A COPULA APPROACH TO SAMPLE SELECTION MODELING OF TREATMENT ADHERENCE AND VIRAL SUPPRESSION AMONG HIV PATIENTS ON ANTIRETROVIRAL THERAPY (ART) IN NAMIBIA

A MINI THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE (BIOSTATISTICS)

OF

THE UNIVERSITY OF NAMIBIA

BY

JASON NAKALUUDHE
201201646

APRIL 2019

SUPERVISOR: PROF. LAWRENCE KAZEMBE (UNIVERSITY OF NAMIBIA, DEPARTMENT OF STATISTICS AND POPULATION STUDIES)

ABSTRACT

Namibia has a generalized human immunodeficiency virus (HIV) epidemic, with HIV mainly transmitted through heterosexual transmission. Although the number of people receiving ART has increased, the achievement of the 90-90-90 strategy on testing, treatment and suppression has not been evaluated. Moreover, examining factors associated with treatment adherence and viral suppression will assist in designing appropriate accelerated interventions. However, modelling treatment adherence and viral suppression may result in biased estimates if sample selection is ignored. The study fit a joint distributional model of ART treatment adherence and viral suppression, to adjust for sample selection bias among HIV patients on ART in Namibia, to examine the presence of tail dependence in sample selection bias, and investigate the factors associated with viral suppression, viral load and ART adherence. The study used two datasets; HIV data of patients, aged above 16 years, on antiretroviral therapy in Erongo region and the selected health facilities in Windhoek, Namibia. A Heckman-type selection analysis using copula were used on the two models: (i) ART adherence with viral suppression margins and (ii) ART adherence with viral load. The families of copulas i.e. Normal, Frank, FGM, AMH, Student-t and the 0, 90, 180 and 270 degrees rotated versions of Joe, Gumbel and Clayton, to capture dependence in the outcomes, were modelled and selected based on the lowest AIC and BIC. The results showed a strong negative correlation between adherence to ART treatment and viral load suppression. The results also showed the dependence structure between ART adherence and viral load margins. The results further showed that Frank and the 180 degrees rotated versions of Gumbel, and Clayton copulas were the best models. Antiretroviral therapy adherence with viral suppression and ART adherence with viral loads exhibit positive dependence structures, even though some demographic and clinical characteristics were not associated with ART adherence, viral suppression and viral load. Similar studies in the future need to consider socio-economic factors in addition to those considered in this study.

LIST OF CONFERENCE PROCEEDINGS/PUBLICATIONS

- A manuscript by Nakaluudhe, J. & Kazembe, L (2018). A Copula Approach to Sample Selection Modeling of Treatment Adherence and Viral Suppression among HIV Patients on Antiretroviral Therapy in Namibia (under review).
- A manuscript by Ntumba, A., Nakaluudhe, J. & Kazembe, L. (2018). Outcomes and Determinants of Treatment Adherence: Evidence from a Retrospective Cohort Study in selected clinics in Windhoek, Namibia (under review).
- 2nd DELTAS Africa SSACAB Research Conference (10th – 12th September 2018 InterContinental Hotel, Nairobi, Kenya).
- 1st National Students Research Symposium (18-19 October 2018 Namibia Institute Public Administration Management (NIPAM), Windhoek, Namibia).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS/ACRONYMS

AAS	African Academy of Sciences
AIC	Akaike Information Criterion
AESA	Alliance for Accelerating Excellence in Science in Africa
AIDS	Acquired Immune Deficiency Syndrome
AMH	Ali-Mikhail-Haq
ART	Antiretroviral Therapy
BIC	Bayesian Information Criterion
CML	Canonical maximum likelihood method
DELTAS	Developing Excellence in Leadership Training and Science
EML	Exact Maximum Likelihood
FGM	Farlie-Gumbel-Morgenstern
GJRM	Generalized Joint Regression Modelling
GLM	Generalized Linear Model
HIV	Human Immunodeficiency Virus
IFM	Inference Functions for Margins
MAR	Missing At Random
MCAR	Missing Completely At Random

MNAR	Missing Not At Random
MoHSS	Ministry of Health and Social Services
NEPAD	New Partnership for Africa’s Development Planning and Coordinating
OR	Odds Ratio
PGSC	Faculty of Science Postgraduate Committee
SSACAB	Sub-Saharan Africa Consortium for Advanced Biostatistics
UK	United of Kingdom
UNAIDS	United Nations Joint Programme on AIDS
UNAM	University of Namibia
WHO	World Health Organization

ACKNOWLEDGEMENT

I would like to thank my supervisor Prof. Lawrence Kazembe for his guidance throughout the research. His invaluable guidance, patience and encouragement despite his busy schedule have allowed me to complete the research.

I am very appreciative to my colleagues, Simon Kashihalwa, Alexander Petrus and Andreas Shipanga, who have been so supportive along the way of doing my thesis. I also would like to express my wholehearted thanks to my late mother for her generous support, she provided me throughout my entire life and particularly through the process of pursuing the master degree.

This study was supported through the DELTAS Africa Initiative SSACAB (Grant No. 107754/Z/15/Z). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and is supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant No. 107754/Z/15/Z) and the UK government. The views expressed in this publication are those of the author and not necessarily those of the AAS, NEPAD Agency, Wellcome Trust or the UK government.

DEDICATIONS

I dedicated this thesis to my late mother, Soini Nambata Ipinge, who always supported and believed in me. I also dedicated this thesis to my beloved daughter, Soini Shawana Nakaluudhe.

DECLARATIONS

I, Jason Nakaluudhe, hereby declare that this study is my own work and is a true reflection of my research, and that this work, or any part thereof has not been submitted for a degree at any other institution.

No part of this thesis may be reproduced, stored in any retrieval system, or transmitted in any form, or by means (e.g. electronic, mechanical, photocopying, recording or otherwise) without the prior permission of the author, or The University of Namibia in that behalf.

I, Jason Nakaluudhe, grant The University of Namibia the right to reproduce this thesis in whole or in part, in any manner or format, which The University of Namibia may deem fit.

………………………..……………. ……………………………… ….…………

Name of Student Signature Date

CHAPTER 1: INTRODUCTION

1.1 Background of the study

Since the discovery of the AIDS epidemic, more than 70 million people have been infected with the HIV virus and about 35 million people have died of HIV attributable illness (World Health Organization, 2017). Globally, 36.7 million people were living with HIV at the end of 2016. An estimated 0.8 percent of adults aged 15 to 49 years worldwide are living with HIV, although the burden of the epidemic continues to vary considerably between countries and regions. Sub-Saharan Africa remains most severely affected, with nearly 1 in every 25 adults (4.2%) living with HIV and accounting for nearly two-thirds of people living with HIV worldwide (World Health Organization, 2017).

The priority strategies for sustainable fight against HIV/AIDS include:` prevention through promotion of sexual abstinence and condom use when having sex, promotion of uptake of antiretroviral therapy (ART), and awareness raising campaigns for behavioural change. The first two are aimed at reducing new infections, while the last two have a goal of improved prognosis, reducing death and prolong longevity. Since the roll-up of ART, globally AIDS-related deaths fell by 45 percent from 2005 to 2015, including a 13 percent decline in the last three years (UNAIDS, 2016).

Global scale-up of ART has been the primary contributor to a 48 percent decline in deaths from AIDS-related cause. According to the United Nations Joint Programme on AIDS (UNAIDS) reported that, in 2016, 51 percent of people living with HIV globally are female, higher treatment coverage and better adherence to treatment among women have driven more rapid declines in AIDS-related deaths among females. Moreover, deaths from AIDS-related illnesses were 27 percent lower among women and girls than they were among men and boys (WHO, 2017).

Achievement of UNAIDS 90–90–90 treatment target by 2020 alongside high coverage of primary HIV prevention interventions can make the end of the AIDS epidemic a reality by 2030 (UNAIDS, 2016). The 90-90-90 treatment target means that 90 percent of people living with HIV know their HIV status, 90 percent of people who know their HIV status are accessing treatment and 90 percent of people on treatment have suppressed viral load. In order to achieve viral suppression, people need to be diagnosed with HIV, linked to care, started on treatment, retained in care and virally suppressed. However, the global gap in achieving the 90-90-90 target in 2015 was around 11.9 million people living with HIV who did not know their HIV status, 12.7 million people in need of antiretroviral treatment and 13.0 million people living with HIV who were not virally suppressed (UNAIDS, 2016).

Strong adherence to antiretroviral therapy suppresses viral load to undetectable levels within people living with HIV, greatly reducing the risk of transmitting the virus to others. When large proportions of people living with HIV within a community are on treatment, it has been shown to have a preventive effect within that community (WHO, 2017).

In Namibia, the proportion of adults presenting with HIV infection and receiving ART have decreased (Namibia Population-Based HIV Impact Assessment [NAMPHIA], 2018). According to NAMPHIA, HIV prevalence, in Namibia, was 12.6% among adults aged 15-64 years in 2017.

The Antiretroviral Therapy (ART) can profoundly inhibit viral replication and delay HIV disease progression, but achieving this potential in clinical practice requires adherence to complex regimens. Antiretroviral adherence and viral suppression go hand in hand. Treatment adherence to ART are considered necessary to achieve viral suppression. Therefore, the goal of ART is to achieve sustained and optimal viral suppression. Nonadherence with antiretroviral therapy may result not only in reduced treatment efficacy but also in the selection of drug-resistant HIV strains (Arnsten et al., 2001).

The optimal designs are required to permit required adherence levels. However, a number of studies have shown that non-adherence is a major issue and a rising problem. There is need to understand factors associated with adherence or non-adherence to enable redesigning health promotions. Similarly, the obtainment of viral load suppression is an important objective that is associated with a variety of socio-demographic and baseline clinical factors (Cescon et al., 2011). However, investigating determinants of viral suppression, one would be challenged with the issue of selection bias. This arises since only those adhering would lead to viral suppression.

Sample selection bias can be addressed with a population-based survey in a clearly defined target population of all people in a community, including those with and without known HIV infection, and with measurement of viral loads in infected people (Miller, Powers, Smith & Cohen, 2013).

Such a random sample would overcome the biased sample associated with engagement in care. As with other cohort analyses, there is the potential for selection bias as a result of the differential losses to follow up at the various clinic sites of those individuals who did not achieve suppression.

There are two types of selection bias for dealing with non-random participation (Alexander, Lopes, Ricchetti-Masterson & Yeatts, 2015). These include self-selection and sample selection bias. However, it is crucial to study both non-random selection bias.

Non-random sample selection arises when individuals select themselves into the sample based on features that are observed and unobserved. In this case, statistical analysis based on commonly known models such as linear and probit regressions will yield biased and inconsistent parameter estimates (Marra & Wyszynski, 2016). Thus, ignoring the non-random selection mechanism induced by the selection equation may result in biased estimation of the coefficients in the outcome equation (Wiesenfarth & Kneib, 2010). Therefore, it is important to avoid the introduction of sample selection bias by analysing both equations simultaneously. One way of addressing this issue is to employ sample selection models based on copula approach.

Copulas are multivariate distribution functions whose one-dimensional margins are uniform on the [0, 1] interval (Nelsen, 1998). Copulas are useful for constructing joint distributions, especially with non-random variables. Copula can be also used to describe dependence between correlated random variables (Sklar, 1959).

Sample selection models or self-selection models have been heavily employed in particular in the econometrical literature but also in the social sciences (see for example Lewis (1974), Strazzera et al. (2004), Smith (2003) or Marra & Wyszynski (2016)). In addition, Brown and Theoharides (2009) provided an example of application in health and Corno (2014) also gave a review on economic development and cultural change. For this purpose, a simple method for constructing treatment adherence and viral suppression based on copulas is suggested.

The design, features, and some applications of copula can be easily extended in multivariate modelling in many fields (Yan, 2006). In finance, copula functions are adopted to handle the interaction between the markets and risk factors in a flexible manner (Bouyé, et al, 2000; Cherubini, Luciano & Vecchiato, 2004). In biology, a Gaussian copula has been applied in quantitative trait linkage. Copulas play an important role in developing an integrated likelihood framework to analyze discrete, continuous, even mixed (discrete and continuous) and censored response data. Sungur (2005) proposed several measures of directional dependence in regression based on copula functions.

The studies of sample selection model based on copula approach are vast and many variants of such models have been proposed. Smith (2003) introduced modelling sample selection using Archimedean copulas which focused on those that fall within the Archimedean class. Strazzera et al. (2004) estimated several copula models, both Archimedean and non-Archimedean, to sample selection model.

1.2 Statement of the problem

The goal of UNAIDS is to reach 90-90-90 treatment target by 2020: 90 percent of people living with HIV know their HIV status, 90 percent of people who tested HIV-positive must be put on treatment, and 90 percent of people on treatment have suppressed viral load. To date, studies in Africa have mainly reported on treatment adherence and viral suppression. However, in Namibia, quantitative studies are needed in order to determine factors related to treatment adherence, viral load and viral suppression among people on ART, there is a need of analyzing treatment adherence, viral load and viral suppression taking into account sample selection bias, and also the needs of developing flexible models that permit unbiased estimates to inform better adjustment of interventions.

1.3 Objectives of the study

1.3.1 The main objective

The main objective of the study was to fit a joint distributional model to adjust for sample selection bias of treatment adherence, viral load and viral suppression among HIV patients on ART in Namibia.

1.3.2 The specific objective were:

a) To simultaneously model treatment adherence with viral suppression;
b) To fit joint model of treatment adherence with viral load;
c) To examine if there is any tail dependence in sample selection bias;
d) To investigate the factors associated with viral suppression and treatment adherence.

1.4 Significance of the study

Although only Erongo region and two selected health facilities in Khomas region from health sector were included in the study, the findings will benefit patients, health systems and societies with a better health outcomes and economic efficiency. The study used as a useful reference for researchers or scientists in their daily work.

1.5 Limitation of the study

The study used only HIV infected patients (aged 16 and above) on ART, who were enrolled in Erongo region and Khomas region (Katutura State Hospital and Okuryongava Clinic) during January 1, 2010 to December 31, 2015. Since the study used secondary data, the data were not collected for the purpose of this study and were not in a certain format for analysis. In addition, some study variables were important but were not captured such as treatment and socio-economic variables.

1.6 Delimitation of the study

The study did not use data from national health facilities because of research limited time, therefore it only used data from Katutura Health Center and Okuryongava Clinic. The data were merely for binary and continuous variables which permitted the study not to consider the count variables.

1.7 Organization of the thesis

The thesis is organized as followed. Chapter 2 provides the methodological Review of Sample Selection model. Chapter 3 gives the methodology and data analysis. Chapter 4 & 5 highlight the results in line with the objective and discussions of the results. Conclusion and recommendations are presented in the last chapter.

CHAPTER 2: METHODOLOGICAL REVIEW OF SAMPLE SELECTION MODEL

2.1 Introduction

This chapter presents the methodological review on sample selection model which is broadly classified into six (6) parts. That is; model definition, linking the model equations, likelihood, model specification, estimation methods and model selection.

2.2 Model definition

In the sample selection context, the aim is to fit a regression model when some observations for the response of interest, W_{2i}, are missing not at random (Wojtyś, Marra & Radice, 2016). There are a number of missing data mechanisms. They are as follows; missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). These mechanisms describe relationships between measured variables and the probability of missing data (Rubin, 1976). In this section, the study handling the missing not at random which is sample selection.

The process of selection can be typically assumed that there is a variable W_{1i} which governs whether or not an observation on the variable of primary interest is generated. Based on Marra Papageorgiou and Radice, (2014); Wojtyś, Marra and Radice, (2016), the latent variable representation can be written as

$$W^*_{1i} = \eta_{1i} + \varepsilon_{1i} = x_{1i}{}^{T}\alpha + \sum_{k_1=1}^{K_1} s_{1k_1}\left(u_{1k_{1i}}\right) + \varepsilon_{1i}, \text{ for } i = 1, 2, \dots, n \qquad (2.1)$$

such that

$$W^{*}_{1i} \sim N(\eta_{1i}, 1), \tag{2.2}$$

$$W_{2i} = \begin{cases} 1 \text{ if } W^{*}_{1i} > 0 \\ 0 \quad \text{otherwise}, \end{cases} \tag{2.3}$$

here n is a sample size, and W^{*}_{1i} is independent and identically distributed, η_{1i} is a linear predictor with the obvious definition, α is a P_1 dimensional coefficient vector of all parametric components, x_{1i} is a vector of (binary and categorical) covariates, $s_{1k_1}(u_{1k_{1i}})$ are unknown smooth functions of the K_1 continuous covariates $u1k_{1i}$ and ε_{1i} is error term with zero mean (Marra, Radice, Bärnighausen, Wood & McGovern, 2017). The outcome equation can be written as (Marra et al., 2017):

$$W_{2i} \sim Ƒ(\mu_i, \sigma, v), \tag{2.4}$$

where $\mu_i = E(W_{2i}) = \exp(\eta_{2i}) = \exp\left(\alpha^T x_i + \sum_{k_1=1}^{K_2} s_{2k_2}(u_{2k_{2i}})\right)$, and Ƒ is a discrete distribution, η_{2i} is a linear predictor, and σ and v are scale and shape distribution specific parameters. The number of parameters that characterize Ƒ depend on the chosen distribution. Vector α has length P_2 which represents the parameters of all parametric components, x_i is a vector of factor variables and $s_{2k_2}(u_{2k_{2i}})$ are unknown smooth functions of the K_2 continuous covariates $u_{2k_{2i}}$ (Marra et al., 2017).

The smooth functions are approximated using regression splines as (Marra and Wyszynski, 2016)

$$\sum_{j=1}^{J_{vkv}} \alpha_{vkvj} b_{vkvj}(u_{vki}) = \mathbf{B}(u_{vkvi})^T \boldsymbol{\alpha}_{vkv}\,, \tag{2.5}$$

where the $b_{vkvj}(u_{vki})$ are known spline basis function, the α_{vkvj} are regression parameters, J_{vkv} is the number of bases used to represent the smooth term, $\mathbf{B}(u_{vki}) = [b_1(u_{vkvi}), b_2(u_{vkvi}), \ldots, b_{J_{vkv}}(u_{vkvi})]^T$ is a vector containing J_{vkv} basis functions and α_{vkv} is the corresponding parameter vector. The subscript indicating which equation each smooth component belongs to has been suppressed for simplicity. The smoothing function are subject to the centering constraint (Wojtyś et al., 2016; Marra & Radice, 2017).

