UNIVERSITY *of the*
WESTERN CAPE

SURVIVAL MODELLING AND ANALYSIS OF HIV/AIDS PATIENTS ON HIV CARE AND ANTIRETROVIRAL TREATMENT TO DETERMINE LONGEVITY PROGNOSTIC FACTORS

By

Innocent Maposa

A THESIS SUBMITTED IN FULFILMENT FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY IN STATISTICS
IN THE DEPARTMENT OF
STATISTICS AND POPULATION STUDIES, FACULTY OF NATURAL SCIENCES
UNIVERSITY OF THE WESTERN CAPE
NOVEMBER 2016

Supervised by Prof. Rénette Blignaut

Declaration

I hereby declare that ***"SURVIVAL MODELLING AND ANALYSIS OF HIV/AIDS PATIENTS ON HIV CARE AND ANTIRETROVIRAL TREATMENT TO DETERMINE LONGEVITY PROGNOSTIC FACTORS"*** is my own work, and that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Signed: Innocent Maposa

Date: November 2016

To Sandra and Paul Mufaro

Table of Contents

List of Tables

List of Figures

Abstract

The HIV/AIDS pandemic has been a torment to the African developmental agenda, especially the Southern African Development Countries (SADC), for the past two decades. The disease and condition tends to affect the productive age groups. Children have also not been spared from the severe effects associated with the disease. The advent of antiretroviral treatment (ART) has brought a great relief to governments and patients in these regions. More people living with HIV/AIDS have experienced a boost in their survival prospects and hence their contribution to national developmental projects.

Survival analysis methods are usually used in biostatistics, epidemiological modelling and clinical research to model time to event data. The most interesting aspect of this analysis comes when survival models are used to determine risk factors for the survival of patients undergoing some treatment or living with a certain disease condition.

The purpose of this thesis was to determine prognostic risk factors for patients' survival whilst on ART. The study sought to highlight the risk factors that impact the survival time negatively at different survival time points. The study utilized a sample of paediatric and adult datasets from Namibia and Zimbabwe respectively. The paediatric dataset from Katutura hospital (Namibia) comprised of the adolescents and children on ART, whilst the adult dataset from Bulawayo hospital (Zimbabwe) comprised of those patients on ART in the 15 years and above age categories. All datasets used in this thesis were based on retrospective cohorts followed for some

period of time. Different methods to reduce errors in parameter estimation were employed to the datasets. The proportional hazards, Bayesian proportional hazards and the censored quantile regression models were utilized in this study.

The results from the proportional hazards model show that most of the variables considered were not significant overall. The Bayesian proportional hazards model shows us that all the considered factors had different risk profiles at the different quartiles of the survival times. This highlights that by using the proportional hazards models, we only get a fixed constant effect of the risk factors, yet in reality, the effect of risk factors differs at different survival time points. This picture was strongly highlighted by the censored quantile regression model which indicated that some variables were significant in the early periods of initiation whilst they did not significantly affect survival time at any other points in the survival time distribution. The censored quantile regression models clearly demonstrate that there are significant insights gained on the dynamics of how different prognostic risk factors affect patient survival time across the survival time distribution compared to when we use proportional hazards and Bayesian propotional hazards models. However, the advantages of using the proportional hazards framework, due to the estimation of hazard rates as well as it's application in the competing risk framework are still unassailable. The hazard rate estimation under the censored quantile regression framework is an area that is still under development and the computational aspects are yet to be incorporated into the mainstream statistical softwares.

This study concludes that, with the current literature and computational support, using both model frameworks to ascertain the dynamic effects of different prognostic risk factors for survival in people living with HIV/AIDS and on ART would give the researchers more insights. These insights will then help public health policy makers to draft relevant targeted policies aimed at improving these patients' survival time on treatment.

Acknowledgements

I would like to thank Professor Rénette Blignaut, my supervisor, for her numerous suggestions and constant support during this research. I am also thankful to her for her guidance through the early days and years of chaos and confusion.

I should also mention that my graduate studies in South Africa were supported in part by my Professor and supervisor Rénette Blignaut. Thank you very much professor.

Of course, I am grateful to my wife Sandra and son Paul for their patience, support and unequalled motivation and *love* during the years of my studies. Without them this work would never have come into existence (literally).

Finally and most importantly, I wish to thank God for giving me the strength and courage to continue with the studies regardless of the difficulties along the way.

Cape Town, South Africa — Innocent Maposa
November, 2016

Acronyms and abbreviations

AIDS	Acquired Immunodeficiency Syndrome
HIV	Human Immunodeficiency Virus
CQR	Censored Quantile Regression
PHM	Proportional Hazards Model
BPHM	Bayesian Proportional Hazards Model
MIS	Management Information Systems
MoHCW	Ministry of Health and Child Welfare
MoHSS	Ministry of Health and Social Services
UNAIDS	United Nations Programme on HIV/AIDS
AFT	Accelerated Failure Time models
ART	Anti-Retroviral Therapy
ARV	Antiretroviral drugs
WHO	World Health Organisation
OR	Odds Ratio
RR	Relative Risk
HR	Hazard Ratio
ANC	Anti-Natal Care clinics
MAR	Missing at Random
MCAR	Missing Completely at Random
MNAR	Missing Not at Random
MDGs	Millenium Development Goals
OLS	Ordinary Least Squares

SAS	Statistical Analysis System
R	R-Statistical Package
IQR	Interquartile Range
CD4	Cluster of Differentiation 4
PEPFAR	President's Emergency Plan for AIDS Relief
LTFU	Loss to Follow-Up
AHR	Adjusted Hazard Ratio
CI	Confidence Interval

Chapter 1

General introduction

Survival modelling and analysis methods are generally used in clinical, biostatistics and epidemiology research to model time until event data. Survival analysis refers to statistical procedures designed to take into account the amount of time an experimental unit contributes to a study period, that is, the time between entry into study and occurrence of event of interest (Nakhaee & Law, 2011).

Survival analysis entails methods that measure the risk of death or progression of a disease and provide predictions that can help clinicians to estimate trends in the patient outcomes. These methods also allow health planners to predict the Human Immunodeficiency Virus/ Acquired Immunodeficiency Syndrome (HIV/AIDS) burden on the health system and to allocate health services resources appropriately (Nakhaee & Law, 2011). Health care planning relies on a good understanding of disease prevalence, which requires an accurate knowledge of survival patterns. Monitoring the length of survival after diagnosis is, therefore, an important component of the surveillance of HIV/AIDS as it provides the basis for evaluating individual prognostic factors (Assefa & Wencheko, 2012). The survival of HIV/AIDS patients

depends on a variety of factors including but not limited to the individual patient's demographic factors, serological baseline factors and presence of co-morbidities.

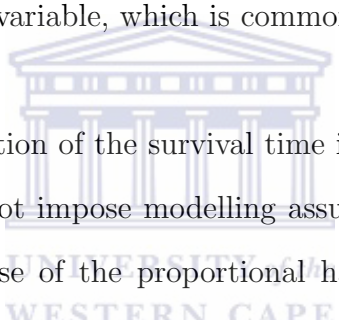
The most widely used survival modelling and analysis techniques include the Cox's proportional hazards models and the accelerated failure time (AFT) models. In the last two to three decades, quantile regression (QR) models as introduced by Koenker and Basset (Koenker & Bassett Jr, 1978) have become an alternative technique to describe and contextualize the distribution of a response variable given a set of explanatory variables. Quantile regression models are very flexible in assessing covariate effects on event times, thereby attracting huge interest in their application to survival modelling.

Survival data is commonly skewed and the marginal distribution of response variable is characterised by marked skewness. Quantile regression methods are robust in characterising and exploring the distribution of skewed data (for example duration or survival data) hence are emerging as popular techniques in survival modelling. According to Fitzenberger and Wilke (2005) quantile regressions are used to model changes of quantiles of the conditional distribution of the survival time in response to changes in the explanatory variables (Fitzenberger & Wilke, 2005).

Whilst in conventional regressions (in this case, the Cox's proportional hazards models), estimators show that the survival times or failure times recognise the presence of some risk factors, quantile regression analysis adds a new dimension to the literature, suggesting that the influence of the risk factor on the survival times varies across the

survival time distribution. For patients with higher duration (survival) times, the survival function will barely recognise some risk factors whilst for patients with the lowest survival times; their survivor function is particularly sensitive to the presence of some risk factors.

Thus, quantile regression techniques can help us get a more complete picture of the underlying relationships between risk factors and the survival time. Censored quantile regression (CQR) like proportional hazards models addresses the issue of right censoring of the response variable, which is common in survival analysis.



By modelling the distribution of the survival time in a flexible semi-parametric way, quantile regression does not impose modelling assumptions that may not be empirically valid like in the case of the proportional hazards assumption for the Cox's proportional hazards regression models. Censored quantile regrssion models are more flexible than the AFT models or the Cox's proportional hazard model because they do not restrict the variation of estimated coefficients over the quantiles (Fitzenberger & Wilke, 2005).

1.1 The conceptual framework and objectives of the study

Long-term sustainable treatment is one choice for people living with HIV/AIDS. Not only can medications slow the progression of the infection, but they can also markedly suppress the virus, thereby restoring the body's immune function and permitting many HIV-infected individuals to lead a normal life. People living with HIV/AIDS

take a number of medications to sustain their survival, that is, to improve their chances of better and longer survival.

Even though HIV/AIDS drugs have become cheaper and more available because of a variety of government and private programmes, millions of others still do not have access to the drugs. The World Health Organisation, WHO, (WHO, 2007) recommends that in resource-limited settings like Zimbabwe, Namibia and other African countries, HIV infected adolescents and adults should start antiretroviral treatment (ART) when the following conditions are met: WHO clinical stage 4 regardless of Cluster of Differentiation 4 (CD4) count, stage 3 disease with consideration of CD4 count below $350/mm^3$ in assisting decision making, stages 1 or 2 diseases with CD4 cell count below $200/mm^3$. In some settings where the CD4 count is not available, the total lymphocyte count (TLC) can be used, and treatment is recommended for WHO stages 3 or 4 (clinical AIDS) irrespective of the TLC, and stage 2 with TLC not exceeding $1200/mm^3$.

With proper and timely use of antiretroviral therapy (ARVs) and good health care support for people living with HIV/AIDS, it has been noted worldwide that the survival time can be improved. The presence of opportunistic infections has however impacted negatively on this prospect. Some of the most lethargic opportunistic illnesses which can drastically reduce HIV/AIDS patients' survival time include Tuberculosis (TB), Pneumonia, and Hepatitis. Of these opportunistic illnesses, TB has been proven to be the most deadly when it combines with HIV/AIDS (Mezzabotta, 2008; Ngwerume, 2008).

A number of medical studies on the epidemiology of HIV/AIDS and co-morbidities within the region have appealed to diverse descriptive and analytic statistical methods. The odds ratios (ORs) and relative risks (RRs) are commonly used to describe and ascertain association between health outcomes and different risk factors. The proportional hazards models (PHM) on the other hand are used to identify and explain the relationship between survival time and the different risk factors such as opportunistic diseases status, HIV status, CD4 count, WHO HIV/AIDS stages, sex, age, and alcohol taking (Mutasa-Apollo et al., 2014; Ngwerume, 2008; Zachariah et al., 2006). All these methods are important as they give a quick picture and appreciation of the dangers posed by the different associations in the variables.

However, the survival times for HIV/AIDS patients vary according to a number of factors including socio-demographic factors, CD4 cell count, HIV viral load and some AIDS defining illnesses. Therefore these covariates need to be tested in a censored quantile regression model in order to allow prediction of these variables' effects on the time to the event of interest, that is death, from ART initiation.

Survival analysis thus represents a more dynamic situation which is not easily captured by commonly used methods like the proportional hazards ratios which only reveal a static situation. For example, these methods can only indicate that a particular variable is a risk factor but they do not highlight at what point of the patient's survival time is the variable a significant risk factor. Thus there is an urgent need to

apply censored quantile regression (CQR) models in conjunction with Cox's proportional hazards models in the analysis of HIV/AIDS patients' survival times in order to gain more insight into the effect of the different prognostic factors at different points of the patients' survival times. In addition, the information derived from the CQR models will also be of significant use in the clinical area with regards to how patients with AIDS defining illnesses and HIV/AIDS can best be managed to improve their survival times.

This study sought to model new survival data from Namibia and Zimbabwe for HIV/AIDS patients who are taking antiretroviral drugs using the Cox's proportional hazards, Bayesian proportional hazards and censored quantile regression models. The study also did a comparative analysis of the inferential power associated with each of the three modelling techniques.

1.2 Aim

The main aim of this study was to model and analyze the survival time distribution for patients on HIV antiretroviral treatment in the presence of serological and clinical risk factors, opportunistic diseases, and some demographic risk factors using different survival modelling techniques.

1.3 Research objectives

The specific objectives for this research study were to:

1. Determine the risk factors for the survival of HIV/AIDS patients on antiretroviral treatment using the Cox's proportional hazards model.

2. Determine the risk factors for the survival of HIV/AIDS patients on antiretroviral treatment using the Bayesian proportional hazards model.

3. Identify risk factors for the survival of HIV/AIDS patients on antiretroviral treatment using the censored quantile regression model. Compare the result to that from the Bayesian proportional hazards model.

4. Compare the effects of the different covariates across different quantiles of the survival time distribution based on censored quantile regression and Cox's proportional hazards models.

5. Characterize the risk factors that affect HIV/AIDS patients' survival in the shorter to medium survival times.

1.4 Significance of the study

This study can give health researchers, biostatisticians and epidemiologists a better picture of the alternative methods that can be used in duration (survival) data analysis. These methods can either be used by themselves or as complementary tools to gain more insight into the risk factor dynamics at any given point of the patients' survival time.

1.5 Thesis layout

The general structure of this thesis is divided into different sections as follows. The first chapter deals with the general introduction to the study and also lays out the different objectives to be met in this research study. The second chapter is devoted to reviewing the literature that underlies the different survival modelling procedures that

are to be employed in this study. The third chapter deals with the data selection and methodologies applied. This chapter gives an explanation of the datasets used in this study and provide details regarding the statistical methods utilized. The fourth chapter presents an analysis and modelling of the paediatric dataset to determine survival prognostic risk factors. The Cox's proportional hazards models are implemented in this chapter and risk factors are determined and discussed. The fifth chapter presents the analysis and modelling of the paediatric dataset using the Bayesian proportional hazards models and the censored quantile regression models. The inferential power of these models are then discussed and compared. Chapters four, five and seven are made up of submitted/published articles in accredited journals. Chapter six fits the Cox's proportional hazards model and the censored quantile regression models and then compares and discusses the strengths and weaknesses of each of the models. The seventh chapter implements the censored quantile regression on the adult ART patients' survival time to determine factors that influence the outcome of death in the early periods (first year) of starting on ART. Chapter eight discusses and concludes on the findings of the study and provides the recommendations.

Chapter 2

Literature review

In this chapter a concise review of the literature that underlies the methodologies that were implemented in this study is covered. Literature is reviewed that surrounds the concept of survival data analysis in relation to the HIV/AIDS pandemic. Brief backgrounds to the history of the disease in each of the countries to be part of this study are also reviewed.

2.1 HIV/AIDS epidemic in Zimbabwe

The first case of HIV/AIDS in Zimbabwe was initially reported in 1985. From that moment, an increasing number of patients were observed to present with different illnesses suggestive of HIV infection. In the years that followed, HIV/AIDS prevalence increased from 2% in 1985 to 8.8% in 1995 (Kerina, Babill, & Muller, 2013) among blood donors. A routine sentinel surveillance of pregnant women attending anti-natal care clinics (ANC) commenced in 1990. Results from the first surveillance indicated that the prevalence exceeded 10% among pregnant women. By the year 2000, prevalence among these women was around 32% but this reduced to about 16% by 2011

(Mbizvo et al., 1996; Gonese et al., 2010; Kerina et al., 2013). Surveillance data from ANCs, for example, has over the years been used to provide estimates of HIV/AIDS prevalence rates for adult populations. The national estimate for HIV/AIDS prevalence amongst adults, that is, 15 years and older, is currently (2014 estimate) estimated at around 14.7%, down from around 21% in 2005 (Mundi, 2015).

Anti-retroviral therapy (ART) has substantially reduced HIV/AIDS related morbidity and mortality worldwide. A growing body of empirical evidence and mathematical modelling suggests that expanded ART use may also prevent population-level transmissions of HIV virus (Novitsky & Essex, 2012). An estimated one million six hundred (1,600,000) people are living with HIV/AIDS in Zimbabwe, of which one hundred and fifty thousand (150,000) are estimated to be children aged from birth to 14 years (UNAIDS, 2014). Access to antiretroviral therapy (ART) in Zimbabwe has been anything but easy. According to the MoHCW (2013), the number of people living with HIV/AIDS disease in 2012 was estimated to be 1.3 million, whilst those in need of ART was estimated to be 621,673 adults and 108,263 children (MoHCW, 2013).

Between 2007 and 2009 Zimbabwe suffered a severe socioeconomic crisis as a result of extreme economic recession and hyperinflation. These conditions coupled with drought significantly affected ART patients in particular as well as the general population at large. The health system was seriously affected with major health facilities experiencing shortages of essential medical supplies and health workers (Meldrum, 2008; Mutasa-Apollo et al., 2014).

The National Opportunistic Infection/Antiretroviral Therapy (OI/ART) programme in Zimbabwe was established in April 2004. From 2008 (Mutasa-Apollo et al., 2014), both ART initiation and follow-up services have been decentralised to most of the lower level health facilities. The rapid expansion of ART programmes in settings with weak health systems is associated with high mortality and loss of patients to follow-up (Nachega, Mills, & Schechter, 2010). A recent study by Mutasa-Apollo and co-researchers on attrition in Zimbabwe for patients on ART indicated that gender, body weight, WHO stage IV and level of health care were significant risk factors of attrition (including mortality) for patients on anti-retroviral treatment (Mutasa-Apollo et al., 2014).

2.2 HIV/AIDS epidemic in Namibia

The government of the republic of Namibia started monitoring HIV/AIDS prevalence since 1992 through the sentinel surveillance of pregnant women at selected antenatal clinics throughout the country. In 1992 prevalence amongst this population group was around 4.2% rising to its peak of 22% in 2002 (MoHSS, 2013).

Over the past two decades the HIV/AIDS pandemic in Namibia has negatively impacted on health and development indicators, and remains a major challenge for the next generations. Regardless of the decline in new HIV infections (incidence rates), HIV prevalence in the country has stabilised at a high level and continues to burden the health care system, fueling new infections and posing serious developmental problems. The epidemic impacts both directly and indirectly on the well being of the

vast majority of the population, the performance of the formal and informal sectors of the economy, the capacity of public and private sectors to provide services, and the attainment of the MDGs (Millennium Development Goals)(MoHSS-UNAIDS, 2011).

Namibia experiences a generalised HIV epidemic with an estimated 260,000 HIV infected adults and children which represents an estimated prevalence of 13.1% (UNAIDS, 2014). However, deaths due to HIV/AIDS reduced from 6310 in 2010 to 5100 in 2014 (UNAIDS, 2014).

Given the impact of HIV/AIDS disease on it's developmental prospects, the Namibian government made significant deliberate efforts on improving the care and treatment of HIV/AIDS. Providing universal access to HIV prevention, treatment, care and support services has been a development priority for the past ten to fifteen years. ART services have been scaled up rapidly in Namibia, with highly successful results. More than 75,000 people were on ART in March 2007 (MoHSS-UNAIDS, 2011), and nearly 50,000 people were on pre-ART in the same year. This roll-out of ART has achieved large reductions in morbidity and mortality among people living with HIV and AIDS.

2.3 Introduction to survival analysis and modelling

In order to understand the concepts that are core in this research, the next section introduce and define the key terms used in this thesis.

2.3.1 Survival analysis

Survival analysis is defined as a branch of statistics that involves the modelling of time to event data. Some of the common areas of application includes biostatistics and epidemiology, where the focus is on observing time to death or some health-related event. Survival analysis explores and models time to event data focusing on the distribution of the survival times. The most profound survival analysis and modelling examines the relationship between survival times and one or more predictors, usually termed *covariates* in survival analysis literature (Fox, 2008).

The standard approaches to survival analysis and modelling are probabilistic (stochastic). The assumption is that the times at which events occur are generated by a random process. This implies that T, the event time, for some particular individual is a random variable having a probability distribution (Allison, 2010).

2.3.2 Censoring

Censoring is the most common phenomenon that is associated with survival data. This occurs when we fail to observe the end point of an event of interest. There are four types of censoring. These are: right censoring, left censoring, interval censoring and informative censoring.

1. Right censoring

   Right censoring occurs when we only observe that a spell has survived until a certain time (duration) (e.g. when the period of observation ends) but we do not know exactly when it ended.

2. Left censoring

   Left censoring occurs when spells observed in the data started before the beginning of the period of observation. Spells which started at the same time and which finished before the beginning of the period of observation are not observed.

3. Interval censoring

   Interval censoring occurs when a spell is known to have occurred within some time interval.

4. Informative censoring

   Informative censoring occurs when a patient is withdrawn from the study for other reasons.

In this study, considerations were made for the case of right censoring only because it is supported in all the model frameworks that were used in this study.

2.3.3 Survival models

Survival modelling is concerned with models for data that have three main characteristics which are:

1. The dependent variable or response is the waiting time until the occurrence of a well defined event.

2. Observations are censored, in the sense that for some units the event of interest has not occurred at the time the data are analysed.

3. There are predictors or explanatory variables whose effect on the waiting time we wish to assess or control.

2.3.4 Quantile regression

Quantile regression is a biostatistics, econometric and biometric tool in which a specified quantile (or percentile) of the conditional distribution of the response variable is regressed on subject characteristics. Thus, while a regression coefficient from a conventional ordinary least squares (OLS) regression model describes how the mean response changes with a one-unit change in the predictor variable, a regression coefficient from the quantile regression model describes how the specified quantile of the response changes with a one-unit change in the predictor variable. This technique is finding its application in survival data analysis due to the skewed nature of the data involved. SAS implements this technique in it's "quantlife" procedure whilst in R it is supported in the "quantreg" package.

2.3.5 Cox's proportional hazards models

Cox's proportional hazards models are the most commonly used modelling frameworks for examining the relationship of the survival time distribution to the different potential risk factors. Most commonly, this examination entails the specification of a linear-like model for the log-hazard (Fox, 2008). This is a semi-parametric survival model with the assumption of proportionality in the hazards. The details of this modelling framework are discussed in the next section.

2.4 Proportional hazards modelling (PHM)

Cox's Proportional Hazards Modelling (PHM) is the basic modelling technique used in exploring the relationship between the survival experience of a patient and potential risk factors in survival data analysis. This model was proposed by Cox (Cox, 1972; Cox & Oakes, 1984) and it has come to be known as the *Cox proportional hazards regression models*. Although the model is based on the assumption of proportional hazards, no particular form of probability distribution is assumed for the survival times. The model is thus semi-parametric in nature.

2.4.1 Model specification for the PHM

Most commonly, this examination of the relationship between survival time and explanatory variables (risk factors) entails the specification of a linear-like model for the log hazard. For example, a parametric model based on the exponential distribution may be written as:

$$log\ h_i(t) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \tag{2.4.1}$$

or equivalently:

$$h_i(t) = exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}). \tag{2.4.2}$$

that is, as a linear model for the log-hazard or as a multiplicative model for the hazard. Here i is a subscript for observation, and the x's are the covariates. The constant α in this model represents a kind of log-baseline hazard, since log $h_i(t) = \alpha$ [*or* $h_i(t) = e^{\alpha}$] when all the x's are zero.

The *Cox proportional hazards model*, in contrast, leaves the baseline hazard function $\alpha(t) = log\ h_0(t)$ unspecified. That is:

$$log\ h_i(t) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \tag{2.4.3}$$

or equivalently:

$$h_i(t) = h_0(t) exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}). \tag{2.4.4}$$

This model is semi-parametric as the baseline hazard can take any form, and the covariates enter the model linearly (Fox, 2008).

This same model can be represented as:

$$h_i(t|x_i) = h_0(t) exp(X_i'\beta). \tag{2.4.5}$$

In this model, $h_0(t)$ is the baseline hazard function that describes the risk for individuals with $x_i = 0$, who serve as the reference cell or pivot, and $exp(X_i'\beta)$ is the relative risk, a proportionate increase or reduction in risk, associated with the set of characteristics X_i'. The increase or reduction in risk is the same at all durations t (Rodrıguez, 2010). This model assumes that the underlying hazard rate (rather than the survival time) is a function of the independent variables (covariates); no assumptions are made about the nature or shape of the hazard function.

Consider a case whereby we are dealing with two groups, that is, say patients are randomised to receive either a standard treatment or a new treatment, and let $h_S(t)$ and $h_N(t)$ be the hazards of death at time t for patients on the standard treatment and new treatment respectively, according to a simple model for the survival times of the two groups of patients, the hazard at time t for a patient on the new treatment is

proportional to the hazard at that same time for a patient on the standard treatment. This proportional hazards model can also be expressed as:

$$h_N(t) = \phi h_S(t) \ ; \ \forall t \geq 0, \tag{2.4.6}$$

where ϕ is a constant. The relative hazard ϕ cannot be negative, and it is thus convenient to set:

$$\phi = exp(\beta). \tag{2.4.7}$$

The β parameters are then the logarithm of the hazard ratio, that is,

$$\beta = log\ \phi, \tag{2.4.8}$$

and any value of β in the range $(-\infty, +\infty)$ will lead to a positive value of ϕ.

The models (2.4.3−2.4.5) can be generalised to the situation where the hazard of death at a particular time depends on the values $x_1,\ x_2, \ldots, x_p$ of p explanatory variables. The values of these variables will be assumed to have been recorded at the time of origin of the study. The set of values of the explanatory variables in the proportional hazards model will be represented by the vector X, so that $X = (X_1,\ X_2, \ldots,\ X_p)'$. Let $h_0(t)$ be the hazard function for an individual for whom the values of all the explanatory variables that make up the vector X are zero. The hazard function for the i^{th} individual can then be written as:

$$h_i(t) = h_0(t)\phi(X_i), \tag{2.4.9}$$

where $\phi(X_i)$ is a function of the values of the vector of explanatory variables for the i^{th} individual.

2.4.2 Fitting the proportional hazards model

Fitting the proportional hazards model given in equation 2.4.5 to an observed set of survival data entails estimating the unknown coefficients of the explanatory variables $(X_1,\ X_2, \ldots,\ X_p)$ in the linear component of the model, $\beta_1,\ \beta_2, \ldots, \beta_p$. The baseline hazard, $h_0(t)$, may also need to be estimated. These two components of the model can be estimated separately. The $\beta^{'}s$ are estimated first and these estimates are then used to construct an estimate of the baseline hazard function. This is an important result, since it means that in order to make inferences about the effects of $p - explanatory$ variables $(X_1,\ X_2, \ldots,\ X_p)$ on the relative hazard, $\frac{h_i(t)}{h_0(t)}$, we do not need an estimate of $h_0(t)$ (Collett, 2015; Allison, 2010).

The β-coefficients in the proportional hazards model, which are the unknown parameters in the model, can be estimated using the *method of partial maximum likelihood*. To utilize this method, we first obtain the *likelihood* of the sample data (Collett, 2015). This is the joint probability of the observed data, regarded as a function of the unknown parameters in the assumed model. For the proportional hazards model, this is a function of the observed survival times and the unknown β-parameters in the linear component of the model. Estimates of the $\beta^{'}s$ are then those values which are the most likely on the basis of the observed data. These maximum likelihood estimates are therefore the values which maximise the likelihood function (Collett, 2015).

