[1.11, 4.61]	1.92	[1.01, 4.54]	1.93	[1.11, 4.59]	-0.25	[-0.62, -0.11]	-0.25	[-0.63, -0.13]
Education										
No	Reference		Reference		Reference		Reference		Reference	
Primary	0.91	[0.10, 0.99]	0.81	[0.36, 0.98]	0.96	[0.42, 0.99]	0.04	[0.32, 0.40]	0.04	[0.37, 0.45]
Secondary	0.34	[0.09, 0.99]	0.31	[0.09, 0.75]	0.36	[0.10, 0.85]	0.42	[0.14, 0.98]	0.46	[0.11, 1.03]
Tertiary	0.00	[0.00, 0.00]	0.00	[0.00, 0.00]	0.00	[0.00, 0.00]	5.898	[1.98, 9.77]	3.49	[0.16, 5.13]
Working Status										
Not Working	Reference		Reference		Reference		Reference		Reference	
Working	0.99	[0.34, 2.85]	1.00	[0.36, 2.82]	0.98	[0.85, 0.99]	-0.01	[-0.45, 0.43]	-0.07	[-0.52, 0.38]

The results from the fitted univariate regression models in Table 4 shows the association of covariates by hazard ratio of continuous variable, weight and categorical variables residence (semi-urban and rural), mothers' education level (no education, primary, secondary and tertiary) and mother's occupation (not working and working).

Table 5: Parametric model coefficient estimates

	Weibull		Exponential		Gompertz		Loglogistic		Lognormal	
Covariate	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**Coeff**	**95% CI**	**Coeff**	**95% CI**
Weight	0.84	[0.71, 0.99]	0.81	[0.69, 0.96]	0.84	[0.71, 0.92]	0.07	[0.01, 0.15]	0.07	[0.01, 0.15]
Residence										
Semi-Urban	Reference		Reference		Reference		Reference		Reference	
Rural	1.94	[1.11, 4.61]	1.92	[1.01, 4.54]	1.93	[1.11, 4.59]	-0.25	[-0.62, -0.11]	-0.25	[-0.63, -0.13]
Education										
No	Reference		Reference		Reference		Reference		Reference	
Primary	0.91	[0.10, 0.99]	0.81	[0.36, 0.98]	0.96	[0.42, 0.99]	0.04	[0.32, 0.40]	0.04	[0.37, 0.45]
Secondary	0.34	[0.09, 0.99]	0.31	[0.09, 0.75]	0.36	[0.10, 0.85]	0.42	[0.14, 0.98]	0.46	[0.11, 1.03]
Tertiary	0.00	[0.00, 0.00]	0.00	[0.00, 0.00]	0.00	[0.00, 0.00]	5.898	[1.98, 9.77]	3.49	[0.16, 5.13]
Working Status										
Not Working	Reference		Reference		Reference		Reference		Reference	
Working	0.99	[0.34, 2.85]	1.00	[0.36, 2.82]	0.98	[0.85, 0.99]	-0.01	[-0.45, -0.43]	-0.07	[-0.52, -0.38]
Cons	0.00	[0.00, 0.01]	0.08	[0.01, 0.44]	0.01	[0.00, 0.06]	2.65	[1.55, 3.74]	2.70	[1.51, 3.90]

Table 5 shows the estimated effect sizes of the covariates through Weibull, Exponential, Gompertz Loglogistic and Lognormal fitted models.

Table 6: Coefficient estimates for parametric model parameters

	Weibull		Exponential		Gompertz		Loglogistic		Lognormal	
Parameter	**Coeff**	**95% CI**	**Coeff**	**95% CI**	**Coeff**	**95% CI**	**Coeff**	**95% CI**	**Coeff**	**95% CI**
/n_p	0.96	[0.63, 1.29]	-	-	-	-	-	-	-	-
p	2.60	[1.87, 3.62]	-	-	-	-	-	-	-	-
1/p	0.38	[0.28, 0.53]	-	-	-	-	-	-	-	-
/gamma	-	-	-	-	0.18	[1.00, 0.25]	-	-	-	-
/ln_gam	-	-	-	-	-	-	-1.00	[-1.32,-0.67]	-	-
gamma	-	-	-	-	-	-	0.37	[0.27, 0.51]	-	-
/ln_sig	-	-	-	-	-	-	-	-	-0.29	[-0.58, -0.00]
sigma	-	-	-	-	-	-	-	-	0.75	[1.00, 0.56]

Table 6 shows the coefficients of parameters for Weibull, Exponential, Gompertz, Loglogistic and Lognormal models except Exponential model. The Exponential model does not have parameters. The parameter estimates provides shapes and location for the graphs.

4.6 Comparison of Cox and parametric survival models

In this study the Akaike Information Criterion (AIC), a statistical criterion was applied.

4.6.1 Akaike Information Criterion (AIC)

The study used the Akaike Information Criterion to select the best fit model among cox and parametric models which were fitted. Weight, residence, mothers' education level and mother's occupation were used in final model. Table 7 shows AIC statistics for the five parametric models and Cox model.

Table 7: AIC values for parametric and Cox models

Model	Observation	ll(null)	ll(Model)	df	AIC
Weibull	186	-77.18	-68.90	6	149.80
Exponential	186	-89.86	-80.13	5	186.39
Gompertz	186	-76.73	-68.76	6	149.53
Loglogistic	186	-77.69	-70.08	6	152.15
Lognormal	186	-78.22	-70.90	6	153.80
Cox	186	-133.39	-125.74	4	259.48

Table 7 illustrates Gompertz model with lowest AIC value of 149.53. With the achievement of the lowest AIC value (149.53), Gompertz is the proposed model for predicting survival of under-five children living with HIV on ART in Chiradzulu district. Gompertz model became:

$$h(t|x_i) = \lambda \exp(\gamma) \exp(\beta_1 x_{1,1} + \beta_2 x_{2,2} + \beta_3 x_{3,2} + \beta_3 x_{3,3} + \beta_5 x_{3,4} + \beta_6 x_{4,2}),$$

where

$\gamma = shape\ parameter, \lambda = location\ parameter, x_{1,1} = Weight, x_{2,2} = Rural, x_{3,2} = Primary\ Education, x_{3,3} = Secondary\ Education, x_{3,4} = Tertiary\ Education, x_{4,2} = Working$

As illustrated in figure 4, a graph of Gompertz survival function shows that 70% of under five children who were on ART survived during the study.