2.3 Linking the model equations

In a sample selection model concept, it is assumed that the selection and outcome equations are linked through unobservable; this link is formulated, in a standard case, using a bivariate normal distribution (Marra & Wyszynski, 2016). The relationship between a genetic pair of variables $(\mathrm{W}_1, \mathrm{W}_2)$ can be represented using a copula function (Sklar, 1959; Heckman, 1976; Heckman, 1979). Let $\mathrm{F}_1(\mathrm{w}_1)$ and $\mathrm{F}_1(\mathrm{w}_1)$ be the cumulative density functions (cdfs) of $\mathrm{w}_1$and w_2, respectively. F_1 relates to the selection equation and F_2 to the outcome equation. Then, the joint cdfs is represented as follows:

$$\mathrm{F}(\mathrm{w}_1, \mathrm{w}_2) = \mathrm{C}(\mathrm{F}_1(\mathrm{w}_1), \mathrm{F}_2(\mathrm{w}_2); \theta), \tag{2.6}$$

for some two-place copula function C which is unique only on closure $\mathrm{Ran}(\mathrm{F}_1) \times \mathrm{Ran}(\mathrm{F}_2)$, where $\mathrm{Ran}(\cdot)$ denotes the range of argument and θ is an association parameter measuring the dependence between the two marginal cdfs. For more details see Marra and Wyszynski (2016), Wojtyś, Marra and Radice (2016) and Smith (2003).

2.4 Copula functions

There are a number of families of copula functions currently implemented in literature. These are summarized in table 2.1. They are normal, Farlie-Gumbel-Morgenstern (FGM), and some belong to Archimedean class of copula. Archimedean copulas form a large family of copulas with a number of convenient properties and they allow for a large number of dependence structures (Manner, 2007). Since in some cases θ cannot be interpreted directly, it can be transformed into Kendall’s τ ranging on the interval [−1, 1] yielding a general interpretation for all copulas. Kendall’s τ is not a margin-free measurement of dependence and should therefore be used with caution especially when the choice of margins is based on dependence (Marra and Wyszynski, 2016). The copulas currently available in literature and the respective conversions from θ to Kendall's τ are summarized in Table 1.

Table 1: Example of the families of bivariate copulas

Name	**Copula $\mathbf{C(u, v; \theta)}$**	**Parameter space of θ**	**Parameter space of Kendall's τ**	**Kendall's τ in terms of θ**
FGM	$uv\big(1+\theta(1-u)(1-v)\big)$	$-1 \le \theta \le 1$	$-2/9 \le \tau \le 2/9$	$\frac{2}{9}\theta$
Normal	$\Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$	$-1 \le \theta \le 1$	$-1 \le \tau \le 1$	$\frac{2}{\pi}\arcsin(\theta)$
AMH	$uv/\big(1-\theta(1-u)(1-v)\big)$	$-1 \le \theta \le 1$	$-0.1817 \le \tau < \frac{1}{3}$	$1-\frac{2}{3\theta^2}(\theta+(1-\theta)^2\log(1-\theta))$
Clayton	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$	$1 \le \theta \le \infty$	$0 < \tau < 1$	$\frac{\theta}{\theta+2}$
Frank	$-\theta^{-1}\log\Big(1+\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)/\left(e^{-\theta}-1\right)\Big)$	$\theta \in \mathbb{R}\{0\}$	$-1 < \tau < 1$	$1-\frac{4}{\theta}[1-D_1(\theta)]$
Gumbel	$\exp\Big(-\left((-\log u)^{\theta}+(-\log v)^{\theta}\right)^{1/\theta}\Big)$	$1 \le \theta \le \infty$	$0 \le \tau < 1$	$1-\frac{1}{\theta}$
Joe	$1-\left((1-u)^{\theta}+(1-v)^{\theta}-(1-u)^{\theta}(1-v)^{\theta}\right)^{1/\theta}$	$1 \le \theta \le \infty$	$0 < \tau < 1$	$1+\frac{4}{\theta^2}D_2(\theta)$

Note: this table presents the example of families of bivariate copulas. u and v represent marginal cdfs. $\Phi_2(\cdot,\ \cdot; \theta)$ denotes bivariate standard normal cdf with correlation coefficient θ, and $\Phi^{-1}(\cdot)$ is the inverse cdf of a standard normal. $D_1(\theta) = \frac{1}{\theta}\int_0^{\theta}\frac{t}{\exp(t)-1}dt$ is the Debye function and $D_2(\theta) = \int_0^1 t\log(t)(1-t)^{\frac{2(1-\theta)}{\theta}}$.

Copula functions have been introduced recently in finance (Bouyé, et al, 2000). They are a general tool to construct multivariate distributions and to investigate dependence structure between random variables (Jouanin, Riboulet & Roncalli, n.d).

The rotation of a copula function makes sense only for ones with an asymmetric dependence structure (Manner, 2007). The Clayton, Joe and Gumbel copulas can be rotated by 90, 180 and 270 degree. These rotations are clearly shown as

$$C_{90} = v - C(1-u, v; \theta), \qquad C_{180} = u + v - 1 + C(1-u, 1-v; \theta),$$
$$C_{270} = u - C(u, 1-v; \theta), \tag{2.7}$$

where u and v are the margins (Brechmann & Schepsmeier, 2013 as cited by Marra & Wyszynski 2016).

2.5 Tail dependence

In this concept, tail dependence measures the strength of dependence in lower and upper tail of a multivariate distribution. According to Joe (1997), tail dependence describes the limits of a proportion that one margin exceeds a certain threshold given that the other margin has already exceeded that threshold.

Let $W_1 \sim F_1$ and $W_2 \sim F_2$ be two random variables with copula C, define the coefficients of tail dependency

$$\lambda_L = \lim_{u \downarrow 0} \mathbb{P}[F_1(W_1) < u | F_2(W_2) < u] = \lim_{u \downarrow 0} \frac{C(u,u)}{u}$$
$$\lambda_U = \lim_{u \uparrow 1} \mathbb{P}[F_1(W_1) > u | F_2(W_2) > u] = \lim_{u \uparrow 1} \frac{1-2u+(u,u)}{1-u} \tag{2.8}$$

where λ_L is the lower tail and λ_U is the upper tail, C is said to have lower or upper tail dependence if and only $\lambda_L \neq 0$ ($\lambda_U \neq 0$).

2.6 Likelihood

The sample selection model occurs when observations on a pair of random variables (R, W) are generated according to the following relationships (Smith, 2003; Marra & Wyszynski, 2016; Wojtyś et al, 2016; Rüschendorf, 2013):

$$R = 1\{W_1{}^* > 0\} \quad \text{and} \quad W = 1\{W_1 > 0\}W_2, \tag{2.9}$$

where $1\{A\}$ denotes the indicator function, taking value 1 if event A holds, and 0 otherwise. In effect, W_2 can be observed only when $W_1 > 0$. The participation mechanism is represented by the Bernoulli variable R, and it derives its properties from those of $W_1{}^*$. Note that when $R = 0,\ W_2$ cannot be observed, and W is assigned a dummy value of 0.

Let $r_1, \dots, r_m$ denote m observations generated on $R\big(r_j \in \{0,1\}, j = 1, \dots, m\big)$, and $w_1, \dots, w_m$ the corresponding m observations generated on $W\big(w_j \in \mathbb{R},\ j = 1, \dots m\big)$. For a random sample of m observations, the likelihood function for the sample selection model is formulated as follows (Smith, 2003 as cited by Marra & Wyszynski, 2016)

$$L = \prod_0 \Pr(W_1 = 0) \prod_1 \Pr(W_2 = w_2, W_1 = 1)$$

$$= \prod\nolimits_0 \Pr\big(W_{1j}{}^* \leq 0\big) \prod\nolimits_1 f_{2|1}\big(w_2 \big| w_{1j}{}^* > 0\big) \Pr\big(W_{1j}{}^* > 0\big), \tag{2.10}$$

where, $\prod_0$ indicate the product over those observations for which $r_j = 0$, and $\prod_1$ the product over those observations for which $r_j = 1$. For convenience, j is the index which can be dropped from now and the function $f_{2|1}$ denotes the probability density function (pdf) of W_2, given event ${W_1}^*$ and it can be derived as follows (Smith, 2003):

$$f_{2|1}(w_2|{w_1}^* > 0) = \frac{\partial}{\partial w_2}\frac{F_2(w_2) - F(0, w_2)}{F_1(1)}$$

$$= \frac{1}{1-F_1(0)}\frac{\partial}{\partial w_2}\left(F_2(w_2) - F(0, w_2)\right) = \frac{1}{1-F_1(0)}\left(f_2(w_2) - \frac{\partial}{\partial w_2}F(0, w_2)\right) \tag{2.11}$$

The general likelihood is the form of the sample selection model. Particularly, likelihood functions arise from specification assumptions. Therefore, the likelihood function is given by

$$L = \prod_0 \Pr\left({W_{1j}}^* \leq 0\right) \prod_1 f_{2|1}\left(w_2\middle|{w_{1j}}^* > 0\right) \Pr\left({W_{1j}}^* > 0\right). \tag{2.12}$$

2.7 Model specification

A copula approach requires the specification of two elements namely; the distribution of the marginal and the copula function. In view of the flexible properties shown elsewhere (Engle & Gallo, 2006). For the first, Gamma probability distribution function has been adopted (but other choices are possible, such as Inverse-Gamma, Weibull, Lognormal, and mixtures of them). For the second, some possible specifications within the class of Elliptical copulas have been discussed. Most importantly, the copula approach permits specifications other than multivariate normality, although it does retain that distribution as a special case.

2.8 Estimation methods

In copula approach, there exist five methods for estimating copula models namely; the one step method (Exact Maximum Likelihood (EML)), the two step method (Inference Functions for Margins (IFM)), the semiparametric estimation (canonical maximum likelihood method (CML)) and the last two methods are nonparametric ways of estimating the copula.

The first method is the one step method (Exact Maximum Likelihood (EML)) which estimates all parameters of the model at the same time (Manner, 2007; Payandeh, Farid-Rohani, & Qazvini, 2013). The EML estimator can be specified by first considering the following representation of the joint density known as the copula decomposition of a joint distribution and the resulting log-likelihood function.

$$\mathrm{h}(\mathrm{w}_{1\mathrm{i}}, \mathrm{w}_{2\mathrm{i}}, \theta) = \mathrm{f}(\mathrm{w}_{1\mathrm{i}}; \varphi) \cdot \mathrm{g}(\mathrm{w}_{2\mathrm{i}}; \gamma) \cdot \mathrm{c}(\mathrm{F}(\mathrm{w}_{1\mathrm{i}}; \varphi), \mathrm{G}(\mathrm{w}_{2\mathrm{i}}, \gamma); \delta)$$

$$\mathrm{L}_{\mathrm{W}_1\mathrm{W}_2} = \sum_{\mathrm{i}=1}^{\mathrm{T}} \ln\big(\mathrm{f}(\mathrm{w}_{1\mathrm{i}}, \varphi)\big) + \sum_{\mathrm{i}=1}^{\mathrm{T}} \ln\big(\mathrm{g}(\mathrm{w}_{2\mathrm{i}}, \gamma)\big) + \sum_{\mathrm{i}=1}^{\mathrm{T}} \ln\big(\mathrm{c}(\mathrm{F}(\mathrm{w}_{1\mathrm{i}}; \varphi), \mathrm{G}(\mathrm{w}_{2\mathrm{i}}, \gamma); \delta)\big) \quad (2.13)$$

where $\theta \in \Theta$ be the parameter vector to be estimated, this parameter vector can be split up into the parameters for the marginal and the copula function as follows $\theta = [\varphi', \gamma', \delta']'$. $\varphi \in \phi$ denotes the parameter(s) of $\mathrm{f}(\mathrm{w}_{1\mathrm{i}})$, $\gamma \in \Gamma$ denotes the parameter(s) of $\mathrm{g}(\mathrm{w}_{2\mathrm{i}})$, and $\delta \in \Delta$ denotes the parameter(s) of $\mathrm{c}\big(\mathrm{F}(\mathrm{w}_{1\mathrm{i}}), \mathrm{G}(\mathrm{w}_{2\mathrm{i}})\big)$. Based on Manner (2007), the EML estimator is then given as

$$\hat{\theta} = \arg\max \mathrm{L}_{\mathrm{W}_1\mathrm{W}_2,} \quad (2.14)$$

The second method is the two step method (Inference Functions for Margins (IFM)), which first estimates the parameters φ and γ of the marginal densities are estimated by MLE, that is

$$\hat{\varphi} = \arg\max \mathrm{L}_{\mathrm{W}_1} \quad (2.15)$$

$$\hat{\gamma} = \arg\max \mathrm{L}_{\mathrm{W}_2} \quad (2.16)$$

Using these estimates to transform the marginals into uniform (0, 1) variables, one can now estimates the copula parameter(s) δ by maximizing the copula density, that is

$$\hat{\delta} = \arg\max \mathrm{Lc}(\hat{\varphi}, \hat{\gamma}, \delta). \tag{2.17}$$

Heckman (1979) proposed an estimation in parametric sample selection models which is typically based on the two-step estimation procedure.

The third method is the semiparametric estimation (canonical maximum likelihood method (CML)) which leaves the marginal densities unspecified and uses the empirical probability integral transform in order to obtain the uniform marginal needed to estimate the copula parameters (Manner, 2007). The empirical distribution function is defined as Payandeh et al, (2013).

$$\hat{F}(\cdot) = \frac{1}{T}\sum_{t=1}^{T} 1_{(X_t \leq \cdot)}, \tag{2.18}$$

where $1_{(X_n \leq \cdot)}$ is the indicator function. The copula parameter δ can be estimated by maximizing the log-likelihood function of the copula density using the transformed variables given by (Araichi, Peretti & Belkacem, n.d.)

$$L_C(\delta) = \sum_{t=1}^{T} \ln\left(c\left(\hat{F}(w_{1t}), \hat{G}(w_{2t})\right)\right)$$

$$= \sum_{t=1}^{T} \ln\left(c(\hat{u}_{1t}, \hat{u}_{2t})\right). \tag{2.19}$$

The semi parametric estimator is then give as

$$\hat{\delta} = \arg\max L_C(\delta) \tag{2.20}$$

Chen, Fan and Tsyrennikov (2006) proposed sieve maximum likelihood estimation procedure for semiparametric multivariate copula models. The proposed estimation achieved efficiency gains in finite samples, especially when prior information of the marginal distribution is incorporated.

The last two methods are nonparametric ways of estimating the copula. The first one is estimating the empirical copula directly from the data, leaving the whole specification nonparametric. The empirical copula is the function C_n given as (Manner, 2007)

$$C_n\left(\frac{i}{T},\frac{j}{T}\right) = \frac{h_i}{T}, \qquad \text{for } i, j = 1, 2, \ldots, T \tag{2.21}$$

where h_i denotes number of pairs (w_{1t}, w_{2t}) in the sample such that $w_{1t} \leq w_{1i}$ and $w_{2t} \leq w_{2i}$.

The other is obtaining a nonparametric estimate for Kendall's tau. The sample of Kendall's tau is then given as

$$\hat{\tau} = \frac{c-d}{c+d} \quad , \tag{2.22}$$

where c denote the number of concordant pairs in the sample and d the number of discordant pairs in the sample. Using this estimate and the relationship between Kendall's tau and the copula parameter to obtain nonparametric estimate if a closed form expression for this relationship exist. On the other hand, if a closed form expression does not exists one can still estimate the copula parameter by using the general form of Kendall's tau for Archimedean copulas and the form is given as

$$\tau_C = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,dt\,. \tag{2.23}$$

2.9 Model Selection

Model selection is a very important issue to compare the competing models, when one has to estimate the best model. The first thing is to assess the goodness-of-fit of the marginal distribution. The second is to test both independent identically distributed (i.i.d.) and the correct specification. The specification of the distribution can be tested by testing the transformed series for uniformity. This can be done by using the well-known Kolmogorov-Smirnov test or a Chi-square test. Many studies such as Chen et al. (2004), Chen and Fan (2005), Genest et al. (2006) and Fermanian (2005) as cited by Manner (2007) proposed for testing the copula specification. Moreover, Chen and Fan (2005) as cited by Huard, Évin and Favre (2005) proposed pseudo-likelihood ratio tests for selecting semi parametric multivariate copula models in which the marginal distributions are unspecified. Furthermore, Vuong test (1989), based on the likelihood ratio to non-nested models can be used. This test is an appropriate when copula models are non-nested to perform for each pair of copula.

Besides the various goodness-of-fit tests, there are a few model selection techniques such as Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), which allow to rank the copulas according to their fit in some way. In copula context, they are then defined as

$$\mathrm{AIC} = -2\mathrm{logL} + 2\mathrm{p} \tag{2.24}$$

$$\mathrm{BIC} = -2\mathrm{logL} + \mathrm{nparlog(N)}, \tag{2.25}$$

where $-2\mathrm{logL}$ is twice the negative log-likelihood value for the model, p is the number of estimated parameters, npar denotes the total number of parameters in the model and N is the total number of observations used to fit the model. Smaller values of AIC and BIC reflect an overall better fit.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

The purpose of the study was to fit a joint distributional model to adjust for sample selection bias of treatment adherence and viral suppression among HIV patients on ART in Namibia. The previous chapter was based on the methodological review on sample selection model. This chapter describes the methods and procedures used in the research including study design, study population, data description, and research ethics.

3.2 Study design

The study followed a retrospective cohort study using quantitative design. The study used data from Khomas region in the selected health facilities in Windhoek and data from health facilities in Erongo region.

3.3 Study Population

The population of the study was all patients initiated on ART from January 2010 to December 2015. The target age group was all patients aged above 16 years.

3.4 Data description

In this retrospective cohort study, the data was recorded from HIV patients on ART treatment at respective hospitals and clinics of Namibia, from January 2010 to December 2015. The data was obtained from Katutura Health Center and Okuryangava Center and the Ministry of Health and

Social Services (MoHSS) at the department of Centers for Disease Control and Prevention (CDC). The data was distributed in "raw" form, as extracted from the data system, meaning that most items are defined in the data set as string or character variables. The data was categorized into two datasets of which the first data was from Khomas region in the selected health facilities in Windhoek and the second data was from Erongo region.

In Windhoek, the data contained the following variables; facility, age, sex, marital status, WHO stage, CD4 at ART Initiation, date at ART Initiation, weight at ART Initiation, weight at 6 months, weight at 12 months, hepatitis B, regimen, adherence, treatment default, viral load at 6 months, viral load suppression at 6 months, viral load at 12 months and viral load suppression at 12 months.