2.4.3 The likelihood function for the model

The basis of the argument used in the construction of a likelihood function for the proportional hazards model is that intervals between successive death times convey no information about the effect of explanatory variables on the hazard of death. This is because the baseline hazard function has an arbitrary form, and so it is conceivable that $h_0(t)$, and hence $h_i(t)$, is zero in those times or intervals where there are no deaths. This in turn means that these intervals give no information about the values of the β-parameters. We therefore consider the probability that the i^{th} individual dies at some time t_j, conditional on t_j being one of the observed set of r ordered death times $t_1, t_2, \ldots, t_r$. If the vector of explanatory variables for the individual who dies at t_j is X_j, this is:

$$P[\textit{individual with variables } X_j \textit{ dies at } t_j | \textit{one death at } t_j]. \tag{2.4.10}$$

Next, from the conditional probability result that the probability of an event A given that an event B has occurred:

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)}. \tag{2.4.11}$$

The expression in 2.4.10 becomes:

$$\frac{P[\textit{individual with variables } X_j \textit{ dies at } t_j]}{P[\textit{one death at } t_j]}. \tag{2.4.12}$$

The numerator of this expression is simply the hazard of death at time t_j for the individual whose vector of explanatory variables is X_j. If it is the i^{th} individual who dies at time t_j, the hazard function can be written as $h_i(t_j)$. The denominator is the sum of the hazards of death at time t_j over all individuals who are at risk of death at this time. This is the sum of the values $h_l(t_j)$ over those individuals indexed by l in

the risk set at time t_j, $R(t_j)$. Consequently, the conditional probability in expression 2.4.11 becomes:

$$\frac{h_i(t_j)}{\sum_{l \in R(t_j)} h_l(t_j)}, \tag{2.4.13}$$

and on using 2.4.12 and 2.4.13, the baseline hazard function in the numerator and denominator cancels out, and we are left with:

$$\frac{exp(\beta' X_j)}{\sum_{l \in R(t_j)} exp(\beta' X_l)}. \tag{2.4.14}$$

Finally, taking the product of these conditional probabilities over the r death times gives the likelihood function:

$$L(\beta) = \prod_{j=1}^{r} \frac{exp(\beta' X_j)}{\sum_{l \in R(t_j)} exp(\beta' X_l)}, \tag{2.4.15}$$

in which X_j is the vector of covariates for the individual who dies at the j^{th} ordered death time, t_j. The summation in the denominator of this likelihood function is the sum of the values of $exp(\beta' X)$ over all the individuals who are at risk at time t_j.

For censored data which is a common phenomenon with survival data, suppose that the data consist of n observed survival times, denoted by $t_1, t_2, \ldots, t_n$ and that δ_i denotes a censoring indicator which is zero if the i^{th} survival time t_i, where $i = 1, 2, \ldots, n$ is right censored, and one if the i^{th} survival time t_i is observed for patient i. The likelihood function in 2.4.15 can then be expressed as:

$$L(\beta) = \prod_{j=1}^{r} \left[\frac{exp(\beta' X_j)}{\sum_{l \in R(t_i)} exp(\beta' X_l)} \right]^{\delta_i}, \tag{2.4.16}$$

where $R(t_i)$ is the risk set at time t_i and δ_i is the indicator variable that takes a value 1 if the outcome of interest is observed at t_i. The corresponding log-likelihood

function is given by:

$$log\ L(\beta) = \sum_{i=1}^{n} \delta_i \left\{ \beta' X_i - log \sum_{l \in R(t_i)} exp(\beta' X_l) \right\}. \tag{2.4.17}$$

The maximum likelihood estimates of the β-parameters in the proportional hazards model can then be found by maximising this log-likelihood function using numerical methods such as the *Newton Raphson Procedure* (Collett, 1994).

2.5 Extensions of the proportional hazards model

In the past few years, the Cox proportional hazards model has been extended to the Bayesian framework and also the generalised additive models (GAMs) framework. Brief reviews of these frameworks are covered in the next sections.

2.5.1 Bayesian proportional hazards model

This model is premised on the proportional hazards modelling framework and adds the Bayesian flavour in estimating the model coefficients. In this approach, the partial likelihood function in equation 2.4.16 is used as the likelihood function in the posterior distribution (Sinha, Chen, & Ibrahim, 2003).

2.6 Quantile regressions and ordinary least squares

Quantile regressions as pioneered by Koenker and Basset (Koenker & Bassett Jr, 1978) seeks to complement classical linear regression methods. The main aim of regression analysis is to analyse the behaviour of a dependent variable Y_i given some

explanatory variables $\mathbf{X_i}$. The ordinary least squares (OLS) methods are used to obtain the parameter estimates through the minimisation of the error sums of squares. The result is an approximation of the mean function of the conditional distribution of the dependent variable. The parameters estimated through the OLS method achieves the property of BLUE (best, linear, and unbiased estimators), if the following assumptions hold, that is: the explanatory variables $\mathbf{X}$ are non-stochastic; the expectation of the error term ϵ_i is zero ($E(\epsilon_i) = 0$); homoscedasticity, that is, the variance of the error terms ϵ_i is a constant ($Var(\epsilon_i) = \sigma^2$) and no auto-correlation of error terms ($Cov(\epsilon_i, \epsilon_j) = 0 \; \forall \; i \neq j$).

However, in most cases, one or more of these assumptions are violated, leading to OLS parameter estimates not being BLUE. Generally quantile regressions perform better than the classical linear regression methods where the error terms are not constant, that is, violating the homoscedasticity assumption; the focus is on the tails of the distribution rather than the mean as a measure of location and where data has extreme outliers (skewed data) because OLSs are too sensitive to extreme outliers leading to significant distortion of results.

2.7 Quantile regression modelling

Quantile regression as introduced by Koenker and Basset (Koenker & Bassett Jr, 1978) may be viewed as a natural extension of the classical least squares estimation of conditional mean models to the estimation of an ensemble of models for conditional quantile functions. To know the relationship between y and x, the classical least-squares regression requires the conditional mean function, that is, the function

that explains how the mean of y changes with the vector of explanatory variables x. However, according to Frisch (1934) and Koopmans (1937), it is the systematic component around which y fluctuates due to an erratic component that is needed to explain this relationship (Frisch, 1934; Koopmans, 1937).

Crucially, the error component is assumed to have precisely the same distribution regardless of the values taken by the vector x. This is referred to as a pure location shift model since it assumes that x affects only the location of the conditional distribution of y, not its scale, or any other aspect of its distributional shape. Given this to be the case, an estimated model of the conditional mean function, supplemented perhaps by an estimate of the conditional dispersion of y around its mean is sufficient. Furthermore, when the requirement that the errors are Gaussian is added, least squares methods deliver the maximum likelihood estimates of the conditional mean function (Koenker & Hallock, 2001b).

However, the argument is that there is more to biometric, biostatistics and epidemiological modelling than what is stipulated in the usual location shift modelling framework. Covariates may influence the conditional distribution of the response in a myriad other ways such as expanding its dispersion as in traditional models of heteroscedasticity, stretching one tail of the distribution, compressing the other tail and even inducing multimodality. An explicit investigation of these effects via quantile regression can provide a more nuanced view of the stochastic relationship between variables, and therefore a more informative empirical analysis.

2.7.1 Fundamentals and features of the quantile regression

Gilchrist (2001) describes a quantile as "simply the value that corresponds to a specified proportion of an (ordered) sample of a population" (Gilchrist, 2001). A commonly used example of a quantile is the median, M, which is the proportion of 0.5 of the ordered data. This relates to a quantile with a 0.5 chance of occurrence. Quantiles hereby mark the boundaries of equally sized, consecutive subsets. Quantile regression therefore generalises the concept of a univariate quantile to a conditional quantile given one or more covariates (Chen, 2013).

For a continuous random variable Y with probability distribution given in the following equation:

$$F(y) = P(Y \leq y), \tag{2.7.1}$$

which states that for the distribution function $F_Y(y)$ one can determine, for a given value y the probability $F(y) = \tau$ of occurrence. Now if one is dealing with quantiles, one wants to do the opposite, that is, one wants to determine for a given probability τ of the sample data set the corresponding value y. A τ^{th}-quantile refers in a sample data to the probability τ for a value y, that is:

$$F_Y(y_\tau) = \tau. \tag{2.7.2}$$

Another form of expressing the τ^{th}-quantile mathematically is as follows:

$$y_\tau = F_Y^{-1}(\tau) \tag{2.7.3}$$

where y_τ is such that it constitutes the inverse of the function $F_Y(y)$ for a probability τ. Note that there are two possible scenarios. The first scenario is, if the distribution

function $F_Y(y)$ is monotonically increasing, quantiles are well defined for every $\tau \in (0;1)$. The second scenario states that if a distribution function $F_Y(y)$ is not strictly monotonically increasing, there are some $\tau's$ for which a unique quantile cannot be defined. In this last case, one uses the smallest value that y can take on for a given probability τ. Then for any $\tau \in (0;1)$, the τ^{th} quantile of Y for both scenarios is defined as follows:

$$Q(\tau) = inf\{y : F(y) \geq \tau\} = y_\tau = F_Y^{-1}(\tau). \tag{2.7.4}$$

That is y_τ is equal to the inverse of the function $F_Y(\tau)$ which in turn is equal to the infinimum of y such that the distribution function $F_Y(y)$ is greater or equal to a given probability τ, that is, the τ^{th}-quantile. In particular, the median is $Q(\frac{1}{2})$, the first decile $Q(\frac{1}{10})$ and the first quarter $Q(\frac{1}{4})$. The quantile function provides a complete characterisation of Y, just like the distribution function F.

2.7.2 Quantile regression model

The linear conditional quantile regression model can be specified as follows:

$$Q(\tau|X = x) = \beta_{0\tau} + \beta_{1\tau}X_{1i} + \cdots + \beta_{p\tau}X_{pi} + \epsilon_i. \tag{2.7.5}$$

where

$\beta_{0\tau}$ is the intercept for the quantile model at the τ^{th} level.

$\beta_{1\tau}, \ldots, \beta_{p\tau}$ are the coefficients for the P covariates at each probability τ.

$X_{1i}, \ldots, X_{pi}$ are the P independent covariates and

$\epsilon_i s$ are the error terms.

This model can also be expressed as a generalised linear model in the form:

$$\boldsymbol{Q(\tau|X = x) = X'\beta(\tau) + \epsilon_i.} \tag{2.7.6}$$

where:

X' is the design matrix and

$\beta(\tau)$ is the vector of parameters.

This model is analogous to the OLS regression model, hence the estimation of the parameters is via the minimisation procedure. That is for a random sample $\{y_1, y_2, \ldots, y_n\}$ of Y it is well known that the sample median is the minimiser of the sum of the absolute deviations:

$$min_{\epsilon \in R} \sum_{i=1}^{n} |y_i - \epsilon|. \tag{2.7.7}$$

2.7.3 Censored quantile regressions

Linear censored quantile regressions (CQR), introduced by Powell in 1984 allows for semi-parametric estimation of quantile regressions for a censored regression model in a robust way (Chernozhukov & Hong, 2011). Censored quantile regression (CQR) is robust against misspecification of the error terms since only fairly weak assumptions on the error terms are required.

Let the possible observed duration be right censored, that is, the observed completed duration T_i be given by $T_i = min\{T_i^*, yc_i\}$, where T_i^* is the true duration of the spell and yc_i is the spell specific threshold value or censoring point, beyond which the spell cannot be observed. For the proportional hazards model (PHM), this can be incorporated in the maximum likelihood estimation (Woodridge, 2002).

In contrast, CQR requires the knowledge of yc_i irrespective of whether the observation is right censored or not. CQR provides consistent estimates of the quantile

regression coefficients β_τ in the presence of fairly general forms fixed censoring. The known censoring points can either be deterministic or stochastic and they should not group in a certain way or around the true quantile regression line. Estimating the linear CQR model involves minimising the following distance function:

$$\sum_{i=1} \rho_\tau(ln(T_i) - min(X_i^{'}\beta_\tau, yc_i)), \tag{2.7.8}$$

with respect to β_τ, where $\rho_\tau(z) = \tau.|z| \; for \; z \geq 0$ and $\rho_\tau(z) = (1-\tau).|z| \; for \; z < 0$ and yc_i denotes the known observation specific censoring points. A quantile regression without censoring is nested as the special case with $yc_i = +\infty$.

The censored quantile regression estimator $\hat{\beta}_\tau$ is $\sqrt{N}$-consistent and asymptotically normally distributed (Powell, 1984). A crucial feature of this result is that the asymptotic distribution depends only upon those observations where the fitted quantiles are not censored.

The actual calculation of the CQR estimator based on individual data is numerically very difficult since the distance function 2.7.7 which is to be minimised is not convex. For heteroscedasticity-consistent inference, bootstrapping is often resorted to through the use of the design-matrix-bootstrapping. The covariance matrix of the CQR estimates across quantiles can also be estimated.

2.7.4 Quantiles via optimisation

Quantiles seem inseparably linked to the operations of ordering and sorting the sample observations that are usually used to define them. So it comes as a mild surprise to observe that quantiles can be defined through a simple alternative expedient as an

optimisation problem or a linear programming problem. Just as we can define the sample mean regression function as the solution to the minimization of the sum of squared residuals, we can define the median regression function as the solution to the problem of minimising the sum of the absolute residuals (Koenker & Hallock, 2001b). The symmetry of the piecewise linear absolute value function implies that the minimisation of the sum of absolute residuals must equate to the number of positive and negative residuals, thus assuring that there are the same number of observations above and below the median. Now for the other quantiles, since the symmetry of the absolute value yields the median, perhaps minimising the sum of asymmetrically weighted absolute residuals, simply giving differing weights to positive and negative residuals would give yield to the quantiles. The case here is of solving the following linear programming problem or minimisation problem:

$$min_{\beta \in R} \sum_{i=1}^{n} \rho_\tau (y_i - \epsilon(x_i, \beta)). \tag{2.7.9}$$

Where ρ_τ serves as a check function and is defined by:

$$\rho_\tau(x) = \left\{ \begin{array}{cc} \tau * x & if\ x \geq 0 \\ (\tau - 1) * x & if\ x < 0 \end{array} \right\}. \tag{2.7.10}$$

This check function ensures that:

1. All ρ_τ are positive, and

2. The scale is according to the probability τ.

The resulting minimisation of equation 2.7.9, when $\epsilon(x, \beta)$ is formulated as a linear function of parameters can then be solved very efficiently by implementing linear programming methods.

2.7.5 Features that characterise quantile regressions

The following features characterise quantile regression methods and differentiate them from other regression methods:

1. The entire conditional distribution of the dependent variable Y can be characterised through different values of τ.

2. Heteroscedasticity can be detected.

3. If the data is heteroscedastic, median regression estimators can be more efficient than the mean regression estimators.

4. The minimisation problem can be solved efficiently by linear programming methods, making estimation easy.

5. Quantile regression functions are robust with regards to outliers.

The quantile regression methods however have their own disadvantages and these include not being able to deal with time-variant covariates, having no support for competing risk framework yet and also not being able to account for unobserved heterogeneity (Fitzenberger & Wilke, 2005).

2.7.6 Quantile regression model estimation

Consider data in the form x_i^T, y_i, for $i = 1,\ 2, \ldots, n$ where y_i are the independent scalar observations of a continuous random variable with common cumulative distribution function F_y, whose shape is not exactly known, x_i^T are row p-vectors of a known design matrix X.

The linear conditional quantile functions are defined as:

$$Q_{y_i}(\tau|x_i) = x_i^T\beta, \tag{2.7.11}$$

where $i = 1,\ 2, \ldots, n$, $0 < \tau < 1$, $Q_{y_i}(.) = F_{y_i}^{-1}(.)$, and $\beta \in R^p$ is a column vector of length p with unknown fixed parameters. We refer to the τ^{th} regression quantile any solution $\beta^*,\ \beta^* \in R^p$, to the minimisation problem:

$$min_{\beta \in R^p} V_n(\beta) \equiv \sum_{i \in (i:y_i \geq x_i^T\beta)} \tau|y_i - x_i^T\beta| + \sum_{i \in (i:y_i < x_i^T\beta)} (1-\tau)|y_i - x_i^T\beta|. \tag{2.7.12}$$

In order to highlight the τ-distributional dependency, the parameter β and the solution β^* should be indexed by τ, that is, $\beta(\tau)$ and $\beta^*(\tau)$. The objective function 2.7.12 presents a minimisation problem otherwise known as a linear programming problem, thus the simplex method for linear programming problems is then used to determine the optimal β solutions for the objective function (Koenker & d'Orey, 1985).

2.8 Missingness and imputations

Missing covariate values is a common problem in survival data research (Satty, 2014). Standard statistical methods have been developed to analyse rectangular data sets. The rows of the data matrix represent units, also called cases, observations, or subjects depending on context, and columns represent variables measured for each unit (Little & Rubin, 1987). The implications of missingness include:

1. Loss of efficiency: we collect and analyze less data than originally planned.

2. Unbalanced datasets: not all subjects have the same number of measurements from the study.

3. Potential bias: missingness may depend on outcome hence affecting the estimates for the different model parameters.

2.8.1 Missingness mechanism

A major challenge for the analysis of survival and longitudinal data outcomes is the fact that these outcomes are often incomplete. Missing data in these studies occur in basically two ways. The first type is when patients are missing at intermittent times, meaning that other measurements are observed following missing values. The second type of missing data occurs when data is not available for a subject after some time point, and the patient is said to have dropped-out of the study. The main concern in longitudinal as well as in survival analysis with missing data arises when there is an association between the longitudinal profile and the missing process (Andrinopoulou, 2014). The appropriateness of different methods of analysis of incomplete data is determined by the missing data mechanism. Specifically, there are three types of mechanisms namely:

1. Missing completely at random (MCAR): when the probability that the responses are missing, is unrelated to the longitudinal or survival outcome. For example if a patient dies from any other cause except the cause of interest in the study (Andrinopoulou, 2014).

2. Missing at random (MAR): when the probability of missingness depends on the set of observed longitudinal responses, but is unrelated to the outcomes that

should have been obtained (Andrinopoulou, 2014).

3. Missing not at random (MNAR): when the probability that the longitudinal or survival responses are missing depends on observed and unobserved data.

According to Little and Rubin (Little & Rubin, 1987), there are three main procedures for treating missing data in data analysis. These procedures are briefly discussed below:

1. Procedures based on completely recorded units (complete case analysis),
2. Imputation based procedures, and
3. Weighting procedures.

2.8.2 Procedures based on completely recorded units

When some variables are not recorded for some of the units, a simple expedient procedure is to discard incompletely recorded units and to analyze only the units with complete data. This procedure is generally easy to carry out and may be satisfactory with small amounts of missing data. However, this method can lead to serious biases and usually it is not very efficient (Little & Rubin, 1987).

2.8.3 Imputation based procedures

The missing values are filled in and the resultant completed data are analyzed by standard methods. Commonly used procedures for imputation include *hot deck* imputation, where recorded units in the sample are substituted using a response from a patient with a similar profile; *mean* imputation, where means from sets of recorded

values are substituted; and *regression* imputation, where the missing variables for a unit are estimated by predicting values from the regression on the known variables for that unit (Little & Rubin, 1987).

Hot deck imputation

This can broadly be defined as a method where an imputed value is selected from an estimated distribution for each missing value; in contrast with mean imputation, where the mean of the distribution is substituted (Little & Rubin, 1987).

Mean imputation

This is the imputation procedure where means from the responding units in the sample are substituted for missing values. The means may be formed within cells analogous to the weighting cells formed for weighting procedures (discussed below). Mean estimates leads to estimates similar to those found by weighting, provided the sampling weights are constant within classes (Little & Rubin, 1987).

Regression imputation

This imputation procedure replaces missing values by predicted values from a regression of the missing item on items observed for the unit, usually calculated from units with both observed and missing variables present. Mean imputation can be regarded as a special case of regression imputation where the predictor variables are dummy indicator variables for the cells within which the means are imputed (Little & Rubin, 1987).

Multiple imputation

Multiple imputation methods (Rubin, 1978, 1987) impute more than one value for the missing items. An important limitation of single imputation methods is that standard variance formulas applied to the filled in data systematically underestimate the variance of estimates, even if the model used to generate the imputations is correct (Little & Rubin, 1987). Multiple imputation procedures or methods allow valid estimates of the variance of estimates to be calculated using standard complete data procedures.

2.8.4 Weighing procedures

Randomisation inferences from sample survey data without nonresponse are commonly based on design weights, which are inversely proportional to the probability of selection. Weighting procedures modify the weights in an attempt to adjust for non-response.

2.9 Summary

This chapter reviewed relevant literature on HIV/AIDS in Zimbabwe and Namibia. Survival analysis and modelling techniques were also reviewed based on the focus of this study. A brief overview on dealing with missing data was also done. The next chapter, that is, Chapter 3, discusses the methodology used in this study.

Chapter 3

Methodology

In this chapter the target populations, samples, variables, data collection and tools used in the collection process are discussed for both countries, that is, Zimbabwe and Namibia. The procedures that were utilized in the modelling of the datasets are highlighted in this chapter. The flowchart in Figure 3.1 indicates the different datasets that were used for the purposes of this study. The analysis methods and chapters in which these analyses were done are also highlighted. The details of each of the datasets are discussed in the sections below.

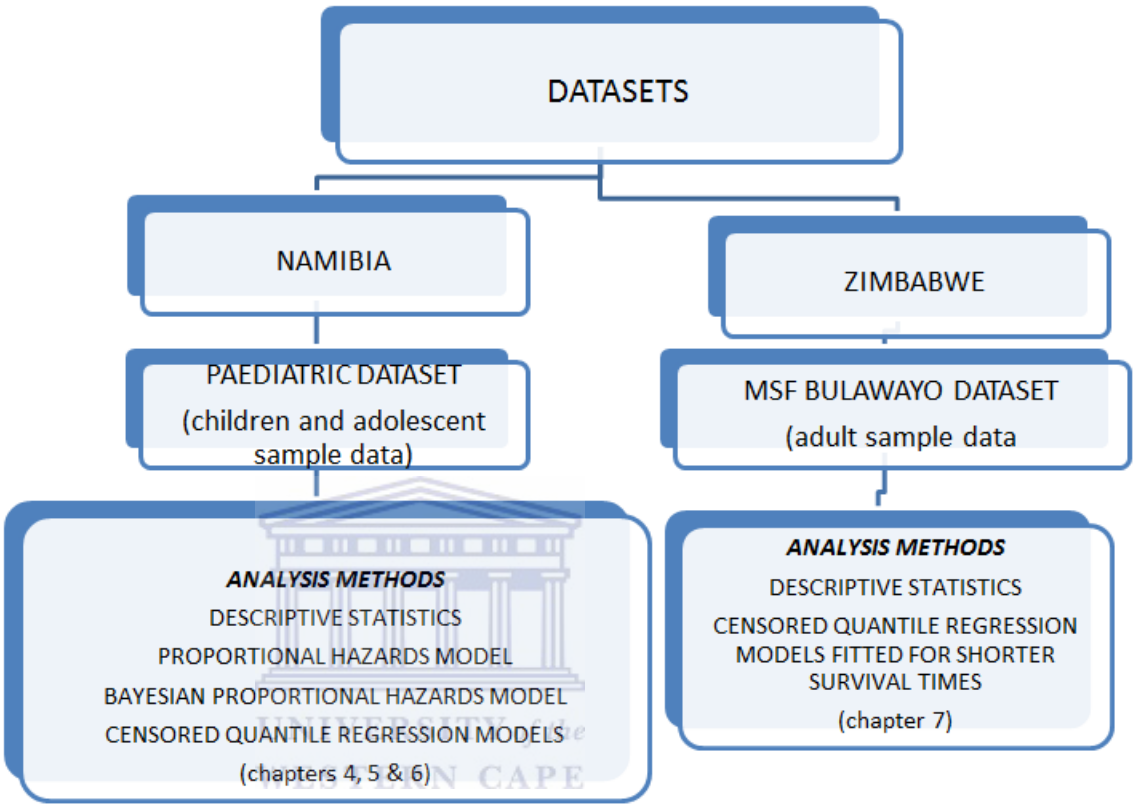


Figure 3.1: Datasets flowchart

3.1 Namibia HIV/AIDS data

A single paediatric dataset from Namibia was used in this research. This dataset was obtained from Katutura hospital, Windhoek.

3.1.1 Katutura hospital paediatric HIV/AIDS dataset

A retrospective cohort study design was conducted for children who initiated ART between 01 January 2006 and 31 December 2010. All children living with HIV/AIDS

below the age of 15 and who started taking ARVs from Katutura Hospital within the period of 2006 - 2010 were included in the sample. The target population was all the children under 15 years who are on Katutura Hospital's records as ART paediatric patients regardless of the period of their ART initiation. The sampling frame constituted about 1605 paediatric patients and from these a final sample of 813 children was selected from those who initiated ART between 01 January 2006 and 31 December 2010. All children who did not have complete information such as missing date of initiation or date of last visit to hospital were dropped from the study. This data was obtained from the hospital's Management Information System database (MIS) with the help of the hospital data clerks. The study subjects were followed from initiation to December 2014. The setting for this study was Katutura State Hospital located in Windhoek, the capital city of Namibia. The hospital has an HIV/AIDS section which deals with all aspects of HIV case management and care, that is, from enrolment, counselling, initiating, follow-up and giving out medication to patients during their scheduled visits to the hospital.

For the purposes of this study and also due to the limitations associated with secondary data collected from the Ministry of Health and Social Services' management information systems database, the study covariates were limited to paediatric patients' gender, baseline World Health Organisation's (WHO) clinical stages of HIV, baseline functional status, baseline age category and absolute CD4 count. Functional status is an individual's ability to perform normal daily activities required to meet basic needs, fulfill usual roles, and maintain health and well-being (WHO, 2004). According to WHO (2004) functional status is defined as:

(a) **Working** if person is able to perform usual work in or out of the house, harvest, go to school or, for children, normal activities or playing.

(b) **Ambulatory** if person or child is able to perform activities of daily living but not able to work or play.

(c) **Bedridden** if person or child is not able to perform activities of daily living.

Absolute CD4 count was classified according to the age of the child and it has two categories, CD4 count below threshold and CD4 count above threshold.CD4 count below threshold was defined for infants as CD4 count less than $1500/mm^3$, for age $12-35$ months as CD4 count less than $750/mm^3$, for age 36-59 months as CD4 count less than $350/mm^3$, and for age >= 60 months as CD4 count less than $200/mm^3$ (Bong, et al., 2007).

Ethical clearance was obtained from the Ministry of Health and Social Services and approval for data collection was granted by the Provincial Medical Doctor (PMD) stationed at Katutura Hospital (See Appendix A.2). The database only captured patient case numbers and not names or any other information that may directly be linked to the patient. The data was also not shared with any third party hence confidentiality was maintained.

Variables of interest

The variables of interest in this dataset are shown in Table 3.1 below and they are all measured at baseline:

Table 3.1: Paediatric dataset variables

Variable	Variable levels
WHO staging	
	Stage 1
	Stage 2
	Stage 3
	Stage 4
Paediatric Gender	
	Male
	Female
Functional status	
	Ambulatory
	Bedridden
	Working
Age category	
	Less than 1 year
	1− < 5 years
	5− < 15 years
Absolute CD4 count	
	CD4 below threshold
	CD4 above threshold

3.2 Zimbabwe HIV/AIDS data

The other dataset that was used for purposes of this study was from Zimbabwe. A retrospective research design was utilized in this study hence secondary data was collected. The dataset was collected from Medicin Sans Frontiers(MSF) Bulawayo hospital. The details and variables to be used in the study are discussed in the section below.