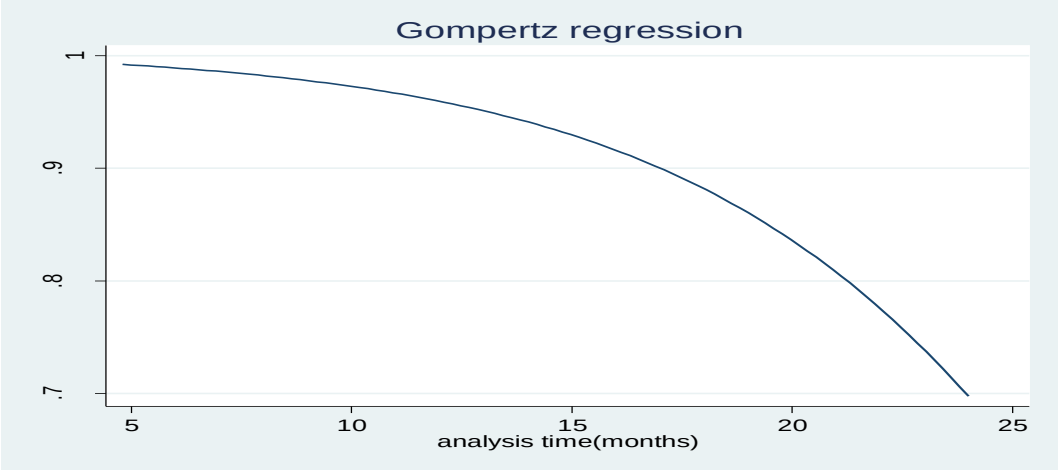


Figure 4: Gompertz survival rate graph for under-five children

4.2 Discussion

Apprehension of the survival patterns of under-five ART patients and determinants for survival is important to the development and implementation of HIV programmes for under five children hence, there is a clear need for local evidence about the burden of disease in childhood as well as determining best models for data analysis.

This study demonstrated that parametric models are best models as compared to the Cox model. This concurs with previous studies conducted by (Nardi et al., 2003). This is on the grounds that parametric models (Exponential, Weibull, Gompertz, lognormal and loglogistic), especially Gompertz model had a smallest AIC value among

Exponential, Weibull, lognormal and loglogistic while Cox model had the largest value. Akaike Information Creterion (AIC) recommends choosing the model with the lower AIC without stating the magnitude of the difference (Akaike, 1974). This result also correlates report by Stanley et al. (2016). The study found that parametric models have better performance than the Cox models. In addition, the adequacy of the Gompertz model was assessed using Cox- Snell residuals as illustrated in figure 3. The results showed that the hazard function followed the 45-degree line very closely. This implies that the Gompertz model is an appropriate and worth model to be used for modelling the survival of under-five children living with HIV on ART in Chiradzulu district.

Another objective of this study was to examine the extent to which risk factors influence the survival of under-five children living with HIV on ART in Chiradzulu district. The results suggest that the hazard of death for under-five children who are on ART depends greatly on a number of socio-economic, demographic and health-related variables, such as place of residence mother educational level, occupation status and weight of child.

The study suggests that mortality rates are higher in rural areas. This concurs with previous studies in and outside Nigeria (Morakinyo et al., 2017). This is on the ground that those living in the urban areas have access to improved water supply, improved sanitation facilities, unlimited access to healthcare as well as other social and economic services (Dejene et al., 2013). This finding also correlates with result reported in previous study (Adekambi et al., 2011). Rural areas are usually far away from health facilities hence poor health care services.

Occupation of mother is also another factor influencing under-five child survival rate in this study. The study found evidence that mortality risk might be lower in group of under-five children with working mothers as compared to the group of under-five children with non-working mothers. This was consistent with the findings by (Bello et al., 2014). The households with working mothers have better housing conditions, better nutrition, more empowerment and hence, they may be able to afford better medical attention and care thus, significantly enhancing the survival rate of all their children (Wegbom et al., 2019).

The study also discovered that under-five children living with HIV with low body weight are associated with increased hazard of death. As found in another study, the greatest risk of mortality occurred in the first 12 months of treatment (Gesesew, 2018) among children who were underweight at ART initiation and among infants (Mutanga, 2019). These under-five children with more weight have better nutrition as compared to those under five children with low body weight (Johannes Sen et al., 2008).

Finally, mother's education had a significant relationship with child survival rate in this study. Higher risks of death for under-five children were identified among the uneducated mothers as compared to educated mothers (mothers with primary, secondary and tertiary education). This is consistent with the findings by Khan and Awan (2017). A possible explanation could be that educated mothers have better socioeconomic status, better knowledge on family health and childcare, preventive care, effective use of modern health services and good management of child illness (Caldwell, 1994). Education additionally changes the customary and social family relationships as regards decision making and engages the mothers in several issues like

childcare which plays a role in reducing child mortality (O'Toole, 1994). Higher mother's education tends to have a lower hazard than those with lower education and mothers with higher education have a higher desire to seek information or knowledge about health care (Nurmalasari et al., 2019).

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The findings from this study have essential district policy implications, particularly in monitoring public health interventions which need to ensure a consistent decline in child mortality rates towards the achievement of the SDG 3. The study analysed and then, identified the socioeconomic, geographical and health-related factors that might influence the survival rate of under-five children on antiretroviral therapy in Chiradzulu. The results from the analyses showed that mother educational level, occupation status, place of residence and weight of a child during antiretroviral therapy might influence the survival rate of under-five children in Chiradzulu. These findings suggest that under-five survival rate is greatly associated with socio-economic, geographical and health related factors. The study determined Gompertz, a parametric model, as the best model for predicting survival rate of under-five children on antiretroviral therapy as compared to Cox model.