In Erongo region, the data contained the following variables; unique number of patients, facility, age at ART Initiation, current age, sex, marital status, clinical stage, beginning of ART, year at ART Initiation, current body weight, CD4 cell count baseline, CD4 cell count at 12 months, viral load, ART adherence and patients who were still on ART.

Sample selection model using joint bivariate copula model was applied for both data. In Windhoek, a sample of 943 patients were selected of which 430 (45.6%) patients from Katutura Health Centre and 513 (54.4%) patients from Okuryangava Clinic. In Erongo region, a total of 34376 HIV patients were included in the study of which 11539 (33.6%) patients from Kuisebmend Hospital Center, 780 (2.3%) patients from Hakhaseb Clinic, 13040 (37.9%) patients from Omaruru Hospital Center and 9017 (26.2%) patients from Tamariskia Clinic.

3.5 Research ethics

The approval was obtained from the Research Ethical Committee of University of University (UNAM) and the Ministry of Health and Social Services (MoHSS) of the Republic of Namibia.

CHAPTER 4: SELF-SELECTION MODEL OF TREATMENT ADHERENCE AND VIRAL SUPPRESSION USING COPULA ON HIV/AIDS DATA IN WINDHOEK, NAMIBIA

4.1 Introduction for Self-Selection model of treatment adherence and viral suppression

This chapter gives a brief review of treatment adherence and viral suppression. It also presents the outcomes of joint bivariate copula based on self-selection model, and binary logistic regression model. The chapter includes methods, the Generalized Joint Regression Modelling (GJRM) package (Marra, & Radice, 2017b) used for analysis, results, discussion, conclusion and reference.

The impact of antiretroviral therapy (ART) adherence in suppressing the HIV virus and improve survival rates for those living with HIV has been well documented (Rougemont, 2009; Peltzer et al, 2010 Reis, Guerra & Lencastre, 2013). However, prosperous antiretroviral treatment is dependent on sustaining high rates of adherence (taking the correct dosage as it is described by health practitioner). Although, more strong ART regimens can allow for effective viral load suppression at moderate level of adherence. Many studies have shown that a very high adherence level required for ART regimens to work effectively is 95% (Castro, 2005; Peltzer et al, 2010; Reda & Biadgilign, 2012).

HIV/AIDS Program plays an important role in reaching viral load suppression and very strong adherence to people who are on antiretroviral therapy (ART) in Namibia. According to 2016 Ministry of Health and Social Services (MoHSS), very strong adherence of taking at least 95% of prescribed doses are required to achieve sustained suppression of HIV replication over time. In particular, the goal of ART is to achieve sustained and optimal viral suppression, both to improve client health and reduce HIV transmission. However, the challenge is to encourage and enable patients to take these medications correctly, in order to achieve their maximum level of adherence and viral load suppression (MoHSS, 2016). Thus, adherence to ART is a key to determine viral suppression, which is the reduction of a person's viral load to an undetectable level. However, sample selection based on ART adherence and viral suppression was determined.

The sample selection model as it is introduced by Heckman (1979) accounts for problems arising when the outcome of the endogenous variable is observed only for a selected sample (Van der Klaauw & Koning, 2003). However, the model is not only very useful when the outcome variable is only observed for a (nonrandom) part of the population, but it can easily be extended to the switching regression model. Therefore, sample selection model require a joint model so-called copula approach.

The copula approach is a modelling strategy whereby a joint distribution is induced by specifying marginal distributions, and a function that links them together (Smith, 2003). However, copula models provide a flexible framework for specifying the joint distribution of two random variables (Nelsen, 1999 as cited by Cunanan & Koopmeiners, 2014). Therefore, the framework allows the copula dependence parameter to be specified as a function of an additive predictor (Marra et al, 2017).

The link between treatment adherence and viral load suppression based on sample selection model has not been explored in Africa. Therefore, this study used a joint bivariate copula based on sample selection to accommodate dependence structure between ART treatment adherence and viral load suppression. A joint bivariate copula is a bivariate distributional margins which separates the choice of dependence between selection and outcome equations (Kuethe et al., 2009).

4.2 Methods for Self-Selection model of treatment adherence and viral suppression

4.2.1 Study setting

Baseline evaluation data were obtained from Katutura Hospital Center (KHC) and Okuryangava Clinic. Katutura Hospital Center and Okuryangava Clinic are the state health care services in Windhoek, the capital city of Namibia. Namibia has 343 health care services, of those only seven state clinics in Windhoek: Khomasdal, Robert Mugabe, Donkerhoek, Black chine, Wanaheda, Okuryangava and Hakahana clinics. These facilities provide free HIV care, including counselling and give ART medications for free. Due to the nature of the evaluation, the study therefore selected only two facilities in Windhoek.

4.2.2 Study design

The design was a retrospective cohort study using quantitative method. The data consisted of 943 patients initiated on antiretroviral therapy (ART) from January 2010 to December 2015 at Khomas region in the selected health facilities in Windhoek. The target age group was all patients aged 16 years and above.

4.2.3 Study Population

The study population consisted of all HIV patients aged 16 and above on treatment attending the ART services from January 2010 to December 2015 at two facilities in Windhoek-capital city of Namibia.

4.2.4 Sample size and Sampling

Two health facilities were purposively selected. In the second stage of sampling, participants in these facilities were sampled systematically whereby every third file in the facilities was selected until the desired sample size was achieved.

The sample size was estimated at 942, to detect a relative risk of 2 between the exposed and non-exposed (Ntumba, 2018). The following assumptions are:

- Two-sided significance of 5%; a power of 80%, ratio of sample equal to 1 between the exposed (less than 95% adherence to ART) and un-exposed (at least 95% adherence to ART);
- Percent of unexposed with outcome equal to 5% and percent of exposed with outcome equal to 10%.

Therefore, the sample size for exposed and unexposed was estimated at 471 each. The sample was selected from two health facilities using proportional stratified sampling based on the client population in each clinic. A total of 943 patients were selected of which 430 (45.6%) patients from Katutura Health Centre and 513 (54.4%) patients from Okuryangava Clinic. They summarized in table 2.

Table 2: Number of patients recruited per facility in Khomas region

Facility	Frequency	Percentage	Cumulative percentage
Katutura Health Center	430	45.6	45.6
Okuryongava Clinic	513	54.4	100
Total	943	100	

Eligibility criteria: Retrospective observational cohort include all HIV patients ever started on ART between January 1, 2010 to December 31, 2015 and aged 16 to 50 years.

Ineligibility criteria: Retrospective observational cohort exclude all HIV patients ever started on ART before January 2010 and those started after December 31, 2015 and aged less than 16 years and greater than 50 years.

4.2.5 Variables

The table 3 presents the description of the outcome and selection variables (response variables) and observed confounders (explanatory variables) for selected facilities in Windhoek.

Table 3: Katutura Hospital Center and Okuryongava Clinic HIV data; description of the outcome and selection variables, and observed confounders

Variable name	**Values of the variable**	**Type**
Selection		
Treatment adherence	1 = yes if >95% adherence, 0 = no if <95% adherence	Categorical
Outcome		
Viral load suppression	1 = yes, 2 = no	Categorical
Viral load	None	Continuous
Demographic		
Age	1 = < 25 years, 2 = 25-49 years , 3 = ≥50 years	Categorical
Sex	1 = male, 2 = female	Categorical
Marital status	1 = single, 2 = married, 3 = others	Categorical
Health-related		
Facilities	1 = KHC, 2 = OC	Categorical
Year of initiation		
WHO stage	1 = stage 1, 2 = stage 2, 3 = stage3, and 4 = stage 4	Categorical
CD4 at ART initiation	1 = <200 CD4 counts, 2 = ≥200 CD4 counts	Categorical
Weight at ART initiation	1 = <50 weights, 2 = 50-70 weights , 3 = >70 weights	Categorical
Weight at 6 months	1 = <50 weights, 2 = 50-70 weights, 3 = >70 weights	Categorical
Weight at 12 months	1 = <50 weights, 2 = 50-70 weights, 3 = >70 weights	Categorical

4.3 Data analysis for Self-Selection model of treatment adherence and viral suppression

In this section, the data analysis were presented in four sub-sections; descriptive analysis, binary logistic regression analysis, Heckman-type model using a joint bivariate copula model and the GJRM package Implementing in R programming.

4.3.1 Descriptive analysis

In the analysis of time to ART treatment adherence and time to viral load suppression, descriptive analysis was used to compare demographic and clinical characteristics by ART adherence and viral load suppression at 12 months. The patient characteristics were compared by whether or not adherence to ART and also compared by whether or not they even achieve viral suppression. Then the patients with at least 95% adherence and those with less than 95% adherence and achieved viral load suppression at 12 months were recorded.

4.3.2 Binary logistic regression analysis

Univariate and multivariate logistic regression analysis were used to assess the factors associated with ART adherence and viral load suppression. In multivariate analysis, adjustments were made for sex, age, facility, WHO stages, CD4 cell count at ART initiation, weight at ART initiation, weight at 6 months, weight at 12 months and ART regimen. The equations for ART treatment adherence and viral suppression were then given as

$$\log \frac{P(\text{adherence} \geq 95\%)}{P(\text{adherence} < 95\%)} = \alpha_{10} + \sum_{i=1}^{n} \alpha_{1i} x_{1i} \qquad (4.1)$$

and

$$\log\frac{\mathrm{P(viral\ suppression)}}{\mathrm{P(no\ viral\ suppression)}} = \alpha_{20} + \sum_{1=i}^{n} \alpha_{2i}x_{2i}\ , \tag{4.2}$$

where α_{10}, α_{20}, α_{1i} and α_{2i} are unknown parameters, x_{1i} and x_{1i} are the explanatory variables mentioned above and Σ indicates the summation of the associated parameters (excluding intercepts) and explanatory variables.

4.3.3 Heckman-type model using a joint bivariate copula model

Heckman-type selection analysis using a joint bivariate copulas was used to investigate the dependence structure between ART treatment adherence and viral suppression. The aim is to adjust estimates for endogeneity in the outcomes. The semi-parametric predictors for the selection (Adherence to ART) and outcome (viral load suppression) equations and for the copula parameter (θ) were then specified as:

$$\log\frac{\mathrm{P(adherence} \geq 95\%)}{\mathrm{P(adherence} < 95\%)} = \alpha_{10} + \alpha_{11}(\mathrm{sex}) + \alpha_{12}(\mathrm{age_i}) + \alpha_{13}(\mathrm{marital\ status_i}) + \alpha_{14}(\mathrm{WHOstage_i}) + \alpha_{15}(\mathrm{CD4counts_i}) + \alpha_{16}(\mathrm{facilities_i})$$

$$\log\frac{\mathrm{P(viral\ suppression)}}{\mathrm{P(no\ viral\ suppression)}} = \alpha_{20} + \alpha_{21}(\mathrm{sex}) + \alpha_{22}(\mathrm{age_i}) + \alpha_{23}(\mathrm{marital\ status_i}) + \alpha_{24}(\mathrm{WHOstage_i}) + \alpha_{25}(\mathrm{CD4counts_i}) + \alpha_{16}(\mathrm{facilities_i})$$

$$\mathrm{Copula\ parameter}\ (\theta) = \alpha_{30} + \alpha_{31}(\mathrm{facilities_i}), \qquad \mathrm{for}\ i = 1, \dots, n \tag{4.3}$$

The equivalence standard selection model that relies on jointly bivariate normal distribution is to use the copula approach (Marra et al. 2015). Therefore, In order to allow for non-linear association between the adherence to ART and viral load suppression equations, we model the dependency of the error terms in the two equations using copulas (McGovern, Bärnighausen, Marra & Radice, 2015).

The copula function is then specified as:

$$\mathrm{F}\left(\log\frac{\mathrm{P(adherence{\geq}95\%)}}{\mathrm{P(adherence{<}95\%)}},\log\frac{\mathrm{P(viral\ suppression)}}{\mathrm{P(no\ viral\ suppression)}}\ \right)=\mathrm{C}\left(\mathrm{F}_1\left(\log\frac{\mathrm{P(adherence\geq 95\%)}}{\mathrm{P(adherence<95\%)}}\right),\mathrm{F}_2\left(\log\frac{\mathrm{P(viral\ suppression)}}{\mathrm{P(no\ viral\ suppression)}}\right);\theta\right),\ (4.4)$$

where θ is an association parameter measuring the dependency between two marginal (Marra & Radice, 2016).

The most commonly known model selection criterions, AIC and BIC (Equations 2.24 & 2.25) suggest a better model with smaller AIC and BIC.

4.3.4 Implementation in R: GJRM package

The R package gjrm implements maximum likelihood inference for copulas, which allows the specification of a wide range of non-random selection models and the dependence structure. The non-random sample selection model is specified as the combination of a distributional assumption and structural assumptions (Wyszynski & Marra, 2017).

The function gjrm ( ) provides an interface to fit sample selection models for binary distributional margins. The function is given as

$$\mathrm{gjrm(formula,\ \ data = list(), \dots)} \qquad (4.5)$$

where formula is a list of two compulsory equations and an optional extra formula for the dependence parameter, and data is a data frame, list containing the variables in the model. These are generalized linear model (glm) formulae. In the context of this study, smooth components can be excluded in the equations because the data contain only categorical variables. The results of copula model specify the mean, variance and shape of the selection, and outcome equations and copula parameter (Marra & Radice, 2017). The families of copula linking the selection and outcome equations are found in gjrm and can be specified through BivD excluding Normal. The families of copula are Normal, Frank, FGM, AMH, Student-t, Joe, Gumbel and Clayton and rotations by 0, 90, 180 and 270 degrees.

4.4 Results for Self-Selection model of treatment adherence and viral suppression

In this section, the results were presented in three sub-sections; descriptive analysis, Univariate and multivariate binary analysis and Heckman-type using bivariate copula analysis.

In the first section, descriptive analysis compared the demographic and clinical characteristics by ART adherence and viral load suppression at 12 months. In the second section, univariate and multivariate logistic regression analysis was used to investigate the factors associated with ART adherence and viral load suppression. In the last section, a simultaneous model of the selection (ART adherence) and outcome (viral suppression) equations was explored based on copula in order to highlight the dependence structure between ART adherence and viral load suppression.

4.4.1 Descriptive analysis

The study used data from Katutura Hospital Center (KHC) and Okuryongava Clinic (OC) both in Khomas region. A total of 943 patients (45.6% from KHC and 54.4% from OC) was included in the study. Baseline demographic characteristics of all infected patients aggregated by adherence to ART and viral load suppression at 12 months were presented in Table 4 and 5. Two hundred and eighty two (72.2%) female and 369 (66.6%) male were adhered to ART treatment.

In terms of ART adherence, table 4 shows that OC received large number of patients 365 (71.2%) who adhered to ART during January 2010 to December 2015. Regarding sex of infected patients, out of total of 943 infected patients 369 (66.6%) were female adherent to ART as compared to male group. The highest age group for those that adhered to ART was aged between 25 to 49 years of age with 545 (70.6%). Table 5 also depicts that 689, 93, 101 and 27 patients were at baseline clinical stage-I, stage-II, stage-III and stage-IV respectively. Totally, 490 (71.1%) of infected patients who adhered to ART occurred in clinical stage-I at baseline time in comparison with other three baseline clinical stages. The majority of those adhered or not adhered to ART were single (table 4).

Table 4: Demographic and clinical characteristics by adherence and viral suppression at 12 months

Variables	Adherence to ART treatment						Viral suppression 12 months					
	Yes		No		Total		Yes		No		Total	
	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
Facility												
KHC	286	66.7	143	33.3	429	100	201	86.3	32	13.7	233	100
OC	365	71.2	148	28.8	513	100	244	85.9	40	14.1	284	100
Sex												
Male	282	72.7	106	27.3	388	100	193	84.3	36	15.7	229	100
Female	369	66.6	185	33.4	554	100	252	87.5	36	12.5	288	100
Age												
< 25 years	79	61.2	50	38.8	129	100	59	92.2	5	7.8	64	100
25-49 years	545	70.6	227	29.4	772	100	366	85.1	64	14.9	430	100
≥ 50 years	26	65.0	14	35.0	40	100	19	86.4	3	13.6	22	100
Marital Status												
Single	536	68.5	246	31.5	782	100	364	86.1	59	13.9	423	100
Married	82	71.3	33	28.7	115	100	58	85.3	10	14.7	68	100
Others	4	66.7	2	33.3	6	100	2	100	0	0.0	2	100
WHO stage												
Stage-I	490	71.1	199	28.9	689	100	354	87.8	49	12.2	403	100
Stage-II	65	69.9	28	30.1	93	100	36	80.0	9	20.0	45	100
Stage-III	65	64.4	36	35.6	101	100	40	80.0	10	20.0	50	100
Stage-IV	15	55.6	12	44.4	27	100	6	66.7	3	33.3	9	100
CD4 at ART initiation												
< 200	236	66.7	118	33.3	354	100	152	79.2	40	20.8	192	100
≥ 200	383	73.8	136	26.2	519	100	271	90.3	29	9.7	300	100

NOTE: KHC = Katutura Hospital Center, OC = Okuryangava Clinic

Table 5: Demographic and clinical characteristics by adherence and viral suppression at 12 months (continues to table 4)

Variable	Adherence to ART treatment						Viral suppression at 12 months					
	Yes		No		Total		Yes		No		Total	
	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
Weight at ART initiation												
< 50	85	60.7	55	39.3	140	100	51	79.7	13	20.3	64	100
50-70	431	72.1	167	27.9	598	100	293	85.4	50	14.6	343	100
> 70	115	71.4	46	28.6	161	100	87	92.6	7	7.4	94	100
Weight at ART 6 months												
< 50	60	65.9	31	34.1	91	100	37	80.4	9	19.6	46	100
50-70	418	77.4	122	22.6	540	100	286	86.7	44	13.3	330	100
> 70	135	81.3	31	18.7	166	100	101	91.7	10	8.3	111	100
Weight at ART 12 months												
< 50	57	76.0	18	24.0	75	100	41	87.2	6	12.8	47	100
50-70	389	84.7	70	15.3	459	100	262	85.9	43	14.1	305	100
> 70	141	84.9	25	15.1	166	100	107	88.4	14	11.6	121	100
Hepatitis B												
Positive	68	73.1	25	26.9	93	100	41	87.2	6	12.8	47	100
Negative	532	69.6	232	30.4	764	100	374	85.6	63	85.6	437	100
ART Regimen												
TDF/FTC/EFV	189	67.5	91	32.5	280	100	146	91.3	14	8.7	160	100
TDF/3TC/EFV	255	70.1	109	29.9	364	100	171	85.9	28	14.1	199	100
TDF/3TC/NVP	186	71.0	76	29.0	262	100	114	81.4	26	18.6	140	100
OTHER	18	69.2	8	30.8	26	100	12	75	4	25	16	100

NOTE: TDF = tenofovir, FTC = emtricitabine, EFV = efavirenz, 3TC = lamivudine, NVP = nevirapine, OTHER = zidovudine (AZT) and stavudine (d4T)

In terms of viral load suppression, a total of 244 (85.9%) infected patients achieve viral load suppression from OC of which the vast majority were female. Of those that achieve suppression, three hundred and sixty six (85.1%) patients were aged between 25 to 49 years, 423 (86.1%) were single, 271 (90.3%) patients have CD4 load more than 200 at ART initiation and 41 (87.2%) patients with available hepatitis B. Table 5 shows that patients with body weights between 50 to 70, regardless of 293 (85.4%) ART initiation, 286 (86.7%) six months and 262 (85.9%) twelve months, have achieved viral load suppression. Prior to receipt ART, 87.8% patients were classified as having WHO clinical stage-I, 80.0% of the patients had stage-II, 80.0% had stage-III, and 66.7%

had stage-IV. Table 5 also shows a combination of ART regimens of which the majority of patients were treated with pill containing tenofovir (TDF) either emtricitabine (FTC) or nevirapine (NVP). In general, the majority of adherent patients had undetectable viral load, which implies that being adherence are more likely to achieve viral load suppression.