3.2.1 MSF Bulawayo dataset

MSF is a non-governmental organisation (NGO) that deals with humanitarian support in developing countries. MSF has been working in Zimbabwe since the year 2000 and it is currently running comprehensive HIV/AIDS programmes in health centers in Beitbridge, Buhera, Bulawayo, Chikomba, Gutu, Gweru and Tsholostho. This research study used data from the Bulawayo hospital MSF health center. The organisation is working with the Ministry of Health and Child Welfare (MoHCW) on efforts to reduce HIV/AIDS infection and mortality through the different HIV/AIDS programmes in Zimbabwe. One such programme is the anti-retroviral (ARV) treatment programme. As of 2010, the MSF programme has provided treatment to more than 34 000 Zimbabweans. The organisation enrols HIV/AIDS patients onto it's ARV programme and the patients are given a schedule of the intervals at which they can come and replenish their ARV medication as well as having check-ups and tests. At every visit the patient information in the ART register is updated on the variables of interest that are measured over time such as weight, CD4 count, adherence and haemoglobin (HB) levels. The baseline patient information is also captured at treatment initiation. Date of current (last) visit is also noted and this was used to

determine the patient's duration on ARV treatment. The data were accessed from an MS Access database where the electronic information from the ART registers was captured and stored. The patients in the database were identified via a unique identification number hence no names or addresses were used.

The target population was adults aged 15 years and above from Bulawayo hospital MSF center who enrolled for ARV treatment in the period 2004 to 2010. For inclusion into the retrospective cohort study sample, a patient had to be alive on 01 January 2004 and receiving ARVs from MSF Bulawayo hospital or started ART treatment at the same hospital in 2004 but before 01 January 2005. A total final sample of 1134 patients was followed over six years. Data were manipulated and new variables were created, that is, the survival time variable was obtained from the difference between the last known hospital visit date and the date the patient was initiated on ART. A censor variable was also created which catered for identifying those patients whose event of interest (death) was observed or failure to follow-up as well as those who were alive at the end of the study period (2010). Demographic, clinical and serological variables all recorded at initiation of treatment were then used in the study to identify prognostic risk factors for survival of ART patients using a censored quantile regression model.

Ethical clearance was granted by the ministry of health and child welfare as well as the MSF operations research unit (See Appendix A.1).

Variables of interest

The variables of interest in this dataset are shown in Table 3.2 below and they are all measured at baseline:

3.3 Data imputation

Missing covariate values are a common problem in survival data and analysis research (Satty, 2014). Since this study analysed survival data, where some covariates had missing values, we applied the multiple imputation procedures to make the data complete for analysis and hence we reduced biases associated with incomplete datasets. For all the datasets, multiple imputation procedures were done, where necessary, on the missing data and then the full dataset after imputation was used in the inferences.

3.4 Proportional hazards regression model parameter estimation

Cox's proportional hazards model was estimated for the paediatric dataset. The Bayesian proportional hazards model was also fitted for the same dataset. The inferential aspects for the Bayesian proportional hazards model were then compared to the CQR model results with a view to determine prognostic risk factors at different survival time points for patients on ART treatment. The SAS statistical package was used for this estimation.

Table 3.2: MSF Bulawayo dataset variables

Variable	Variable levels
WHO staging	
	Unspecified
	Stage 1
	Stage 2
	Stage 3
	Stage 4
Gender	
	Male
	Female
Profession	
	Unemployed
	Formally employed
	Self employed
	Other
Marital status	
	Single
	Married
	Widowed
	Divorced
Age category	
	Lower than 35 years ($15- < 35\ years$)
	Between 35 and 50 years
	Above 50 years
Weight category	
	Lower than 45 kgs
	Between 45 and 60kgs
	Above 60kgs

The global test for the proportionality assumption and the the Schoenfeld residual plots were used to assess whether there was any violation of the proportional hazards assumption.

3.5 Censored quantile regression model parameter estimation

Censored quantile regression (CQR) models were estimated for different quantile values, or τ values. These τ values covered the spectrum on which the quartiles are described. The standard errors that were used in the 95% confidence interval estimates of these parameters were obtained through using the bootstrap resampling methods. The p-values were used to determine the significant prognostic risk factors that influences survival time at different time points. The graphical representations were used to highlight the effect of the risk factors across all the quantile values of the survival time distribution. The CQR model was fitted to the paediatric dataset to identify the effect of each risk factor at different quantile points of the patients' survival times. The results from this model fitting were compared to those from the Bayesian proportional hazards model. The SAS statistical package was used for this estimation.

The paediatric dataset was also used for the comparison between the Cox's proportional hazards and the CQR models. This was done to highlight that the effects of the risk factors as estimated by the CQR method vary across the survival time distribution whilst they seem constant from the Cox's proportional hazards model.

The CQR model was estimated for the MSF Bulawayo hospital adult dataset to determine risk factors for shorter survival times. The results are discussed and conclusions given.

3.6 Summary

This chapter discussed the two datasets from Namibia and Zimbabwe that were utilized in this study as well as the research design. The study populations were also highlighted. The modelling techniques implemented were highlighted together with the purposes they intended to achieve. The next chapter, that is, Chapter 4, implements the Cox's proportional hazards model on the paediatric dataset in order to determine survival time risk factors.

Chapter 4

Survival analysis of HIV positive adolescents and paediatrics on antiretroviral treatment at Katutura hospital, Windhoek, Namibia

This chapter has been submitted for publication as:

Maposa, I. and Blignaut, R.J. (2016). Survival Analysis of HIV Positive adolescents and Paediatrics on Antiretroviral Treatment at Katutura Hospital, Windhoek, Namibia. *Journal of Social Aspects of HIV/AIDS*.

Submission ID:RSAH-2016-0002

Status: Under review.

4.1 Summary

The scourge of HIV/AIDS has been around for the past three decades but still there is no cure in sight, hence care and therapy are the best way of dealing with the disease. Children have not been spared from the wrath of this epidemic and millions in Southern African Development Countries (SADC) and the African continent are in dire need of treatment. Namibia also has not been spared from this either.

The purpose of this study is to analyze and determine the risk factors that negatively affect survival of children younger than 15 years who are on antiretroviral treatment (ART) in Windhoek. The study also seeks to identify the survival prospects of these children on ART, given the different risk factors. A retrospective cohort study design was conducted for children who initiated ART between 01 January 2006 and 31 December 2010 at Katutura Hospital in Windhoek, Namibia. All HIV/AIDS positive children below 15 years who started ART at Katutura Hospital between January 2006 and December 2010 were included in the study.

The Kaplan-Meier survival functions were used to assess survival chances of these adolescent and paediatric patients given different risk factors. The log-rank tests were used to identify any significant differences in survival for different factor groups. The proportional hazards model was used to determine the significant risk factors for survival of patients on antiretroviral treatment. The log-rank tests indicate the likelihood of significant differences in the chances of survival for children who were initiated on ART treatment in different WHO stages and also in different functional statuses. Age category was found to be a significant risk factor and being an infant

was associated with a significant negative effect on survival time.

Improvements and targeted interventions are recommended on paediatric ART patients follow-up programmes especially on infants. Special attention should be given to children who initiate in stage IV as well as in the bedridden functional status so that their survival prospects may be improved.

4.2 Introduction and background

According to the World Health Organisation (WHO) report on treatment of children living with HIV, an estimated 3.2 million children were living with HIV at the end of 2013, and most of these were in Sub-Saharan Africa. The report also indicates that the majority of these children acquire HIV from their HIV-infected mothers during pregnancy, birth or breastfeeding. The WHO believes that with efficacious interventions, the risk of mother-to-child HIV transmission can be reduced to as low as 2%. Unfortunately, according to the report, such interventions are still not widely accessible or available in most resource-limited countries like those in the SADC region where the burden of HIV infection is highest. Thus these children who are already infected with HIV would need to be enrolled in effective treatment for them to stay healthy (WHO, 2014).

The WHO report on treatment of children living with HIV also highlights that the number of children (younger than 15 years) receiving ART in low- and middle-income countries more than doubled from 2009 to 2013, from 355 000 to 740 000. At the end of 2013, less than one quarter (23%, range $21 - 25\%$) of children living with HIV

were receiving ART in low and middle income countries compared with more than one third (37%, range $35 - 39\%$) of adults living with HIV (WHO, 2014).

In the absence of antiretroviral therapy (ART), over 50% of HIV infected infants progress to AIDS and death by 2 years (Newell et al., 2004). The introduction of ART has dramatically improved health outcomes of HIV-infected children (Gibb & Giaquinto, 2000). However, according to the UNAIDS 2013 report, at the end of 2012, only 35% of eligible children in low and middle-income countries had initiated ART (UNAIDS, 2013).

Paediatric HIV infection is still a leading cause of death among children in developing countries, with mortality rates highest during infancy (Thorne et al., 2007). In Namibia, HIV/AIDS is one of the leading causes of death and is estimated to cause 23% of the mortality (PEPFAR, 2012). However, among children under 15 years, HIV/AIDS is estimated to be responsible for 8% of the deaths ranking it among the top three causes of deaths in Namibia (MoHSS, 2008).

Paediatric HIV care started in 2002 and has gained traction over the years. Around 14 000 children (aged 0-14) are estimated to be living with HIV (PEPFAR, 2012; WHO, 2014) and the antiretroviral therapy coverage for these children in Namibia was estimated at 87% in 2010 (UNICEF, 2011). According to Leroy and co-researchers, characteristics at ART initiation and short-term outcomes (mortality, retention in programme) are evolving in paediatric ART programmes (Leroy et al., 2013). Many reports show decreased mortality in children after ART initiation but despite this,

some children are started on ART with an advanced disease. In addition, programmes that have rapidly scaled-up ART have, as in adults, high loss to follow-up (LTFU) rates. In adults, high rates of LTFU correspond to underreported mortality, suggesting that in children a significant proportion of child mortality may be unseen as well. In Namibia, there is insignificant documentation on the factors that are associated with mortality and loss to follow-up in children initiated on ART treatment and this needs investigation.

The effects of socio-demographic, clinical and immunological factors on survival of HIV positive children after initiation of ART were studied in other low to medium income countries. However, according to Kedir and co-researchers, the findings of the studies are not consistent. For instance, in studies conducted on cohorts of HIV positive children in Cote dIvoure, Malawi and Zambia, only low weight-for-height and a low CD4% were found to have a significant negative effect on survival of children after ART initiation (Kedir, Desta, & Fesseha, 2014). Also in another large cohort of HIV positive children of Zambia, besides a low CD4% and a low weight-for-height z-score (WHZ), younger age and low haemoglobin level were also found to have a significant negative effect on survival after initiation of ART (Kedir et al., 2014).

The scarcity of this kind of information in the Namibian context was another reason for conducting this study. In this study the children on ART at Katutura Hospital were followed for a period between four and nine years to estimate their survival status. Though ART services for children started in 2002 in Namibia, there are no studies on survival analysis and factors that affect the survival of children on ART

in the country as well as the study area. This further highlights the need for local evidence to estimate the survival patterns and also determine the factors affecting the survival of children on ART at Katutura Hospital in Windhoek.

4.3 Materials and methods

4.3.1 Study design, sample and setting

A retrospective cohort study design was conducted for children who initiated ART between 01 January 2006 and 31 December 2010. All children living with HIV/AIDS below the age of 15 and started taking ARVs from Katutura Hospital within the period of 2006 - 2010 were included in the sample. The target population was all the children under 15 years who are on the Katutura Hospital records as ART paediatric patients regardless of the period of their ART initiation. All children who did not have complete information such as missing date of initiation or date of last visit to hospital were dropped from the sample. The population constituted about 1605 paediatric patients and from these a final sample of 813 children was selected from those who initiated ART between 01 January 2006 and 31 December 2010. This sample size was established through the use of EpiCalc for sample size calculation and then adjusting the value for covariate item missingness. This dataset was obtained from the hospital's Management Information System database (MIS), with the help of the hospital data clerks. The study subjects were followed from initiation to December 2014.

The setting for this study was Katutura State Hospital located in Windhoek, the

capital city of Namibia. The hospital has an HIV/AIDS section which deals with all aspects of HIV case management and care, that is, from enrolment, counselling, initiating, follow-up and giving out of the drugs to patients during their scheduled visits to the hospital.

4.3.2 Statistical analysis and data quality

Survival analysis was used in this study. The term survival analysis pertains to a statistical approach designed to take into account the amount of time an experimental unit contributes to the study before experiencing an event. The analysis consists of following the subject until death or lost to follow-up. For this study, the time to death was the outcome variable. The time to death was assessed for subjects from the date of initiation to the date of death before 31 December 2014 (the end of the study period). Those who did not experience the event of interest before the end of the study were considered censored together with those who were lost to follow-up. For the lost to follow-up paediatric patients, their time at risk was taken up to the last visit date. Data on HIV related deaths for these children on ART was captured in the database after being informed by physicians and/or relatives. The deaths which the hospital could not confirm as HIV related were not recorded, thus such children were considered as censored in this study. This is one of the main challenges with data from resource limited countries like Namibia as this tends to underestimate the actual HIV related deaths. The baseline socio-economic, immunological and clinical variables were measured and recorded at initiation. These were then used in the analysis either as they were collected or used to create indicator variables that are recommended when dealing with paediatric survival analysis. The Kaplan-Meier test

was used to estimate survival after the initiation of ART, and the log-rank tests were done to compare the survival prospects of different groups. The variables with p-value less than 20 percent in log-rank tests were selected as potential risk factors for use in the multivariate Cox proportional hazards model. The variables were also tested against the proportionality assumption and all that fulfilled the assumption through the Kaplan Meier survivor curves were included in the Cox's model. The significant variables in the multiple Cox proportional hazards model (equation 2.4.5) were included in the final model. Goodness-of-fit tests were performed on the final model using the Cox-Snell method. A significance level of five percent was used. The proportional hazards assumption was also checked for the statistically significant variables included in the final model using the re-estimation based method and none of the variables in the final model violated the proportionality assumption. Data quality for this study was enhanced by focussing on the paediatric patients who fulfilled the defined study design and who were under the age of 15 at initiation. Quality was also assured by dropping all the cases that did not have complete information especially for the measurement of survival time. The collected data was cleaned, coded and then analysed using SAS version 9.4 (SAS Institute Inc, 2013).

4.3.3 Covariate description

For the purposes of this study and also due to the limitations associated with secondary data collected from the Ministry of Health and Social Services management information systems database, the study covariates were limited to the paediatric patient's gender, baseline World Health Organisations (WHO) clinical stages of HIV, baseline functional status, baseline age category and absolute CD4 count. Absolute

CD4 count was classified according to the age of the child and it has two categories, CD4 count below threshold and CD4 count above threshold. For these variables and their categories see Table 3.1 and paragraph 3.2.1 for CD4 threshold definitions.

4.3.4 Ethical clearance

Ethical clearance was obtained from the Ministry of Health and Social Services and approval for data collection was granted by the provincial medical doctor (PMD) stationed at Katutura Hospital. As the study was done using secondary data, individual paediatric patients' consent was not necessary since the database only captured patient identification numbers and not names or any other information that may directly be linked to the patient. The data was also not shared with any third party hence confidentiality was maintained (See Appendix A.2).

4.4 Results and analysis

4.4.1 Descriptive results and analysis

A total of eight hundred and thirteen (813) records of children under the age of fifteen who initiated antiretroviral treatment (ART) between 01 January 2006 and 31 December 2010 were included for this study. Table 4.1 and Table 4.2 show the distribution and median survival times for paediatrics and adolescents on ART in each category.

From Table 4.1, two hundred (24.6%) were infants below one-year, three hundred and three (37.3%) were between one and five years, whilst three hundred and ten (38.1%)

Table 4.1: Distribution of paediatric and adolescent ART patients

Risk factor variable	n	Percent(%)
Clinical stage		
Stage 1	184	25.5
Stage 2	194	26.9
Stage 3	283	39.3
Stage 4	60	8.3
Absolute CD4 count		
CD4 count below threshold	430	53
CD4 count above threshold	382	47
Age category		
$< 1 \ \ year$	200	24.6
$1- < 5 \ \ years$	303	37.3
$5- < 15 \ \ years$	310	38.1
Functional status		
Ambulatory	85	11.2
Bedridden	12	1.6
Working	659	87.2
Gender		
Female	412	50.7
Male	401	49.3

were five and above but below fifteen. Four hundred and twelve (50.7%) of these children were female whilst four hundred and one (49.3%) were male.

The majority of children, two hundred and eighty-three (39.3%), initiating on ART at Katutura Hospital during the period under study were in clinical stage III of HIV infection whilst only sixty (8.3%) initiated with a clinical stage IV condition. One hundred and eighty-four (25.5%) and one hundred and ninety-four (26.9%) initiated in clinical stages I and II respectively. Four hundred and thirty (53%) initiated ART with CD4 count below threshold whilst three hundred and eighty-three (47%) had

their CD count above the threshold values (refer to section 3.2.1 for the threshold definitions or (Bong et al., 2007)). In terms of the functional status at which these children started treatment on, the majority, six hundred and fifty-nine (87.2%), started in a working functional condition whilst just twelve (1.6%) and eighty-five (11.2%) were in the bedridden and ambulatory functional conditions respectively.

Table 4.2 shows that the median survival time for children who initiated ART in

Table 4.2: Median survival times by baseline clinical stage, functional status and age category

Risk factor variable	n	Percent(%)	Median(days)	IQR in days
Clinical stage				
Stage 1	184	25.5	1591	336.5-2206
Stage 2	194	26.9	2134	609-2867
Stage 3	283	39.3	1612	391-2469
Stage 4	60	8.3	1373.5	195-2466.5
Age category				
< 1 *year*	200	24.6	831	181-2001
$1- < 5$ *years*	303	37.3	1673	444-2707
$5- < 15$ *years*	310	38.1	2032.5	887-2657
Functional status				
Ambulatory	85	11.2	1908	473-2785
Bedridden	12	1.6	401.5	85.5-1916.5
Working	659	87.2	1677	437-2465

clinical stage I of HIV infection, is 1591 days, with an interquartile range (IQR) of (336.5-2206) days. The children who initiated while in stage II had a much longer median survival time of 2134 days with an IQR of (609-2867) days. The least median survival time of 1373.5 days with IQR (195-2466.5) days was observed for children who initiated ART in the disease clinical stage IV of HIV infection. The children who initiated ART in a bedridden condition had a median survival time of 401.5 days with

IQR (85.5-1916.5) days whilst those who started ART in an ambulatory and working condition had median 1908 with IQR (473-2785) days and 1677 with IQR (437-2465) days respectively.

The median survival time for infants, that is, children below one year was 831 with IQR (181-2001) days while those $1- < 5$ years and $5- < 15$ years was 1673 with IQR (444-2707) days and 2032.5 with IQR (887-2657) days, respectively.

Figure 4.1 shows the survival estimates for the children who initiated ART between

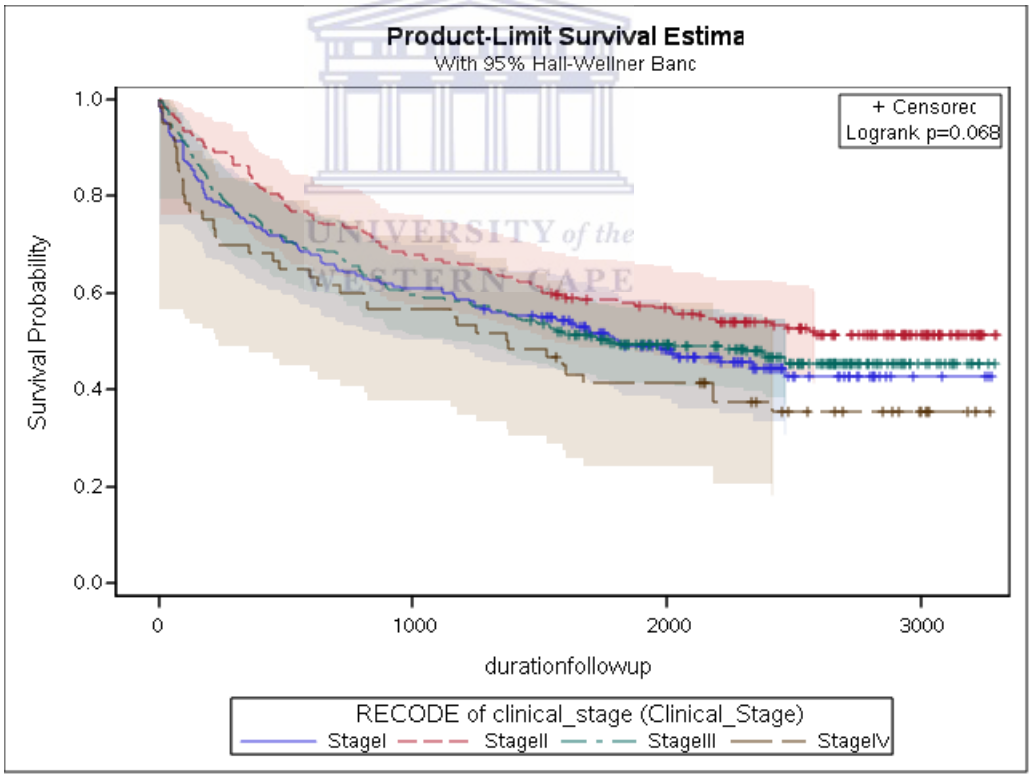


Figure 4.1: Kaplan-Meier survival estimates curve by clinical stages

2006 and 2010 at Katutura Hospital. The children's survival probability is estimated

based on the baseline clinical stage and from Figure 4.1, children who initiate ART in stage IV has the lowest survival probability curve over the whole period under study. Children who initiate ART at earlier disease clinical stages have higher survival probabilities. A log-rank statistic of 7.1005 and a p-value of 0.0688 indicates that there is no statistical difference in the survivor probabilities for the different clinical stages. However, a low p-value close to 0.05 may be indicative of a possible difference in survival prospects of children who initiate at different clinical stages.

Figure 4.2 shows that children who initiated on ART at Katutura Hospital with

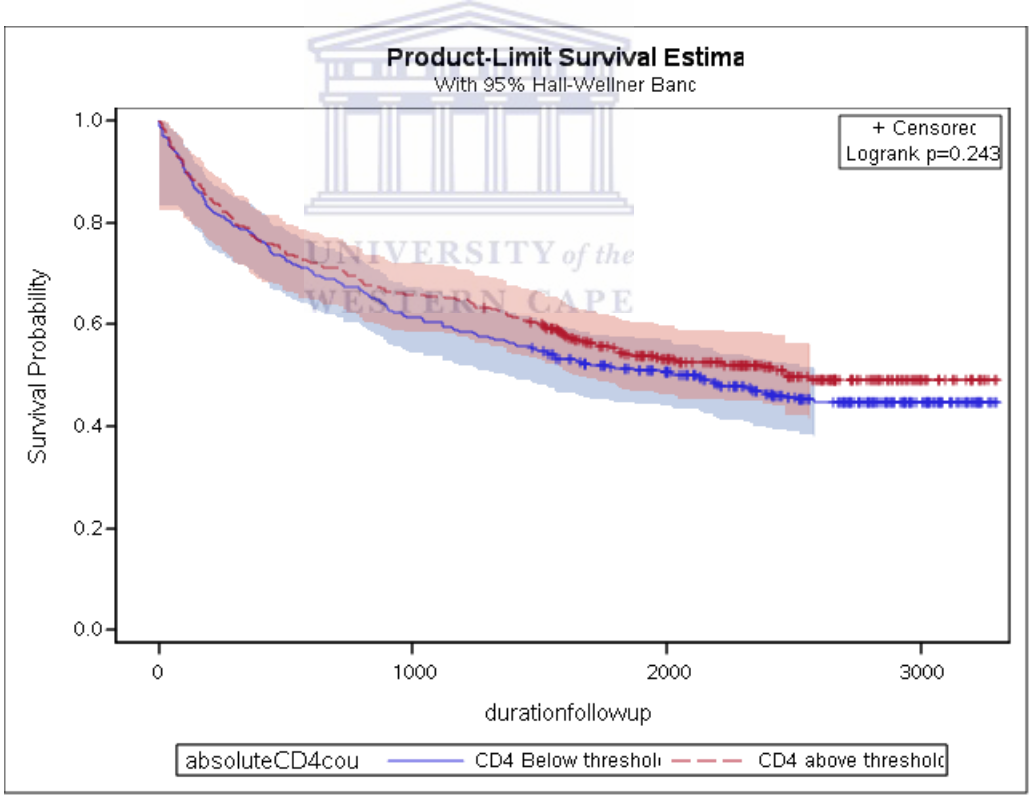


Figure 4.2: Kaplan-Meier survival estimates curve by absolute CD4 count

a CD4 count below threshold seem to have lower survival probability estimates as

compared to the other children of the same age who initiate on ART with their CD4 count level above the threshold value. A log-rank value of 1.3581 and a corresponding p-value of 0.2439 indicate that there is no significant difference in the survival prospects of these groups of children, that is, those that start with their CD4 count below threshold and those that start when their CD4 is above threshold.