5.2 Recommendations

Recommendations from this study are that, it is necessary for the readers to understand the identified factors influencing the survival rate of under five children to be directions for future studies with plans to design study procedures that can explain confounders. Further still, this study's data is from Chiradzulu district hence, it is recommended that these results should be substantiated by similar survival studies from other parts of the district to generalize the results to other individuals in the country. Finally, researchers

should check the underlying assumptions of Cox model before using it in order to use a proper models during analysis.

REFERENCE

Adekambi, V.T., Kayode, G.A., & Uthman, O.A. (2011). *Individual and Contextual factors associated with Childhood stunting in Nigeria: a Multilevel analysis. Maternal and child nutrition http://*doi/10.1111/j.1740-8709.2011. 00361.

Aho, K. D., Derryberry, D., & Teri, P.T. (2014). Model selection for ecologists: *the worldviews of AIC and BIC.* (Ecology, 95(3), 2014, pp. 631–636 2014 by the Ecological Society of America).

Akaike, H. (1974). *A new look at the statistical model identification*. *IEEE Transactions on automatic Control*, 19(6), 716 -23.

Altiman, D.G. (1998). Suboptimal analysis using 'optimal' cut points. *Br J Cancer* 78: 556 – 557.

Andreas, J., Harries, A. D., Schouten, E.J., Libamba, E., Ford, N., Maher, D., & Chimbwandira, F. (2016). Scaling up antiretroviral therapy in Malawi. Bulletin of the World Health Organisation. 2016 Oct 1, 94(10), 772-776.

Ata, N., & Sozer, M. (2007). *Cox regression models with non-proportional hazards applied to lung cancer survival data.* Hacettepe Journal of Mathematics and Statistics, 36 (2), 157{167}.

Atmab, D.G., & Anderson P.K. (1989). *Bootstrap investigation of the stability of a Cox regression model.* Stat Med, 8,771-783.

Ayalu, A.R., & Biadgilign, S. (2012). Determinants of adherence to antiretroviral therapy among HIV-Infected patients in Africa. AIDS research and treatment, 2012. https//doi:10.1155/2012/574656.

Bello, R.A., & Joseph, A.I. (2014). Determinants of Child Mortality in Oyo State,Nigeria. African Research Review, 8(1), 252-272. DOI: http://dx.doi.org/10.4314/afrrev.v8i1.17.

Byers, R.H.J.R., Morgan, W.M., Darrow, W.W.W., Doll, L., Jaffe, H.W., & Rutherford, G. (1988). *Estimating AIDS infection rates in the Sanfransisco Cohort*. *AIDS*, 2(3), 207 10.

Caldwell, J.C. (1994). *How is greater maternal education translated into lower child mortality? Health Transition Review.* 4(2), 224–9.

Cambridge, C. (2015). Regional Variations in Infant and Child Mortality in Nigeria: *A Multilevel Analysis*. J. Biosoc. Sci. 2015, 47, 165–187. 13.

Chiradzulu District Council, (2017). *Chiradzulu Social Economic Profile (2017/22).*

Cleves, M.A., Gould, W.W. & Gutierrez, R.G. (2010). *An introduction to survival analysis using STATA*. Place of publication. A press.

Cox, D. R., & Oakes, D. (1984). *Analysis of survivor data.* London: Chapman and hall.

Cox, D.R., & Oakes, D. (1996). *Analysis of survivor data.* London: Chapman and hall.

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of Royal Statistical Society, 34, 187- 220.*

Cox, D.R., (1975). Partial likelihood. *Biometrika* 62, 269-276.

Fox, M.P., & Rosen, S. (2015). Systematic review of retention of pediatric patients on HIV treatment in low and middle-income countries 2008-2013. *AIDS.* London: England. 2015, 29(4), 493–502.

Dejene, T., & Girma, E. (2013). Social determinants of under-five mortality in Ethiopia: *Event history analysis using evidence from Ethiopian Demographic and Health Survey* (EDHS). Health (Irvine. Calif), 5, 879–84.

Gesesew, H.A., Ward P., Woldemichael K., & Mwanri, L. (2018). Early mortality among children and adults in antiretroviral therapy programs in Southwest Ethiopia, 2003 –15. PLoS One. 2018; 13(6), e0198815.

Getachew, Y., & Bekele, S. (2016). Survival Analysis of Under-Five Mortality of Children and its Associated Risk Factors in Ethiopia. J. Biosens Bioelectron 7, 213. doi:10.4172/2155-6210.1000213.

Goeman, J.J., Oosting, J., Jansen, A.M.C., Anninga, J.K., & Houwelinigen, H.C. Testing association of a pathway with survival using gene expression data. Bioinformatics 2005; 21, 1950–1957.

Isingo, R., Zaba, B., Marston, M., Ndenge, M., Wringe, A., Beckles, D., Changalucha, J., & Urassa, M. (2007). *Survival after HIV infection in the pre antiretroviral therapy era in rural Tanzania cohort. AIDS,* 21(Suppl.6), 55-513.

Jessica, L., Schue, P., Janneke, H., Dijk, V., Hamangaba, F., Hamahuwa, M., Moyo, N., Philip, E., Thuma, P.E., Moss, W.J., & Sutcliffe, C.G. (2021). Treatment outcomes among children younger than five years living with, HIV in rural Zambia, 2008–2018: *a cohort study. BMC Pediatrics* (2021)21, 315. https://doi.org/10.1186/s12887-021-02793.

Johannessen, A., Naman, E., Ngowi, B. J., Sandivik, L., Matee, M.I., Aglen, H.E., Gunersen, S.G., & Bruun, J.N. (2008). Predictors of mortality in HIV-infected patients starting antiretroviral therapy in a rural hospital in Tanzania. *BMC infectious diseases,* 2008, 8:52. https//doi:10.1186/1471-2334-8-52.