4.4.2 Binary regression analysis of ART adherence

This sections provide the results of unadjusted univariate and adjusted multivariate odd ratios (OR) of the factors associated with adherence to ART treatment. In adjusted multivariate analyses, table 6 revealed that patients with age between 25 to 49 years of age (OR 1.201, 95% CI 0.320-4.510), female (OR 1.676, 95% CI 1.029-2.730) and those who initiated ART in Okuryongava Clinic (OR 1.405, 95% CI 0.870-2.267) were more likely to adhere to ART. Patients with CD4 count greater than or equal to 200 at ART initiation (OR 1.402, 95% CI 0.846-2.321) were more likely to be adherent to ART treatment than those individual with less than 200 CD4 count at ART initiation. Patients on WHO stage-II (OR 0.771, 95% CI 0.191-3.119), stage-III (OR 0.580, 95% CI 0.121-2.782) and stage-IV (OR 0.739, 95% CI 0.167-3.266) were less likely to adhere to ART treatment than individual on stage-I.

Table 6 further revealed that patients on baseline antiretroviral 50 to 70 weight (OR 0.960, 95% CI 0.258-3.565) and those with greater than 70 weight (OR 0.721, 95% CI 0.265-1.962) were less likely to adhere to ART treatment than those individual with less than 50 weight. Patients on ART six (6) months 50 to 70 weight (OR 1.238, 95% CI 0.252-6.087) and those with greater than 70 weight (OR 1.595, 0.497-5.115) were less likely to adhere to ART treatment than those individual with less than 50 weight.

Table 6: Univariate and multivariate odd ratios (OR) of the factors associated with adherence

Factors	**Unadjusted univariate OR (95% CI)**	**P-value**	**Adjusted multivariate OR (95% CI)**	**P-value**
Sex (ref. male)				
Female	0.75 (0.564 , 0.997)	0.047	1.676 (1.029 , 2.730)	0.038
Age (ref. < 25 years)				
25-49 years	1.175 (0.561 , 2.464)	0.669	1.201 (0.320 ,4.510)	0.786
≥50 years	0.774 (0.397 , 1.509)	0.451	0.852 (0.266 , 2.728)	0.787
Facility (ref. KHC)				
OC	1.233 (0.935 , 1.627)	0.138	1.405 (0.870 ,2.267)	0.164
WHO stage (ref. stage-I)				
stage-II	0.508 (0.233 , 1.104)	0.087	0.771 (0.191 , 3.119)	0.715
stage-III	0.538 (0.224 , 1.297)	0.167	0.580 (0.121 , 2.782)	0.496
stage-IV	0.692 (0.293 , 1.638)	0.403	0.739 (0.167 , 3.266)	0.690
CD4 at ART Initiation (ref. < 200)				
≥ 200	1.408 (1.048 , 1.892)	0.023	1.402 (0.846 , 2.321)	0.190
Weight at ART Initiation (ref. < 50)				
50-70	1.618 (0.999 , 2.618)	0.050	0.960 (0.258 , 3.565)	0.951
> 70	0.969 (0.659 , 1.425)	0.872	0.721 (0.265 , 1.962)	0.523
Weight at 6 months (ref. < 50)				
50-70	2.250 (1.256 , 4.032)	0.006	1.238 (0.252 , 6.087)	0.793
> 70	1.271 (0.819 , 1.972)	0.285	1.595 (0.497 , 5.115)	0.432
Weight at 12 months (ref. < 50)				
50-70	1.781 (0.903 , 3.514)	0.096	1.645 (0.453 , 5.974)	0.449
> 70	1.015 (0.618 , 1.666)	0.953	0.969 (0.384 , 2.446)	0.947
Regimen (ref. TDF/FTC/EFV)				
TDF/3TC/EFV	1.083 (0.454 , 2.585)	0.857	0.412 (0.139 , 1.218)	0.109
TDF/3TC/NVP	0.962 (0.406 , 2.278)	0.929	0.509 (0.181 , 1.434)	0.201
Other	0.919 (0.383 , 2.204)	0.851	0.488 (0.169 , 1.416)	0.187

Note: this table presents the univariate and multivariate odd ratios (OR) of the factors associated with adherence to ART for 943 patients receiving antiretroviral therapy (ART) at Katutura Hospital Center (KHC) and Okuryongava Clinic (OC), Namibia. CI = Confidence interval, “ref.” = reference, “P-value” = probability value

Patients on ART twelve (12) months greater than 70 weight (OR 0.969, 95% CI 0.384-2.446) were less likely to adhere to ART treatment than those individual with weight less than 50 weight). Finally, patients on baseline ART regimens having tenofovir (TDF), emtricitabine (FTC) and nevirapine (NVP) (OR 0.412, 95% CI 0.139-1.218) and tenofovir (TDF), lamivudine (3TC) and efavirenz (EFV) (OR 0.509, 95% CI 0.181-1.434) were less likely to adhere to ART treatment than those whose first regimen contained tenofovir (TDF), emtricitabine (FTC) and efavirenz (EFV).

4.4.3 Binary regression analysis of viral suppression

Table 7 shows the unadjusted univariate and adjusted multivariate odd ratios (OR) of the factors associated with viral load suppression. In adjusted multivariate analyses, patients greater than 50 years (OR 4.798, 95% CI 0.540-42.618) and female (OR 1.003, 95% CI 0.523-1.922) were less likely to ever achieve viral load suppression (Table 7). Higher CD4 cell counts (greater than or equal to 200) at ART initiation was associated with viral suppression (OR 2.734, 95% CI 1.397-5.352, p-value = 0.003). Patients receiving treatment in Okuryongava clinic (OR 0.790 95% CI 0.408-1.527) were less likely to achieve suppression than those individual in Katutura hospital center. Patients on baseline clinical stage-II (OR 0.424, 95% CI 0.070-2.554), stage-III (OR 0.468, 95% CI 0.065-3.366) and stage-IV (OR 0.683, 95% CI 0.100-4.678) were less likely to achieve viral suppression than individual on baseline clinical stage-I (table 7).

Table 7: Univariate and multivariate odd ratios of the factors associated with viral suppression

Factors	**Unadjusted univariate OR (95% CI)**	**P-value**	**Adjusted multivariate OR (95% CI)**	**P-value**
Sex (ref. male)				
Female	1.306 (0.793 , 2.150)	0.294	1.003 (0.523 , 1.922)	0.994
Age (ref. < 25 years)				
25-49 years	0.537 (0.117 , 2.459)	0.423	1.498 (0.108 , 20.727)	0.763
> 50 years	1.107 (0.318 , 3.851)	0.872	4.798 (0.540 , 42.618)	0.159
Facility (ref. KHC)				
OC	0.971 (0.588 ,1.603)	0.909	0.790 (0.408 , 1.527)	0.483
WHO stage (ref. stage-I)				
stage-II	0.277 (0.067 , 1.143)	0.076	0.424 (0.070 , 2.554)	0.349
stage-III	0.500 (0.104 , 2.395)	0.386	0.468 (0.065 , 3.366)	0.450
stage-IV	0.500 (0.106 , 2.355)	0.381	0.683 (0.100 , 4.678)	0.698
CD4 at ART Initiation (ref. <200)				
≥ 200	2.459 (1.465 , 4.127)	0.001	2.734 (1.397 , 5.352)	0.003
Weight at ART Initiation (ref. <50)				
50-70	3.168 (1.187 , 8.456)	0.021	3.641 (0.520 , 25.510)	0.193
> 70	2.121 (0.928 , 4.846)	0.075	3.650 (0.739 , 18.025)	0.112
Weight at 6 months (ref. <50				
50-70	2.457 (0.926 , 6.521)	0.071	6.105 (0.813 , 45.864)	0.079
>70	1.554 (0.754 , 3.202)	0.232	1.804 (0.392 , 8.303)	0.449
Weight at 12 months (ref. <50)				
50-70	1.118 (0.403 , 3.108)	0.830	0.192 (0.031 , 1.201)	0.078
>70	1.254 (0.659 , 2.388)	0.490	0.571 (0.173 , 1.881)	0.357
Regimen (ref. TDF/FTC/EFV)				
TDF/3TC/EFV	0.288 (0.082 , 1.012)	0.052	0.441 (0.088 , 2.212)	0.320
TDF/3TC/NVP	0.491 (0.148 , 1.631)	0.246	0.900 (0.197 , 4.103)	0.892
Other	0.684 (0.204 , 2.292)	0.538	1.158 (0.250 , 5.362)	0.851

Note: this table presents the univariate and multivariate odd ratios (OR) of the factors associated with viral load suppression for 943 patients receiving antiretroviral therapy (ART) at Katutura Hospital Center (KHC) and Okuryongava Clinic (OC), Namibia. CI = Confidence interval, KHC = Katutura hospital center, OC = Okuryongava clinic, “ref.” = reference, “P-value” = probability value.

Patients’ body weight at ART initiation between 50 to 70 (OR 3.641, 95% CI 0.520-25.510), greater than 70 body weight (OR 3.650, 95% CI 0.739-18.025), body weight at six months between 50 to 70 (OR 6.105, 95% CI 0.813-45.864) and those with greater than 70 weight (OR 1.804, 95% CI 0.392-8.303) were less likely to achieve viral load suppression than those individual with less than 50 weight. Patients’ body weight at twelve months between 50 to 70 weights (OR 0.192, 95%

CI 0.031-1.201) and those with greater than 70 weight (OR 0.571, 95% CI 0.173-1.881) were less likely to achieve viral load suppression than those individual with less than 50 weight. Finally, patients on initial antiretroviral regimens containing tenofovir (TDF), emtricitabine (FTC) and nevirapine (NVP) (OR 0.441, 95% CI 0.088-2.212) and tenofovir (TDF), lamivudine (3TC) and efavirenz (EFV) (OR 0.900, 95% CI 0.197-4.103) were less likely to achieve viral load suppression than those whose first regimen contained tenofovir (TDF), emtricitabine (FTC) and efavirenz (EFV).

4.4.4 Heckman-type model of ART adherence and viral suppression

Table 8 summarizes the estimation of the association parameters together with 95% confidence interval bounds from the selection model based on the Normal, Frank, FGM, Ali-Mikhail-Haq (AMH), Student-t and the rotated versions of Joe, Clayton and Gumbel models. The copula models are summarized in Table 1 (Chapter 2). There is support for the hypothesis of selection bias, with a positive association for copula models such as Normal, Frank, AMH and Student-t, and the 95% CI for τ includes zero. These copula models were statistically significant at 5% level of significance which implies that there is dependence structure between the two distributional margins. FGM, and the rotated copula models indicate negative association, which implies that there is a strong negative association between adherence to ART treatment and viral load suppression. In fact, Joe, Gumbel and Clayton exhibit negative skew distribution. For more details see table 8 below.

Table 8: Measure of the association between ART adherence and viral load suppression in Heckman-type model

Copula Model	**Theta ($\hat{\theta}$)**	**95% CI**	**Kendall's Tau ($\hat{\tau}$)**	**95% CI**
Normal	0.049	(-0.857 , 0.851)	0.034	(-0.671 , 0.669)
Frank	0.645	(-7.570 , 7.560)	0.060	(-0.583 , 0.544)
FGM	-0.076	(-0.985 , 0.955)	-0.017	(-0.219 , 0.212)
AMH	0.217	(-0.872 , 0.952)	0.082	(-0.161 , 0.307)
Student-t	0.074	(-0.833 , 0.863)	0.052	(-0.631 , 0.711)
Joe 90	-1.460	(-24.20 , -1.090)	-0.173	(-0.734 , -0.048)
Joe 270	-2.690	(-39.10 , -1.230)	-0.356	(-0.943 , -0.100)
Gumbel 90	-1.370	(-42.70 , -1.100)	-0.222	(-0.860 , -0.078)
Gumbel 270	-1.640	(-39.50 , -1.080)	-0.308	(-0.924 , -0.068)
Clayton 90	-1.740	(-58.10 , -0.240)	-0.953	(-0.953 , -0.096)
Clayton 270	-0.532	(-46.90 , -0.130)	-0.179	(-0.809 , -0.057)

Note: this table presents the estimation of the association parameters with 95% confidence intervals (in brackets) from the selection models based on the normal, Frank, FGH, AMH, Student-t and Clayton, Joe and Gumbel copulas rotated by 90, 270 degrees with Bivariate outcome margins.

To examine the presence of tail dependence in sample selection bias. The results further show that Joe, Gumbel and Clayton rotated 90 and 270 degrees indicate lower tail dependence by looking at 95% confidence interval for Kendall's Tau ($\hat{\tau}$). The Normal, Frank FGM, AMH, Student-t copula model indicate asymmetric dependence between ART adherence and viral load suppression margins.

Table 9 shows the AIC and BIC for the sample selection models with different copulas. These are Normal, Frank, FGM, AMH, Student-t and the 90 and 270 degrees rotated versions of Joe, Gumbel and Clayton. The copula with lowest AIC and BIC values in which viral suppression was used as outcome and adherence to ART treatment was used as selection correspond to Frank followed by Normal. Although AIC and BIC revealed the best model as Frank, Vuong test and Clarke test did not indicate any preferred model. See table 9 below.

Table 9: Akaike and Bayesian information criteria (AIC and BIC) for copula functions in Heckman-type model

Model	AIC	BIC	Model	AIC	BIC
Normal	778.693	895.675	Joe 270	779.072	896.055
Frank	**778.567**	**895.549**	Gumbel 90	779.203	896.186
FGM	778.864	895.847	Gumbel 270	779.163	896.146
AMH	779.247	896.229	Clayton 90	779.073	896.056
Student-t	778.900	895.882	Clayton 270	779.228	896.211
Joe 90	779.220	896.202			

Note: this table presents the Akaike and Bayesian information criteria (AIC and BIC) from the sample selection models based on Normal, Frank, FGM, AMH, Student-t, Joe, Gumbel and Clayton and rotations by 90 and 270 degrees.

Table 10 presents the results of the Frank copula model which claim to be a better model. For both logits, the results revealed that patients who received ART treatment at Okuryangava clinic were associated with ART adherence but did not associated with viral suppression. The results also revealed that having CD4 counts of more than 200 were associated with both ART treatment adherence (p-value = 0.008) and viral suppression (p-value = 0.022). The results further revealed that patients who adhered at least 95% were not associated with viral suppression (p-value = 0.609). Baseline clinical stages for both margins regardless stage-II, stage-III and stage-IV and ART regimens containing a combination of TDF/3TC/EFV, TDF/3TC/NVP and other regimens were not statistical significant. Patients who have age between 25 to 49 years and those great than 50 years of age were no associated with ART treatment adherences and viral suppression in both margins.

Table 10: Frank copula parameter estimates of the Heckman-type model

Factors	**ART treatment adherence**			**Viral suppression**		
	Estimate	**Z-value**	**P-value**	**Estimate**	**Z-value**	**P-value**
Intercept	-0.429	-1.520	0.129	-0.557	-2.863	0.004
Sex (ref. Male)						
Female	-0.019	-0.125	0.901	-0.035	-0.211	0.833
Age (ref. <25 years)						
25-49 years	-0.360	-1.677	0.094	0.330	1.099	0.272
≥50 years	0.062	0.168	0.866	0.231	0.539	0.597
Facility (ref. KHC)						
OC	-0.273	-1.810	0.070	0.177	0.701	0.484
WHO Stage(ref. Stage-I)						
Stage-II	-0.097	-0.377	0.706	-0.032	-0.121	0.903
Stage-III	0.041	0.175	0.861	0.093	0.391	0.696
Stage-IV	-0.250	-0.413	0.679	0.626	1.267	0.205
Regimen (ref. TDF/FTC/EFV)						
TDF/3TC/EFV	0.047	0.265	0.791	0.189	0.942	0.346
TDF/3TC/NVP	0.055	0.281	0.778	0.298	1.460	0.144
Other	0.265	0.693	0.488	0.625	1.468	0.142
CD4 at ART Initiation(ref. <200)						
≥200	-0.406	-2.640	0.008	-0.384	-2.299	0.022
ART adherence (ref. <95%)						
≥95%				0.473	0.511	0.609

Note: this table presents the parameter estimates of the Heckman-type for the best model

4.5 Discussion for Self-Selection model of treatment adherence and viral suppression

In this retrospective cohort study, the main objective was to fit a joint distributional model to adjust for sample selection bias of treatment adherence and viral load suppression among HIV patients on ART in Windhoek, Namibia. Precisely, to examine if there is any tail dependence in sample selection bias and to investigate the factors associated with viral load suppression and treatment adherence.