Figure 4.3 shows the 95% Hall-Wellner Bands which are overlapping and thus indi-

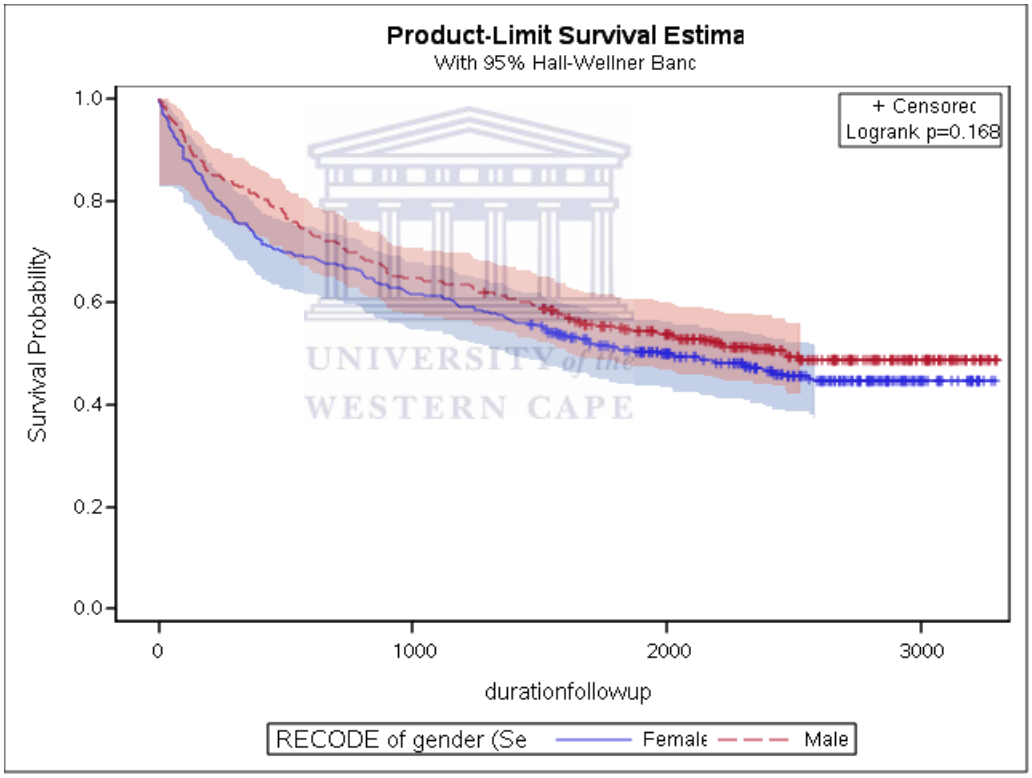


Figure 4.3: Kaplan-Meier survival estimates curve by gender

cating that there is no significant difference in the prospects of survival for the male or female children on treatment. The log-rank statistic of 1.8930 and a corresponding p-value of 0.1689 confirm that gender has no significant effect on the survival prospects

for the children on treatment at the hospital. Figure 4.4 shows the survival estimates

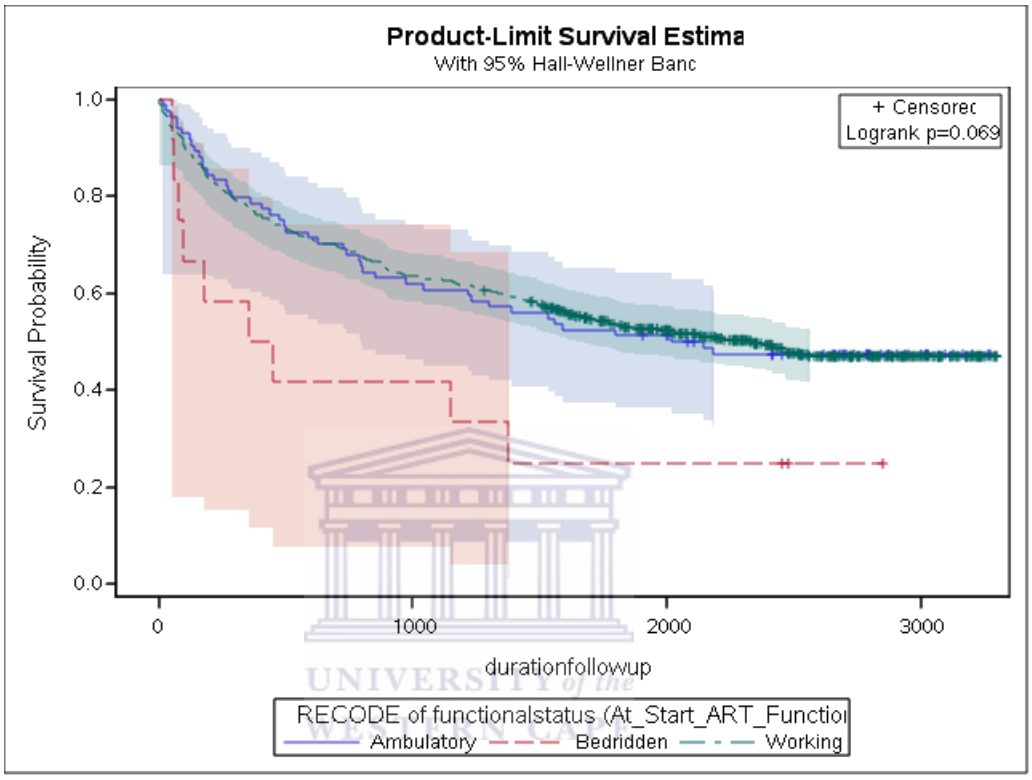


Figure 4.4: Kaplan-Meier survival estimates curve by functional status

for the children who initiated ART at Katutura Hospital between 2006 and 2010. The children's survival probability is estimated based on the baseline functional status and from Figure 4.4, the children who initiated ART in a bedridden condition have the lowest survival probability curve over the whole period under study. Children who initiate with a working or ambulatory condition have higher survival probabilities. A log-rank statistic of 5.3293 and a p-value of 0.0696 indicate that there is no statistical difference in the survivor probabilities for the different functional status, however, a low p-value close to 0.05 as this may be indicative of a possible difference in survival

prospects of children who initiate at these different functional status conditions.

Figure 4.5 shows that for the children that were initiated on ART at the hospi-

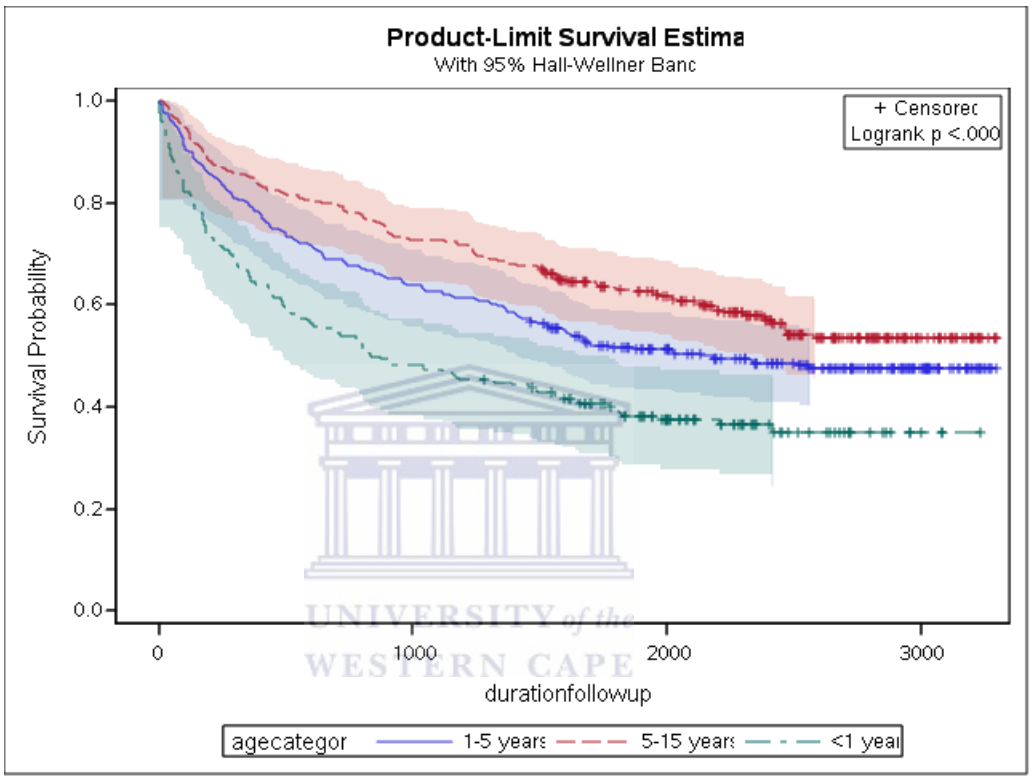


Figure 4.5: Kaplan-Meier survival estimates curve by age category

tal between 2006 and 2010, the infants had significantly lower survival prospects as compared to the other age groups. A log-rank value of 31.3799 and a corresponding p-value of < 0.0001 is indicative of a significant difference in the effect of age group in affecting the child's survival chances.

4.4.2 Proportional hazards model results and analysis

For the proportional hazards model (PHM), all the risk factor variables that had a p-value less than 20% from the log-rank test were included in the model. According to this criterion, all the variables except absolute CD4 count were included in the adjusted Cox proportional hazards model.

From the adjusted Cox regression analysis, Table 4.3, only age category was signifi-

Table 4.3: Cox's proportional hazards model estimates

Risk factor variable	AHR (95% CI)	P-Value
Clinical stage		
Stage 1	1	
Stage 2	1.01 (0.733-1.390)	0.9530
Stage 3	1.142 (0.860-1.515)	0.3596
Stage 4	1.387 (0.888-2.164)	0.1503
Age category		
< 1 *year*	1	
$1- < 5$ *years*	0.699 (0.534-0.915)	0.0093*
$5- < 15$ *years*	0.546 (0.414-0.721)	0.00001*
Functional status		
Ambulatory	1	
Bedridden	1.640 (0.741-3.629)	0.2225
Working	0.965 (0.691-1.347)	0.8321
Gender		
Female	1	
Male	0.939 (0.761-1.158)	0.5548

*significant at 5% level

cantly associated with the risk of death for the children on ART at Katutura Hospital for the period under study. Children in the age category $1- < 5$ years had an adjusted hazard ratio of 0.699 with a 95% CI of (0.534-0.915) showing that this age group is significantly around 30% less at risk as compared to the infants(< 1 year) age group. The children who start ART at a much older age group of $5- < 15$ years

were also at low risk of death as compared to the infants. An adjusted hazard ratio of 0.546 with a 95% CI of (0.414-0.721) was observed for this age group and it shows that these children are significantly around 45% less at risk when compared to the infants. All children who started ART in clinical stages II, III and IV were at a higher risk of death, 1%, 14% and 39% respectively, as compared to those who started in clinical stage I. However, the data could not highlight any significant difference in initiating at different clinical stages. Although not significantly different, the children who initiate on ART with a bedridden condition were around 64% at higher risk as compared to those who initiate in an ambulatory condition. Those who initiated with a working functional status were around 4% less at risk of death as compared to those who initiated with an ambulatory functional status. The child's gender did not have a significant effect on the risk of death but being male had less than a 7% protective effect from death.

4.4.3 Discussion

This study showed that children under the age of one year have a shorter follow-up (survival) time as compared to the other age groups. Children who started ART in a bedridden condition had the shortest median survival time amongst all who started ART in different functional statuses. Most children who were part of this study started ART when they were already in stage III of the disease, with only a few starting in the most advanced stage, that is, stage IV. This study also showed that there was a gradual decrease in the probability of survival for most of the risk factors at different levels except for WHO stage IV and bedridden conditions where the survival curves have a sharp decrease. From the Cox proportional hazards model analysis, this study

found out that the infants who initiated on ART during the period under study were at high risk of death, that is, being an infant had a significant negative effect on survival after initiation, as compared to the other age groups of children who were on treatment. This result highlights the need for close monitoring of infants who start on ART to make sure that all the complications or reactions that may result due to the administered regimens may be attended to as quickly as possible. This result is comparable to a study that was done in a large cohort of HIV positive children in Zambia (Bolton-Moore et al., 2007). This study could not find any significant effect for the absolute CD4 count and other clinical variables such as WHO clinical stages and functional status. A similar study that was done in Ethiopia (Kedir et al., 2014) found that children who initiated ART with advanced WHO clinical staging, CD4 count below threshold, underweight and low haemoglobin levels were predisposed to death more than the other children who initiated with different conditions. This study could not include as many risk factors as these other studies due to the limitations with the number of variables that the ministry records. However, the results from the log-rank tests for WHO staging and functional status showed the possibility of a significant difference in how the different levels of these factors impacted negatively on the survival prospects of the children under study. This is notwithstanding the results of the Cox proportional hazards model analysis which did not show these variables to be significantly affecting the survival prospects of the children.

4.4.4 Conclusions

The findings of this study indicate that the children who start ART in a bedridden condition, though few, are at a higher risk of death when compared to those who

start with a working and ambulatory functional status. The same can be said for those children who start ART at an advanced HIV/AIDS condition, that is, in stage IV. Targeted close monitoring as well as other interventions should be implemented for infants that are on ART treatment in order to reduce deaths among this group of children. Extensive research should also be carried out to identify the reasons why infants' survival times are negatively affected by initiating on ART when the treatment is supposed to have a positive effect.

This chapter utilized some descriptive survival analysis techniques to describe the paediatric HIV/AIDS patient data and then implemented the Cox's proportional hazards model to determine the prognostic risk factors for HIV/AIDS patients' survival time. The next chapter, that is, Chapter 5, further explores the same dataset using the Bayesian proportional hazards and the censored quantile regression models. Chapter 6 also uses the same dataset.

Chapter 5

Bayesian Cox proportional hazards model and insights from the censored quantile regression model for paediatric and adolescent HIV/AIDS patients on antiretroviral treatment

This chapter has been submitted for publication as:

Maposa, I. and Blignaut, R.J. (2016). Bayesian Cox Proportional Hazards model and insights from the censored quantile regression model for paediatric and adolescent HIV/AIDS patients on antiretroviral treatment. *African Journal of AIDS Research.*
Submission: Manuscript ID RAAR-2016-0097
Status: Under review

5.1 Summary

Survival analysis techniques are often used in biostatistics, epidemiological and clinical research to model time until event data. The purpose of this study was to fit a Bayesian proportional hazards model and a censored quantile regression (CQR) model to paediatric data and then compare the results in relation to inferences on the effect of the different prognostic risk factors on anti-retroviral treatment patients' survival times.

A retrospective cohort study design was conducted for children who initiated anti-retroviral treatment between 01 January 2006 and 31 December 2010. The sampling frame constituted 1605 paediatric patients and from these a final sample of 813 children was selected.

The results from a Bayesian proportional hazards model indicate that not being an infant had a positive effect on survival time. Patients initiating treatment in clinical stage II instead of stage IV had a significant positive effect on survival time.

The results from the censored quantile regression model are more revealing, highlighting that initiating in clinical stage II had a significant positive effect on survival time during the early periods of initiation compared to initiating in clinical stage IV. The effect reduces towards the 80th quantile, that is, towards the end of the follow-up period. The results also reveal that the patient's gender has a significant effect during the early periods of starting ART but this is not significant at any other point during treatment.

The conclusion from this study is that more insights are obtained from using censored quantile regression models as compared to the Bayesian proportional hazards models framework. However, it is our belief that in applying both models together, we can get hazard ratios and insights into how the prognostic risk factors affect survival time at different time points.

5.2 Introduction

Survival analysis examines and models the time it takes for events to occur. Survival analysis focuses on the distribution of survival times. Although there are well known methods for estimating unconditional survival distributions, most insightful survival modelling techniques examine the relationship between survival and one or more predictors (Fox, 2008).

Survival analysis techniques are often used in clinical and epidemiological research to model time until event data (Nakhaee & Law, 2011). The most common approach that is used to model survival data is through proportional hazards models as introduced by Cox in 1972. This model examines the relative hazard rates for a condition between covariates or predictors.

In the analysis of regression models for censored survival data, one often wishes to assess the importance of certain prognostic factors such as age, gender, or race in predicting survival outcome. This is a general problem which is encountered in most clinical trials research in cancer and AIDS (Ibrahim, Chen, & MacEachern, 1999).

In recent years, thanks to modern computational advancements, there have been several developments in the literature on modelling survival data. Of interest is the Bayesian perspective that has been introduced and added within the proportional hazards modelling framework. Censored quantile regression models have also recently found some traction in survival data analysis. Koenker and Billias (2001), Koenker and Geling (2001) and Portnoy (2003) all concur that non-linear quantile regression models have recently emerged and they are an attractive alternative to the popularly used models such as the proportional hazards and the accelerated failure time models (Koenker & Bilias, 2001; Koenker & Geling, 2001; Portnoy, 2003).

The proportional hazards model, or Cox regression model, is widely used in the analysis of time-to-event data to explain the effect of explanatory variables on hazards rates (Stokes, Chen, & Gunes, 2014). The Cox model is fitted by maximising the partial likelihood function. In the Bayesian approach, the partial likelihood function is used as the likelihood function in the posterior distribution (Sinha et al., 2003).

On the other hand, censored quantile regression models provide a direct and flexible approach to modelling survival times without the proportionality hazard constraint of the proportional hazards model. In clinical studies, quantile regression is helpful for identifying and distinguishing important prognostic factors for patient subpopulations that are characterised by short or long survival times (Lin & Rodriguez, 2013). Because lifetime distributions are generally skewed, the quantiles of the lifetimes, for example the 25th, 50th and 75th percentiles, are more informative than the mean for

summarising the lifetime distribution (Lin & Rodriguez, 2013).

Quantile regressions as introduced by Koenker and Basset (1978) provide a direct approach for modelling the quantiles of a response variable as a function of covariates (Koenker & Bassett Jr, 1978). By modelling these quantiles of a lifetime, insights can be gained about the dependence of the response distribution on its covariates. The quantile regression approach has two advantages, that is, the specific parametric form does not need to be specified for the lifetime distribution and a regression model can be fitted for each quantile (Lin & Rodriguez, 2013).

In this study paediatric HIV/AIDS survival data is used to implement the Bayesian proportional hazards model and the censored quantile regression model. Using both models contributes to additional information being obtained in identifying prognostic risk factors and their effects on prospects of survival for paediatric HIV/AIDS patients who are taking antiretroviral drugs.

5.2.1 Aim and objectives

The main aim of this study is to show that more insights can be gained on the dynamics of covariates in relation to patient survival times when a combination of models are used, in this case, the Bayesian proportional hazards model and then the censored quantile regression model. The objectives are:

1. Determining the median survival times for patients starting treatment in different levels of the risk factors;

2. Fitting a Bayesian Cox proportional hazards model on the paediatric dataset;

3. Fitting a censored quantile regression model on the same dataset;

4. Comparing and discussing the prognostic inferences obtained from the Bayesian proportional hazards and the censored quantile regression models.

5.3 Materials and methods

5.3.1 Study design, sample and setting

A retrospective cohort study design was conducted for children who initiated anti-retroviral treatment (ART) between 01 January 2006 and 31 December 2010. All children living with HIV/AIDS below the age of 15 and who were on anti-retroviral (ARV) treatment administered from Katutura Hospital within the period of 2006 - 2010 were included in the sample. All children who did not have complete information such as missing date of initiation or date of last visit to hospital were dropped from the sample.

The target population was all the children under 15 years who are on Katutura Hospital's records as ART paediatric patients regardless of the period of their ART initiation. The population constituted 1605 paediatric patients who initiated ART between 01 January 2006 and 31 December 2010. Of these, a final sample of 813 was selected.

Missing data from all the variables included in the study were replaced by imputed values. The multiple imputation (MI) procedures in SAS were implemented. Logistic regression was used to impute binary and categorical missing values whilst a

predictive mean matching regression model was used to impute missing values for continuous variables. These data were obtained from the hospital's Management Information System database (MIS) with the help of the hospital data clerks.

The study subjects were followed from initiation to December 2014. The setting for this study was Katutura State Hospital located in Windhoek, the capital city of Namibia. The hospital has an HIV/AIDS section which deals with all aspects of HIV case management and care, that is, from enrolment, counselling, initiating, follow-up and giving out of the drugs to patients during their scheduled visits to the hospital.

5.3.2 Variables of interest

For the purposes of this study and also due to the limitations associated with secondary data collected from the Ministry of Health and Social Services' MIS database, the study covariates were limited to paediatric patients' gender, baseline World Health Organisations (WHO) clinical stages of HIV, baseline functional status, baseline age category and absolute CD4 count. Absolute CD4 count was classified according to the age of the child and it had two categories; CD4 count below threshold and CD4 count above threshold. CD4 count below threshold was defined as:

- for infants, less than 12 months, CD4 count less than $1500/mm^3$;
- for age $12 - 35$ months, CD4 count less than $750/mm^3$;
- for age $36 - 59$ months, CD4 count less than $350/mm^3$;
- for age $>= 60$ months (5 years) up to 180 months (15 years), CD4 count less than $200/mm^3$ (Bong, et al., 2007) .

The categories for each of the variables were as follows: (also see Table 3.1).

- WHO staging: (1) Stage I (2) Stage II (3) Stage III (4) Stage IV;
- Absolute CD4 count: (1) CD4 below threshold (2) CD4 above threshold;
- Age category: (1) < 1 year (2) $1- < 5$ years (3) $5- < 15$ years;
- Functional status: (1) Ambulatory (2) Bedridden (3) Working;
- Paediatric gender: (1) Female (2) Male.

5.3.3 Ethical clearance

Ethical clearance was obtained from the Ministry of Health and Social Services and approval for data collection was granted by the Provincial Medical Doctor (PMD) stationed at Katutura Hospital. As the study was done using secondary data, individual paediatric patients' consent was not necessary since the database only captured patient numbers (created for the extracted database to link patient observations over time) and not names or any other information that may directly be linked to the patient. The data was also not shared with any third party hence confidentiality was maintained (See Appendix A.2).

5.3.4 Kaplan-Meier median survival time and log-rank test

The Kaplan-Meier median survival time is based on the Kaplan-Meier survival estimator method. This method is the best for determining median survival times for follow-up patients as it incorporates the censoring aspect in the computation of the

estimates. The log-rank tests are used to assess whether there are significant differences between the survival patterns amongst different levels in which patients initiate treatment.

5.3.5 Bayesian proportional hazards model

This model is premised on the proportional hazards modelling framework and adds the Bayesian flavour in estimating the model coefficients. In this approach, the partial likelihood function 5.3.1 is used as the likelihood function in the posterior distribution (Sinha et al., 2003). The likelihood is given as:

$$L(\beta) = \prod_{j=1}^{r} \left[\frac{exp(\beta' X_j)}{\sum_{l \in R(t_i)} exp(\beta' X_l)} \right]^{\delta_i}, \tag{5.3.1}$$

where $R(t_i)$ is the risk set at time t_i and δ_i is the indicator variable that takes a value 1 if the outcome of interest is observed at t_i. Also the X_j is the vector of covariates for the individual who dies at the j^{th} ordered death time, t_j. The summation in the denominator of this likelihood function is the sum of the values of $exp(\beta' X)$ over all the individuals who are at risk at time t_j and the $\beta' s$ are the coefficients to be estimated.

5.3.6 Censored quantile regression model

A censored quantile regression model is estimated for the quantiles up to the 80th percentile. This estimation of coefficient effect variation up to the 80th percentile gives us a picture of how the different risk factors affect the patients at different survival times. The model being fit is:

$$Q(\tau|X = x) = X^{'}\beta(\tau) + \epsilon_i, \tag{5.3.2}$$

where $Q_{y_i}(x)$ is the τth quantile of follow-up time y_i at $\mathbf{X}$ (Lin & Rodriguez, 2013). The quantile function is invariant under a monotone transformation. The following minimisation equation is used to determine the optimal parameter estimates.

$$min_{\beta \in R^p} V_n(\beta) \equiv \sum_{i \in (i:y_i \geq x_i^T \beta)} \tau|y_i - x_i^T\beta| + \sum_{i \in (i:y_i < x_i^T \beta)} (1-\tau)|y_i - x_i^T\beta|. \tag{5.3.3}$$

This function represents a linear programming problem which implements the simplex method to find the optimal β solutions for the objective function (Leng & Tong, 2013).

5.4 Results and analysis

The data were analysed using SAS 9.4 (SAS Institute Inc, 2013) and the results are shown in the following tables and figures. Table 5.1 shows the patient numbers as well as the median survival times for children in each category of the risk factor variables that were considered in this study. The children who started ART in WHO stage IV of the disease had the lowest median survival time. Low median survival times were also observed for children that initiated ART in the bedridden health condition as well as those children who initiated in the age group below one-year. The log-rank test shows that the survival probabilities were significantly different among children who initiated ART in different WHO stages (p=0.038) and also different age groups (p=0.000).

The next sections fit the Bayesian proportional hazards and then the censored quantile regression models. The inferential insights from these models are then discussed.

Table 5.1: Kaplan-Meier median survival times and log-rank tests

Variable	n	Median survival time(days)	Log-rank p-value
WHO stage			
Stage 1	206	2020	
Stage 2	220	1792	
Stage 3	314	1726	
Stage 4	73	1530	p=0.038*
Absolute CD4 count			
CD4 count below threshold	431	2021	
CD4 count above threshold	382	2469	p=0.2102
Age category			
less than 1 year	200	851.50	
$1- < 5$ *years*	303	2134	
$5- < 15$ *years*	310	2465	p=0.000*
Functional status			
Ambulatory	96	1712.50	
Bedridden	13	447	
Working	704	2338	p=0.0502
Gender			
Female	412	1977	
Male	401	2467	p=0.1508

*significant at 5% level

5.4.1 Bayesian proportional hazards model results

Further analyses on the effect of the different risk factors on children's survival times are in the tables and figures below. Table 5.2 shows the different risk factors that were considered in this study with the corresponding adjusted hazard ratios (AHR) linked to each risk factor. From the cohort of children under study, those who initiated ART in the clinical stage I, were at a slightly high risk of death (AHR=1.1237) on average as compared to those who initiated in the clinical stage II. Those who initiate ART in stage I have a better chance of survival when compared to those who

Table 5.2: Mean adjusted hazard ratios (AHR) and quartiles' specific hazard ratios

Variables	Hazard ratios (95% HPD Interval)	Hazard ratios at quartiles		
		25%	50%	75%
WHO staging				
WHO stage I vs II	1.1237 (0.8151-1.4371)	1.0092	1.1119	1.2265
WHO stage I vs III	0.8453 (0.6389-1.0389)	0.7671	0.8386	0.9163
WHO stage I vs IV	0.8363 (0.5350-1.1858)	0.7119	0.8176	0.9343
WHO stage II vs III	0.7597* (0.5720-0.9474)	0.6906	0.7543	0.8219
WHO stage II vs IV	0.7521 (0.4822-1.0672)	0.6424	0.7326	0.8421
WHO stage III vs IV	0.9955 (0.6401-1.3676)	0.8617	0.9721	1.1060
Absolute CD4 count				
Below vs above	1.0914 (0.8870-1.3239)	1.0131	1.0860	1.1625
Age category				
$1- < 5$ vs 5-15 yrs	1.2560 (0.9821-1.5538)	1.1523	1.2478	1.3471
$1- < 5$ vs < 1 yrs	0.6479* (0.4969-0.8099)	0.5915	0.6422	0.6996
$5 - 15$ vs < 1 yrs	0.5198* (0.3909-0.6573)	0.4720	0.5153	0.5616
Functional status				
Ambul vs bedridden	0.6317 (0.2353-1.1307)	0.4533	0.5829	0.7498
Ambul vs working	0.9808 (0.6910-1.2780)	0.8768	0.9742	1.0760
Bedridden vs working	1.7547 (0.6956-2.9892)	1.3116	1.6728	2.1014
Paediatric gender				
Female vs male	1.1529 (0.9450-1.3915)	1.0739	1.1466	1.2248

*significant at 5% level

initiate the ART treatment in stages III and IV, with AHR=0.8453 and AHR=0.8363 on average, respectively. Initiating ART in stage II also implied a better chance of survival when compared to stages III and IV, with AHR=0.7597 and AHR=0.7521 on average, respectively. Those who started ART in stage III were at a better chance of survival when compared to those who initiated in stage IV with AHR=0.9955 on average. The effect of clinical stages seem to decrease with time on treatment as shown by the reducing adjusted hazard ratios (AHR) from 25th to 75th percentiles on all the different combinations of clinical stages being compared. The risk of death

among these different children enrolling for ART at different clinical stages was only significant for those who initiated in stage II compared to those who initiated in stage III. This indicates that the clinical stage with which the child initiated on ART had some effect on the child's chances of survival at any point in the follow-up period.

Initiating ART with a CD4 count below threshold implied a higher risk of death as compared to those who initiated with an absolute CD4 above threshold (AHR=1.0914). The risk seems higher in the longer term under treatment when compared to early periods of initiation. The 95% confidence intervals show that the risk difference is not significant on average.

The risk of death for the children between 1 and 5 years was higher in this cohort as compared to the age group between 5 and 15 (AHR=1.2560). This shows that on average, all other variables being the same, the [1-5) years age group has a 25.6% higher chance of dying as compared to the [5-15) year age group. The different quantiles reveals more; the risk of death is much higher for this age group in the later periods (75th percentile) of taking ART as compared to when they just begin. The risk profile is not significant since the 95% confidence interval includes 1. The infants ($<$ 1 year) are at a significantly higher risk of death as compared to both the [1-5) and the [5-15) year age groups. The risks are much higher during the early initiation period as compared to the later parts of ARV treatment. Both the 95% confidence intervals do not include 1 signifying significant risk at these points.

The infants and adolescents who initiated on ART in an ambulatory condition were

at a lesser risk of death (AHR=0.6317) as compared to those who initiated in a bedridden health condition. Moreover, the risk of death is high (54.7%) and probably significant in the early days of initiation, that is, at the 25th percentile of survival time after initiation (AHR=0.4533) for bedridden children as compared to children in an ambulatory condition. The risk of death reduces with longer periods spent on taking ART, that is, at 75th percentile the hazard of death is reduced to 25.02% compared to that for the first quantile. The variation for the adjusted hazard ratio of those children who initiate in the ambulatory condition against those in working condition is small across the three quantiles, that is, the AHR25= 0.8768, AHR50= 0.9742 and AHR75= 1.0760. Bedridden condition puts a high risk of death on children as compared to initiating in the working health condition. A higher difference is noted in the 75th percentile where those children who initiated in a bedridden condition are at a much higher risk of death as compared to those children who initiates with a working functional condition. The 95% confidence intervals indicate that the different risk levels are on average not significantly different.