Kalbfleisch, J.D., & Prentice, R.L. (2011). *The statistical Analysis of Failure Data,* (2nd ed.). New York: Wiley.

Kaplan, E.L., & Meier, P.L. (2018). *Non parametric estimation from incomplete observation.* Journal of the American statistical association, 53, 457-481.

Khan, J.R., & Awan, N. (2017). A comprehensive analysis on child mortality and its determinants in Bangladesh using frailty models Archives of Public Health. 75, 58. DOI: 10.1186/s13690-017-0224-6.

Kleibaum, D.G., & Klein, M. (2010). *Survival analysis: a Self –learning text* (second edition*).* New Jersey: John Willey & Sons.

Klein, J.P., Houwelingen, H.C., Ibrahim, J.G. & Scheike, T.H. (2016). Handbook of survival analysis. CRC Press.

Lawless, J.F. (2011). *Statistical Models and Methods for Life Time Data.* John Wiley & Sons, Hoboken.

Lee, E.T., & Wang, J.W. (2003). *Some well-known parametric survival distribution and their applicability. Statistical methods for survival data analysis (3rd ed).* Hoboken, New Jersey: *John Willey.*

Machin, D., Cheung, Y.B. & Parmar, M. (2006). Survival Analysis: *A Practical Approach* (2nd ed). West Sussex, England: John Wiley & Sons.

Malawi Child mortality rate. *Accessed on 17th April, 2021 from: https://knoema.com /atlas /Malawi/Child-mortality-rate.*

Malawi Ministry of Health (2009). *Guidelines for ART provision (3rd).* Lilongwe. Ministry of Health.

National Statistical Office (2019). Malawi Population and Housing Census 2018 main Report.

Malawi Population based HIV Impact Assessment (2016). A Final survey report.

Mantel, N. (1966). *Evaluation of survival data and two new rank statistics arising its consideration.* Cancer – thermotherapy Report, 50, 163-170.

Marinda, E., Humphrey, J.H., Iliff, P.J., Mutasa, K., Nathoo, K.J., Piwoz, E.G., Moulton, L.H., Salama, P., & Ward, B.J. (2007). *Child Mortality According to Maternal and Infant HIV Status in Zimbabwe.* Pediatric Infectious Diseases, 26**,** 519-52 610.10 97/01.inf. 0000 26 4527.69954.4c.

MDHS (2016*). Malawi Demographic and Health Survey 2016*. Zomba: National Statistical Office.

MoH, HIV Unit (2004). *Malawi ART Progress Report for 2003*. Lilongwe: Ministry of Health.

MoH, HIV Unit (2011). *Malawi ART Progress Report for 2010*. Lilongwe: Ministry of Health.

Morakinyo, O.M., & Fagbamigbe, A.F. (2017). Neonatal, infant and under-five mortalities in Nigeria: *An examination of trends and drivers* (2003-2013). PLoS ONE 12(8), e0182990. https://doi.org/10.1371/journal.pone.0182990.

Mutanga, J.N., Mutembo, S., Ezeamama, A.E., Song, X., Fubisha, R.C., Kapembwa, K.M., Sialondwe, D., Simuchembu, B., Chinyonga, J., Thuma, P.E., & Whalen, C.C. (2019).Long-term survival outcomes of HIV infected children receiving antiretroviral therapy: *an observational study from Zambia* (2003- 2015). BMC. Public Health. 2019; 19(1), 115.

Mykhailo, A., Dmytro, Z. (1213). *Effect of antiretroviral therapy on survival of HIV/tb – infected patients in Ukraine*. Our Dermatol online, 4(2).

Nardi, A., & Schemper, M. (2003). Comparing Cox and parametric models in clinical studies. Statist Med, 22, 3597-610.

National Statistical Office (2018). *Malawi Population and Housing Census Report 2018*.

National Statistical Office (2017). *Malawi Demographic and Health Survey 2015/16*. Zomba, Malawi.

Newell, M.L., Brahmbhatt, H., & Ghys, P. D. (2004). *Child mortality and HIV infection in Africa*. 18**,** p S27-S34.

Nurmalasari, M., Ifada, L., & Pramana, S. (2019). Survival Analysis of Under-five Mortality in West Sulawesi Indonesia using Cox Regression. *In Proceedings of the 1st International Conference on Health* (ICOH 2019), pages 92-96 ISBN. 978-989 758454-1. DOI: 10.5220/0009566900920096.

Oakes, D. (1977). The asymptotic information in censored survival data. Biometrika, 64,441-8.

O'Toole, J., & Wright, R.E. (1991). Parental education and child mortality in burundi. J Biosoc Sci. 23(3), 255–62.

Obimbo, E. M., Ngacha, M.D.A., Ochieng, J. O., Richardson, B. A., Otieno, P.A., Bosire, R., Farquhar, C., Overbaugh, J. & Stewart, J.G.C. (2004). *Predictors of early mortality in a cohort of human immunodeficiency virus type 1-infected african children. Pediatr Infect Dis J,* 23**,** 536-43.

Rinne, H. (2020). The Weibull Distribution: *A Handbook.* Chapman and Hall/CRC. 808 Pages. ISBN 9780367577469.

Robert, Z., Gilbert, R., & Jean, P. (2012). Survival differences in Australian Patients with acquired Immunodeficiency Syndrome. *European Journal of Epidemiology* 1995, 11:519-526.*AScand J infect Dis* mar: 44(3), 201-8. Do10.3109/00365548.2011.631937. Epub 2011 nov 28.

Stanley, C., Molyneux, E., & Mukaka, M. (2016). *Comparison of performance of exponential, Cox proportional hazards, Weibull and frailty survival models for analysis of small sample size data.* Journal of Medical Statistics and Informatics 4(2), 456-71.

Therneau, T. M., Grambsch, P.M., & Fleming, T.R. (1990). *Martingale-based residuals for survival models.* Biometrika, 77,147-160.

United Nations Programme on HIV and AIDS, (2019). *Communities at the centre*: Global AIDS Update 2019.