With regard to ART adherence, the results found that 71.2% patients treated in Okuryongava clinic, 70.6% patients' age between 25 to 49 years, 66.6% female and 68.5% never married were adherence to ART treatment. This is consistent with the study conducted by Peltzer, Friend-du Preez, Ramlagan and Anderson (2010) where 427 patients (82.9%) were 95% adherent in the month prior to the survey in KwaZulu-Natal, South Africa, nearly three-quarters (73.5%) of the 519 patients who had initiated ARVs were female. However, 62.2% patients' age between 30 and 49 years old and nearly three-quarters (73.3%) were never married. These findings are inconsistent with the study conducted by Noens, et al (2009) where 202 patients were recruited from 34 centers in Belgium, of whom 169 were evaluable and the majority of patients were male (55.0%) considered to be non-adherent. The results further found that ART adherence was less likely with currently preferred ART regimens having tenofovir (TDF) either emtricitabine (FTC) or nevirapine (NVP) and tenofovir (TDF) either lamivudine (3TC) or efavirenz (EFV) than those first regimen contained tenofovir (TDF) either emtricitabine (FTC) or efavirenz (EFV). These findings in line with those found by Sethi et al. (2003) and Reis et al. (2013).

With regard to viral load suppression, the results found that patients' age between 25 to 49 years old, female sex, never married and having CD4 more than 200 at ART initiation, patients who have hepatitis B and treated in Okuryongava clinic rather than in Katutura hospital center achieved viral load suppression. It also found that suppression was less likely with baseline clinical stage-II, stage-III and stage-IV than stage-I.

In the adjusted multivariate analyses (ART adherence), the results found that female patients were less likely to be associated with ART adherence than male. It also found that ART adherence has not been associated with age group, facility, WHO stage, CD4 load more than 200 at ART initiation, body weight at ART initiation between 50-70 weights and body weight at 6 months, and ART regimen. These results in line with those found by Adiga, Adiga and Malawadi, (2016) where adherence differ from CD4 counts but did not find significant between four treatment regimens and CD4 counts.

In the adjusted multivariate analysis (viral load suppression), the results found that Higher CD4 cell counts (greater than or equal to 200) at ART initiation was associated with viral suppression. It also found that patients treated with a combination ART regimens containing tenofovir (TDF) either emtricitabine (FTC) or nevirapine (NVP) and tenofovir (TDF) either lamivudine (3TC) or efavirenz (EFV) were less likely to achieve viral load suppression than those whose first regimen contained tenofovir (TDF) either emtricitabine (FTC) or efavirenz (EFV). These findings are consistent with those found by other studies in demonstrating the importance of the initial ART regimen in the rate of achieving viral load suppression at 12 months (Noens et al, 2009; Heestermans et al., 2016; Olowookere et al., 2016).

In the Heckman-type analysis, the framework allows the user to specify the dependence structure between the margins as documented by Marra and Radice, (2017). The results found that Normal and Frank models were statistically significant which implies that there is a positive dependence structure between the two distributional margins. It also found that FGM, AMH and Student-t and Joe, Gumbel and Clayton models rotated by 90 and 270 indicated negative association, which implies that there is a strong negative association between adherence to ART treatment and viral load suppression margins. These agreed with the results found by McGovern et al. (2015) where nine of copula models (normal, Frank, Student-t and Clayton, Joe and Gumbel copulas rotated by 90, 270 degrees) employed and show negative association. A study done by Cunanan (2014) disagreed with the present results of which the copula models show a very strong degrees of positive association.

4.6 Conclusions for Self-Selection model of treatment adherence and viral suppression

Some demographic and clinical characteristic factors have not been associated with ART adherence and viral load suppression. ART adherence and viral suppression were dependent, even though some demographic and clinical characteristics were not associated with ART adherence and viral suppression. The Normal, Frank, FGM, AMH and Student-t indicate a very positive dependence structure between the two distributional margins. The rotated models indicate a strong negative association between adherence to ART treatment and viral load suppression margins.

CHAPTER 5: JOINT MODEL OF TREATMENT ADHERENCE AND VIRAL LOAD USING COPULA ON HIV/AIDS DATA IN ERONGO REGION, NAMIBIA

5.1 Introduction for Joint Model of treatment adherence and viral load

This chapter gives a brief review of treatment adherence and viral load. It also presents the outcomes of joint bivariate copula based on self-selection model, multiple regression, proportional odds logistic with a ordered category of viral load and binary logistic regression model. The chapter includes methods, results, discussion, conclusion and reference.

Access to viral load testing is not only the right of every person living with HIV, but regular viral load testing also aids in achieving viral suppression by alerting providers to early signs of adherence problems (UNAIDS, 2016). However, national viral load strategies are urgently needed to bring these essential viral load diagnostic technologies to scale.

According to UNAIDS (2017), approximately 1.7 million newly HIV infected adults did not know their status and were not on ART treatment. Globally, there needs to be better access to viral load assays at affordable price, combined with effective laboratory systems and robust health services. A recent study found that measuring viral load at three months after treatment initiation, combined with an adherence intervention, reduced the risk of virologic failure by 22% compared to patients whose viral load was monitored six months after starting therapy.

Adherence to antiretroviral therapy (ART) is essential to ensure that viral load decreases in human body. According to Rougemont, Stoll, Elia and Ngang, (2009), the risk of disease progression and drug resistance may decrease. However, sustaining higher levels of adherence to ART over the long term requires accurate and consistent monitoring. In particular, according to Nachega, Mills and Schechter (2010), and Reda and Biadgilign (2012), sustaining adherence to ART is a challenge for countries in sub-Saharan Africa.

The link between treatment adherence and viral load based on self-selection model has not been explored in Africa. Therefore, this study used a joint bivariate copula based on self-selection to accommodate dependence structure between ART treatment adherence and viral load margins. A joint bivariate copula is a bivariate distributional margin which separates the choice of dependence between selection and outcome equations (Kuethe et al., 2009).

5.2 Methods for Joint Model of treatment adherence and viral load

5.2.1 Study setting

Erongo is one of the 14 regions of Namibia. It comprises the Swakopmund magisterial district up to the Ugab River and includes Walvis Bay, Omaruru and Karibib magisterial districts. This region is named after Erongo Mountain, a well-known landmark in this area and Namibia at large. The region contain four health facilities namely: Kuisebmund Hospital Center (KHC), Hakhaseb Clinic (HC), Omaruru Hospital Center (OHC) and Tamariskia Clinic (TC). All the facilities are offering ART services.

5.2.2 Study design

The design was a retrospective cohort study using quantitative method. The data consisted of 34376 patients initiated on antiretroviral therapy (ART) from January 2010 to December 2015 at Erongo region health facilities. The target age group was all patients aged 16 years and above.

5.2.3 Study population

The population group were all patients initiated on ART from January 2010 to December 2015, who were aged above 16 years.

5.2.4 Sample

The sample was all patients initiated on ART treatment from January 2010 to December 2015 Erongo region health facilities.

5.2.5 Variable

The table 11 presents the description of the outcome and selection variables (response variables) and observed confounders (explanatory variables) for health facilities in Erongo region.

Table 11: Erongo region HIV data; description of the outcome and selection variables, and observed confounders

Variables	Definition
Selection	
ART adherence	1 = at least 95% adherence to ART treatment
Outcome	
Viral load	Viral load
Demographic	
AgeIn	Age at ART initiation
AgeC	Age currents
Sex	1 = male, 2 = female
Marital status	1 = single, 2 = married and 3 = others
Health-related	
Facilities	1 = Kuisebmend Hospital, 2 = Hakhaseb Clinic, 3 = Omaruru Hospital and 4 = Tamariskia Clinic
Year of initiation	1 = 2010, 2 = 2011, 3 = 2012. 4 = 2013, 5 = 2014 and 6 = 2015
WHO Clinical stage	1 = stage 1, 2 = stage 2, 3 = stage3, and 4 = stage 4
CD4_cell_count	CD4 cell counts current
Weight Current	Weight current

5.3 Data analysis for Joint Model of treatment adherence and viral load

In this section, the data analysis were presented in five sub-sections; descriptive analysis, multiple regression of log transformation of viral load, multiple regression using normal-gamma of viral load, binary logistics regression of ART adherence, Heckman-type model using a joint bivariate copula model and model selection criterions.

5.3.1 Descriptive analysis

In the analysis of time to ART treatment adherence and viral load, simple descriptive analysis was used to describe demographic and clinical characteristics of the study population. For the purposes of this analysis the viral load values were used regardless of whether patients have lower or higher viral loads.

5.3.2 Multiple regression of log-transformed viral load

The viral load was transformed using log transformation. Multivariate multiple regression analysis was used to assess the factors associated with log transformation viral load. In multivariate analysis, adjustments were made for sex, age at ART initiation, current age at ART initiation, facility, WHO stages, CD4 cell count at ART initiation, and current body weight.

5.3.3 Gamma regression model of viral load

The viral load was used as a continuous variable. Therefore, the study also used gamma regression model to assess the factors associated with viral load. The gamma regression model is a class of generalized linear models (GLM) used to model a positive-valued dependent variable. Thus, gamma regression models allow for a monotone and have the reproductive property that the sums of independent gamma distributions are also gamma distributed (Cepeda-Cuervo et al., 2016).

The gamma GLM is of the form

$$w \sim G(\mu, v), \quad g(\mu) = x'\beta. \tag{5.1}$$

The canonical link for the gamma distribution is the inverse function. Since parameters from a model with inverse link are difficult to interpret, the log link is usually regarded as more useful (De Jong & Heller, 2008). The gamma is often a reasonable fit for variables such as viral load. Gamma random variables are continuous, non-negative and skewed to the right, with the possibility of large values in the upper tail.

The $\mathrm{G}(\mu, \mathrm{v})$ probability function is

$$\mathrm{f(w)} = \frac{\mathrm{w}^{-1}}{\Gamma(\mathrm{v})}\left(\frac{\mathrm{wv}}{\mu}\right)^{\mathrm{v}} \mathrm{e}^{-\mathrm{wv}/\mu}, \qquad \mathrm{w} > 0, \tag{5.2}$$

with

$$\mathrm{E(w)} = \mu, \qquad \mathrm{Var(w)} = \frac{\mu^2}{\mathrm{v}}$$

Small values of v result in a distribution with a long tail to the right, i.e. a more right-skewed distribution (De Jong & Heller, 2008).

5.3.4 Binary logistics regression of ART adherence

Binary logistic regression analysis was used to assess the factors associated with ART adherence. In the analysis, adjustments were made for sex, age at ART initiation, facility, WHO stages, CD4 cell count at ART initiation, and current body weight.

5.3.5 Heckman-type selection model of ART adherence and log-transformed viral load

A Heckman-type selection analysis using a joint bivariate copula was used on the two outcomes: ART adherence and log transformation of viral load. In the context of this study, semi-parametric predictors for the selection (Adherence to ART) and outcome (log transformation of viral load) equations and for the sigma and copula parameters (σ, θ) copula function are specified as (Marra & Radice, 2017b):

$$\log\frac{\mathrm{P(adherence \geq 95\%)}}{\mathrm{P(adherence < 95\%)}} = \alpha_{10} + \mathrm{x_{1i}}^{\mathrm{T}}\alpha_{11} + \mathrm{s_{11}(AgeIn_i)} + \mathrm{s_{12}(AgeC_i)} + \mathrm{s_{13}(WeightC_i)} + \mathrm{s_{14}(CD4cell\ count_i)}$$

$$\mathrm{log(Viral\ load_i)} = \alpha_{20} + \mathrm{x_{2i}}^{\mathrm{T}}\alpha_{21} + \mathrm{s_{21}(AgeIn_i)} + \mathrm{s_{22}(AgeC_i)} + \mathrm{s_{23}(WeightC_i)} + \mathrm{s_{24}(CD4cell\ count_i)}$$

$$\mathrm{Sigma}(\sigma) = 1$$

$$\text{Copula parameter } (\theta) = \alpha_{30} + \alpha_{31}(\mathrm{sex_i}) + \alpha_{32}(\mathrm{facilities_i}), \tag{5.3}$$

where parameters $\alpha_{10}, \alpha_{20}, \alpha_{30}$ are constant containing the overall levels of the predictors, vectors x_{1i} *and* x_{2i} contain binary and categorical variables that are associated with the selection and outcome equations, α_{11}, and α_{21} are respective parameter vectors, and the s_{uk} *for* $u = 1, 2$ *and* $k = 1, 2, 3, 4$ are smooth functions of age at ART initiation, current age, current body weight and CD4 cell counts. The variables included in x_{1i} *and* x_{2i} are: sex, marital status, clinical stage, and facilities.

The equivalence standard selection model that relies on jointly bivariate normal distribution is to use the copula approach (Marra et al. 2017). Therefore, to allow for nonlinear association between the adherence to ART and viral suppression, copula function can then be used as:

$$F\left(log\frac{P(adherence\geq 95\%)}{P(adherence<95\%)}, log(Viral\ load_i)\right) = C\left(F_1\left(log\frac{P(adherence\geq 95\%)}{P(adherence< 95\%)}\right), F_2\left(log(Viral\ load_i)\right); \theta\right), \qquad (5.4)$$

where θ is an association parameter measuring the dependency between two marginal (R$\ddot{u}$schendorf, 2013).

5.3.6 Heckman-type selection model of ART adherence and gamma of viral load

Heckman-type selection model was also used to fit a joint bivariate copula distribution to assess the dependence structure between ART adherence and viral load margins. The first selection equation is a binary distributional margin which is specified as:

$$\begin{aligned}\log\frac{\mathrm{P(adherence \geq 95\%)}}{\mathrm{P(adherence < \ 95\%)}} = {} & \alpha_{10} + \ x_{1i}{}^{T}\alpha_{11} + \ s_{11}(\mathrm{AgeIn_i}) + s_{12}(\mathrm{AgeC_i}) + s_{13}(\mathrm{WeightC_i}) \\ & + s_{14}(\mathrm{CD\ cell\ count_i}), \qquad (5.5)\end{aligned}$$

The second outcome equation is a gamma GLM which is of the form

$$\mathrm{Viral\ load} \sim G(\mu, v), \quad g(\mu) = x'\beta. \qquad (5.6)$$

$$\begin{aligned}\log(\mathrm{Viral\ load_i}) = {} & \beta_{20} + \ x_{2i}{}^{T}\beta_{21} + \ s_{21}(\mathrm{AgeIn_i}) + s_{22}(\mathrm{AgeC_i}) + s_{23}(\mathrm{WeightC_i}) \\ & + s_{24}(\mathrm{CD4cell\ count_i}) \qquad (5.7)\end{aligned}$$

The canonical link for the gamma distribution is the inverse function. Since parameters from a model with inverse link are difficult to interpret, the log link is usually regarded as more useful (De Jong & Heller, 2008).

5.3.7 Information criterions

There are a number of model selection criterions to select the best model for the copula models. In this context, the most commonly known model selection criterions; Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were considered for this study. These criterions referred to those in equations 2.24 and 2.25 in chapter 2.

5.4 Results for Joint Model of treatment adherence and viral load

5.4.1 Descriptive analysis

The study used data from Kuisebmend Hospital Center (KHC), Hakhaseb Clinic (HC), Omaruru Hospital Center (OHC) and Tamariskia Clinic (TC) both in Erongo region. Out of 34376 patients, 33.6% of patients from KHC, 2.3% HC, 37.9% OHC, and 26.9% TC included in the study. Demographic and clinical characteristics were described in table 12. Eleven thousand five hundred and thirty nine treated in KHC, 19762 (57.5%) were female, 29946 (93.6%) were single, and 33160 (96.9%) were at baseline clinical stage-I and stage-II. The median and interquartile (IQR) of age at ART initiation, age current, current body weight and the baseline CD4 cell count were 39 (33 – 45), 40 (34 – 46), 64 (56 – 74.2), and 339 (221 – 480), respectively. Of those included in the study, 20012 (86.9%) were adhered at least 95% and 9635 (28.0%) were initiated ART treatment at the beginning of January 2011.

Table 12: Demographic and clinical characteristics

Characteristics	Subjects (n = 34376)
Health Facility	
KHC	11539 (33.6)
HC	780 (2.3)
OHC	13040 (37.9)
TC	9017 (26.2)
Sex	
Male	14614 (42.5)
Female	19762 (57.5)
Age at ART initiation, median (IQR)	39 (33 - 45)
Current Age, median (IQR)	40 (34 - 46)
Current body weight, median (IQR)	64 (56 - 74.2)
Current CD4 cell count, median (IQR)	339 (221 - 480)
Marital Status	
Single	29946 (93.6)
Married	1888 (5.9)
Others	146 (0.5)
WHO Clinical Stage	
Stage-I vs II	33160 (96.9)
Stage-III vs IV	1047 (3.1)
ART adherence	
At least 95%	20012 (86.9)
Less than 95%	3013 (13.1)
Year at ART Initiation	
2010	7329 (21.3)
2011	9635 (28.0)
2012	5135 (14.9)
2013	4622 (13.4)
2014	4596 (13.4)
2015	3059 (8.9)

Note: this table presents the demographic and clinical characteristics for 34376 patients receiving antiretroviral therapy (ART) during 2010-2015. KHC = Kuisebmend Hospital Center, HC = Hakhaseb Clinic, OHC = Omaruru Hospital Center and TC = Tamariskia Clinic.

5.4.2 Multiple regression model of log-transformed viral load

Table 13 presents the parameter estimation calculated by use of multiple regression model. The results show that factors such as baseline clinical stage, and weights were statistically significant between viral loads. Patients with CD4 cell count greater than 200 were more likely to be associated with viral load. Patients whom adhered at least 95% were more likely to be associated with viral load. The results further show that female and age at ART initiation in ART treatment were not statistically significant at 5% level of significance.