Adjusting for other factors, female children on ART are at a slightly higher risk of death as compared to their male counterparts. The 95% confidence intervals indicate that the risk is not significantly different for this cohort of children. The quantiles hazard ratios however indicate that the risk for female children is only 0.074% higher compared to male children at the 25th percentile but in the upper survival quantiles, 75%, the risk of dying goes up to 22.48% for female children as compared to male children of the same age group.

5.4.2 Censored quantile regression model results

Table 5.3: The estimated parameters for variables at each quantile level

Variable	τ_s (Quantile levels)							
	0.05	0.2	0.25	0.4	0.5	0.6	0.75	0.8
WHO stage								
Intercept	-30	150.3	284	581*	904.3*	1304*	2115*	2186*
WHO stage I	62	156.3	144	9	-16.8	110.4	-275	-262
WHO stage II	120*	325*	320	284	396	500.6*	132	97
WHO stage III	61.5	68.7	-55	-106.5	-79.8	-17.4	-288	-206
WHO stage IV	Ref							
Absolute CD4 count								
Below threshold	2.5	-38	-42	-34	77.5	198.6	125	154**
Above threshold	Ref							
Age category								
$1- < 5$ yrs	40.5	171.7*	310*	794.5*	860.5*	584.6*	593*	606*
$5- < 15$ yrs	76.5*	442.7*	636*	1168.5*	1037.8*	759.6*	578*	594*
< 1 yr	Ref							
Functional status								
Ambulatory	19	61.7	116	-31.5	338.8	365.2	191*	91
Bedridden	55	-41.7	-196	-191	-495.5	-857	-1139	-435
Working	Ref							
Gender								
Female	-34.5*	-182.3*	-257*	-231.5	-198.8	-82.2	-28	-22
Male	Ref							

*significant at 5% level

Table 5.3 shows the results from the censored quantile regression model. The quantiles 0.05, 0.2, 0.25, 0.4, 0.5, 0.6, 0.75, and 0.8 were used and censored quantile regressions were fitted for each of these quantiles. The results in this table indicate that there is a significant positive effect on survival time of initiating in stage II compared to stage IV in the early periods of initiating on ART. Initiating with CD4 below threshold has no significant effect on survival prospects of patients in the early periods of initiation. However, there is a significant effect in the later periods of being on treatment when

compared to those who initiates with a CD4 count above threshold. Age category effect is significant in all quantiles, that is, from start to the end of the follow-up period. The other interesting result from this table shows that being a female child has a significant negative effect on survival time in the early periods of initiating on ART compared to being a male child. Initiating with an absolute CD4 count below threshold had a significant effect on survival time only in the 80th percentile showing that compared to starting ART with an absolute CD4 count above threshold; children have a similar survival path until later on in the treatment process. Functional status at initiation had no significant effect on survival prospects of the children in this study. Table 5.4 shows the overall effect test for each of the risk factors at the different

Table 5.4: The overall variable effect at each quantile level (p-values)

Variable	τ_s (Quantile levels)							
	0.05	0.2	0.25	0.4	0.5	0.6	0.75	0.8
WHO stage	0.027*	0.026*	0.008*	0.156	0.054	0.000*	0.00*	0.00*
Absolute CD4 count	0.900	0.554	0.572	0.803	0.587	0.076	0.053	0.008*
Age Category	0.005*	0.000*	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*
Functional Status	0.843	0.760	0.508	0.878	0.305	0.048	0.004*	0.291
Gender	0.045*	0.003*	0.000*	0.067	0.128	0.473	0.657	0.703

*significant at 5% level

quantiles. A p-value less than 0.05 indicates that the variable effect is significant. The projected picture in this table is much more interesting. The overall effect of WHO staging indicates that the clinical stage that a child initiates on ART has a significant effect in affecting survival time at both ends of the survival time distribution but it does not have a significant effect in the intermediate survival time periods. The implication for this is that by using methods that are based on reflecting the average effect, such patterns may not be noticed yet they are vital in devising policies for

effective interventions. Also of interest in this table is the effect of gender, which highlights that gender has a significant effect on influencing survival times in the early periods of initiating on ART but this has no effect later on in the duration of taking treatment. Furthermore, age category at initiation had a significant effect across the whole survival time distribution. Functional status and absolute CD4 count at initiation had slight effects in the later periods of taking antiretroviral treatment. The figures below show the treatment effects across all the quantiles and highlight the results from Tables 5.3 and 5.4. Figure 5.1 shows that initiating in stage II instead

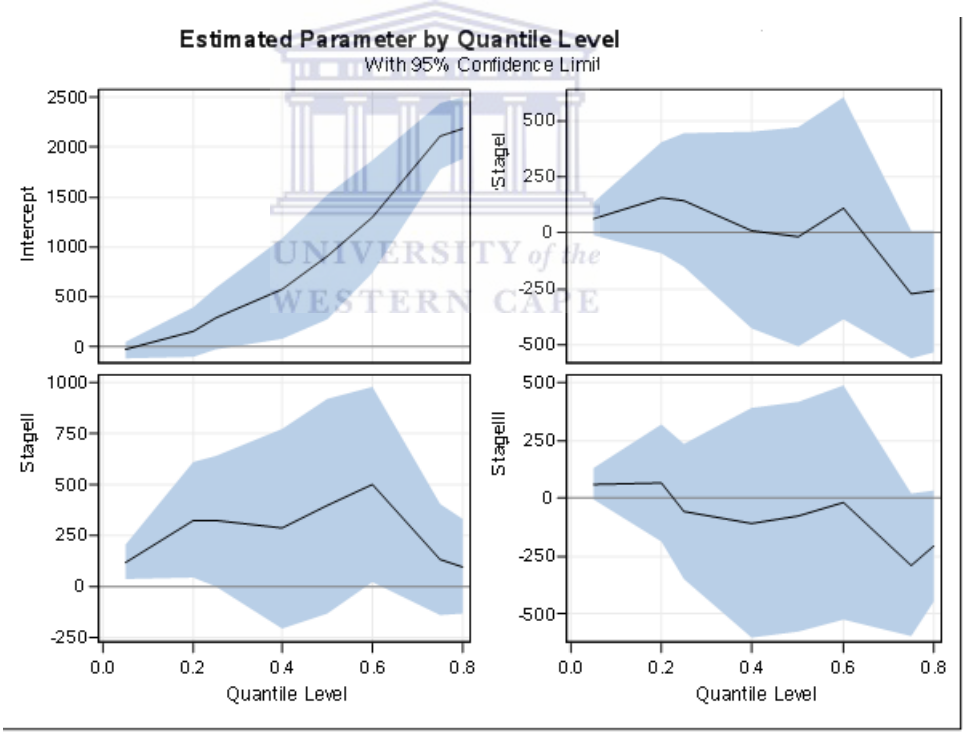


Figure 5.1: The effects of stage I, II and III compared to IV on paediatric patients' survival

of stage IV has a positive significant effect in the early periods of initiation but the effect becomes insignificant the longer the time spent on treatment. Initiating

in stage III as compared to stage IV is not significantly different across the whole survival distribution. Figure 5.2 shows that the paediatric and adolescent patients

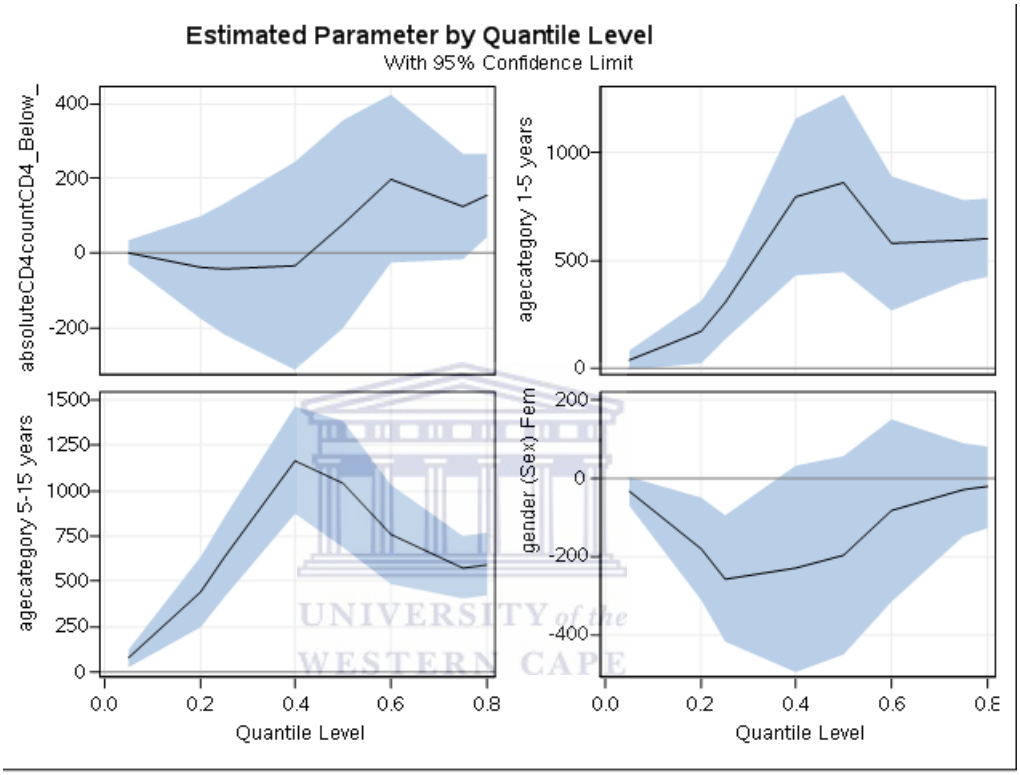


Figure 5.2: The effects of CD4 count below threshold (reference: above threshold), age category (reference: < 1 year) and gender (reference: male) on paediatric patients' survival time

who initiated with their absolute CD4 count being below threshold had a smaller chance of survival when compared to those who initiated with absolute CD4 count above threshold. The pattern changes however around the 50th quantile and the effect becomes positive though insignificant. Compared to initiating as an infant (< 1 year), the other age groups have a positive significant effect on the prospects of patients' survival across all quantiles. Being female had a significant negative effect

on survival time in the early periods of enrolling on treatment. The effect remains in the negative territory across all quantiles though in the upper quantiles the effect is no longer significant. Figure 5.3 shows that those patients who initiated in the

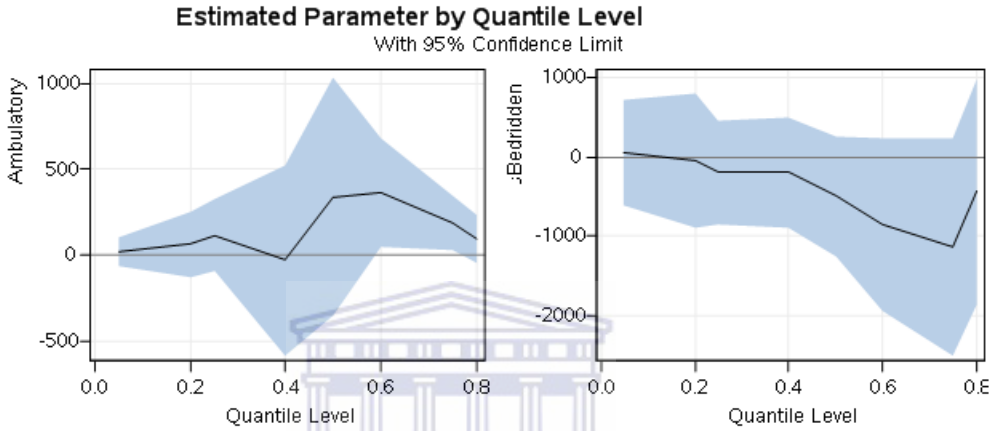


Figure 5.3: The effects of functional status (reference: working status) on paediatric patients' survival time

bedridden health condition had a smaller chance of survival compared to those who initiated in the working health status across all the quantiles. Moreover, being in an ambulatory state has a significant effect on survival time of patients only at a later point during treatment, that is, from around the 60th quantile for this study.

5.4.3 Discussion and conclusion

This study highlights that age categories and WHO clinical stages at start of ART have significantly different effects on patient survival times. This study also demonstrates that the use of Bayesian proportional hazards models in assessing the effects of different risk factors on survival of patients on ART can be enlightening due to the estimation of the hazard ratios at different quartiles. However, the details of how the different covariates affect the survival of patients across the survival time distribution can be augmented through the use of quantile regression models. This study shows that variables such as WHO staging had a significant effect on survival time at both extremes of the survival time distribution, that is, during the early periods of initiation as well as later towards the end of the follow-up period. The effect was shown not to be significant in-between. This insight is only possible if the censored quantile regression models are used in determining the covariate effects on survival times for patients. The patterns and dynamics of variable effect on survival times revealed by this modelling strategy are easily missed when the Bayesian proportional hazards model is used. The other variable which showed the same patterns was the gender of the patient; being female had a negative effect on the survival time of a patient during the early periods of initiation only. From the Bayesian proportional hazards model, the gender of the child, absolute CD4 count and functional status did not have a significant effect on survival time. In conclusion, the use of these methods, that is, the censored quantile regression model and the Bayesian proportional hazards model gives a rich inferential framework on the dynamics of the different prognostic risk factors' effects on the survival time of patients on ART.

This chapter implemented the Bayesian proportional hazards and the censored quantile regression models to identify and assess the different risk factors. The next chapter, that is, Chapter 6, compares the Cox's proportional hazards and the censored quantile regression models' estimated parameter effects across the survival time distribution.

Chapter 6

Comparison of the Cox's proportional hazards model and the censored quantile regression model: An analysis of the effects of different risk factors on paediatric survival time

This chapter was a presentation in a departmental seminar at the Namibia University of Science and Technology, Faculty of Health and Applied Sciences:

Maposa, I. and Blignaut, R.J. (2016). Comparison of the Cox's proportional hazards model and the censored quantile regression model: An analysis of the effects of different risk factors on paediatric survival time. *Department of Mathematics and Statistics: 24 March 2016.*

Windhoek,

Namibia

6.1 Summary

The fitted proportional hazards model quantifies the hazard or risk associated with each variable on patient survival time. The hazard is estimated for the whole duration of the survival time (average hazard). The fitted censored quantile regression (CQR) model reflects the effect of each risk factor on survival time at each quantile. The censored quantile regression model thus characterizes the effect of risk factors at each point of the patient survival. The two models reflect relatively similar information in terms of risk factors in some situations. The censored quantile regression model gives more insight on the effects of each risk factor at the different quantiles, showing that the effect is not constant over the duration of patient survival. The CQR model reveals that a risk factor can significantly affect the survival of patients in the early periods after initiation or in the later periods of survival on treatment. When survival time is skewed, as is usually the case, the censored quantile regression models provide a better picture on the dynamics of how the risk factors affect patient survival at any point during treatment.

6.2 Background and purpose of study

In the past two to three decades, quantile regression (QR) has become a more widely used statistical technique to describe the distribution of a response variable given a set of covariates (Koenker & Bassett Jr, 1978). This regression approach offers great flexibility in assessing covariate effects on event times (at quantile levels), thereby attracting considerable interest in it's applications in survival analysis and modelling. In some data types, for example, clinical trials, the marginal distribution of response is characterized by marked skewness and quantile regression methods are emerging as popular techniques for exploring the distribution of the event time data. Quantile regression models the change of quantiles of the conditional distribution of the duration in response to changes of the covariates (Fitzenberger & Wilke, 2005).

Quantile regression is a statistical method that is used to estimate and conduct inference about conditional quantile functions. Quantile regression methods, as with classical linear regression which implements the ordinary least squares approach in estimating models for conditional mean function, offers a mechanism to estimate models for the conditional median function and a range of other conditional quantile functions (Koenker, 2005). By estimating an entire family of conditional quantile functions instead of a single conditional mean function, quantile regression techniques are capable of providing a more complete statistical analysis of the stochastic relationships among random variables (Koenker, 2005).

Quantile regression techniques are used in a broad range of applications some of which includes ecology, econometrics and survival analysis. In survival analysis, and

event history analysis more generally, there is often a desire to focus attention on particular segments of the conditional distribution, for example survival prospects in the early periods after starting ART treatment (Koenker, 2005).

While conventional regressions (in this case, the proportional hazards regression model) estimators show that the survival times or failure times recognise the presence of some risk factors, quantile regression adds a new dimension to the literature, suggesting that the effect of the risk factors varies from quantile to quantile of the survival time distribution. For patients with the higher survival times, the survival function will barely recognise some risk factors, yet for those with lowest survival times, their survival function is particularly sensitive to the presence of some risk factors (Fitzenberger & Wilke, 2005). The implication being that some risk factors may be significant for shorter survival times whilst some may be significant for the longest survival times of patients on ART treatment.

Thus, quantile regression techniques can help us obtain a more complete picture of the underlying relationships among risk factors and survival times. Censored quantile regression addresses the issue of right censoring of the survival times, which is common in duration analysis.

The current study was carried out in order to compare the use of censored quantile regression and the Cox's proportional hazards model in assessing the association among patients' clinical and demographic variables with their corresponding survival times. Interestingly, the censored quantile regression methods are nonparametric,

while the Cox's proportional hazards are semi-parametric, with the proportionality assumption being integral.

6.2.1 Purpose of study

The purpose of this study was to compare the effects of covariates on the survival time distribution using the Cox proportional hazards model and the censored quantile regression model. The aim being to determine the method that provides a better characterisation of the covariate effect across the survival time distribution between these two modelling frameworks.

6.2.2 Objectives

The objectives of the study were to:

- Estimate a proportional hazards model;
- Estimate a censored quantile regression model;
- Make comparisons of the two models in terms of their robustness in inference;
- Determine whether the censored quantile regression method provides a better characterisation of the effects of the risk factors across the survival time distribution compared to the Cox's proportional hazards model.

6.3 Significance of the study

This study illustrates the characterisation of the risk factor effects across the chosen quantile points using both the censored quantile regression and Cox's proportional

hazards models. Public health researchers and biostatisticians may find the information from this study useful in terms of designing interventions at the right time for the correct risk factors.

6.4 Materials and methods

6.4.1 Censored quantile regressions

Linear censored quantile regression (CQR) as derived from quantile regression (QR), introduced by Powell (1984), allow for semi-parametric estimation of quantile regressions for a censored model in a robust way (Powell, 1984). Censored quantile regression (CQR) is robust against mis-specifications of the error term since only fairly weak assumptions are required for the error term. Duration data are often censored. Right censoring occurs when we only observe that a spell has survived until a certain duration (e.g. when the period of observation ends) but we do not know exactly when it ends. Left censoring occurs when spells observed in the data did start before the beginning of the period of observation. Spells who started at the same time and who finished before the beginning of the period of observation are not observed. Quantile regression can not be used with left censored data. Left censoring is also difficult to handle for Cox's proportional hazards models (PHM) since strong assumptions have to be invoked to estimate the model (Fitzenberger & Wilke, 2005). In this analysis, consideration is made on the case of right censoring which both PHM and CQR are well suited for. The theoretical framework for the CQR models and how they are estimated are laid out in sections 2.6 and 2.7.

6.4.2 Cox's proportional hazards model

Cox's proportional hazards modelling (PHM) is the basic modelling technique used in exploring the relationship between the survival experience of a patient and the covariates (risk factors) in survival data analysis. The details of the modelling framework and the estimation procedures are laid out in section 2.4.

6.4.3 Data and study design

A retrospective cohort study design was conducted for children who initiated antiretroviral treatment (ART) between 01 January 2006 and 31 December 2010. A sample of 813 was used. Missing data from all the variables in the study were replaced by imputed values. Multiple imputation (MI) methods were used. The setting for this study was Katutura State Hospital located in Windhoek, the capital city of Namibia. The variables included are as in Table 3.1, with the exception that the staging variable was made binary, that is, stages 1 and 2 were combined as well as stages 3 and 4. This was done also to assess the effect of starting ART in the HIV condition (stages I and II) compared the the AIDS disease condition (stages III and IV).

6.4.4 Parameter estimation

Cox's proportional hazards and censored quantile regression models were estimated. For the censored quantile regression model, the following quantile (τ) values were used: $\tau = (0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, \ and \ 0.45)$. The joint model fit of the proportional hazards and the censored quantile regression models was estimated on the same quantile values. The joint fit combines the two set of results onto the same graph for comparative purposes. The estimated effects from both models can

then be observed on the graph. The R package was used in this study as it provides the framework for combining the estimated effects from the Cox proportional hazards model and the censored quantile regression model.

6.5 Results and analysis

This section presents the results and analysis on the comparisons of the two models using the paediatric dataset. The models were initially estimated individually and then combined for the comparison using R-package.

6.5.1 Descriptive analysis of the covariates

Table 6.1: Patient distribution

Risk factor	Level	n(%)
Gender	Male	401 (49.32)
	Female	412 (50.68)
clinical stages	stages I and II	426 (52.4)
	stages III and IV	387 (47.6)
Absolute CD4count	Below threshold	431 (53.01)
	Above threshold	382 (46.99)
Age category	$< 1yr$	200 (24.6)
	$1- < 5yr$	303 (37.27)
	5-15yr	310 (38.13)
Functional status	Ambulatory	96 (11.81)
	Bedridden	13 (1.60)
	Working	704 (86.59)

Table 6.1 shows the distribution of the patients in each covariate category. Gender distribution was even and reflective of the national subgroup population. The number of children starting treatment in the AIDS condition (clinical stages III and IV) are slightly lower than those who start in earlier HIV stages, that is, clinical stages I

and II. The age distribution indicates that the dataset has almost all paediatric and adolescent age groups well represented. The results in Table 6.1 also reveal that there were very few children in the bedridden functional status (1.6%), compared to those who starts treatment in the working functional status (86.59%). Table 6.2 presents the Cox's proportional hazard model estimated results from the data.

6.5.2 Cox's proportional hazards estimates

Table 6.2: Proportional hazards model parameter estimates

Risk factor	Level	Coef	HR=exp(coef)	p-value
Gender	Male	-0.1374	0.8716	0.1643
	Female	Ref		
clinical stages	stages III and IV	0.2328	1.2621	0.0241*
	stages I and II	Ref		
Absolute CD4count	Above threshold	-0.0791	0.9240	0.4343
	Below threshold	Ref		
Age category	$1- < 5yr$	-0.4596	0.6316	0.00017*
	$5- < 15yr$	-0.6870	0.5031	0.0000*
	$< 1yr$	Ref		
Functional status	Bedridden	0.5942	1.8116	0.0872
	Working	0.0248	1.0251	0.8711
	Ambulatory	Ref		

*significant at 5% level

Table 6.2 indicates that there are two significant risk factors, that is, WHO clinical stage and age category of the child at treatment initiation. Starting treatment in clinical stage III/IV exposes the children to a high risk of death, 26% risk compared to initiating in clinical stages I/II. Age category below 1 year (infants) are at a high risk of death when compared to the other age groups, that is, $5- < 15$ years and $1- < 5$ years. Censored quantile regression was then used on the same dataset to

identify the risk factors and the results are in Table 6.3.

6.5.3 Censored quantile regression estimates

Table 6.3: The estimated coefficients for quantile levels

Variable	τ_s (Quantile levels)								
	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
WHO stage									
Intercept	43	129	198	231*	338*	377*	354	393	789
Stage III/IV	-34	-79*	-151*	-178*	-274*	-332*	-359*	-316	-434
Stage I/II	Ref								
CD4count									
Above thres	3	17	7	21	25	46	92	136	103
Below thres	Ref								
Age category									
$1- < 5\ yrs$	64*	101*	170*	214*	307*	450*	656*	763*	1070*
$5- < 15\ yrs$	84*	149*	291*	451*	677*	960*	1245*	1578*	2030*
$< 1yr$	Ref								
Funct status									
Working	-31	-77*	-109*	-115	-143	-86	-45	106	-120
Bedridden	-19	-91*	-157	-206*	-281*	-335*	-349	-278	-796
Ambulatory	Ref								
Gender									
Male	20	27	129*	219*	265*	233*	275	231	325
Female	Ref								

*significant at 5% level

Table 6.3 indicates that the different risk factors significantly affect the survival time of the paediatrics in different ways at different quantile points. Only one variable, that is, absolute CD4 count (CD4count), does not indicate significant differences in the effects of starting ART either with an above or below threshold CD count. All the other variables highlight that at some point in the follow-up, the variables will have a significant effect on paediatric survival times depending on baseline initiation levels.

The two models which were estimated as shown in Tables 6.2 and 6.3 were then combined in R for the graphical assessment of the coefficient effects on the survival time distribution. To observe the estimated coefficient effects for the Cox's proportional hazards model, refer to Figure C.1

6.5.4 Comparison of the Cox's proportional hazards model and censored quantile regression model

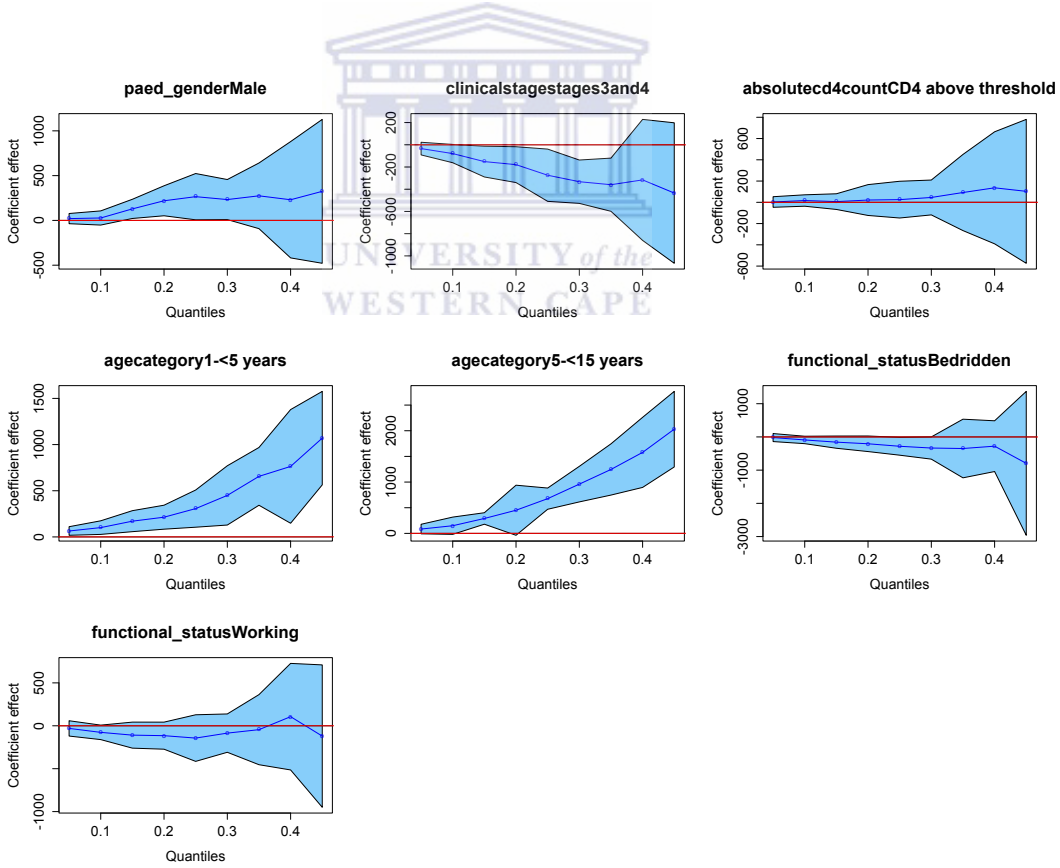


Figure 6.1: CQR and PHM comparison on risk factor effects

Figure 6.1 combines and assesses the coefficient effects as estimated through the Cox proportional hazards model (solid red line) and the censored quantile regression model (dotted blue line). The figure also shows the 95% bootsrap confidence intervals for the CQR model coefficient effects. Figure 6.1 indicates that male paediatric patients' survival time is significantly more than females in the quantiles 0.1 to around 0.3. Cox proportional hazards model effects highlight that the effect of being male compared to female is not significantly different from zero ($paedgenderMale$ plot). The effect of the Cox's proportional hazards estimates are clearly illustrated in the zoomed graph, that is, Figure C.1. Starting ART in clinical stages III/IV compared to I/II has a significant negative effect on survival time between quantiles 0.1 and 0.35. The Cox's proportional hazards model reflects that the effect is negative overall. The effects are shown not to be significantly different for $absolutecd4count$, $functionalstatusBedridden$ and $functionalstatusWorking$.