United Nations Programme on HIV and AIDS (2014), AIDS information update 2014.

United Nations International Children Emergency Fund (2021). *Global HIV and AIDS update* 2021.

United Nations International Children Emergency Fund (2018). Coverage of Antiretroviral Treatment among Children aged 0-14 living with HIV.

UNICEF (2017). *Levels and Trends in Child Mortality*: Report 2017. Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation. Available from: http://www.childmortality.org/files_v20/download/igme%20report%202015%20c.

United Nations (2015). Transforming our World: the 2030 Agenda for Sustainable Development. United Nations. Retrieved April 28, 2018, from http://www.un.org/ga/search/viewdoc.asp?symbol=A/RES/70/1&Lang=E.

United Nations Inter-Agency Group for Child Mortality Estimation (2021). '*Levels & Trends in Child Mortality*: Report 2021, Estimates Developed by the United Nations Inter-Agency Group for Child Mortality Estimation', United Nations Children's Fund, New York.

Venkatesh, K. K., DeBruyn, G., Marinda, E., Otwombe, K., Van Niekerk, R., Urban, M., Triche, E. W., Mcgarvey, S.T., Lurie, M.N. & Gray, G.E. (2011). *Morbidity and mortality among infants born to HIV-infected women in South Africa: implications for child health in resource-limited settings*. J Trop Pediatric, 57**,** 109-19.

Viscomi, S., Pastorore, G., Damma, E., Zuccolo, L., & Peace, N. (2006). *Life expectancy as an indicator in Follow up of population-based Cancer registries. The example of childhood leukemia*. Ann Oncol, 17 (1), 167-71.

Wegbom, A.I., Essi I.D., & Kiri, V.A. (2019). Survival Analysis of Under-five Mortality and its Associated Determinants in Nigeria: *Evidence from a Survey Data. International Journal of Statistics and Applications* 2019, 9(2): 59-66 DOI: 10.5923/j.statistics.20190902.03.

Williams, B.K., Nichols, J.D., & Conroy, M.J. (2002). Analysis and management of animal populations. Academic Press.

Wilson, D.L. (1994). The analysis of survival (mortality) data: fitting Gompertz, Weibull and logistic functions. 74(1-2), 15-33. Doi: 10.1016/00476374 (94)90095-7.

World Health Organization, (2021). HIV and AIDS in Malawi 2021.

World Health Organization (2018). Malawi Clinical HIV Guidelines 2018.

Zhou, Y.Y. (2007). *Estimating of parameters in logistic and log-logistic distribution with grouped data.* Lifetime Data Anal, 13(3), 421-9.

Zhou, J., & Nkumarasamyi. (2005). *Predicting short-term disease progression among HIV-infected patients in Asia and Pacific region:* Preliminary results from the treat Asia HIV observation Database CTAHOOD.

APPENDICES

APPENDIX 1: Commands for data analysis

```
use "F:\New folder (7)\RESEARCH
cONCEPT\WORKING\THESIS\Revised_Dataset_14Nov2019.dta"
stset Time_months , failure( Died )
summarize Time_months
summarize Time_months, detail
sts graph, na
**Distribution graphs
histogram Age_0, freq normal
histogram Weight_0, freq normal
histogram Height_0, freq normal

stvary //checking time varying variables
**Descriptive
tab Sex
tab Residence
tab M_Education
tab M_Occupation
tab Marital_Status
tab Died
***
tabstat Age_0, statistics (mean median sd iqr)
tabstat Weight_0, statistics (mean median sd iqr)
tabstat Height_0, statistics (mean median sd iqr)
*****
sum Age_0, detail
sum Height_0, detail
sum Weight_0, detail
**Defining variables
label define Sex 1 "Male" 2 "Female"
label values Sex Sex
```

```
label define Residence 1 "Semi Urban" 2 "Rural"
label values Residence Residence

label define Mother_Education 1 "No Education " 2 "Primary Education" 3
"Secondary Education" 4 "Tertiary"
label values M_Education Mother_Education

label define Mother_Occupation 1 "Not working" 2 "Working"
label values M_Occupation Mother_Occupation

label define Marital_Status 1 "Single" 2 "Married" 3 "Divorced"
label values Marital_Status Marital_Status

**Table 4.1 Descriptive Summary of Explanatory Variables
tab Sex Died, row
tab Residence Died, row
tab M_Education Died, row
tab M_Occupation Died, row
tab Marital_Status Died, row
sum Age_0, detail //report mean and standard deviation
sum Height_0 , detail //report mean and stsndard deviation
sum Weight_0 , detail //report mean and stsndard deviation

**Table 4.2 Association between Death and Covariates
tab Sex Died, chi2
tab Residence Died, chi2
tab M_Education Died, chi2
tab M_Occupation Died, chi2
tab Marital_Status Died, chi2
ttest Age_0, by(Died)
ttest Height_0, by(Died)
ttest Weight_0, by(Died)
tab Age_0 Died, chi2
sts test Residence, logrank
```

```
sts graph, by(Residence)

sts test Sex, logrank
sts graph, by(Sex)

sts test M_Education, logrank
sts graph, by(M_Education)

sts test M_Occupation, logrank
sts graph, by(M_Occupation)

sts test Marital_Status, logrank
sts graph, by(Marital_Status)

//Cox Proportional Hazard Models with one predictor variables
stcox Age_0, nohr
stcox Height_0, nohr
stcox Weight_0, nohr

//Model Building
//UNIVARIATE COX PROPORTIONAL HAZARD MODELS /
stcox Sex
stcox Residence
stcox Age_0
stcox Height_0
stcox Weight_0
stcox M_Education
stcox M_Occupation
stcox Marital_Status

stcox i.Sex
stcox i.Residence
stcox Age_0
stcox Height_0
```

```
stcox Weight_0
stcox i.M_Education
stcox i.M_Occupation
stcox i.Marital_Status

stcox Residence, nohr
stcox M_Education, nohr
stcox M_Occupation, nohr

//MULTIVARIATE COX PROPORTIONAL HAZARD MODELS
stcox i.Residence i.M_Education i.M_Occupation Weight_0, nolog
stphplot, by(Residence)
stcoxkm, by(Residence) separate legend(cols(1))
stcoxkm, by(M_Education) separate legend(cols(1))
stphplot, by(M_Education)
stphplot, by(M_Occupation)
stcoxkm, by(M_Occupation) separate legend(cols(1))

//Tests of proportionalhazards assumption
stcox  Weight_0 Residence, nolog
estat phtest, detail

stcox  Weight_0 M_Education, nolog
estat phtest, detail

stcox  Weight_0 M_Occupation, nolog
estat phtest, detail

//Cox AIC test
stcox  Weight_0 Residence M_Education M_Occupation, nolog
estat ic

//multilevel GLOBAL TEST
stcox  Weight_0 i.Residence i.M_Education i.M_Occupation, nolog
```