Table 13: Estimation of the factors associated with log transformation of viral load

Coefficients parameter	**Unstandardized Coefficients**				**95% CI for estimate**	
	Estimate	**Standard error**	**t value**	**P-value**	**LB**	**UB**
Intercept	1.882	0.031	60.953	<0.0001	1.822	1.943
Sex (ref. male)						
Female	0.002	0.010	0.164	0.870	-0.019	0.022
Facility (ref. KHC)						
HC	0.017	0.024	0.710	0.478	-0.031	0.065
OHC	-0.023	0.017	-1.370	0.171	-0.056	0.010
TC	0.095	0.011	8.439	<0.0001	0.073	0.117
Age (ref. <25 years)						
25 - 49 years	-0.014	0.025	-0.565	0.572	-0.063	0.035
≥50	0.005	0.029	0.171	0.864	-0.052	0.062
Who stage(ref. stage-I vs II)						
Stage-III vs IV	0.245	0.032	7.605	<0.0001	0.182	0.308
Body weight (ref. <50)						
50-70 weight	-0.049	0.021	-2.353	0.019	-0.089	-0.008
>70 weight	-0.005	0.022	-5.300	<0.0001	-0.158	-0.072
CD4 count (ref. <200)						
≥200	-0.121	0.012	-9.860	<0.0001	-0.145	-0.097
ART adherence (ref. <95%)						
≥95%	2.202	0.012	178.105	<0.0001	2.177	2.226

Note: this table presents the estimates of the factors associated with viral load for 34376 patients receiving antiretroviral therapy during 2010 – 2015, LB = Lower Bound and UB = Upper Bound

5.4.3 Gamma regression model of viral load

Table 14 presents the results of parameter estimation for the gamma regression model. The results revealed that the relationship between log viral load and ART adherence is roughly linear. For one unit increase in ART adherence, we expect about 6.705 increase in the expected value of viral load in the log scale, given that all other variable in the model are held constant. The results further revealed that facilities, age, WHO stage, body weight and CD4 cell counts were statistical significant at 5% level of significance. It also demonstrated that there is not association between log viral load and females (p-value = 0.428).

Table 14: Gamma fit for viral load

Dependent Variable	Viral Load			
Probability Distribution	Gamma			
Link Function	Log			
Deviance	28728.871			
Degree of freedom	12584			
Parameters	**Estimate**	**Std. Error**	**Chi-Square**	**P-value**
(Intercept)	4.884	.0732	4454.609	<0.001
Sex (ref. Male)				
Female	-.020	.0249	.629	0.428
Age at ART initiation (ref. <25 Years)				
25-49 Years	.158	.0597	7.017	0.008
>=50 Years	.179	.0697	6.605	0.010
Facility (ref. KHC)				
HC	.456	.0586	60.629	<0.001
OHC	.155	.0413	14.106	<0.001
TC	.324	.0271	143.216	<0.001
WHO Stage (ref. Stage-I)				
Stage-II	.310	.0758	16.713	<0.001
Stage-III	.338	.0880	14.770	<0.001
Stage-III	1.062	.1542	47.423	<0.001
Body weight (ref. <50)				
50-70 weight	-.157	.0505	9.695	0.002
>70 weight	-.382	.0533	51.277	<0.001
CD4 cell counts (ref. <200)				
>=200	-.244	.0294	69.361	<0.001
ART adherence (ref. < 95%)				
>= 95%	6.705	.0300	49911.643	<0.001
Scale (.)	1.821	.0191		

Note: this table presents the estimation parameters from the gamma regression model of viral load.

5.4.4 Binary logistic regression model of ART adherence

Table 15 presents the parameter estimation calculated by use of logistic regression models. These models include the underlying ART adherence term. The results show that patients whom treated in health facilities in Erongo region and whom on clinical stage-III against IV were associated with ART adherence. The results further show that female patients, age group regardless of whether age between 25 to 49 and greater than or equal to 50 were associated with ART adherence. Moreover, the results indicate that patients whose body weight between 50 to 70 weight (p-value<0.001) and whose body weight greater than 70 weight (p-value<0.001) were associated with ART adherence. Patients having CD4 count greater than or equal to 200 (p-value<0.0001) were associated with ART adherence.

Table 15: Parameter estimates of the factors associated with ART adherence

Coefficients parameter	Estimate	Std error	t-value	P-value
Intercept	0.331	0.014	23.213	<0.0001
Sex (ref. male)				
Female	-0.023	0.005	-4.876	<0.001
Facility (ref. KHC)				
HC	0.087	0.013	6.682	<0.001
OHC	0.129	0.009	14.245	<0.0001
TC	0.076	0.005	15.891	<0.0001
Age (ref. <25 years)				
25 - 49 years	-0.116	0.012	-9.880	<0.0001
≥50	-0.136	0.013	-10.217	<0.0001
Who stage (ref. stage-I vs II)				
Stage-III vs IV	0.143	0.017	8.539	<0.0001
Body weight (ref. <50)				
50-70 weight	-0.053	0.009	-5.617	<0.001
>70 weight	-0.074	0.010	-7.490	<0.001
CD4 count (ref. <200)				
≥200	-0.078	0.006	-13.627	<0.0001

Note: this table presents estimation of the factors associated with ART adherence for 34376 patients receiving antiretroviral therapy (ART) during 2010 -2015.

5.4.5 Heckman-type selection model of ART adherence and log-transformed viral load

Table 16 presents the parameters estimation of sample selection model with probit and log-transformed distributional margins for some of the copula model such as Normal, Frank, FGM, AMH and Gumbel, and Clayton rotated by 0 and 180 degrees. The results show that the rotated copula models indicate a strong upper tail dependence structure between ART adherence and viral load margins. By looking at the intervals of θ and τ reported in table 16 below we can see a very small significant dependence parameter obtained for the copula models except FGM model since the interval includes zero. These indicate that there exists some positive association between the unstructured terms of the model equations. This suggests that individuals with at least 95% ART adherence are more likely to have more viral load as compared to those with less than 95% ART adherence.

Table 16: Measure of the association between ART adherence and transformed viral load

Copula Model	**Sigma2 (95% CI)**	**Theta (95% CI)**	**Kendall's Tau (95% CI)**
Normal	1413(1357 , 1472)	-0.200(-0.218 , -0.185)	-0.129(-0.140 , -0.118)
Frank	1473(1425 , 1533)	-4.270(-4.640 , -3.870)	-0.398(-0.424 , -0.371)
FGM	1417(1377 , 1466)	-0.675(-0.736 , -0.598)	-0.150(-0.164 , -0.133)
AMH	1403(1356 , 1464)	-0.622(-0.685 , -0.543)	-0.116(-0.127 , -0.101)
Gumbel 0°	1213(1177 , 1253)	1.020(1.010 , 39.80)	0.016(0.006 , 0.421)
Gumbel 180°	1635(1568 , 1691)	1.010(1.010 , 1.010)	0.006(0.005 , 0.008)
Clayton 0°	1499(1461 , 1546)	0.008(0.006 , 96.80)	0.004(0.003 , 0.954)
Clayton 180°	864(843 , 887)	0.020(0.012 , 46.10)	0.009(0.006 , 0.466)

Note: this table presents the estimation of the association parameters with 95% confidence intervals (in brackets) from the selection models based on the Normal, Frank, FGH, AMH and Gumbel, and Clayton Copulas rotated by 0, 180 degrees with bivariate outcome margin.

The presence of tail dependence in sample selection bias was also checked. By looking at the table 16, there is support for the hypothesis of selection bias, with a positive association for some of the copula models, and the 95% CI for τ excludes zero in some case. Table 16 also revealed that the Nornal, Frank, FGM and AMH copula models indicate lower tail dependence structure between ART adherence and viral load by looking at the association parameter and 95% CI for τ which includes zero.

Table 17: Akaike and Bayesian information criterions (AIC and BIC) for copula functions in Heckman-type model

Copula Model	**AIC**	**BIC**
Normal	360049.1	360639.9
Frank	359306.2	359896.3
FGM	361991.1	362578.8
AMH	362123.7	362708.4
Gumbel 0°	360126.3	360747.3
Gumbel 180°	363244.1	363812.5
Clayton 0°	362151.6	362734.3
Clayton 180°	**355282.7**	**355835.0**

Note: this table presents the Akaike and Bayesian information criteria (AIC and BIC) from the sample selection models based on Normal, Frank, FGM, AMH and Gumbel and Clayton rotated by 0 and 180 degrees.

Table 17 shows the AIC and BIC for the sample selection models with different copulas. These includes Normal, Frank, FGM, AMH and Gumbel and Clayton rotated by 0 and 180 degrees. The copula model with lowest AIC and BIC values in which viral load was used as outcome and adherence to ART treatment was used as selection correspond to Clayton copula rotated by 180 degree followed by Frank copula model.

Table 18 presents the results of Clayton copula rotated by 180 degree for the two margins. The results revealed that ART adherence and female were associated between viral loads. It is also shows that some demographics and clinical characteristics indicate a significant regardless ART adherence and viral load. The small yet significant dependence parameter obtained for the Clayton copula indicates that there exists some positive association between the unstructured terms of the model equations. The results also revealed that age at ART initiation, current age, current body weight and CD4 cell counts indicate a very significant between ART adherences. The results further revealed that age at ART initiation, current age and CD4 cell counts were statistically significant between viral loads. Current body weight was not significant between viral loads.

Table 18: Model parameter estimates of the Heckman-type for 180° Clayton copula

Parametric coefficients	**ART adherence**				**Viral load**			
	Estimate	**Std Err**	**z value**	**P-value**	**Estimate**	**Std Err**	**z value**	**P-value**
Intercept	-1.487	0.024	-61.506	<0.0001	15.994	0.486	32.890	<0.0001
Sex (ref. male)								
Female	-0.056	0.025	-2.251	0.024	-2.399	0.498	-4.820	<0.0001
Facility(ref. KHC)								
HC	1.138	0.096	11.897	<0.0001	10639.978	3162.278	3.365	0.0007
OH	0.544	0.043	12.637	<0.0001	18.143	1.041	17.432	<0.0001
TC	0.402	0.026	15.606	<0.0001	4.890	0.496	9.252	<0.0001
WHO stage (ref. Stage-I vs II)								
Stage-III vs IV	0.302	0.070	4.305	<0.0001	27164.693	3162.278	8.590	<0.0001
ART adherence (ref. <95%)								
≥95%					56121.418	20.798	2698.44	<0.0001
Smooth components' approximate significance								
	Edf	**Ref.df**	**Chi.sq**	**P-value**	**edf**	**Ref.df**	**Chi.sq**	**P-value**
Age at ART initiation	7.916	8.521	1330.98	<0.0001	8.990	9.000	4758.34	<0.0001
Current age	3.851	5.017	1264.88	<0.0001	8.992	9.000	1727.97	<0.0001
Current weight	2.265	2.924	17.140	0.0006	6.135	7.221	2.917	0.898
CD4 cell counts	7.847	8.277	55.900	<0.0001	3.898	4.596	38.120	<0.0001

Note: this table presents the estimation of the association parameters for the selection and outcome equations for Clayton copula rotated by 180 degrees.

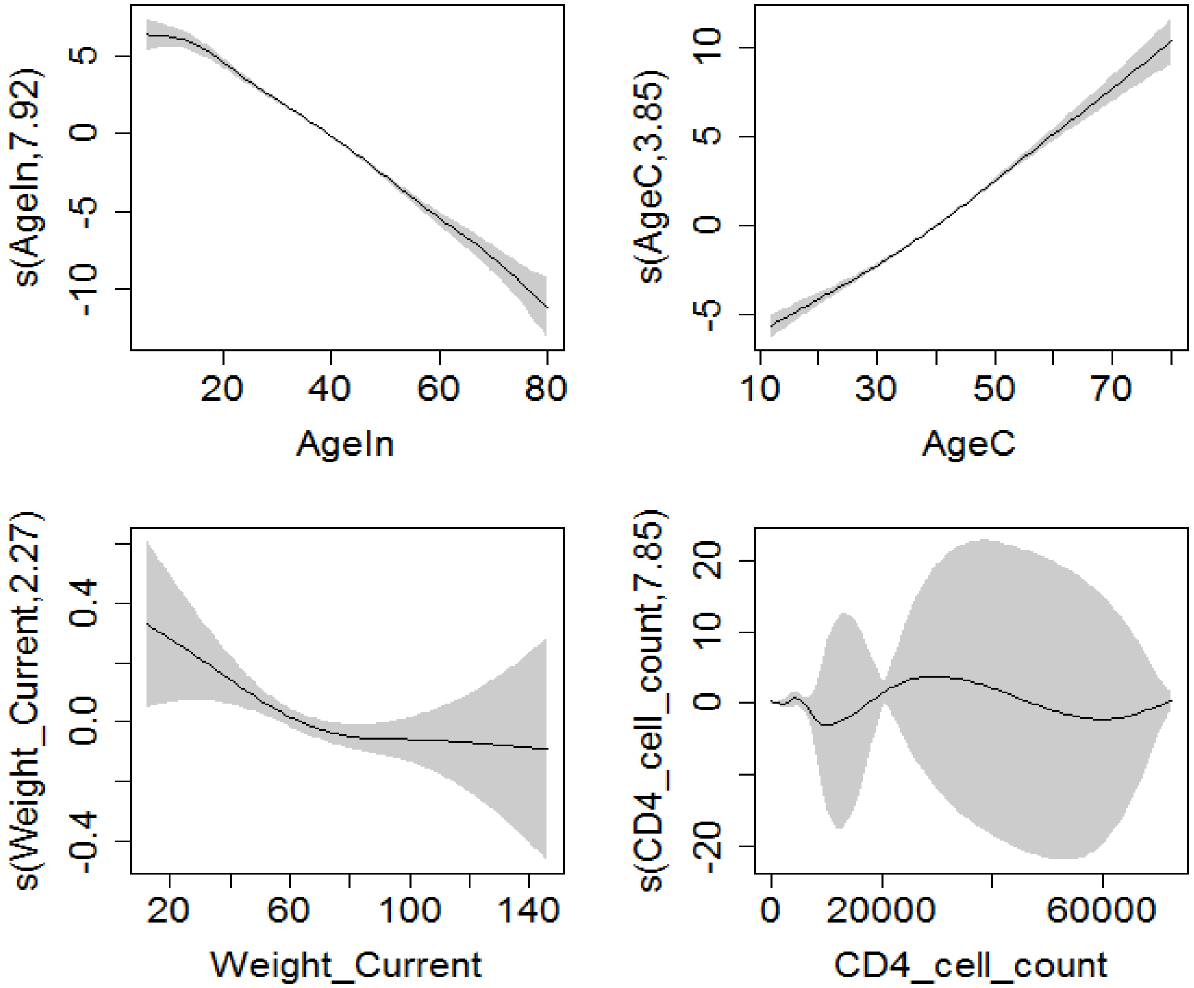

Figure 1: The smoothing function estimates of selection equation

Note: this figure presents the smoothing function estimates of selection equation and associated 95% point-wise confidence intervals obtained by fitting the best Clayton copula model rotated by 180 degree on the 2010-2015 Erongo HIV data. Results are plotted on the scale of the additive predictor. The jittered rug plot, at the bottom of each graph, shows the covariate values. The numbers in brackets in the y-axis captions are the effective degrees of freedom (edf) of the smooth curves.

The effects of age at ART initiation, age current, weight current and CD4 cell counts in the selection equation show different degrees of non-linearity. The point-wise confidence intervals of the smooth functions for age at ART initiation and current age in the selection equation did not contain the zero line in the range of the covariate values. The point-wise confidence intervals of the smooth functions for weight current in the selection equation contain the zero line for the some range of the covariate values. The intervals of the smooth for CD4 cell counts in the selection equation contain the zero line for the whole range of the covariate values (figure 1). This suggests that age at ART initiation, current age and current body weight were the weak predictors of ART adherence, and that CD4 cell counts might be a very important determinant of ART adherence.

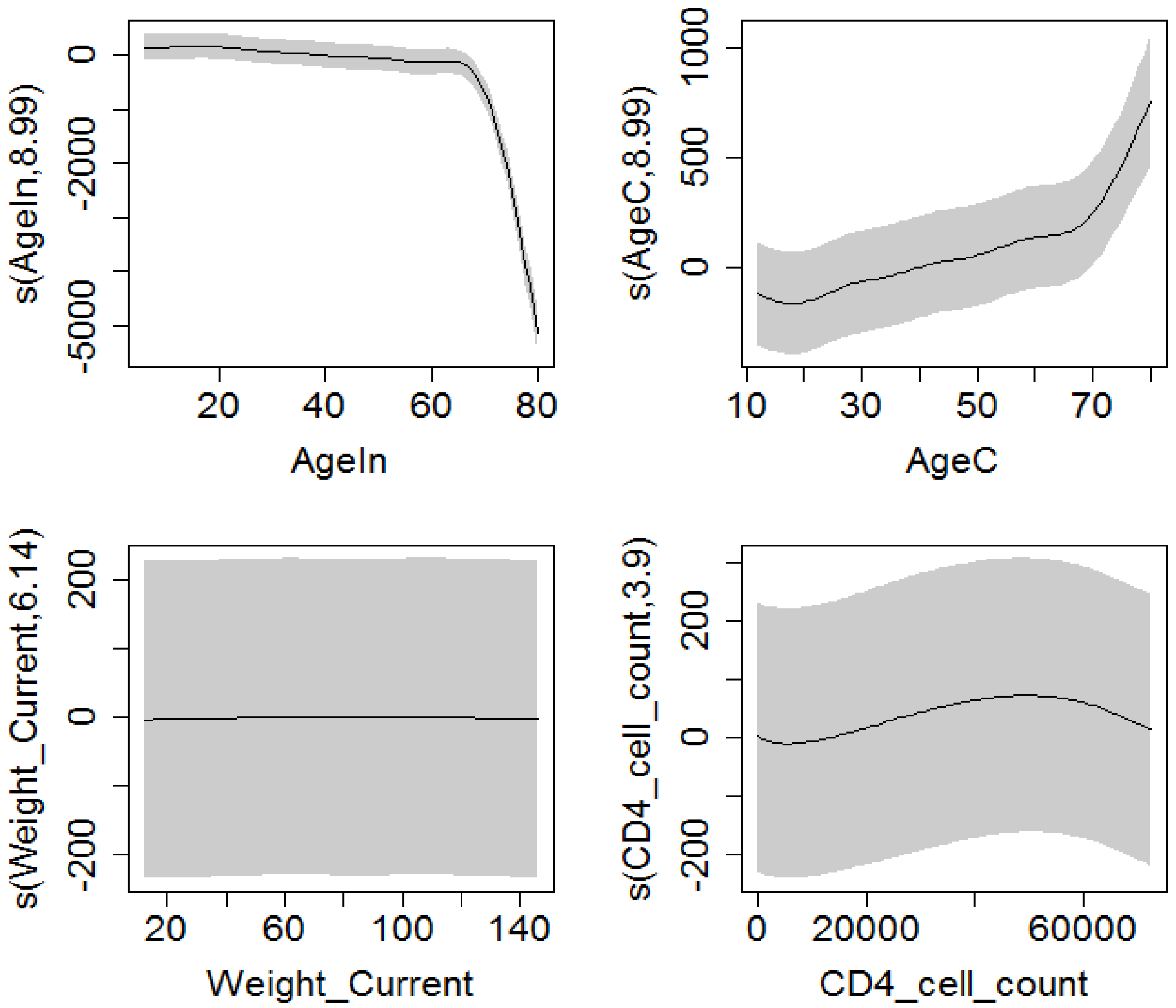

Figure 2: The smoothing function estimates of outcome equation

Note: this figure presents the results from outcome equation: smooth function estimates and associated 95% point-wise confidence intervals obtained by fitting Clayton copula model rotated by 180 degree on the 2010-2015 Erongo HIV data. Results are plotted on the scale of the additive predictor.