6.6 Discussion and conclusion

This study notes that for survival data analysis, these two methods are both essential and can augment each other. Overall, they reflect the same information about the variable effects but with different details and insights.

As noted from Figure 6.1, the Cox proportional model indicates an almost constant effect for the risk variables (see also Figure C.1) for this result, whilst the censored quantile regression model captures and better characterises the variability in effect of different risk factor variables on survival time across the distribution. Given that the Cox proportional hazards are widely used, the question is, can censored quantile

regression models replace the proportional hazards models? A number of studies such as Chernozhukov and Hong (2011); Koenker and Geling (2001); Leng *et al.,* (2013); Wang and Feygenson (2009) among others have demonstrated that censored quantile regression models can be used successfully in modelling survival (duration) data in econometrics and other application areas (Chernozhukov & Hong, 2011; Leng & Tong, 2013; Wang & Fygenson, 2009; Koenker & Geling, 2001). Koenker in his work in 2008 suggested that quantile regression for censored survival (duration) data offers a more flexible alternative to the Cox proportional hazard model for some applications (Koenker, 2008).

This study shows that censored quantile regression methods can be implemented as a biostatistics modelling tool that can augment the Cox proportional hazards model as it gives more insights on the effects of risk factors. The previous chapter, that is, Chapter 5, implemented the censored quantile regression and the Bayesian proportional hazards models and revealed that the risk factor effects are not constant over the survival time distribution. The result in this chapter confirms that observation as can be noted from Figures 6.1 and C.1.

The next chapter, that is, Chapter 7, implements the censored quantile regression model to an adult MSF Bulawayo dataset for lower survival times (early deaths). The result will be compared to other studies that were done focusing on similar survival time bands.

Chapter 7

Determinants of shorter survival times for HIV/AIDS patients on antiretroviral therapy using censored quantile regression models

This chapter has been submitted for publication as:

Maposa, I. and Blignaut, R.J. (2014). Determinants of shorter survival times for HIV patients on Antiretroviral Therapy using Censored Quantile Regression models. *Journal of Social Aspects of HIV/AIDS*.

Resubmission ID:RSAH-2016-0023

Status: Under review.

7.1 Summary

Survival analysis methods are often used to measure the risk of death and identify the prognostic risk factors for patients' survival times. Censored quantile regression has emerged as a powerful survival analysis method as it links the quantiles of patients' survival times to their demographic and clinical profiles, facilitating the identification of important prognostic factors. HIV/AIDS has been one of the major epidemic diseases for the past decades and it is still currently the major cause of deaths in many African countries. The introduction of antiretroviral treatment (ART) has been a positive development with regards to improving survival prospects of HIV patients. The purpose of this study is to implement the censored quantile regression model and use the results to explore and characterize the effect of demographic factors at different lower quantiles of the HIV patients survival time. A retrospective cohort study was used in this study. The target population was all adults in Bulawayo who started ART in 2004. Medecins Sans Frontieres (MSF) Bulawayo, Zimbabwe, was purposely approached for the cohort type data suitable for this study and after receiving the anonymised patient database; the researchers selected the variables of interest and selected a final sample of 1134 patients based on the inclusion criterion. Censored quantile regression modelling was used to estimate the quantile estimates for each demographic and clinical factors at the first five chosen quantile points. These five quantile points were selected such that they cover the defined shorter survival time of first year on ART treatment. The results show that all the risk factors have different effects at different quantile points. In conclusion, the use of censored quantile regressions revealed some patterns and hence rare insights in the effect of different risk factors on the patients' survival time.

7.2 Introduction

Survival analysis techniques are often used in clinical and epidemiologic research to model time until event data (Nakhaee & Law, 2011). Survival analysis entails methods that measure the risk of death or progression of a disease and it provides predictions that can help clinicians to estimate trends in their patient outcomes. Quantile regression has emerged as a powerful survival analysis method as it directly links the quantiles of patients' survival times to their demographic, clinical and serological profiles, facilitating the identification of important prognostic factors (Wang, Zhou, & Li, 2013). These methods also allow health planners to predict the HIV/AIDS burden on the health system and to allocate health services resources appropriately (Nakhaee & Law, 2011). Health care planning depends on good knowledge of prevalence, which requires an accurate understanding of survival patterns. Monitoring the length of survival after diagnosis is, therefore, an important component of the surveillance of HIV/AIDS. It provides a basis for evaluating individual prognostic factors (Assefa & Wencheko, 2012). The survival of HIV/AIDS patients depends on a variety of factors including but not limited to individual patient's demographic factors, serological baseline factors and presence of co-morbidities (Bellamy, 2007).

Long-term sustainable treatment is one choice for people living with HIV/AIDS. Not only can medication slow the progression of the infection, but it can also markedly suppress the virus, thereby restoring the body's immune function and permitting many HIV-infected individuals to lead a normal life and thus increase their survival time (Grover, Swain, & Ravi, 2014).

Even though HIV/AIDS drugs have become cheaper and more available because of a variety of government and private programmes, millions of others still do not have access to the drugs. The World Health Organisation (WHO) recommends that in resource-limited settings like Zimbabwe, Namibia and other African countries, HIV infected adolescents and adults should start antiretroviral treatment (ART) when the minimum conditions as in the guidelines are met (WHO, 2007).

With proper and timely use of antiretroviral therapy (ARVs) and good health care support for people living with HIV/AIDS, it has been noted worldwide that the survival time can be improved (Zachariah et al., 2006). The presence of opportunistic infections has however impacted negatively on this prospect (Mezzabotta, 2008). Some of the most lethargic opportunistic illnesses which can drastically reduce HIV/AIDS patients' survival time include Tuberculosis (TB), Pneumonia, and Hepatitis (Mezzabotta, 2008). Of these opportunistic illnesses, TB has been proven to be the most deadly when it combines with HIV/AIDS (Ngwerume, 2008). However, ART is known to dramatically improve the survival chances of patients living with HIV/AIDS (Zachariah et al., 2006).

Since April 2004, Zimbabwe established a National Opportunistic Infections/Antiretroviral Therapy (OI/ART) programme (Mutasa-Apollo et al., 2014). Early outcomes of patients initiated on ART in Zimbabwe and other developing countries show that despite ART, between 10 and 15% of individuals die within a median follow-up period of about 15 months (Harries, Schouten, & Libamba, 2006). Substantial proportions

(about 70%) of these deaths occur very early after starting ART. In the case of attrition, Mutasa-Apollo and co-researchers highlight that the high attrition occurring among HIV-infected patients in ART care in Sub-Saharan Africa has been widely documented, with patient retention declining from 86% at 6 months to 77% by 36 months after initiation (Mutasa-Apollo et al., 2014).

Several factors have been attributed to HIV and AIDS mortality after initiation of ART. In Vietnam, Cuong and co-researchers (Cuong et al., 2012) found that the predictive baseline factors for AIDS-related death were age > 35 years, clinical stage III or IV, body mass index (BMI) $< 18\ kg/m^2$, CD4 count $< 100/l$, haemoglobin level < 100 g/l, and plasma viral load $> 100,000$ copies/ml. Tuberculosis (TB) was the most common cause of death (40%). Zachariah *et al.,* (2006) discovered in their Malawi study that significant risk factors associated with mortality in the first 3 months and 6 months after ART initiation included WHO clinical stage IV disease, a CD4 cell count under 50cells/l and increasing grades of malnutrition (Zachariah et al., 2006). Mutasa-Apollo and co-researchers (Mutasa-Apollo et al., 2014) in their study in Zimbabwe found that gender (male sex), baseline WHO clinical stage IV were associated with an increased rate of attrition whilst baseline CD4 cell count < 50 cells/μl was associated with higher attrition as compared with CD4 cell count > 200cells/μl. Low weight was also observed as a risk factor for increase in attrition (Mutasa-Apollo et al., 2014).

Identifying the individual patients initiating in the ART programme who are at risk

of early deaths in Bulawayo, Zimbabwe would be useful for targeting potential interventions to prevent such deaths. Factors that are associated with early deaths in patients initiating on ART are not yet well characterised based on exploring the survival time distribution, because most of the current literature indicates the use of usual survival models which estimate fixed effects across the whole survival distribution. In this study we hypothesise that the risk factors impact the HIV patients' survival time differently from quantile to quantile of the survival time distribution. We thus implement the censored quantile regression (CQR) model for identifying and characterising the effects of risk factors at different survival time points.

7.3 Aim and objectives

The main purpose of this study is to use a censored quantile regression model in exploring the survival time distribution of HIV patients who initiated on ART in 2004 through characterising the prognostic risk factor effects at different quantile points within the first year of starting therapy.

The objectives for this study are to:

- Describe the patients' characteristics at initiation for the whole period of follow-up and;
- Characterize the effect of each of the risk factors for the different quantiles of the survival time distribution in the first year of ART initiation.

7.4 Materials and methods

7.4.1 A brief introduction to quantile regression models

Quantile regression (QR) is a common way to investigate the possible relationships between covariates, **X**, and a response variable, Y. Unlike the mean regression method that relies only on the central tendency of the data, the quantile regression approach allows the analyst to estimate the functional dependence between variables for all portions of the conditional distribution of the response variable. In other words, quantile regression extends the framework of estimating only the behaviour of the central part of a cloud of data points onto all parts of the conditional distribution (El Ghouch & Van Keilegom, 2009).

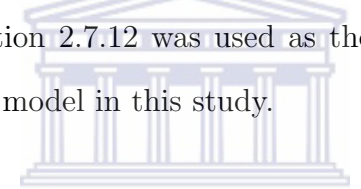
7.4.2 Quantile regression model parameter estimation

A censored quantile regression (CQR) model was estimated using the following quantiles, or τ values; $\tau = (0.05,\ 0.1,\ 0.15,\ 0.2,\ and\ 0.25)$. These τ values were chosen arbitrarily but making sure that they cover the survival time frame for deaths within the first year of ART initiation. The ordinary least squares method was not used for the estimation of parameters for this model because the CQR model assumes that the error term variances are not homogeneous but have the property of heteroscedasticity (Koenker & Bassett Jr, 1978). Because of this reason, linear programming methods are used to estimate the parameters of the linear conditional quantile regression model (Koenker & Hallock, 2001a).

Linear censored quantile regression (CQR) allows for semi-parametric estimation of

quantile regressions for a censored regression model (Koenker & Bassett Jr, 1978). CQR is robust against misspecification of the error term since only fairly weak assumptions on the error terms are required. CQR provides consistent estimates of the quantile regression coefficients β_τ, in the presence of fairly general forms of fixed right censoring.

The Monte Carlo bootstrap resampling method is used to estimate the variances and standard errors for the estimated parameters (Bilias, Chen, & Ying, 2000). The SAS Quantlife procedure (SAS Institute Inc, 2013) was used in the estimation of these parameters. Equation 2.7.12 was used as the objective function for the censored quantile regression model in this study.



7.4.3 Data source, design and sample

Anonymous secondary data for this research were provided by MSF. MSF's ART patient follow-up database was used in capturing the follow-up data at every patient's visit and this database was shared with the researcher courtesy of the MSF Monitoring and Evaluation team after the research was endorsed by the Ministry of Health in Zimbabwe. The data suppliers had no role in the data manipulation, analysis and decision to submit the manuscript for publication. Furthermore, all individuals who qualified for ARVs according to WHO guidelines after diagnosis were entered into the database and at every visit to the hospital the database was updated with the current patient information, that is: CD4 count; weight and weight gain; and next date for visiting.

The database had some patients who started on ART well before 2004 but for inclusion into this cohort study sample, a patient had to be alive on 01 January 2004 and receiving ARVs from MSF hospital in Bulawayo or started ART treatment at the same hospital in 2004 but before 1 January 2005. A retrospective cohort study design was used and a sample of 1134 HIV patients on ART followed for almost 6 years to July of 2010 from 2004 was considered.

Variables of interest

The data were exported to MS Excel from the MS Access database. The survival time and a censoring indicator were generated from the data. The censoring variable was used to identify those patients whose event time, that is, time to death was observed and also those whose event time was not observed due to loss to follow-up and the study ending. The risk factors include the baseline variables such as weight category at first initiation (weight above 60kgs, weight between 45kg and 60kgs, and weight lower than 45kgs), age category at initiation (age above 50 years, age between 35 years and 50 years, and age lower than 35 years, that is, (15-35 years)), gender (male and female), profession as at first initiation (unemployed, formally employed, self-employed, and other), marital status at initiation (single, married, widowed, and divorced) and WHO stage at initiation (WHO stage: unspecified, stage I, stage II, stage III, and stage IV). Also see Table 3.2.

Ethics clearance

Ethical clearance was obtained from the Ministry of Health and Child Welfare through the HIV/TB Research unit, (See Appendix A.1), and approval for data collection

was granted by the MSF Operational research unit. As the study was done using secondary data, individual patients' consent was not necessary since the database only captured patient identification numbers and not names or any other information that may directly be linked to the patient. The data was also not shared with any third party hence confidentiality was maintained.

Analysis

Descriptive statistics for the sample data were done and then a censored quantile regression model was fitted on the survival times to explore the risk factors' effect at different quantiles of patients' survival time in the early periods after initiation.

7.4.4 Results and analysis

A single data set obtained from MSF was used in the estimation of the censored quantile regression model (CQR). Some descriptive statistics for the patients' survival times were done to ascertain median survival times for different categories of the risk factor variables for the entire period of follow-up study. The log-rank tests were also done to assess differences in survival probabilities for patients who initiated in different levels of the risk factor variables throughout the entire study period. The CQR model was then estimated for shorter survival times, that is, up to the 0.25 quantile and after its estimation, it was used to do an analysis on the risk factor effects in different quantiles of the survival time distribution for patients on ART. The different risk factors which were investigated include gender, profession, marital status, age category, WHO stage and weight category at initiation. All these factors are baseline risk factors for HIV/AIDS mortality.

Descriptive analysis

Table 7.1 shows the sample sizes in each category, median survival times and log rank test p-values. Table 7.1 shows that the male survival time for the sample is lower

Table 7.1: Median survival times and log-rank tests for risk factor variables for entire period of follow-up

Variable	Level	n(%)	Median (days)	Log-rank
WHO stage	Not classified	15(2)	70	
	Stage 1	146(13)	1386	
	Stage 2	251(22)	1394	
	Stage 3	365(32)	1374	
	Stage 4	357(31)	1309	p=0.000*
Gender	Male	359(32)	1322	
	Female	775(68)	1372	p=0.0004*
Profession	Unemployed	543(48)	1128	
	Formally employed	352(31)	1706	
	Self employed	190(17)	1387	
	Other	45(4)	1121	p=0.0026*
Marital status	Single	207(18)	1689	
	Married	545(48)	1691	
	Widowed	298(26)	1719	
	Divorced	84(8)	1104	p=0.0324*
Age category	Lower than 35 years	405(36)	1843	
	Between 35 and 50 years	626(55)	1677	
	Above 50 years	103(9)	985	p=0.0094*
Weight category	Lower than 45kg	107(9)	1148	
	Between 45 and 60kg	553(49)	1775	
	Above 60kg	474(42)	1436	p=0.000*

*significant at 5% level

than that of females. The median survival time for males is 1322 days (3.6 years), which is less than the overall median survival time for females at 1362.5 days (3.7 years). The table also shows that the median survival time for HIV/AIDS patients

who initiated onto ART with clinical stage I is 1386 days (3.79 years) whilst that for those who initiated with clinical stage II is 1394 days (3.82 years). Patients who started ART with clinical stage III and IV had median survival times of 1374 days (3.76) and 1309 days (3.58 years) respectively. The median survival time is lower for HIV/AIDS patients who start ART in clinical stage IV. The patients who started ART in the age group above 50 years had a lower median survival time compared to those who initiated in the age groups lower than 50 years. Also patients who initiated on ART with weight below 45kgs had a lower survival time compared to those who started with a weight in categories above 45kgs.

With respect to patient profession and marital status at initiation, the formally employed had higher median survival times overall compared to those who initiated in any other stated profession. The widowed patients had the largest median survival time compared to those patients who started ART as single, divorced or married.

Overall, there were significant differences in the survival probabilities or prospects (as reflected in the log rank tests p-values) for patients who started ART in different WHO clinical stages (p=0.000), age categories (p=0.0094), weight categories (p=0.000), gender (p=0.0004), profession (p=0.0026) and marital status (p=0.0324).

Defining shorter survival times

To determine the effects of these different prognostic risk factors in early deaths (shorter survival times) we considered the patients' first year on treatment. The first year in this study constitutes the first 365 days (approximately 25th percentile of

the patients' survival time in days) after initiation on ART. The identification of prognostic factors that are associated with these shorter survival times helps in the designing of care programmes and intervention measures that can then prolong ART patients' survival times. Adding the dimension of the characterisation of these prognostic factors across the survival time quantiles will highlight some subtle trends in the risk factors' effect at different points of the survival time. Table 7.2 and Figure 7.1 highlight the period under consideration. Table 7.2 shows that 10% of the people

Table 7.2: Percentiles for patient survival time (n=1134)

	Percentiles and corresponding survival times								
Percentiles (%)	10	20	30	40	50	60	70	80	90
Survival time (days)	42	192	584	1116	1362.5	1448	1558	2075	2196

who initiated on ART at the hospital only managed up to 42 days of survival, 20% were dead by the 192nd day whilst the median survival time is 1362.5 days. This indicates that 50% of the people who initiated with HIV/AIDS on ART survived for approximately four (4) years.

Figure 7.1 shows the distribution of survival times for the 1134 ART patients who were followed from 2004 to July 2010. The highlighted section of early deaths constitutes the deaths that occurred in the first year of ART initiation. In this study, we use quantiles 0.05, 0.1, 0.15, 0.2 and 0.25 in the censored quantile model for assessing the risk factor effects in the early days after ART initiation. Figure 7.1 read together with Table 7.2 highlights that the one-year period of follow-up for the implementation of the censored quantile regression model is around the 25th percentile which is the 0.25 quantile point. The next section fits the censored quantile regression model on

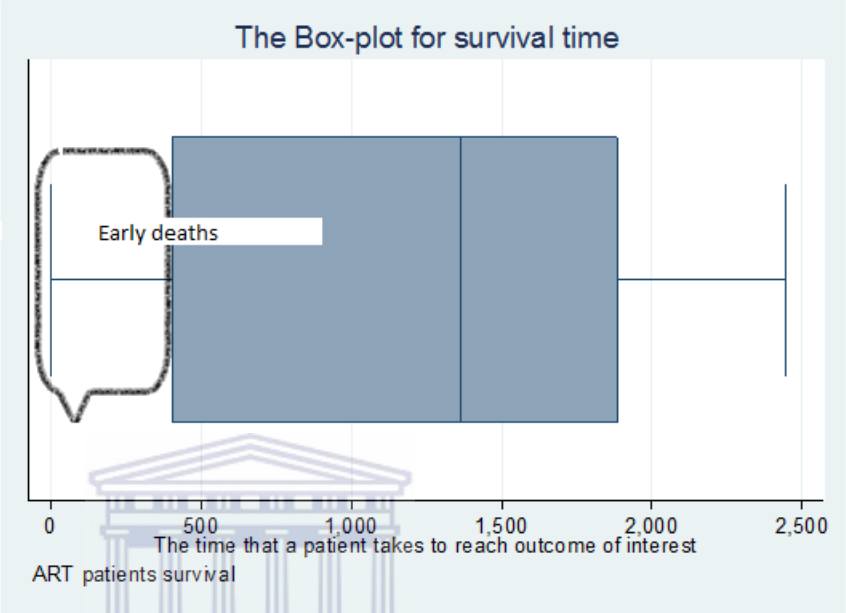


Figure 7.1: The box-plot for survival time (in days) showing the early deaths

the adult MSF Bulawayo dataset up to the 0.25 quantile point.

Assessing the risk factor effects using censored quantile regression model

Table 7.3 below shows the estimates of the effects of different risk factors at different quantile levels. This table shows that initiating in clinical stage I had a positive increasing effect on survival time compared to initiating in stage IV. The effect was significantly different from initiating in WHO stage IV at the 0.15 quantile (p=0.0439), at the 0.2 quantile (p=0.0193) and at the 0.25 quantile (p=0.0085). Starting ART in WHO stage II also had a positive effect on patient survival time which became significant at the 0.25 quantile (p=0.0036). Moreover, being formally employed had a

Table 7.3: CQR parameter estimates for quantiles within first-year of ART treatment

Risk factor variables	Parameter	τ_s (Quantile levels)				
		0.05	0.1	0.15	0.2	0.25
WHO stage	Intercept	4.3	-22.6	-33.3	37	66.9
	Unspecified	-0.3	-16.9	-34.5	-186.5	-124.5
	WHO stage I	-8.7	14	217.3*	302*	355*
	WHO stage II	-7.7	43.4	91.3	167.7	378.1*
	WHO stage III	0.7	14	44.5	87.3	158.4
	WHO stage IV	Ref	-	-	-	-
Profession	Employed	11.3	41.6	113.5*	208.5*	215.5*
	Other	99.3	340.6	491*	497.5*	736.3*
	Self-employed	8.3	34.8	47.8	104.7	54.1
	Unemployed	Ref	-	-	-	-
Age category	Above 50 yrs	-5.7	-21.7	-87	-140.7*	-243.6*
	35 to 50 yrs	4.3	12.5	-8.5	-5.7	5.5
	Lower than 35 yrs	Ref	-	-	-	-
Weight category	Above 60kg	22.7*	72.2*	158.8	260.2*	363.1*
	45 to 60kgs	10.3	39.1	76.5	142.3*	213.3*
	Lower than 45kg	Ref	-	-	-	-
Gender	Female	6	20.2	69.8	82.8	130.5
	Male	Ref	-	-	-	-
Marital status	Divorced	-9.3	3.4	-21	-118.8	-196.4
	Married	-4.3	-11.7	-30.5	-149.5*	-167.6*
	Single	-6.7	-12.3	-60.5	-188.5*	-262.2*
	Widowed	Ref	-	-	-	-

*significant at 5% level

positive effect on survival time of patients compared to being unemployed. The effect became increasingly significant from the 0.15, 0.2 and 0.25 quantile with p-values 0.0318, 0.0026 and 0.0116 respectively. Initiating in the age group above 50 years was associated with having a negative effect on survival time compared to the age group below 35 years. The effect was significant at the 0.2 quantile (p=0.0343) and at the 0.25 quantile (p=0.0085).

Furthermore, starting ART with weight above 60kgs was associated with a significant positive effect within the first year period in comparison to starting with a weight below 45kgs. Starting with a weight between 45 and 60kgs was only significantly different from initiating with weight below 45kgs towards the end of the first year period on treatment. In addition, being female on ART was associated with a positive effect on survival time compared to being male, this effect is however not significant. The state of being married or single was negatively associated with survival time compared to those who initiated as widows. The effects of each of these variables as well as the level of effect across the different quantiles of the survival time distribution are highlighted in the following figures.

Figure 7.2 below presents the effects of each employment category in comparison to the state of being unemployed. The formally employed plot suggests that the effect of being formally employed at initiation of ART on the survival time is positive and gradually increasing in relation to initiating in the state of being unemployed. Being self-employed also follows the same pattern but its not significantly different from being unemployed. Both these effects are not constant but vary from quantile to quantile.

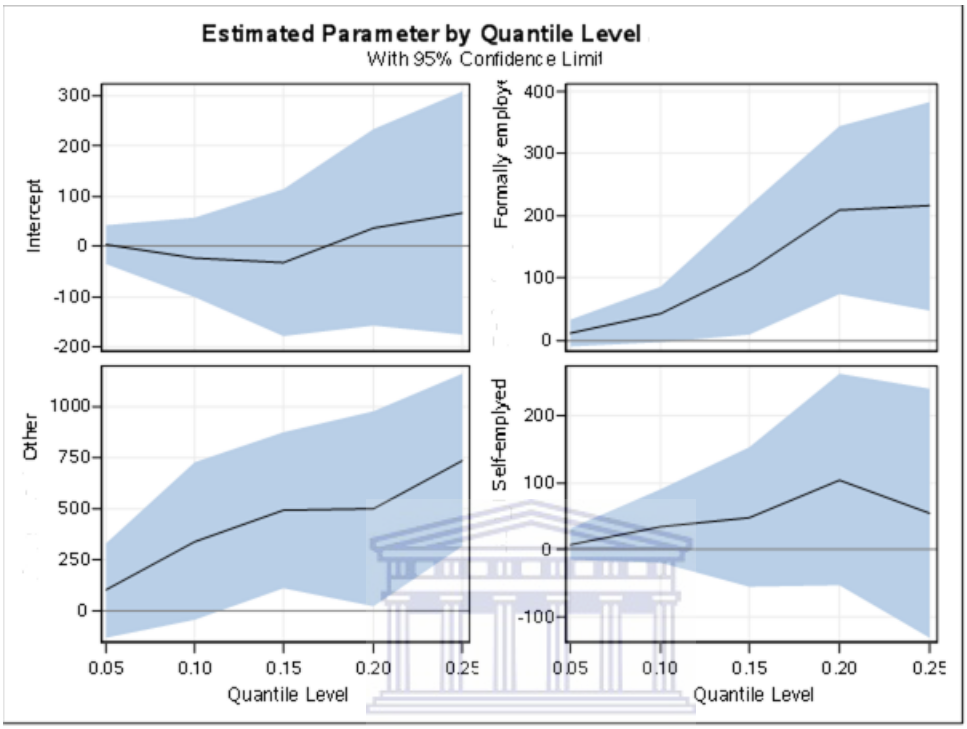


Figure 7.2: Estimated parameter effects by quantile level for survival time for profession (reference: unemployed)

Figure 7.3 below presents the effects of age categories and weight categories in comparison to lower than 35 years age group and weight lower than 45kg respectively on survival time within the first year of antiretroviral treatment. Starting ART with age above 50 years is associated with a negative effect on survival time and this negative effect is slightly stronger in the upper tails, that is, towards the end of the first year on treatment. Initiating in the age group between 35 and 50 years has a constant effect across all quantiles and the effect is not significantly different from that of initiating with age group lower than 35 years. Having weight above 60kgs has a significant positive effect on survival time across all quantiles of the first year treatment period

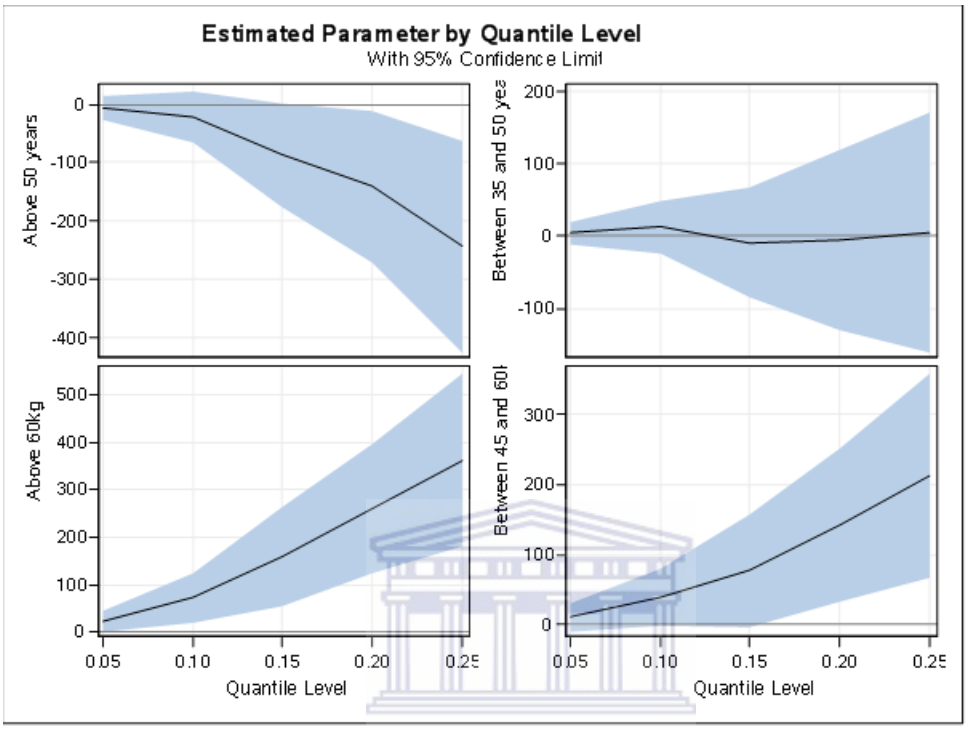


Figure 7.3: Estimated parameter effects by quantile level for survival time on age (reference: below 35 years) and weight (reference: below 45kgs) categories

and the effect becomes stronger towards the end of this period.