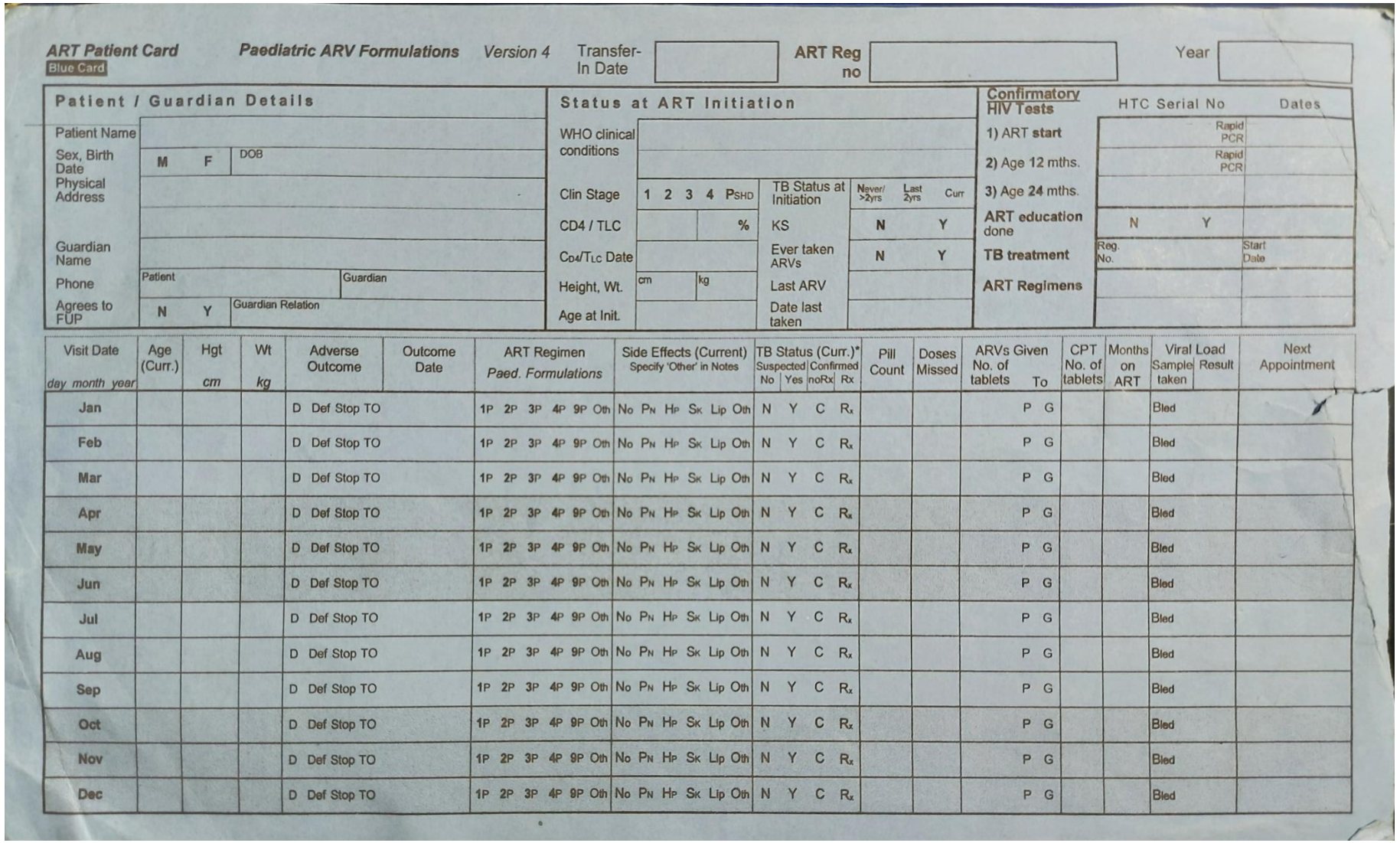
```
estat phtest, detail

//COMPARING PARAMETRIC SURVIVAL MODELS UNIVARIATE
streg Weight_0, distribution(weibull)
streg Weight_0, distribution(exponential)
streg Weight_0, distribution(gompertz)
streg Weight_0, distribution(loglogistic)
streg Weight_0, distribution(lognormal)

streg Age_0, distribution(weibull)
streg Age_0, distribution(exponential)
streg Age_0, distribution(gompertz)
streg Age_0, distribution(loglogistic)
streg Age_0, distribution(lognormal)

streg Height_0, distribution(weibull)
streg Height_0, distribution(exponential)
streg Height_0, distribution(gompertz)
streg Height_0, distribution(loglogistic)
streg Height_0, distribution(lognormal)

streg i.Sex, distribution(weibull)
streg i.Sex, distribution(exponential)
streg i.Sex, distribution(gompertz)
streg i.Sex, distribution(loglogistic)
streg i.Sex, distribution(lognormal)

streg i.Residence, distribution(weibull)
streg i.Residence, distribution(exponential)
streg i.Residence, distribution(gompertz)
streg i.Residence, distribution(loglogistic)
streg i.Residence, distribution(lognormal)
streg i.M_Education, distribution(weibull)
streg i.M_Education, distribution(exponential)
```

```
streg i.M_Education, distribution(gompertz)
streg i.M_Education, distribution(loglogistic)
streg i.M_Education, distribution(lognormal)

streg i.M_Occupation, distribution(weibull)
streg i.M_Occupation, distribution(exponential)
streg i.M_Occupation, distribution(gompertz)
streg i.M_Occupation, distribution(loglogistic)
streg i.M_Occupation, distribution(lognormal)

streg i.Marital_Status, distribution(weibull)
streg i.Marital_Status, distribution(exponential)
streg i.Marital_Status, distribution(gompertz)
streg i.Marital_Status, distribution(loglogistic)
streg i.Marital_Status, distribution(lognormal)

//COMPARING PARAMETRIC SURVIVAL MODELS MULTIVARIATE //
streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(weibull)
nolog
estat ic

streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(exponential)
nolog
estat ic

streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(gompertz)
nolog
estat ic
streg Weight_0 Residence i.M_Education i.M_Occupation, distribution(loglogistic)
nolog
estat ic

streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(lognormal)
nolog
```

```
estat ic

//with coefficient estimates
streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(weibull)
nohr
streg Weight_0 i.Residence i.M_Education i.M_Occupation,
distribution(exponential)nohr
streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(gompertz)
nohr

streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(loglogistic)
nolog
streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(lognormal)
nolog

//to select the best model fit the model and type
estat ic

//schonfeld test
quietly stcox Weight_0 Residence M_Education M_Occupation, schoenfeld(sch*)
scaledsch(sca*)
stphtest, detail
stphtest, plot(Weight_0) msym(oh)
stphtest, plot(Residence) msym(oh)
stphtest, plot(M_Education) msym(oh)
stphtest, plot(M_Occupation) msym(oh)

//Goodness of fit cox model
quietly stcox Weight_0 Residence M_Education M_Occupation, nohr mgale(mg)
predict cs, csnell
stset cs, failure(Died)
sts generate H = na
line H cs cs, sort xlab(0 1 to 4) ylab(0 1 to 4)
drop mg
```

```
// survival function
stcox  Weight_0 Residence M_Education M_Occupation, nolog
stcurve, survival
stcurve, survival at1(Weight_0=1) at2(Residence=2) at3(M_Education=3)
at4(M_Occupation=4)
lpattern(solid dash dot)
stcurve, cumhaz

streg Weight_0 i.Residence i.M_Education i.M_Occupation, distribution(gompertz)
stcurve, survival
stcurve, hazard
stcurve, cumhazard
```