The effects of age at ART initiation, age current, weight current and CD4 cell counts in the outcome equations show different degrees of non-linearity. The point-wise confidence intervals of the smooth functions for weight current in outcome equation contain the zero line for the whole range of the covariate values. The intervals of the smooth for CD4 cell counts in the outcome equation contain the zero line for most range of the covariate values (figure 2). This suggests that weight current is a weak predictor of viral load, and might not be a very important determinant of viral load.

5.4.6 Heckman-type selection model of ART adherence and gamma of viral load

Table 19 presents the parameters estimation of Heckman-type selection model with probit and gamma distributional margins for some of the copula models such as Normal, Frank, FGM, AMH and Gumbel, and Clayton rotated by 0 and 180 degrees. The results show that there were a positive association between ART adherence and log viral load by looking at the association parameter Kendall’s Tau for all copula models presented in the table 19. The results further show that the copula models indicate strong upper tail dependence structure between ART adherence and log viral load margins.

Table 19: Measure of the association between ART adherence and log viral load

Copula Model	Sigma2 (95% CI)	Theta (95% CI)	Kendall's Tau (95% CI)
Normal	12.40(10.80 , 14.30)	0.981(0.976 , 0.985)	0.875(0.861 , 0.890)
Frank	13.40(12.60 , 14.40)	61.50(51.90 , 71.60)	0.937(0.925 , 0.945)
FGM	2.080(1.950 , 2.180)	0.948(0.925 , 0966)	0.211(0.206 , 0.215)
AMH	2.130(2.010 , 2.240)	0.963(0.940 , 0.977)	0.311(0.298 , 0.319)
Gumbel 0°	1.060(0.993 , 1.150)	1.190(1.180 , 1.220)	0.161(0.150 , 0.179)
Gumbel 180°	9.140(8.220 , 9.880)	7.310(6.300 , 8.130)	0.863(0.841 , 0.877)
Clayton 0°	15.80(14.10 , 16.50)	61.80(51.90 , 73.50)	0.969(0.963 , 0.973)
Clayton 180°	1.100(1.030 , 1.170)	0.329(0.304 , 0.361)	0.141(0.132 , 0.153)

Note: this table presents the estimation of the association parameters with 95% confidence intervals (in brackets) from the selection models based on the Normal, Frank, FGH, AMH and Gumbel, and Clayton Copulas rotated by 0, 180 degrees with bivariate outcome margin.

Table 20 presents the AIC and BIC for the Heckman-type selection models with different copulas. These includes Normal, Frank, FGM, AMH and Gumbel and Clayton rotated by 0 and 180 degrees. The copula model with lowest AIC and BIC values in which viral load was used as outcome and adherence to ART treatment was used as selection correspond to Gumbel copula rotated by 180 degree followed by Normal copula model.

Table 20: Akaike and Bayesian information criterions (AIC and BIC)

Copula Model	AIC	BIC
Normal	79720.1	80235.4
Frank	79838.7	80441.7
FGM	81862.2	82352.1
AMH	81735.2	82215.6
Gumbel 0°	80044.7	80606.0
Gumbel 180°	**79588.9**	**80085.0**
Clayton 0°	79825.7	80302.0
Clayton 180°	79995.5	80514.8

Note: this table presents the Akaike and Bayesian information criteria (AIC and BIC) from the sample selection models based on Normal, Frank, FGM, AMH and Gumbel and Clayton rotated by 0 and 180 degrees.

Table 21 presents the results of Gumbel copula rotated by 180 degree for the two margins. In both distributional margins, the results show that patients who treated in Erongo health facilities and whom were at baseline clinical stage-III verses stage-IV were associated to ART adherence and viral load. Female patients were not statistically significant between ART adherence (0.144) and also viral load (p-value = 0.128). The significant dependence parameter obtained for the Gumbel copula indicates that there exists positive association between the unstructured terms of the model equations. To adjust for non-linearity in both distributional margin, the table also shows that Age at ART initiation, current age, current body weight and CD4 cell counts indicate a very significant between ART adherence and also log viral load.

Table 21: 180° Gumbel copula parameter estimation of the Heckman-type selection

Parametric coefficients	**ART adherence**				**Viral load**			
	Estimates	**Std.Error**	**z value**	**P-value**	**Estimates**	**Std.Error**	**z value**	**P-value**
Intercept	-1.438	0.023	-62.513	<0.0001	6.964	0.063	110.803	<0.0001
Sex (ref. Male)								
Female	-0.033	0.022	-1.462	0.144	-0.084	0.055	-1.523	0.128
Facility(ref.KHC)								
HC	0.975	0.057	17.247	<0.0001	2.118	0.178	12.085	<0.0001
OH	-0.926	0.037	-24.727	<0.0001	-1.759	0.075	-23.512	<0.0001
TC	0.348	0.025	13.868	<0.0001	0.488	0.067	7.305	<0.001
WHO stage (ref. Stage-I vs II)								
Stage-III vs IV	0.186	0.059	3.130	0.002	0.554	0.161	3.452	0.0006
Smooth components' approximate significance								
	edf	**Ref.df**	**Chi.sq**	**P-value**	**edf**	**Ref.df**	**Chi.sq**	**P-value**
Age at ART initiation	2.231	2.983	752.35	<0.0001	1.000	1.000	382.04	<0.0001
Current age	7.598	8.293	729.29	<0.0001	8.019	8.684	431.31	<0.0001
Current weight	6.375	7.331	15.92	0.033	7.061	8.028	46.07	<0.001
CD4 cell counts	6.504	7.068	44.45	<0.001	6.308	7..043	59.85	<0.001

Note: this table presents the estimation of the association parameters for the selection and outcome equations for Gumbel copula rotated by 180 degrees

5.5 Discussion for Joint Model of treatment adherence and viral load

In this retrospective cohort study, the aim was to fit a joint distributional model to adjust for sample selection bias of treatment adherence and viral load among HIV patients on ART in Erongo region, Namibia.

The results for Erongo region found that some demographic characteristics were more likely to be affected by selection bias. Using this modelling framework, the results found that HIV patients' response to ART was good, with a median increase in age at ART initiation, current age on ART, current body weight on ART and the baseline CD4 cell count on ART. Although current body weights on ART treatment were found to be associated with viral load, the magnitude of the changes was very small. Moreover, those with age at ART initiation, current age on ART treatment and current CD4 cell count on ART treatment were found to be not associated with viral load. These findings in line with those found in other studies (Bisson et al. 2008; Abrogoua et al. 2012; Seyoum, Ndlovu & Temesgen, 2017).

Further analysis consider that, HIV patients whom treated in Hakhaseb Clinic and were on baseline clinical stage-II were found to be associated with ART adherence. The results also found that patients whom treated in Hakhaseb Clinic and female were more likely to be associated with ART adherence. Moreover, calendar year was also found to be associated with ART adherence, with those start ART during 2011, 2012, 2013, and 2014 increasing the change of ART adherence. Several factors have to explain this changes. Adherence could be assumed to have improved; as the importance of adherence has become established, awareness of ways of improving it among those receiving ART has increased (Smith et al, 2004).

In the analysis of sample selection, the results found that copula models (Normal, Frank, FGM, AMH and Gumbel and Clayton rotated by 0 and 180 degrees) exhibit dependence structure between ART adherence and viral load margins. It also found that there exists some positive association between the unstructured terms of the model equations. It further found that the FGM model show asymmetric dependence structure between ART adherence and viral load. It is therefore suggested that individuals with at least 95% ART adherence are more likely to have more viral load as compared to those with less than 95% ART adherence. These results agree with those found by Marra and Radice (2017) where the bivariate distributions are chosen so that positive dependence is allowed.

5.6 Conclusion for Joint Model of treatment adherence and viral load

Some Erongo baseline demographic and clinical characteristics were more likely to be affected by selection bias. Although current body weight vary significant difference between viral loads, the magnitude of the changes were very small. Based on the copula models that were in the study, there were a very positive significant dependence structure between ART adherence and viral load margins. This suggests that weight current is a weak predictor of ART adherence, and that income might not be a very important determinant of ART adherence. It also suggested that weight current is a weak predictor of viral load, and that CD4 cell counts might not be a very important determinant of viral load.

CHAPTER 6: DISCUSSION, CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

The last part of the thesis is dedicated to summarizing the key discussion in line with the study objectives. The chapter also indicates the conclusion and the recommendations drawn from these discussion.

6.2 Discussion

The overall study discussion were based on the four objectives of the study which were:

- To simultaneously model treatment adherence with viral suppression;
- To fit a joint model of treatment adherence with viral load;
- To examine if there is any tail dependence in sample selection bias;
- To investigate the factors associated with viral suppression, viral load, and treatment adherence.

Although major research studies of viral load, viral suppression and ART treatment adherence have been reported, the simultaneous models of treatment adherence with viral suppression and treatment adherence with viral load at end of 12 months have not been included in those studies. The results found that there was a very positive dependence structure between ART adherence and viral suppression margins. Even though only a small negative dependence present, the magnitude charges. It also found that there was positive association between ART adherence and the viral loads. It further found that some demographic and clinical characteristics were not associated with ART adherence and viral suppression. Furthermore, the findings also revealed that the copula models show asymmetric dependence structure between ART adherence and viral load. Many studies agreed with these findings (Marra and Radice, 2017; McGovern et al., 2015;).

6.3 Conclusions

The main objective of the study was to fit a joint distributional model to adjust for sample selection bias of treatment adherence and viral suppression among HIV patients on ART in Namibia. In the retrospective cohort, there is evidence to support the hypothesis of sample selection bias that there were negative dependence between ART adherence and viral suppression. On the other hand, ART adherence and viral load indicate a very positive association. In addition, some demographic and clinical characteristics were not associated with ART adherence and viral suppression although female were more likely to be associated with ART adherence. This study was only for few facilities and it will be extended to other facilities. However, similar studies need to consider socio-economic factors in addition to those considered in this study.

The adjustment has been evaluated for sample selection bias of ART adherence and viral suppression and this would be to control the tail dependence, although the association parameters may lose their interpretation

6.4 Recommendations

The study recommended that socio-economic factors should be instituted in order to provide the association between those factors with ART adherence, viral load and viral suppression. The study is also recommended that the effect of smoothing function should be conducted on the individual data. The study further recommended that stronger efforts should be put in place to ensure that all people living with HIV have access to treatment as soon as they are diagnosed. The study was carried out in Erongo region and selected health facilities in Khomas region only. However, similar studies could be carried out in other parts of the country to gather adequate information on the subject to be able to generalize.

REFERENCES

Abrogoua, D. P., Kablan, B. J., Kamenan, B. A. T., Aulagner, G., N'Guessan, K. & Zohoré, C. (2012). Assessment of the impact of adherence and other predictors during HAART on various CD4 cell responses in resource-limited settings. Patient Preference and Adherence Journal. 23 March 2012 Volume 2012:6 Pages 227—237. Retrieved from http://dx.doi.org/10.2147/PPA.S26507.

Adiga, S., Adiga, U. & Malawadi, B. N. (2016). Comparison of Adherence and CD4 counts of HIV patients on various ART Regimens. International Journal of Medical Research and Review. July, 2016/Vol 4/Issue 7.

Alexander, L. K., Lopes, B., Ricchetti-Masterson, K. & Yeatts, B. K. (2015). Selection Bias. 2nd Edition Eric notebook. Retrieved from: https://sph.unc.edu/files/2015/07/nciph_ERIC13.pdf.

Araichi, S., Peretti, C. & Belkacem, L. (n.d.). Time Varying Copula Model for claims reserving in Non-life insurance. Retrieved from: https://www.em-lyon.com/minisiteen/content/download/166676/6553128/version/1/file/Christian+de+Perretti.pdf.

Arnsten, H. J., Demas, A. P., Farzadegan, H., Grant, W. R., Gourevitch, N. M., Chang, J. C., Buono, D., Eckholdt, H., Howard, A. A. & Schoenbaum, E. E. (2001). Antiretroviral Therapy Adherence and Viral Suppression in HIV-Infected Drug Users: Comparison of Self-Report and Electronic Monitoring. Infectious Diseases Society of America 33, 1417–23.

Bisson, P. G., Gross, R., Bellamy, S., Chittams, J., Hislop, M., Regensberg, L., Frank, I., Maartens, G. & Nachega, B. J. (2008). Pharmacy Refill Adherence Compared with CD4 Count Changes for Monitoring HIV-Infected Adults on Antiretroviral Therapy. PLoS Med 5(5): e109. Doi:10.1371/journal.pmed.0050109

Bouyé, E., Durrleman, A., Nikeghbali, A., Riboulet, G., & Roncalli, T. (2000). Copulas for finance—a reading guide and some applications. Technical Report, Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Castro, A. (2005). Adherence to antiretroviral therapy: Merging the clinical and social course of AIDS. PLoS Med 2(12): e338.

Cepeda-Cuervo, E., Corrales, M., Cifuentes, V. M. & Zarate, H. (2016). On Gamma Regression Residuals. DOI: 10.7508/jirss.2016.01.002. Vol. 15, No. 1, pp 29-44.

Cescon, A. M., Cooper, C., Chan, K., Palmer, K. A., Klein, M. B., Machouf, N., Loutfy, M, R., Raboud, J., Rachlis, A., Ding, E., Lima, V. D., Montaner, J. S. G., Rourke, S. B., Smieja, M., Tsoukas, C. & Hogg, R. S. (2011). Factors associated with virological suppression among HIV-positive individuals on highly active antiretroviral therapy in a multi-site Canadian cohort. HIV Medicine (2011), 12, 352–360.

Chen, X. & Fan, Y. (2005). Pseudo-Likelihood Ratio Tests for Semiparametric Multivariate Copula Model Selection. Canadian Journal of Statistics, vol. 33, 389-414

Chen, X., Fan, Y. & Tsyrennikov, V. (2006). Efficient Estimation of Semiparametric Multivariate Copula Models. Journal of the American Statistical Association.101:1228–1240.

Chen, X., Fan, Y. & Patton, A. (2004). Simple Tests for Models of Dependence Betweeen Multiple Financial Time Series: with Applications to U.S, Equity Returns and Exchange rates. London Economics Financial Markets Group Working Paper No. 483.

Cherubini, U., Luciano, E. & Vecchiato W (2004). Copula Methods in Finance. New York: John Wiley & Sons.

Corno, L. (2014). Learning (or Not) in Health-Seeking Behavior: Evidence from Rural Tanzania. Economic Development and Cultural Change, vol. 63, no. 1, 2014, pp. 27–72. JSTOR, www.jstor.org/stable/10.1086/677724.

Cunanan, K. & Koopmeiners, S. J. (2014). Evaluating the performance of copula models in phase I-II clinical trials under model misspecification. Retrieved from: http://www.biomedcentral.com/1471-2288/14/51

De Jong, P. & Heller, G. Z. (2008). Generalized Linear Models for Insurance Data. Cambridge University Press; New York. Retrieved from: http://www.cambridge.org/9780521879149

Engle, R. F. & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. Journal of Econometrics, 131, 3–27.

Fermanian, J. D. (2005). Goodness of Fit Tests for Copulas. Journal of Multivariate Analysis, 95, 119-152.

Genest, C., Quessy, J. F. & Rémillard, B. (2006). Goodness-of-Fit Procedures for Copula Models Based on the Probability Integral Transformation. Scandinavian Journal of Statistics, 33, 337-366.

Heckman, J.J. (1979). Sample selection bias as a specification error. Econometric a 47, 153-162.

Heckman, J.J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement 5(4), 475-492.

Heestermans, T., Browne, L. J., Aitken, C. S., Vervoort, C. S. & Klipstein-Grobusch, K. (2016) Determinants of adherence to antiretroviral therapy among HIV-positive adults in sub-Saharan Africa: a systematic review. BMJ Glob Health 2016; 1:e000125. Doi: 10.1136/bmjgh-2016-000125.

Huard, D., Évin, G. & Favre, A. C. (2005). Bayesian copula selection. Computational Statistics & Data Analysis 51 (2006) 809 – 822.

Joe, H. (1997). Multivariate Models and Dependence Concepts, Chapman & Hall/CRC. New York. Retrieved from: https://doi.org/10.1201/b13150.

Jouanin, J. F., Riboulet, G. & Roncalli, T. (n.d). Financial Applications of Copula Functions. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1032588.

Kuethe, T. H., Hubbs, T., & Waldof, B. (2009). Copula models for spatial point patterns and processes. Department of Agricultural Economics, Purdue University, West Lafayette, Indiana, USA. Retrieved from: http://www.ub.edu/sea2009.com/Paper/162.pdf

Lewis, H.G. (1974). Comments on selectivity biases in wage comparisons. Journal of Political Economy 82(6), 1145-1155.

Manner, H. (2007). Estimation and Model Selection of Copulas with an Application to Exchange Rates. Department of Quantitative Economics, Maastricht University, Netherlands.

Miller, C. W., Powers, A. K., Smith, K. M. & Cohen, S. M. (2013). Community viral load as a measure for assessment of HIV treatment as prevention. Department of Medicine: Division of Infectious Diseases, University of North Carolina, Chapel Hill, 13: 459–64.

Marra, G. & Radice, R (2017). A joint regression modeling framework for analyzing bivariate binary data in R. De Gruyter Open. Retrieved from https://doi.org/10.1515/demo-2017-0016

Marra, G., & Radice, R. (2017b). Generalized Joint Regression Modelling (GJRM). R package version 3.4.1. https://CRAN.R-project.org/package=GJRM.

Marra, G., Papageorgiou, G. & Radice, R. (2014). Estimation of a Semiparametric Recursive Bivariate Probit Model with Nonparametric Mixing. Australian & New Zealand Journal of Statistics; Stat.55 (3), 2014, 321–342.

Marra, G., Radice, R., Bärnighausen, T., Wood, S. N., & McGovern, M. E. (2017). A Simultaneous Equation Approach to Estimating HIV Prevalence with Non-Ignorable Missing Responses. Journal of the American Statistical Association, 112(518), 484-496. DOI: 10.1080/01621459.2016.1224713.