Figure 7.4 below presents the effect of initiating in different marital statuses compared to initiating in the state of being widowed. In the lower quantiles, the effect on patient survival time due to all the different states is relatively constant and not significantly different from the widowed patients but becomes more negative in the upper quantiles. Being a female patient on ART had a positive effect on the survival time compared to being male. In addition, there is a larger positive effect in the upper quantiles.

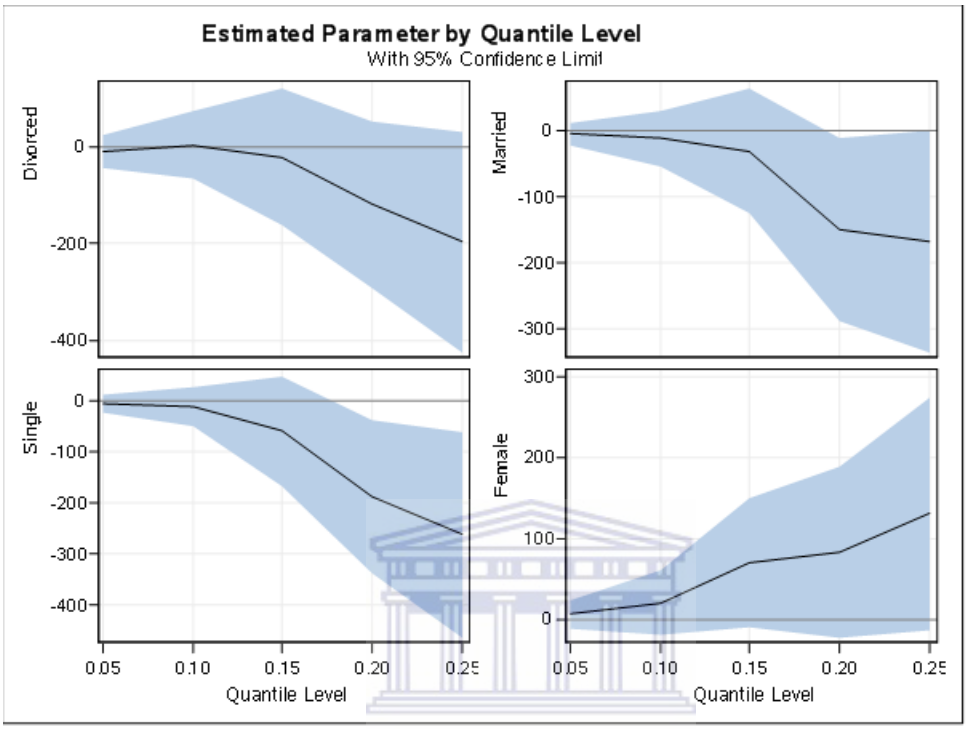


Figure 7.4: Estimated parameter effects by quantile level for survival time on marital status (reference: Widowed) and gender (reference: Male)

Figure 7.5 below presents the parameter quantile function estimates for WHO staging. The figure shows that there is a progressively negative effect to the unclassified group of patients as compared to those who are classified as being in WHO stage IV. The effect is non-significant initially but it becomes significantly different as survival time increases towards the 25th percentile. There is also a significant positive effect of being in WHO stage I as compared to being in WHO stage IV. For the WHO stage II, there is an initial insignificant positive effect on survival time but the positive effect becomes larger with increasing survival time and it becomes significant as compared

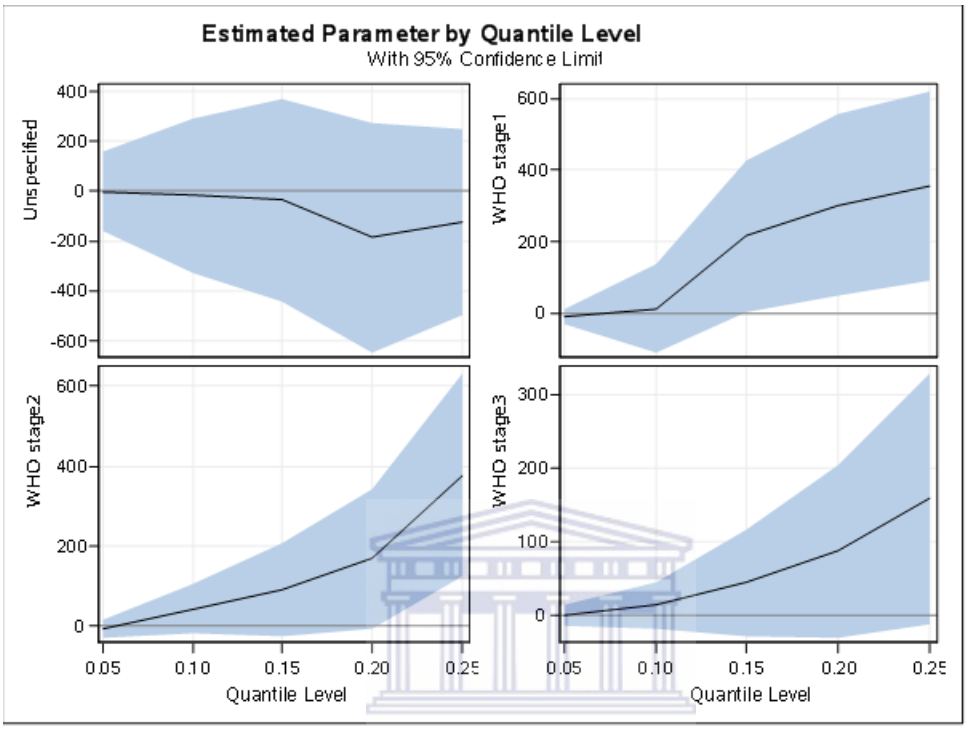


Figure 7.5: Estimated parameter effects by quantile level for survival time on WHO staging (reference: WHO stage IV)

to WHO stage IV effects around the 0.25 quantile. Moreover, there is no significant effect difference between WHO stage III and WHO stage IV effects at all quantiles.

These results show that regardless of the fact that those patients receive the same treatment; their survival time will be strongly affected by the baseline weight category and WHO clinical stages as well as age group category and profession. The effects of these factors are not constant over time reflecting on the need to put correct interventions at different points of patients' survival on treatment. As noted from the analysis above, being in a certain category of age, weight, clinical stage, gender

or marital status is associated with different varying effects on patients' survival time at different quantile points within the first year period of treatment.

7.4.5 Discussion

This study determined that being female patient was associated with a positive effect on survival time, implying that being male was a risk factor for reduced survival time in the first year of ART treatment. This result relates with a national study observation from Mutasa-Apollo *et al.,* (2014) who concluded that the male sex is a risk factor for increasing the attrition rate of this cohort of people. This is also supported by findings from Koenig *et al.,* (2009) which indicated that males with HIV/AIDS and on ART have a shorter survival time as compared to their female counterparts (Koenig et al., 2009).

Mutasa-Apollo *et al.,* (2014) concluded from their national study that baseline weights of 45 to 60kgs and less than 45kgs had an increased risk of attrition as compared to patients with weights greater than 60kgs (Mutasa-Apollo et al., 2014). It is interesting to note that from this study, it could also be shown that the weight category of above 60kgs at initiation, has a positive significant effect on survival time across the first-year ART treatment survival time distribution, highlighting some additional critical information that may help in designing interventions. This effect is more pronounced towards the end of the first year of treatment compared to the beginning.

The effect of being in the age group above 50 years at initiation compared to being below 35 years is significantly negative towards the end of the first year treatment

phase but not significant in the first few months after starting on treatment.

For early deaths, that is, deaths within the first year of ARV uptake, WHO stage 1 has a positive effect on survival time as compared to WHO stage 4. The effect is not significant during the early phase after starting on ART, implying that at the start of treatment, HIV patients in different clinical stages are all equally exposed. A significant positive effect of being in clinical stages 1 and 2 was evident from approximately the middle of the first year to the end of first year on treatment. This result is consistent with many other studies (Mutasa-Apollo et al., 2014; Zachariah et al., 2006) which concluded that patients who start ART in the early stages of HIV have a greater chance of surviving longer on treatment compared to those who start in the later stages of HIV.

7.4.6 Conclusion

In conclusion, this study shows that overall the results are similar to those observed by other scholars. However, when considering early deaths and using quantile regression models, it was found that the effect of different risk factors on the survival time during the time period of interest varies from point to point of the survival time. It can be deduced from this study that additional information can be gained from using quantile regression models than just concluding with the traditional methods of analysing survival time of patients on therapy. Therefore these results might help in finding appropriate measures against different risk factors at different points of patients' survival times during their uptake of antiretroviral drugs.

The next chapter, that is, Chapter 8, summarizes the conclusions from different research objectives that were stipulated in Chapter 1 and the chapter also discusses the contributions of this study to the literature on survival modelling and analysis.

Chapter 8

General discussion and conclusions

This thesis reflected on the general information about HIV/AIDS care through ART programmes, the literature on survival analysis and modelling as well as using new datasets from Zimbabwe and Namibia.

Throughout this thesis, survival analysis and modelling using different methods such as the Kaplan-Meier survival estimators, log-rank tests, the proportional hazards, the Bayesian proportional hazards and the censored quantile regression modelling techniques were presented. The general thrust of this thesis was to highlight the important insights that can be gained by using the novel censored quantile regression procedures in survival modelling. The strategy for this exploration was to start off with the usual basic analysis of survival data and then to fit the proportional hazards model, identify prognostic risk factors and finally to make inferences. The next step was to fit a Bayesian proportional hazards model to the same dataset as well as a censored quantile regression model to identify risk factors. Inferences were then made in terms of making comparisons to insights that were obtained in applying each of the modelling procedures. Having noted the important insights that were obtainable

from using censored quantile regressions, this modelling framework was then applied to another dataset to identify risk factors for shorter survival times. The following questions based on this thesis' objectives were responded to as follows:

What are the risk factors for survival of HIV/AIDS patients on antiretroviral treatment using the Cox's proportional hazards models?

From Chapter 4, only age category was identified as a significant risk factor for paediatric and adolescents on ART at 5% significant level. All the other variables were not significant. Being an infant was found to be associated with a significant negative effect on survival time. A large study in Zambia showed the same outcome (Bolton-Moore et al., 2007).

What are the risk factors for survival of HIV/AIDS patients on antiretroviral treatment using the Bayesian proportional hazards model and censored quantile regression models? Can more information be gained from these models on how risk factors affect patient survival at different points of the survival time distribution?

The Bayesian proportional hazards and censored quantile regression models were then implemented on the data. The results from the Bayesian model indicated that not being an infant had a positive significant effect on survival time compared to being an infant. Also initiating treatment in clinical stage II was determined to have a positive significant effect on survival time compared to initiating in clinical stage

IV. The Bayesian model also revealed that the effect of the different risk factors on survival time was not fixed across the survival time distribution. The effect was different for the different quartiles. From the censored quantile regression model, it was determined that initiating in clinical stage II had a significant positive effect on survival time during the early periods of initiation compared to clinical stage IV. Patient gender was also shown from the CQR model as having a significant effect during the early periods of initiation but not significant at any other point during the follow-up period. Being female was shown to have a significant negative effect on survival time during these early periods of initiation on treatment. These results highlighted the insights that can possibly be gleaned from using CQR models in modelling survival data with intentions to identify risk factors for survival.

What are the risk factors for shorter patient survival times of HIV/AIDS patients on antiretroviral treatment using the censored quantile regression models? Are the risk factor effects constant over the period?

Chapter 7 applied the CQR model to determine the risk factors for shorter survival times in patients on ART treatment. Results showed that the effect of risk factors was varied at different points of the first year ART treatment survival time distribution. For this Zimbabwean adult cohort, it was determined that being a female patient was associated with a positive insignificant effect on survival time within the first year of ART treatment. Weight category of above 60kg at initiation was shown to have a positive significant effect on survival time across the first year ART treatment

survival time distribution compared to weight category of below 45kgs. For the determination of risk factors that are associated with early deaths, the CQR highlights that some risk factors, such as weight category, are significant towards the end of the first year ART treatment period. This result is consistent with many other studies especially relating to which factors were significant in causing deaths to patients on ART after three or six months (Mutasa-Apollo et al., 2014; Zachariah et al., 2006).

How does the Cox's proportional hazards model parameter estimate effects compare to censored quantile regression model estimate effects?

Chapter 6 fitted both the Cox proportional hazards model and the censored quantile regression model over the survival time distribution to assess and compare the characterisation of the risk factors across the survival time distribution on the paediatric dataset. The coefficient effects were shown to vary remarkably when the censored quantile regression model was used as compared to the proportional hazards model. The Cox's proportional hazards model indicated that the risk factor effects are almost constant throughout the survival time distribution. The censored quantile regression model indicated that a risk factor can be significant at some quantile points in the patients' survival time distribution and not significant in others. This result is similar to what Koenker and Billias (2001) found in their study on reappraisal of the Pennsylvania reemployment bonus experiments (Koenker & Bilias, 2001). The implication of this result is that the general proportional hazards model, despite retaining the applause from many researchers in the field, fails to capture the dynamic effects of the risk factors throughout the survival time distribution.

Findings, conclusions and limitations

The thesis showed that when quantile regression models are used in modelling survival data, in this case, the HIV/AIDS survival times for patients on ART, the effect of different risk factors on survival times for patients are shown to vary from one time point to the other. Whereas the Bayesian proportional hazards model implies to this dynamism, by estimating the hazard rates at quartiles, also highlighting that the estimates are not constant or fixed throughout the survival time distribution, the CQR models go a step further to highlight significant risk factors at each chosen time point.

As this study has shown, where the proportional hazards model could only indicate that age category was significant in affecting survival times, the CQR model showed at what point of the survival time distribution were these risk factors significant. The CQR model also highlighted the dynamic nature by which the risk factors can affect survival times through showing that the effect of a risk factor can be significant in the early periods of initiation and yet after a certain period of continuous treatment, the risk factor ceases to affect survival time in a significant manner. The advantage of such an insight is that decision makers can plan based on the pattern of risk factor effectiveness, thereby helping positively those that are on treatment to aid in their quest to survive longer.

This thesis clearly shows that additional information is gained from using censored regression models in modelling survival data. This additional information can be

invaluable to public health practitioners working in the health care delivery sector for people living with HIV/AIDS. Koenker *et al.,* (2008) in their work suggest that censored quantile regression can be a more flexible alternative to the proportional hazards model as it gives more information about how risk factors impact survival time (Koenker, 2008).

The contributions of this thesis include a clear illustration of the merits of using CQR models in survival analysis and modelling when identifying risk factors for patient survival on treatment. Using these methods enables an exploration of the dynamics of risk factors throughout the survival time distribution. The thesis managed to show that the effect of risk factors on patient survival time was not constant throughout the follow-up period, that is, from initiation to the end of the study period. This is expected looking at the skewed nature of survival times data. The implication of this finding is that risk factors should not be treated as a constant risk throughout a patient's treatment period but attention should be directed towards limiting the effects at particular time points. The other significant contribution is that the CQR methods can be implemented to augment the proportional hazards model in a very effective and insightful way. These methods can also be used to analyse the effect of different risk factors at different points in causing early deaths.

From a public health perspective, this thesis posits that these methods and findings can be very useful and effective in designing care intervention programmes for patients on ART in Namibia and Zimbabwe as well as the region at large.

Furthermore, this study contributes positively to the literature on survival modelling by implementing the Bayesian proportional hazards and CQR models in assessing risk factor effects on ART patient survival times as well as utilizing the CQR models in determining prognostic risk factors for early deaths among people on ART treatment.

Recent work by Xue and co-researchers (Xue, Xie, & Strickler, 2016) suggest that the commonly used statistical model for studying time to event data, the Cox proportional hazards model is limited by the assumption of a constant hazard ratio over time (i.e., proportionality assumption), and the fact that it models the hazard rate rather than the survival time directly. The censored quantile regression model, defined on the quantiles of time to event, provides an alternative that is more flexible and interpretable. This thesis has highlighted the strengths of censored quantile regression models in relation to determining prognostic risk factors for patients that are on ART treatment.

A few challenges and limitations were noted in this study and these include: The reliance on secondary data meant that the study had to focus on the available collected data rather than the initially anticipated. This was a limitation on assessing the effect of opportunistic infections for people on ART treatment. Most hospital records were not complete hence the randomisation process at design of the study had to be altered in order to find the appropriate records to include in the study. Variables such as baseline Body Mass Index (BMI), TB status and viral load were dropped from the study due to heavy missingness.

Further research could look at implementing the CQR models to prospective cohort designs as well as in clinical trials where the researcher has some semblance of control on the measurement of variables as well as monitoring of subjects. The development of literature on determining hazard rates at each quantile point in the survival time distribution is another area of research that is still ongoing and needs focus. Development of adequate validation methods for the models is also another area that needs attention.

References

Allison, P. D. (2010). *Survival analysis using SAS: a practical guide. (2^{nd} ed.)*. Cary, NC: Sas Institute.

Andrinopoulou, E.-R. (2014). *Introduction to joint models of longitudinal and survival data.* Retrieved from `http://www.repub.eur.nl/pub/77159/141118` (Doctoral thesis, Erasmus University Medical Center, Department of Biostatistics)

Assefa, T., & Wencheko, E. (2012). Survival analysis of patients under chronic hiv-care and antiretroviral treatment at tikur anbessa specialized hospital, addis ababa, ethiopia. *Ethiopian Journal of Health Development*, *26*(1), 22–29.

Bellamy, R. (2007). Managing Tuberculosis in People with HIV. *British Medical Journals*, *15*(2), 20-35.

Bilias, Y., Chen, S., & Ying, Z. (2000). Simple resampling methods for censored regression quantiles. *Journal of Econometrics*, *99*(2), 373-386.

Bolton-Moore, C., Mubiana-Mbewe, M., Cantrell, R. A., Chintu, N., Stringer, E. M., Chi, B. H., ... others (2007). Clinical outcomes and cd4 cell response in children receiving antiretroviral therapy at primary health care facilities in zambia. *Jama*, *298*(16), 1888–1899.

Bong, C.-N., Yu, J. K.-L., Chiang, H.-C., Huang, W.-L., Hsieh, T.-C., Schouten, E. J., ... Harries, A. D. (2007). Risk factors for early mortality in children

on adult fixed-dose combination antiretroviral treatment in a central hospital in malawi. *AIDS*, *21*(13), 1805-1810.

Chen, C. L. (2013). *An introduction to quantile regression and the quantreg procedure*. Cary, NC: SAS Institute Inc.

Chernozhukov, V., & Hong, H. (2011). Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association*, *97*, 1-11.

Collett, D. (1994). *Modelling survival data in medical research*. London: Chapman and Hall.

Collett, D. (2015). *Modelling survival data in medical research*. London, UK: CRC press.

Cox, D. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, *34*, 187-220.

Cox, D., & Oakes, D. (1984). *Analysis of survival data*. London, UK: Chapman and Hall.

Cuong, D. D., Thorson, A., Sönnerborg, A., Hoa, N. P., Chuc, N. T. K., Phuc, H. D., & Larsson, M. (2012). Survival and causes of deaths among hiv-infected patients starting antiretroviral therapy in north-eastern vietnam. *Scandinavian Journal of Infectious Diseases*, *44*(3), 201-208.

El Ghouch, A., & Van Keilegom, I. (2009). Local linear quantile regression with dependent censored data. *Statistica Sinica*, *1*(1), 1621-1640.

Fitzenberger, B., & Wilke, R. A. (2005). Using quantile regresion for duration analysis. *Springer*, *2*(3), 103-118.

Fox, J. (2008). Cox proportional-hazards regression for survival data. *Appendix: R and S-PLUS Companion to Applied Regression*, *12*(4), 1-18.

Frisch, R. (1934). Statistical confidence analysis by means of complete regression systems. *Oslo*, *48*, 749-755.

Gibb, D. M., & Giaquinto, C. (2000). Children with hiv infection: special cases. *Lancet: The Lancet Perspectives*, *356*, S34.

Gilchrist, W. (2001). *Statistical modelling with quantile functions*. London, UK: Chapman and Hall.

Gonese, E., Dzangare, J., Gregson, S., Jonga, N., Mugurungi, O., & Mishra, V. (2010). Comparison of hiv prevalence estimates for zimbabwe from antenatal clinic surveillance (2006) and the 2005-2006 zimbabwe demographic and health survey. *PLoS one*, *5*(11), e13819.

Grover, G., Swain, P. K., & Ravi, V. (2014). A competing risk approach with censoring to estimate the probability of death of hiv/aids patients on antiretroviral therapy in the presence of covariates. *Statistics Research Letters*, *3*, 7-16.

Harries, A. D., Schouten, E. J., & Libamba, E. (2006). Scaling up antiretroviral treatment in resource-poor settings. *The Lancet*, *367*(9525), 1870-1872.

Ibrahim, J. G., Chen, M.-H., & MacEachern, S. N. (1999). Bayesian variable selection for proportional hazards models. *Canadian Journal of Statistics*, *27*(4), 701–717.

Kedir, A. A., Desta, A., & Fesseha, G. (2014). Factors affecting survival of hiv positive children taking antiretroviral therapy at adama referral hospital and medical college, ethiopia. *Journal of AIDS and Clinical Research*, *2014*, 1-6.

Kerina, D., Babill, S.-P., & Muller, F. (2013). Hiv/aids: The zimbabwean situation and trends. *American Journal of Clinical Medicine Research*, *1*(1), 15-22. Retrieved from `http://pubs.sciepub.com/ajcmr/1/1/5`

Koenig, S. P., Riviere, C., Leger, P., Joseph, P., Severe, P., Parker, K., ... Fitzgerald, D. W. (2009). High mortality among patients with aids who received a diagnosis of tuberculosis in the first 3 months of antiretroviral therapy. *Clinical Infectious Diseases*, *48*(6), 829-831.

Koenker, R. (2005). *Quantile regression* (No. 38). Cambridge, UK: Cambridge University Press.

Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software*, *27*(6), 1–25.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, *46*, 33-50.

Koenker, R., & Bilias, Y. (2001). Quantile regression for duration data: a reappraisal of the pennsylvania reemployment bonus experiments. In *Economic applications of quantile regression* (p. 199-220). Springer.

Koenker, R., & d'Orey, V. (1985). Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *36*(3), 383–393.

Koenker, R., & Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association*, *96*(454), 458-468.

Koenker, R., & Hallock, K. (2001b). Quantile regressions: An introduction. *Journal of Economic Perspectives*, *1*(4), 43-56.

Koenker, R., & Hallock, K. F. (2001a). Quantile regression. *Journal of Economic Perspectives*, *15*(4), 144-146.

Koopmans, T. C. (1937). *Linear regression analysis of economic time series*. Cornell University: De erven F. Bohn.

Leng, C., & Tong, X. (2013). A quantile regression estimator for censored data. *Bernoulli*, *19*(1), 344–361.

Leroy, V., Malateste, K., Rabie, H., Lumbiganon, P., Ayaya, S., Dicko, F., ... Aka, E. (2013). Outcomes of antiretroviral therapy in children in asia and africa: a comparative analysis of the iedea paediatric multiregional collaboration. *Journal of Acquired Immune Deficiency Syndromes*, *62*(2), 208-209.

Lin, G., & Rodriguez, R. N. (2013). Using the quantlife procedure for survival analysis. In *Proceedings of the SAS Global Forum 2013 Conference, Cary, NC: SAS Institute Inc. Available at http:/support. sas. com/resources/papers/proceedings11/421-2013. pdf.*

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. NY, USA: John Wiley and Sons.

Mbizvo, M. T., Mashu, A., Chipato, T., Makura, E., Bopoto, R., & Fottrell, P. (1996). Trends in hiv-1 and hiv-2 prevalence and risk factors in pregnant women in harare. *The Central African Journal of Medicine*, *42*(1), 14-21.

Meldrum, A. (2008). Zimbabwe's health care system struggles on. *The Lancet*, *371*, 1059-1060.

Mezzabotta, G. (2008). A global and regional perspective of hiv and tb co-infection and related morbidity. *WHO Report*, *1*, 89-129.

MoHCW (Ed.). (2013). *Zimbabwe HIV estimates* (Tech. Rep.). Retrieved from `http://www.nac.org.zw`

MoHSS. (2008). *HIV/AIDS in Namibia* (Tech. Rep.). Retrieved from `http://data.unaids.org/pub/Report/2008/20080915`

MoHSS. (2013). *HIV/AIDS in Namibia: Behavioral and Contextual Factors Driving the Epidemic* (Tech. Rep.). Retrieved from http://www.safaids.net/files/HIVAIDS/Behavioural/Report

MoHSS-UNAIDS. (2011). *Hiv/aids in namibia: Behavioral and contextual factors driving the epidemic* (Tech. Rep.). Retrieved from http://www.safaids.net/files/HIVAIDS/Behavioural/Report

Mundi, I. (2015). Zimbabwe HIV/AIDS - adult prevalence rate. Retrieved from http://www.indexmundi.com/zimbabwe/hiv aids adult prevalence rate.html

Mutasa-Apollo, T., Shiraishi, R. W., Takarinda, K. C., Dzangare, J., Mugurungi, O., Murungu, J., ... Woodfill, C. J. (2014). Patient retention, clinical outcomes and attrition associated factors of hiv-infected patients enrolled in zimbabwe's national antiretroviral therapy programme, 2007-2010. *PloS one*, *9*(1), 1-9.