APPENDIX 2: ART patient card, front



ART Patient Card Blue Card ***Paediatric ARV Formulations*** *Version 4* Transfer-In Date ART Reg no Year

Patient / Guardian Details

Patient Name
Sex, Birth Date — M F — DOB
Physical Address
Guardian Name
Phone — Patient — Guardian
Agrees to FUP — N Y — Guardian Relation

Status at ART Initiation

WHO clinical conditions
Clin Stage — 1 2 3 4 PSHD
CD4 / TLC — %
CD4/TLC Date
Height, Wt. — cm — kg
Age at Init.

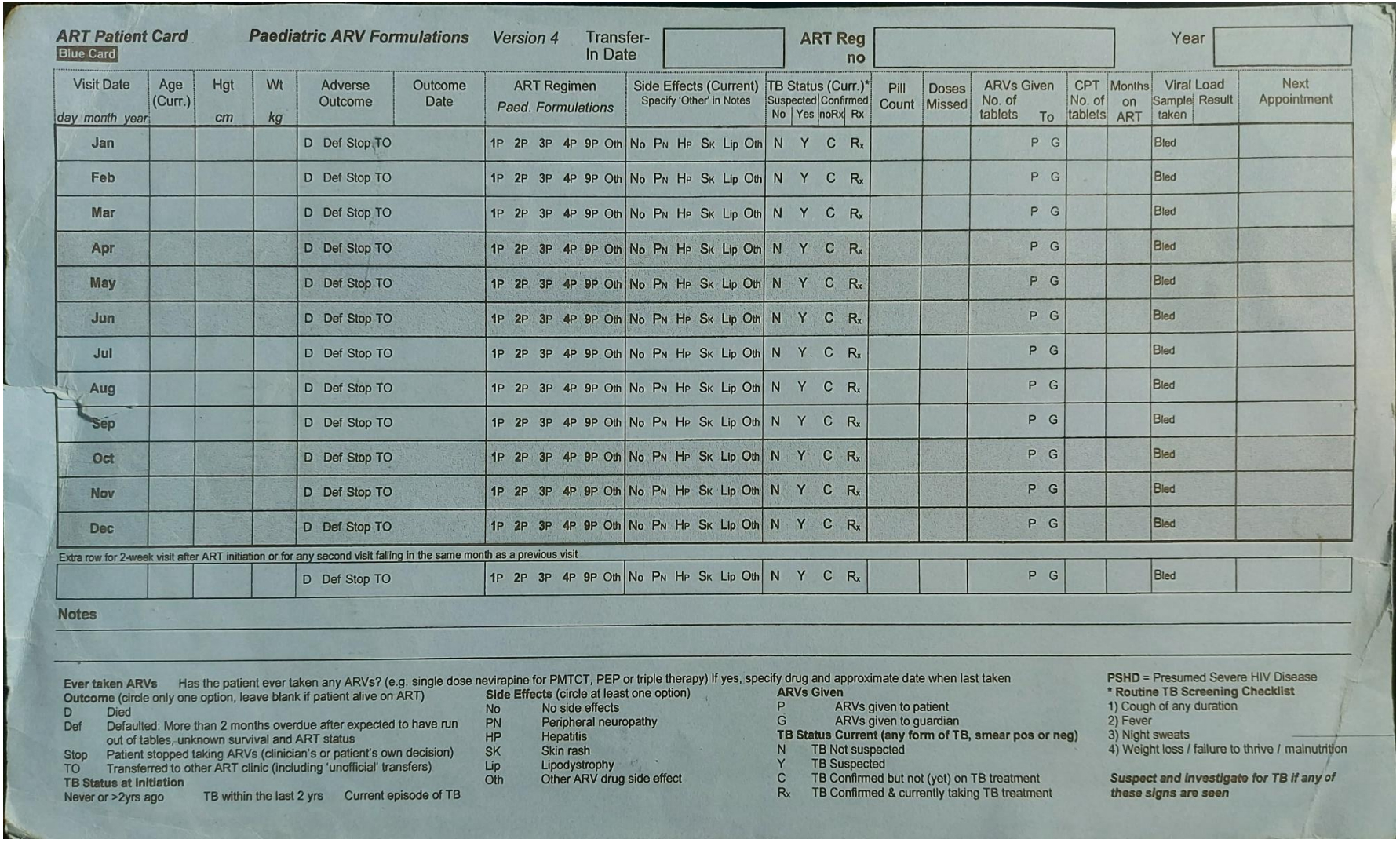
TB Status at Initiation — Never/ >2yrs — Last 2yrs — Curr
KS — N Y
Ever taken ARVs — N Y
Last ARV
Date last taken