Marra, G. & Wyszynski, K. (2016). Semi-parametric copula sample selection models for count responses. Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK.

Ministry of Health and Social Services (2016). National Guidelines for Antiretroviral Therapy. Fifth edition 2016. MoHSS. Windhoek. Namibia.

McGovern, M. E., Barnighausen, T., Marra, G., & Radice, R. (2015). On the Assumption of Bivariate Normality in Selection Models: A Copula Approach Applied to Estimating HIV Prevalence. Epidemiology, 26(2), 229-237. DOI: 10.1097/EDE.0000000000000218.

Nachega, B. J., Mills, E. J. & Schechter, M. (2010). Antiretroviral therapy adherence and retention in care in middle-income and low-income countries: current status of knowledge and research priorities, Current Opinion in HIV and AIDS. vol. 5, no. 1, pp. 70–77.

Namibia Population-based HIV Impact Assessment (NAMPHIA) (2018). Preliminary results. Retrieved from: https://www.cdc.gov/globalhealth/countries/namibia/pdf/Namibia_FactSheet.pdf.

Noens, L., Van Lierde, M., De Bock, R., Verhoef, G., Zache´e, P., Berneman, Z., Martiat, P., Mineur, P., Van Eygen, K., MacDonald, K., De Geest, S., Albrecht, T. & Abraham, I. (2009). Prevalence, determinants, and outcomes of nonadherence to imatinib therapy in patients with chronic myeloid leukemia: the ADAGIO study. Republished online as Blood First Edition paper; DOI: 10.1182/blood-2008-12-196543.

Nelsen, R. B. (1998). An Introduction to Copulas. New York: Springer-Verlag.

Nelsen, R. B. (1999). An Introduction to Copulas. Lecture Notes in Statistics. New York: Springer.

Ntumba, A. (2018). Evaluation of the adherence and Viral Load Completion Accomplishments after Antiretroviral Therapy Initiation at Okuryangava Clinic and Katutura Health Centre-Windhoek-Khomas Region-Namibia. PhD Thesis. Texla International University.

Olowookere, S. A., Fatiregun, A. A., Ladipo, M. M. A., Abioye-Kuteyi, E. A. & Adewole, I. F. (2016). Effects of adherence to antiretroviral therapy on body mass index, immunological and virological status of Nigerians living with HIV/AIDS. Alexandria University Faculty of Medicine; Alexandria Journal of Medicine (2016) 52 51-54.

Payandeh, T. A., Farid-Rohani, R. M., & Qazvini, M. (2013). A GLM-Based Method to Estimate a Copula's Parameter(s). JIRSS (2013). Vol. 12, No. 2, pp 321-334.

Peltzer, K., Friend-du Preez, N., Ramlagan, S. & Anderson, J. (2010). Antiretroviral treatment adherence among HIV patients in KwaZulu-Natal, South Africa. Retrieved from http://www.biomedcentral.com/1471-2458/10/111; BMC Public Health 2010, 10:111.

Reda, A. A. & Biadgilign, S. (2012). Determinants of Adherence to Antiretroviral Therapy among HIV-Infected Patients in Africa. Hindawi Publishing Corporation AIDS Research and Treatment Volume 2012, Article ID 574656, doi:10.1155/2012/574656.

Reis, A., Guerra, M. P. & Lencastre, L. (2013). Treatment Adherence, Quality of Life and Clinical Variables in HIV/AIDS Infection. Retrieved from http://dx.doi.org/10.4236/wja.2013.33032 Published Online September 2013 (http://www.scirp.org/journal/wja).

Rougemont, M., Stoll, E. B., Elia, N. & Ngang, P. (2009). Antiretroviral treatment adherence and its determinants in Sub-Saharan Africa: a prospective study at Yaounde Central Hospital, Cameroon.

Rüschendorf, L. (2013). Copulas, Sklar's Theorem, and Distributional Transform. Mathematical Risk Analysis, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-642-33590-7 1, © Springer-Verlag Berlin Heidelberg.

Sethi, K. A., Celentano, D. D., Gange, J. S., Moore, D. R. & Gallant, E. J. (2003). Association between Adherence to Antiretroviral Therapy and Human Immunodeficiency Virus Drug Resistance. Retrieved from: https://academic.oup.com/cid/article-abstract/37/8/1112/435290; Clinical Infectious Diseases Society of America 2003; 32:1112-3

Seyoum, A., Ndlovu, P. & Temesgen, Z. (2017). Joint longitudinal data analysis in determinants of CD4 cell count change and adherence to highly active antiretroviral at Felege Hiwot Teaching and Specialized Hospital, North-west Ethiopia (Amhara Region). AIDS Res Ther (2017) 14:14. DOI 10.1186/s12981-017-0141-3.

Sklar, A. (1959). Fonctions de répartition àndimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8, pp. 229–231.

Smith, M. D. (2003). Modelling sample selection using Archimedean copulas. Wiley on behalf of the Royal Economic Society. The Econometrics Journal, Vol. 6, No. 1 (2003), pp. 99-123

Smith, J. C., Sabin, A. C., Youle, S. M., De Loes, K., S., Lampe, C. F., Madge, S., Cropley, I., Johnson, A. M., & Phillips, N. A. (2004). Factors Influencing Increases in CD4 Cell Counts of HIV-Positive Persons Receiving Long-Term Highly Active Antiretroviral Therapy. Retrieved from https://academic.oup.com/jid/article-abstract/190/10/1860/2191801.

Strazzera, E. & Genius, M. (2004). The Copula Approach to Sample Selection Modelling: An Application to the Recreational Value of Forests, Nota di Lavoro, Fondazione Eni Enrico Mattei, No.73.2004

Sungur, E. A. (2005). A Note on Directional Dependence in Regression Setting. Communications in Statistics: Theory and Methodology. 34:1957–1965.

UNAIDS (2016). Global AIDS update. Retrieved from http://www.unaids.org/en/resources/ documents/2016/Global-AIDS-update-2016, accessed 19 May 2017.

UNAIDS (2016). 90-90-90 On the right track towards the global target. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/90_90_90_Progress_ReportFINAL pdf.

UNAIDS (2017). UNAIDS data 2017. Retrieved from http://www.unaids.org/sites/default/files/media_asset/ 20170720_Data_book_2017_en.pdf

UNAIDS/WHO. AIDS epidemic update. Retrieved from http://www.unaids.org; December 2009 [accessed 01.09.10].

Van der Klaauw, B. & Koning, R. H. (2003). Testing the normality assumption in the sample selection model with an application to travel demand. Journal of Business and Economic Statistics, 21, 31-42. DOI: 10.1198/073500102288618739.

Vuong, Q. H., (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis, Econometrica, 57(2): 307-333.

Wiesenfarth M, Kneib T (2010) Bayesian geoadditive sample selection models. J R Stat Soc C 59(3):381–404

WHO (2016). Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV.

WHO (2017). National Guidelines for Antiretroviral Therapy, Fifth Edition 2016. Retrieved from https://aidsfree.usaid.gov/sites/default/files/na_national_guidelines_art.pdf

Wojtyś, M., Marra, G., Radice, R. (2016). Copula Regression Spline Sample Selection Models: The R Package SemiParSampleSel. Journal of Statistical Software.

Yan, J. (2006). Enjoy the Joy of Copulas. Technical Report #365, Department of Statistics and Acturial Science, University of Iowa.

APPENDIX

Appendix A: R codes using GRJM package for Windhoek data analysis

```
## We load GJRM, read the data-set and specify the selection and outcome equations
 #by including smooth functions for age, marital status, WHO stage and facilities


library("GJRM")


##load data from Excel
Jason1 <- read.table("file:///C:/Users/me/Documents/UNAM/MSc Biostatistics/2018/
                    Jason Dataset Windhoek/Jason1.csv", header = TRUE, sep ="," )
str(Jason1)


# discribe age group
Age <- table(Jason1$Age, Jason1$Sex)
Age


# modeling the selection equation
sel.eq <- Adherence ~ as.factor(Sex) + as.factor(Age) + as.factor(Facility) +
  as.factor(WHOstage) + as.factor(Regimen) + as.factor(CD4ARTIn)


# modeling the outcome equation
out.eq <- VLS12 ~ as.factor(Sex) + as.factor(Age) +  as.factor(Facility) +
  as.factor(WHOstage) + as.factor(Regimen) + as.factor(CD4ARTIn) + as.factor(Adherence)


# modeling copula parameter
theta <- ~ as.factor(Sex) + as.factor(Facility)
```

```
f.list <- list(sel.eq, out.eq, theta)
mr <- c("probit", "probit")
bpN <- gjrm(f.list, data = Jason1, Model= "B", margins = mr)
bpF <- gjrm(f.list, data = Jason1, BivD = "F", Model = "B", margins = mr)
bpC90 <- gjrm(f.list, data = Jason1, BivD = "C90", Model= "B", margins = mr)
bpC270 <- gjrm(f.list, data = Jason1, BivD = "C270", Model= "B", margins = mr)
bpT <- gjrm(f.list, data = Jason1, BivD = "T", Model= "B", margins = mr)
bpG90 <- gjrm(f.list, data = Jason1, BivD = "G90", Model= "B", margins = mr)
bpG270 <- gjrm(f.list, data = Jason1, BivD = "G270", Model = "B", margins = mr)
bpJ90 <- gjrm(f.list, data = Jason1, BivD = "J90", Model = "B", margins = mr)
bpJ270 <- gjrm(f.list, data = Jason1, BivD = "J270", Model = "B", margins = mr)
bpAMH <- gjrm(f.list, data = Jason1, BivD = "AMH", Model = "B", margins = mr)
bpFGM <- gjrm(f.list, data = Jason1, BivD = "FGM", Model = "B", margins = mr)

## we then need to check for convergence by using conv.check ()
conv.check(bpN)

## Based on the AIC and BIC the preferred model is the survival Clayton copula
AIC(bpN, bpF, bpFGM, bpAMH, bpT, bpJ90, bpJ270, bpG90, bpG270, bpC90, bpC270)
BIC(bpN, bpF, bpFGM, bpAMH, bpT, bpJ90, bpJ270, bpG90, bpG270, bpC90, bpC270)


## let look at the results now
set.seed(1)
summary(bpN)
summary(bpF)
summary(bpFGM)
summary(bpAMH)
```

```
summary(bpT)
summary(bpG90)
summary(bpG270)
summary(bpJ90)
summary(bpJ270)
summary(bpC90)
summary(bpC270)

# Vuong test to compare to model
#....................................................................................
VuongClarke( bpG90, bpN)
```

Appendix B: R codes using GRJM package for Erongo region data analysis

```
## Preliminary results for Erongo region


## load data from Excel

Erongo <- read.table("file:///C:/Users/me/Documents/UNAM/MSc Biostatistics/2018/Jason
Dataset Erongo/Erongo.csv", na.strings = ".", header = TRUE, sep =",")

str(Erongo)


## Factors associated with log transformation of viral load

LogVL <- lm(Log_Viral_Load ~ as.factor(Sex) + as.factor(Facility)+ as.factor(AgeGroupIn) +
as.factor(Stage_group) +

            as.factor(Weightgroup) + as.factor(CD4_Group) + as.factor(ARTadherence), data =
Erongo)

summary(LogVL)

confint(LogVL, level = 0.95)



## Factors associated with ART adherence

ARTmodel <- lm(ARTadherence ~ as.factor(Sex) + as.factor(Facility) + as.factor(AgeGroupIn) +
as.factor(Stage_group) +

              as.factor(Weightgroup) + as.factor(CD4_Group), data = Erongo)

summary(ARTmodel)

confint(ARTmodel, level = 0.95)



####################Sample selection model using copula approach###############

## We load GJRM, and specify the selection and outcome equations

##including sigma and theta functions

library("GJRM")
```

```
##modeling the selection, outcome and copula parameters (sigma and theta)
sel.eq <- ARTadherence ~ as.factor(Sex) + as.factor(Stage_group) + as.factor(Facility) +
  s(AgeIn) + s(AgeC) + s(Weight_Current) + s(CD4_cell_count)
out.eq <- Viral_Load ~ as.factor(Sex) + as.factor(Stage_group) + as.factor(Facility) +
  as.factor(ARTadherence) + s(AgeIn) + s(AgeC) + s(Weight_Current) + s(CD4_cell_count)
sigma.eq <- ~ 1
theta.eq <- ~ as.factor(Sex) + as.factor(Facility)

## we then use list() function in order to join the marginals
f1 <- list(sel.eq, out.eq, sigma.eq, theta.eq)

##fitting the different types of copula
#....................................................................................................
bpN <- gjrm(f1, data = Erongo, margins = c("probit","N"), gamlssfit = TRUE, Model = "B")
bpF <- gjrm(f1, data = Erongo, margins = c("probit","N"), gamlssfit = TRUE, BivD = "F", Model = "B")
bpFGM <- gjrm(f1, data = Erongo, margins = c("probit","N"), gamlssfit = TRUE, BivD = "FGM", Model = "B")
bpAMH <- gjrm(f1, data = Erongo, margins = c("probit","N"), gamlssfit = TRUE, BivD = "AMH", Model = "B")
bpG0 <- gjrm(f1, data = Erongo, margins = c("probit","N"), gamlssfit = TRUE, BivD = "G0", Model= "B")
bpG180 <- gjrm(f1, data = Erongo, margins = c("probit","N"), gamlssfit = TRUE, BivD = "G180", Model = "B")
bpC0 <- gjrm(f1, data = Erongo, margins = c("probit","N"), gamlssfit = TRUE, BivD = "C0", Model= "B")
bpC180 <- gjrm(f1, data = Erongo, margins = c("probit","N"), gamlssfit = TRUE, BivD = "C180", Model= "B")
```

```
## we then need to check for convergence by using conv.check () function
conv.check(bpN)

## We then need to estimate the dependence parameter
## of the copula coefiecients corresponding 1.
## let look at the results now the above distributions
set.seed(1)
summary(bpN)
summary(bpF)
summary(bpFGM)
summary(bpAMH)
summary(bpG0)
summary(bpG180)
summary(bpC0)
summary(bpC180)

## Checking the lowest model by using AIC and BIC
#..........................................................................
AIC(bpN, bpF, bpFGM, bpAMH, bpG0, bpG180, bpC0, bpC180)
BIC(bpN, bpF, bpFGM, bpAMH, bpG0, bpG180, bpC0, bpC180)

##plots of the best model using selection equation
#.......................................................................
par(mfrow=c(2,2),mar=c(4.5,4.5,2,2), cex.axis = 1.6, cex.lab = 1.6)
plot(bpC180, eq = 1, select = 1, seWithMean = TRUE, scale = 0, shade = TRUE, jit = TRUE)
plot(bpC180, eq = 1, select = 2, seWithMean = TRUE, scale = 0, shade = TRUE, jit = TRUE)
plot(bpC180, eq = 1, select = 3, seWithMean = TRUE, scale = 0, shade = TRUE, jit = TRUE)
plot(bpC180, eq = 1, select = 4, seWithMean = TRUE, scale = 0, shade = TRUE, jit = TRUE)
```

```
##plots of the best model using outcome equation
#..........................................................................
par(mfrow=c(2,2),mar=c(4.5,4.5,2,2), cex.axis = 1.6, cex.lab = 1.6)
plot(bpC180, eq = 2, select = 1, seWithMean = TRUE, scale = 0, shade = TRUE, jit = TRUE)
plot(bpC180, eq = 2, select = 2, seWithMean = TRUE, scale = 0, shade = TRUE, jit = TRUE)
plot(bpC180, eq = 2, select = 3, seWithMean = TRUE, scale = 0, shade = TRUE, jit = TRUE)
plot(bpC180, eq = 2, select = 4, seWithMean = TRUE, scale = 0, shade = TRUE, jit = TRUE)

# Vuong test to compare to model
#............................................................................
VuongClarke( bpG0, bpN)



########Heckman-type model with ART adherence (probit) and Viral load (Gamma)#########


library("GJRM")
## load data from Excel
Erongo <- read.table("file:///C:/Users/me/Documents/UNAM/MSc Biostatistics/2018/Jason
Dataset Erongo/Erongo.csv", na.strings = ".", header = TRUE, sep =",")
str(Erongo)
##modeling the selection, outcome and copula parameter theta
sel <- ARTadherence ~ as.factor(Sex) + as.factor(Stage_group) + as.factor(Facility) +
  s(AgeIn) + s(AgeC) + s(Weight_Current) + s(CD4_cell_count)
out <- Viral_Load ~ as.factor(Sex)  + as.factor(Stage_group) +
  as.factor(Facility) + s(AgeIn) + s(AgeC) + s(Weight_Current) + s(CD4_cell_count)
```

```
## we then use list() function in order to join the marginals
f1 <- list(sel, out)

##fitting the different types of copula
#..................................................................
N <- gjrm(f1, data = Erongo, BivD = "N", margins = c("probit", "GA"), Model = "BSS")
MF <- gjrm(f1, data = Erongo, BivD = "F", margins = c("probit", "GA"), Model = "BSS")
FGM <- gjrm(f1, data = Erongo, BivD = "FGM", margins = c("probit", "GA"), Model = "BSS")
AMH <- gjrm(f1, data = Erongo, BivD = "AMH", margins = c("probit", "GA"), Model = "BSS")
G <- gjrm(f1, data = Erongo, BivD = "G0", margins = c("probit", "GA"), Model = "BSS")
G180 <- gjrm(f1, data = Erongo, BivD = "G180", margins = c("probit", "GA"), Model = "BSS")
C <- gjrm(f1, data = Erongo, BivD = "C0", margins = c("probit", "GA"), Model = "BSS")
C180 <- gjrm(f1, data = Erongo, BivD = "C180", margins = c("probit", "GA"), Model = "BSS")

## we then need to check for convergence by using conv.check () function
conv.check(G)

## We then need to estimate the dependence parameter of the copula coefiecients corresponding 1.
## let look at the results now the above distributions
set.seed(1)
summary(N)
summary(MF)
summary(FGM)
summary(AMH)
summary(G)
summary(G180)
summary(C)
summary(C180)
```

```
## Checking the lowest model by using AIC and BIC
AIC(N, MF, FGM, AMH, G, G180, C, C180)
BIC(N, MF, FGM, AMH, G, G180, C, C180)

#The best model using selection equation and outcome equation
#....................................................................................

set.seed(1)
summary(G180)
```