Nachega, J. B., Mills, E. J., & Schechter, M. (2010). Antiretroviral therapy adherence and retention in care in middle-income and low-income countries: current status of knowlwdge and research priories. *Current Opinion in HIV and AIDS*, *5*(1), 70-77.

Nakhaee, F., & Law, M. (2011). Parametric modelling of survival following HIV and AIDS in the era of Highly Active Antiretroviral Therapy: Data from Australia. *Eastern Mediterranean Health Journal*, *7*, 232-233.

Newell, M.-L., Coovadia, H., Cortina-Borja, M., Rollins, N., Gaillard, P., & Dabis, F. (2004). Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: A pooled analysis. *The Lancet*, *364*(9441), 1236-1243.

Ngwerume, M. (2008). The impact of tb and hiv co-infection. experiences of seke rural home based care. *Safaids files*, 15-16. Retrieved from `http://www.safaids.net/files/The deadly marriage -Mrs Ngwerume -Seke Rural Home Based Care.pdf`

Novitsky, V., & Essex, M. (2012). Using hiv viral load to guide treatment-for-prevention interventions. *Current Opinion in HIV and AIDS*, *7*(2), 117-124.

PEPFAR. (2012). *Namibia operational plan report- fy 2011* (Tech. Rep.). Retrieved from `http://www.pepfar.gov/documents/organization/199714.pdf`

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, *98*(464), 1001-1012.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, *25*(3), 303-325.

Rodrıguez, G. (2010). *Notes on Survival Models* (Tech. Rep.). Retrieved from `http://data.princeton.edu/wws509/notes/c7.pdf` (Lecture Notes, Princeton University, Department of Demography, & Statistics)

SAS Institute Inc. (2013). SAS Software. SAS 9.4 [Computer software manual]. Cary, NC. Retrieved from `http://www.sas.com/`

Satty, A. (2014). A simulation study comparing two methods of handling missing covariate values when fitting a cox proportional hazards regression model. *STATISTIKA*, *94*(1), 64-65.

Sinha, D., Chen, M.-H., & Ibrahim, J. G. (2003). Bayesian inference for survival data with a surviving fraction. *Lecture Notes-Monograph Series*, 117–138.

Stokes, M., Chen, F., & Gunes, F. (2014). An introduction to Bayesian analysis with SAS/STAT® software. In *Proceedings of the SAS Global Forum 2014*

Conference, SAS Institute Inc, Cary, USA (available at https:// support. sas. com/ resources/ papers/ proceedings14/ SAS400-2014. pdf).

Thorne, C., Little, K., , Luo, C., Bunders, M., Ngongo, N., ... Newell, M.-L. (2007). Disease progression in children with vertically-aquired HIV infection in Sub-Saharan Africa: Reviewing the need for HIV treatment. Current HIV Research. *Current HIV Research*, *5*(2), 139-153.

UNAIDS. (2013). *UNAIDS Report on the Global AIDS Epidemic* (Tech. Rep.). Retrieved from `http://www.unaids.org/sites/default/files/media asset/UNAIDSGlobal Report 2013en1.pdf`

UNAIDS. (2014). *HIV AND AIDS estimates (2014)* (Tech. Rep.). Retrieved from `http://www.unaids.org/en/regionscountries/countries/namibia`

UNICEF. (2011). *Children and HIV/AIDS in Namibia* (Tech. Rep.). Retrieved from `http://www.unicef.org/namibia/UNICEF Nam 2011 Children and AIDS Namibia.pdf`

Wang, H. J., & Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *The Annals of Statistics*, 756–781.

Wang, H. J., Zhou, J., & Li, Y. (2013). Variable selection for censored quantile regression. *Statistica Sinica*, *23*(1), 145-167.

WHO. (2004). *Patient monitoring guidelines for HIV care and antiretroviral therapy (ART)* (Tech. Rep.). Retrieved from `http://www.who.int/hiv/pub/guidelines/patientmonitoring.pdf`

WHO. (2007). *WHO case definitions of HIV for surveillance and revised clinical staging and immunological classifications of HIV related* (Tech. Rep.). Retrieved from `http://www.who.int/hiv/pub/guidelines/HIVstaging150307.pdf`

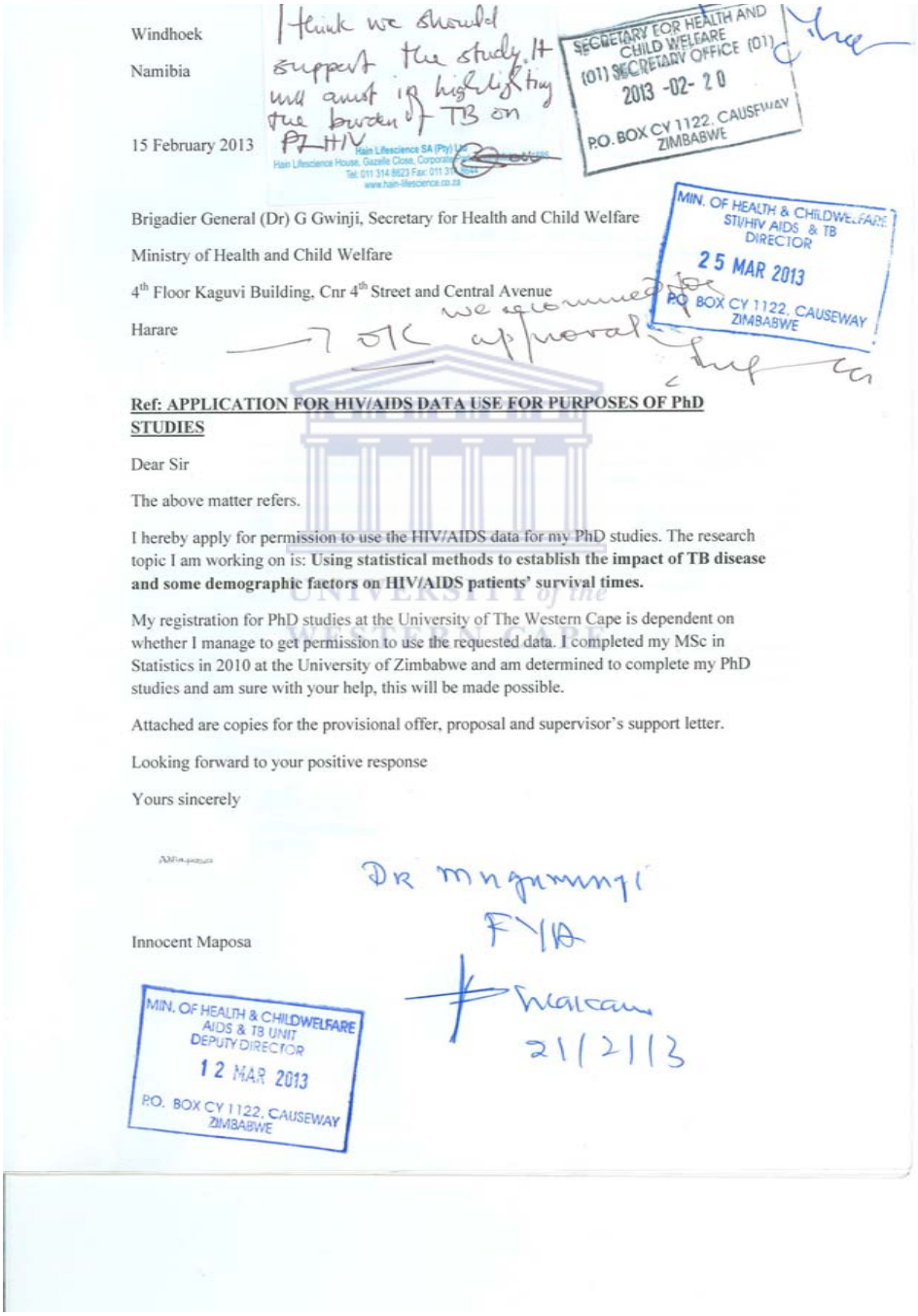
WHO. (2014). *Treatment of Children Living with HIV* (Tech. Rep.). Switzerland. Retrieved from `http://www.who.int/hiv/topics/paediatric/en/`

Xue, X., Xie, X., & Strickler, H. D. (2016). A censored quantile regression approach for the analysis of time to event data. *Statistical Methods in Medical Research*, *21*(6), 21-35.

Zachariah, R., Fitzgerald, M., Massaquoi, M., Pasulani, O., Arnould, L., Makombe, S., & Harries, A. D. (2006). Risk factors for high early mortality in patients on antiretroviral treatment in a rural district of malawi. *AIDS*, *20*(18), 2355-2360.

Appendix A

Approved ethical clearance letters



Windhoek

Namibia

I think we should support the study. It will assist in highlighting the burden of TB on PLHIV

15 February 2013

Hain Lifescience SA (Pty) Ltd
Hain Lifescience House, Gazelle Close, Corporate Park
Tel: 011 314 8623 Fax: 011 314 8637
www.hain-lifescience.co.za

SECRETARY FOR HEALTH AND CHILD WELFARE
SECRETARY OFFICE (01)
2013 -02- 20
P.O. BOX CY 1122, CAUSEWAY
ZIMBABWE

Brigadier General (Dr) G Gwinji, Secretary for Health and Child Welfare

Ministry of Health and Child Welfare

4th Floor Kaguvi Building, Cnr 4th Street and Central Avenue

Harare

MIN. OF HEALTH & CHILDWELFARE
STI/HIV AIDS & TB
DIRECTOR
25 MAR 2013
P.O. BOX CY 1122, CAUSEWAY
ZIMBABWE

we recommend for approval

→ 7 DIC

Ref: APPLICATION FOR HIV/AIDS DATA USE FOR PURPOSES OF PhD STUDIES

Dear Sir

The above matter refers.

I hereby apply for permission to use the HIV/AIDS data for my PhD studies. The research topic I am working on is: **Using statistical methods to establish the impact of TB disease and some demographic factors on HIV/AIDS patients' survival times.**

My registration for PhD studies at the University of The Western Cape is dependent on whether I manage to get permission to use the requested data. I completed my MSc in Statistics in 2010 at the University of Zimbabwe and am determined to complete my PhD studies and am sure with your help, this will be made possible.

Attached are copies for the provisional offer, proposal and supervisor's support letter.

Looking forward to your positive response

Yours sincerely

Innocent Maposa

Dr Mngqunngi
FYI
21/2/13

MIN. OF HEALTH & CHILDWELFARE
AIDS & TB UNIT
DEPUTY DIRECTOR
12 MAR 2013
P.O. BOX CY 1122, CAUSEWAY
ZIMBABWE

Figure A.1: Zimbabwe ethical clearance

9 – 0/0001

REPUBLIC OF NAMIBIA

Ministry of Health and Social Services

Private Bag 13198
Windhoek
Namibia

Ministerial Building
Harvey Street
Windhoek

Tel: 061 – 203 2510
Fax: 061 – 222558
E-mail: eshaama@mhss.na

OFFICE OF THE PERMANENT SECRETARY

Ref: 17/3/3
Enquiries: Ms. E. Shaama

Date: 29 October 2014

Mr. Innocent Maposa
Polytechnic of Namibia
School of Health and Applied Sciences
Private Bag 13388
Windhoek

Dear Mr. Maposa

Re: Survival Analysis of HIV/AIDS Patients on HIV Care and Antiretroviral Treatment to Determine Longevity Prognostic Factors in Namibia

1. Reference is made to your application to conduct the above-mentioned study.

2. The proposal has been evaluated and found to have merit.

3. **Kindly be informed that permission to conduct the study has been granted under the following conditions:**

3.1 The data to be collected must only be used for operational purpose;
3.2 No other data should be collected other than the data stated in the proposal;
3.3 A quarterly report to be submitted to the Ministry's Research Unit;
3.4 Preliminary findings to be submitted upon completion of the study;
3.5 Final report to be submitted upon completion of the study;
3.6 Separate permission should be sought from the Ministry for the publication of the findings.

Yours sincerely,

PERMANENT SECRETARY
Republic of Namibia
MINISTRY OF HEALTH AND SOCIAL SERVICES

03 NOV 2014

Andrew Ndishishi (Mr)
Permanent Secretary

"Health for All"

Figure A.2: Namibia ethical clearance

Appendix B

Presentations at conferences and other research platforms

B.1 Oral presentations

1. A Survival analysis of HIV positive children on antiretroviral treatment at Katutura hospital, Windhoek, Namibia.

   This was an oral presentation during the institutional research week at the Nambia University of Science and Technology.

   Maposa, I. and Blignaut, J.R. (2016). *Institutional research week, 21-23 November 2015*
   Windhoek, Namibia.

2. Bayesian PHM and insights from the CQR model in determining risk factors for survival in paediatric HIV/AIDS patients on antiretroviral treatment.

   This was an oral presentation during the faculty of health and applied sciences research day at the Namibia University of Science and Technology:

   Maposa, I. and Blignaut, J.R. (2016). *Faculty of health and applied sciences, 13 May 2016*
   Windhoek, Namibia.

B.2 Conferences and workshops attended and to be attended

1. Bayesian Biostatistics short course, 4 to 8 November 2013, SACEMA, South Africa.

2. Clinic on Meaningful Modelling of Epidemiological Data (MMED) International

Conferences, African Institute of Mathematical Sciences (AIMS) in collaboration with International Clinics on Infectious Disease Dynamics Data (ICI3D) programme, 02 June to 13 June 2014.

3. Introduction to the Joint Modelling of Longitudinal and Survival Data, with Applications in R Course, 19 to 21 October 2015, SACEMA, South Africa.

4. Australian Statistical Conference 2016, 5-9 December 2016, Canberra, Australia.

   Presentation titled: Inferential insights from the censored quantile regression and bayesian proportional hazards models: a case of paediatric and adolescent hiv/aids patients on antiretroviral treatment at katutura hospital, windhoek, namibia.

5. The 58th Annual Conference of the South African Statistical Association, 28 November -01 December 2016, Cape Town, South Africa.

   Presentation titled: Bayesian Cox Proportional Hazards model and insights from the censored quantile regression model for paediatric and adolescent HIV/AIDS patients on antiretroviral treatment, Windhoek, Namibia.

B.3 Poster presentations

Figure B.1: Meaningful modelling of epidemiological data (MMED) international clinic poster, 2-13 June 2014, AIMS, South Africa

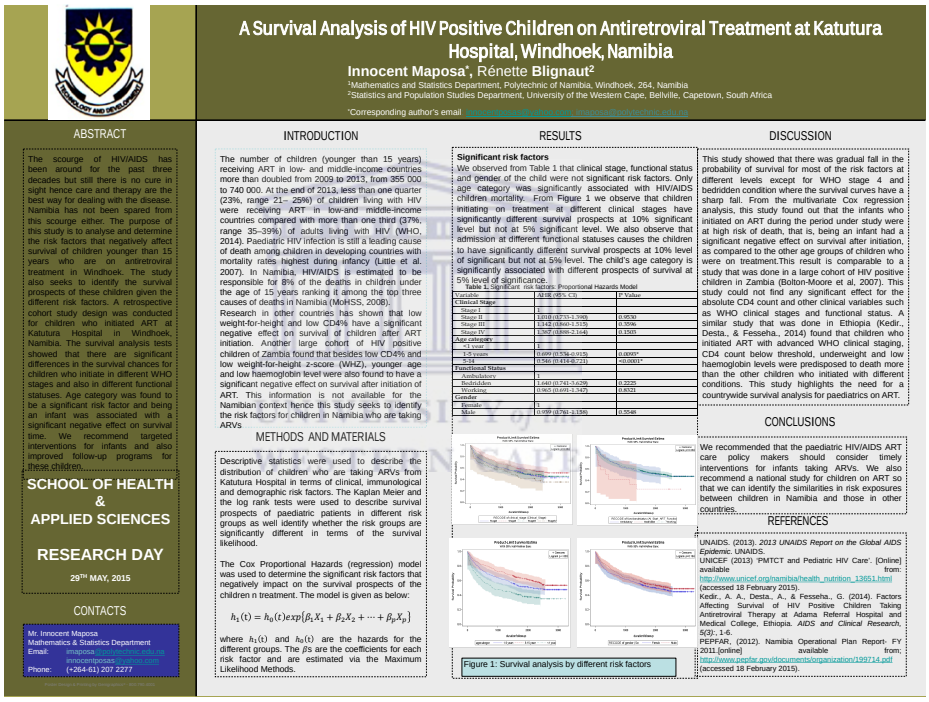


Figure B.2: Namibia University of Science and Technology, Institutional research day, 21-23 November 2015, poster: Windhoek, Namibia

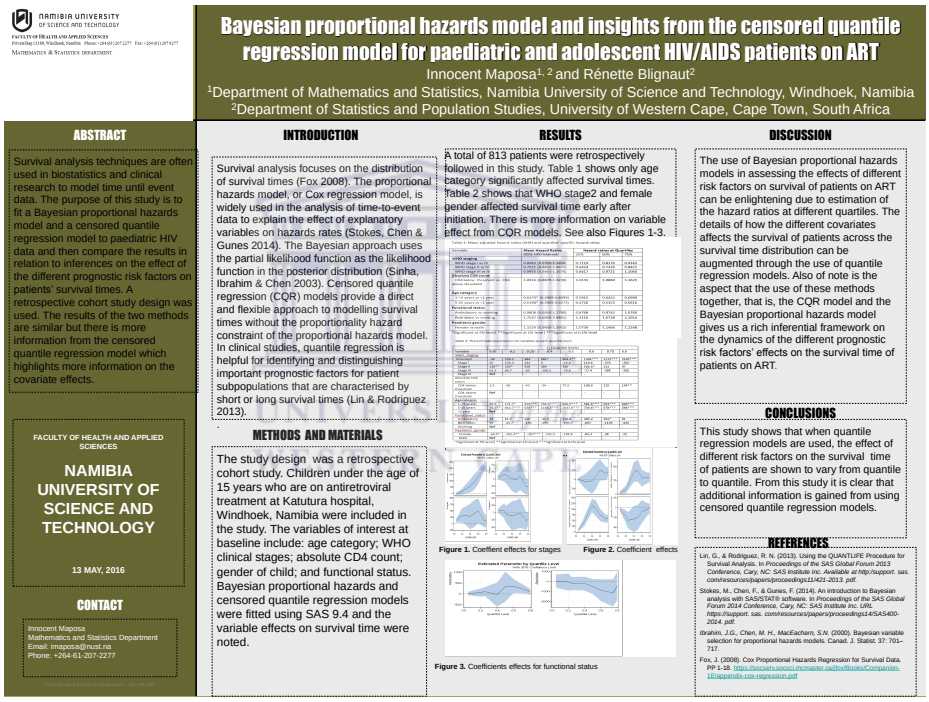


Figure B.3: Namibia University of Science and Technoloy, Faculty research day, 13 May 2016, poster: Windhoek, Namibia

Appendix C

R and SAS codes used in analysis and PHM effects graph

C.1 SAS code

```
proc contents data=zimdataset;
run;
proc mi data=zimdataset nimpute=5 seed=135782 out=Imputeddata;
var survivaltime status Gender baselineage clinicalstages baseline_weight ;
class clinicalstages Gender status;
FCS logistic (clinicalstages);
FCS logistic (Gender);
FCS logistic (status);
FCS regpmm (baselineage);
FCS regpmm (baseline_weight);
FCS regpmm (survivaltime);
run;
proc print data=Imputeddata;
run;
proc freq data=Imputeddata;
tables clinicalstages;
by _Imputation_;
run;
ods rtf;
ods graphics on;
proc quantreg data=Imputeddata outest=cqrestimates ci=resampling plots=ALL;
   Title ' Kaplan-Meier-Type Estimator for Censored Quantile Regression and Censor
   class Gender clinicalstages;
```

```
   model survivaltime*status(0)= Gender baselineage baseline_weight clinicalstages
                        /quantile=(0.2 0.4 0.6 0.8) seed=1268;
   Gender_Effect: test Gender;
   clinicalstages_Effect: test clinicalstages;
   baselineage_Effect: test baselineage;
   baseline_weight_Effect: test baseline_weight;
  ODS OUTPUT PARAMETERESTIMATES=cqrsparms;
by _Imputation_;

run;
ods graphics off;
ods rtf close;

ods rtf;
ods graphics on;
PROC MIANALYZE PARMS(CLASSVAR=FULL)=cqrsparms;
modeleffects Intercept Gender baselineage baseline_weight clinicalstages;
ODS OUTPUT PARAMETERESTIMATES=mian_ cqrsparms;
RUN;
ods graphics off;
ods rtf close;

ods rtf;
```

```
proc phreg data=Imputeddata outest=phmestimates covout;
     class Gender clinicalstages;
     model survivaltime*status(0)=Gender clinicalstages baselineage baseline_weigh
  by _Imputation_;
 run;

proc mianalyze data=phmestimates;
modeleffects GenderFemale clinicalstages1__WHO_Stage_1 clinicalstages2__WHO_Stage_
run;
ods rtf close;

proc print data=phmestimates;
run;
```


Imputed results code

```
proc contents data=paedimpute;
run;
proc mi data=paedimpute nimpute=5 seed=135782 out=Imputeddata_paediatrics;
var durationfollowup status paed_gender functional_status agecategory absoluteCD4c
class WHO_staging paed_gender status functional_status absoluteCD4count agecategor
FCS logistic (paed_gender);
FCS logistic (functional_status);
FCS logistic (status);
```

```
FCS logistic (absoluteCD4count);
FCS logistic (agecategory);
FCS regpmm (durationfollowup);
run;
proc print data=Imputeddata_paediatrics;
run;
proc freq data=Imputeddata_paediatrics;
tables WHO_staging;
by _Imputation_;
run;
ods rtf;
ods graphics on;
proc phreg data=Imputeddata_paediatrics;
class WHO_staging absoluteCD4count agecategory functional_status paed_gender;
model durationfollowup*status(0)= WHO_staging absoluteCD4count agecategory functio
bayes seed=1 diagnostic=all plots=all outpost=out;
hazardratio WHO_staging;
hazardratio absoluteCD4count;
hazardratio agecategory;
hazardratio functional_status;
hazardratio paed_gender;
by _Imputation_;
run;
ods graphics off;
```

```
ods rtf close;



ods rtf;
ods graphics on;
proc quantreg data=Imputeddata_paediatrics outest=crq_paed_estimates ci=resampling
   Title ' Kaplan-Meier-Type Estimator for Censored Quantile Regression and Censor
   class WHO_staging absoluteCD4count agecategory functional_status paed_gender;
   model durationfollowup*status(0)= WHO_staging absoluteCD4count agecategory func
                        /quantile=(.05 .2 .25 .4 .5 .6 .75 .8);
   WHO_staging_Effect: test WHO_staging;
   absoluteCD4count_Effect: test absoluteCD4count;
   agecategory_Effect: test agecategory;
   functional_status_Effect: test functional_status;
   paed_gender_Effect: test paed_gender;
   by _Imputation_;
run;
ods graphics off;
ods rtf close;
```


C.2 R code

```
library(survival)
library(quantreg)
```

```
library(foreign)
paeddata<-read.dta("C:\\Users\\imaposa\\Desktop~~
\\IMaposa\\Desktop\\research\\IRPC-poly\\ARV study~~~
\\paediatric HIV stuff\\paediatrics_imputed_R2.dta")

head(paeddata)

fit<-crq(formula=Surv(durationfollowup, status)~
paed_gender+clinicalstage+ absolutecd4count+
agecategory+functional_status, taus=taus1,
data=paeddata, method="Portnoy")

fit

summary(fit, 1:19/20)

plot(summary(fit, 1:19/20))

Sfit<-summary(fit, 1:19/20)

PHit<-coxph(Surv(durationfollowup, status)~paed_gender+
clinicalstage+ absolutecd4count+agecategory+
functional_status, data=paeddata)
```

```
PHit

myplot<-plot.summary.crqs

myplot

fix(myplot)

myplot(Sfit, CoxPHit=PHit)
```

C.3 Cox's proportional hazards effects graph

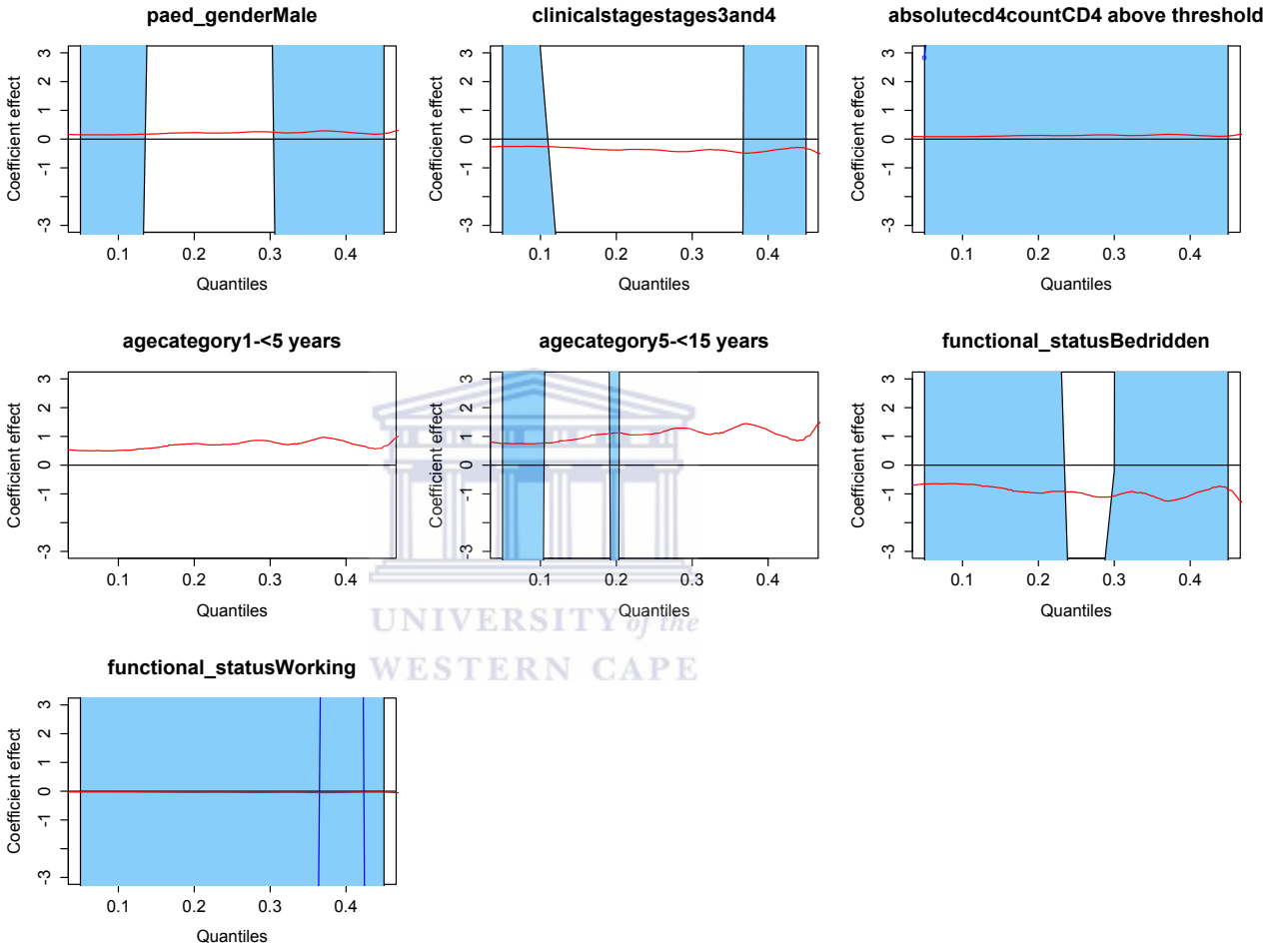


Figure C.1: Comparison zoomed PHM parameter effects