Confirmatory HIV Tests — HTC Serial No — Dates

1) ART **start** — Rapid PCR
2) Age 12 mths. — Rapid PCR
3) Age 24 mths.
ART education done — N Y
TB treatment — Reg. No. — Start Date
ART Regimens

Visit Date *day month year*	Age (Curr.)	Hgt *cm*	Wt *kg*	Adverse Outcome	Outcome Date	ART Regimen *Paed. Formulations*	Side Effects (Current) Specify 'Other' in Notes	TB Status (Curr.)* Suspected No / Yes	Confirmed noRx / Rx	Pill Count	Doses Missed	ARVs Given No. of tablets P	ARVs Given No. of tablets G	CPT No. of tablets	Months on ART	Viral Load Sample taken	Viral Load Result	Next Appointment
Jan				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Feb				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Mar				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Apr				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
May				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Jun				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Jul				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Aug				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Sep				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Oct				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Nov				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		
Dec				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N	Y C Rx			P	G			Bled		

APPENDIX 3: ART patient card, back



ART Patient Card Blue Card — **Paediatric ARV Formulations** *Version 4* — Transfer-In Date — **ART Reg no** — Year

Visit Date (day month year)	Age (Curr.)	Hgt (cm)	Wt (kg)	Adverse Outcome	Outcome Date	ART Regimen *Paed. Formulations*	Side Effects (Current) Specify 'Other' in Notes	TB Status (Curr.)* Suspected: No / Yes; Confirmed: noRx / Rx	Pill Count	Doses Missed	ARVs Given No. of tablets: P / G	CPT No. of tablets	Months on ART	Viral Load Sample taken	Viral Load Result	Next Appointment
Jan				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Feb				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Mar				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Apr				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
May				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Jun				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Jul				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Aug				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Sep				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Oct				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Nov				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Dec				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		
Extra row for 2-week visit after ART initiation or for any second visit falling in the same month as a previous visit																
				D Def Stop TO		1P 2P 3P 4P 9P Oth	No PN HP SK Lip Oth	N Y C Rx			P G			Bled		

Notes

Ever taken ARVs Has the patient ever taken any ARVs? (e.g. single dose nevirapine for PMTCT, PEP or triple therapy) If yes, specify drug and approximate date when last taken

Outcome (circle only one option, leave blank if patient alive on ART)
- D Died
- Def Defaulted: More than 2 months overdue after expected to have run out of tables, unknown survival and ART status
- Stop Patient stopped taking ARVs (clinician's or patient's own decision)
- TO Transferred to other ART clinic (including 'unofficial' transfers)

TB Status at Initiation
Never or >2yrs ago TB within the last 2 yrs Current episode of TB

Side Effects (circle at least one option)
- No No side effects
- PN Peripheral neuropathy
- HP Hepatitis
- SK Skin rash
- Lip Lipodystrophy
- Oth Other ARV drug side effect

ARVs Given
- P ARVs given to patient
- G ARVs given to guardian

TB Status Current (any form of TB, smear pos or neg)
- N TB Not suspected
- Y TB Suspected
- C TB Confirmed but not (yet) on TB treatment
- Rx TB Confirmed & currently taking TB treatment

PSHD = Presumed Severe HIV Disease

*** Routine TB Screening Checklist**
1) Cough of any duration
2) Fever
3) Night sweats
4) Weight loss / failure to thrive / malnutrition

Suspect and investigate for TB if any of these signs are seen

APPENDIX 4: Support letter from Chancellor College

PRINCIPAL
Prof. Richard Tambulasi, B.A (Pub Admin), BPA (Hons), MPA, Ph.D
Our ref: CC/PF/AC
Your ref:

CHANCELLOR COLLEGE
P.O. Box 280, Zomba, Malawi
Telephone: (265) 524 222
Fax: (265) 524 046
E-mail: principal@cc.ac.mw

2nd July 2019

The District Health Officer
Chiradzulu District Hospital
CHIRADZULU.

Dear Sir/Madam

REFERENCE LETTER FOR MR AUBREY JAZZA (MSC/STAT/02/14)

I write this reference letter for Mr Aubrey Jazza who is our Master of Science in Biostatistics student in the Mathematical Sciences Department, under the Faculty of Science, Chancellor College.

He is currently writing a research proposal titled, "Survival of Under - Five HIV/Aids infected Children taking antiretroviral therapy at Chiradzulu District Hospital: a comparison of stratefied Cox Regression and Extended Cox Regression Models" so he would like to collect data in your organization. Please note that this research is for academic purpose only and therefore any research ethics of confidentiality shall be honoured by the student and the department.

Any assistance rendered to him will therefore be highly appreciated.

Yours faithfully

pp Dr Mwawi Nyirenda-Kayuni
HEAD, MATHEMATICAL SCIENCES DEPARTMENT